J Microbiol Immunol Infect
2000;33:69-78

Online identification of viruses

A.S. Kolaskar[1,2], P.S. Naik[1]

[1]Bioinformatics Centre, University of Pune, Pune, India; and [2]Bioinformatics, American Type Culture Collection, Manassas, USA

Received: May 4, 1999 Revised: February 17, 2000 Accepted: February 24, 2000

A computerized animal virus information system is developed in the Sequence Retrieval System (SRS) format. This database is available on the Word Wide Web (WWW) at the site http://bioinfo.ernet.in/www/avis/avis.html. The database has been used to generate large number of identification matrices for each family. The software is developed in C.Unix shell scripts and Hypertext Marked-up Language (HTML) to assign the family to an unknown virus deterministically and to identify the virus probabilistically. It has been shown that such web based virus identification approach provides results with high confidence in those cases where identification matrix uses large number of independent characters. Protein sequence data for animal viruses have been analyzed and oligopeptides specific to each virus family and also specific to each virus species are identified for several viruses. These peptides thus could be used to identify the virus and to assign the virus family with high confidence showing the usefulness of sequence data in virus identification.

Key words: Virus identification, web based

In the microbial world, viruses the obligate parasites have established themselves as very important biological species. Viruses require hosts for their growth and reproduction, which could be bacteria, plants and animals. Thus, the nature and characteristics of viruses differ. Based on the hosts, viruses are categorized as animal viruses, plant viruses and phages. Further, the distinctive features among the macro characters as well as differences at molecular level have helped the taxonomists to classify known viruses into 184 genera and 54 families [1]. Most of the classification of viruses is carried out at the species level and macro character values are generally used for their classification. In very rare cases, molecular properties, particularly the sequences of oligopeptides/oligonucleotides are used for classification and identification [2]. One of the main reasons for such a lacuna is nonavailability of comprehensive data on viruses at one place that the experts can use to formulate rules of classification.

Identification and classification of viruses is of paramount importance not only because of its esoteric nature, but because these viruses cause serious damage to crops, animal and human health as well as environment. Design of control strategy is possible only if the identification is done at an early stage. Though viruses have played an important role in early development of molecular biology, the role/function of several viruses is still unknown. For example, Herpesvirus 6 is present in most Indian population in an inactivated form and causes no damage. Activated Herpesvirus 6 in an human immunodeficiency virus (HIV) positive mother seems to help in transmitting HIV. The coexistence of more than one virus and thus the function of these viruses in the presence of each other are still not understood and studied.

In the light of the above mentioned facts a computerized database on animal viruses called the Animal Virus Information System (AVIS) was created at our center in the University of Pune, India and is available at the web site: http://bioinfo.ernet.in/www/avis/avis.html [3]. This database was used to develop a computerized method to identify the virus on the World Wide Web (WWW). This approach is discussed in succeeding sections.

Materials and Methods

AVIS database

The information on viruses in AVIS is at two main levels: alphanumeric information and pictorial information. Pictorial information contains mostly high-resolution electron micrographic pictures of the viruses. The alphanumeric information is broadly divided into 17 different categories as given in table 1. The experts

Corresponding author: Dr. A.S. Kolaskar, Bioinformatics Centre, University of Pune, Pune 411007, India.

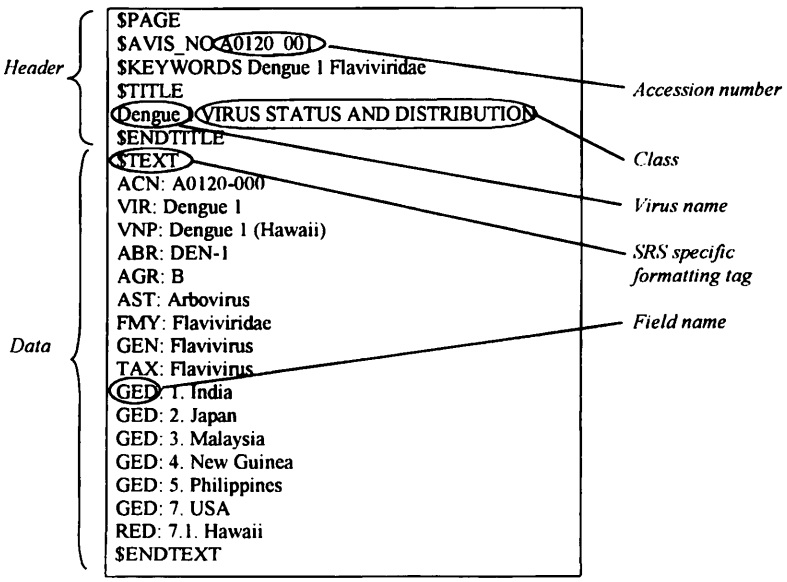

Fig. 1. Record of AVIS formatted for SRS. The example for the VIRUS STATUS AND DISTRIBUTION record of the Dengue 1 virus.

verify the data and thus the correctness of the data is maintained. Data is updated regularly by searching the literature using keywords in an automatic fashion.

The information on about 1700 animal viruses is organized in the Sequence Retrieval System (SRS) format and implemented on the web using the SRS search engine. SRS was developed at the European Bioinformatics Institute, UK [4]. SRS is considered as one of the standard formats to access biological databases on the web and several databases including the nucleic acid and protein sequence databases are available in this format. SRS uses the "Icarus" scripting language to parse flat file databases. SRS produces

Table 1. Classes created for the virus database

No.	Classes
1	Virus status and distribution
2	Original source of the virus
3	Method of isolation and validity
4	Physicochemical properties
5	Stability of infectivity and virulence
6	Virion morphology
7	Morphogenesis
8	Hemagglutination
9	Antigenic relationship
10	Susceptibility of cell systems
11	Natural Host range
12	Experimental viremia
13	Histopathology
14	Human disease
15	Links with other data banks
16	Pictorial information
17	References

almost all of the query reports dynamically for the web. SRS also allows linking of other databases loaded on the computer system. The SRS formatted database file of AVIS was indexed with the indexing programs available with the SRS package and was loaded on the computer. The record structure of the AVIS database in the SRS format is given in figure 1.

The AVIS also contains two subdatabases: (i) Database containing all viral nucleic acids sequences. This information is extracted from Genbank and European Molecular Biology Laboratory (EMBL) databases of nucleic acids sequences. The data is organized in the Genbank format. (ii) Database consists of virus protein sequences. These sequences are extracted from NBRF-PIR and Swiss-Prot and organized in the NBRF-PIR format. The protein sequence information is organized under the virus family. Conversion of sequence data into the NBRF-PIR format allows the data analysis to be carried out using PSQ and NAQ software that is available in the public domain, and also available along with the ATLAS CDs from PIR International [5]. In order to use these software along with the virus sequence data in the web environment, necessary utilities were written using the Unix shell script language.

Virus identification

The software has been developed to identify viruses online through the web. The software developed for virus identification uses the schema given in figure 2. As can be seen from figure 2, the process of

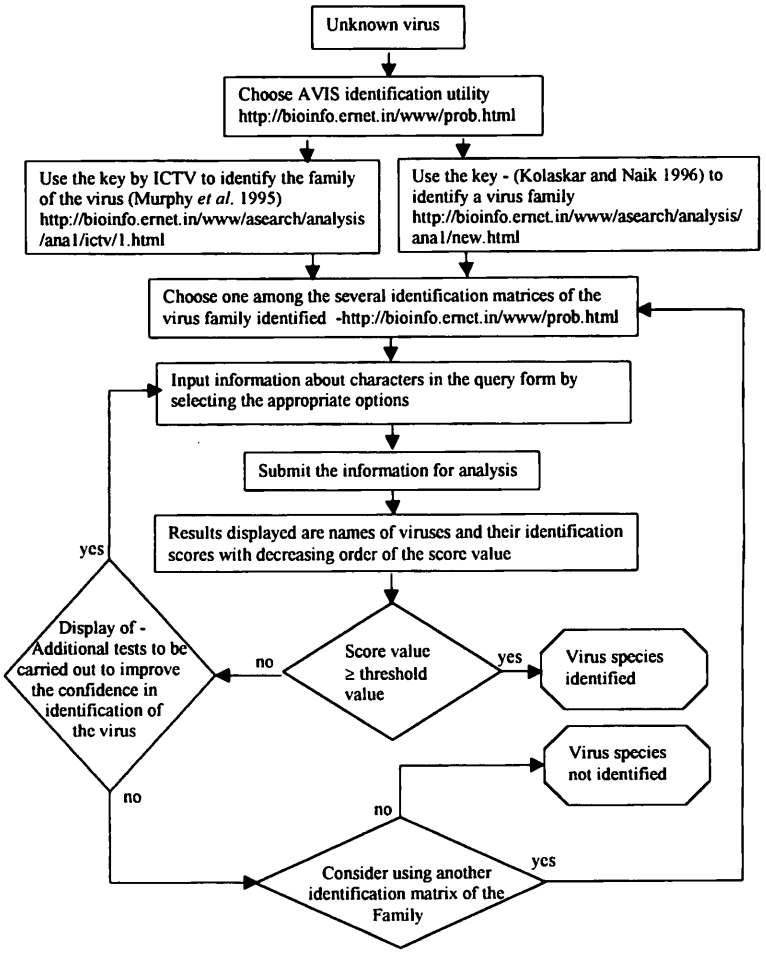

Fig. 2. Diagram illustrating the computer aided identification of viruses on the WWW.

identification has been divided into two parts: (i) assignment of the family uses a deterministic approach and (ii) identification of a virus species using a probabilistic approach.

The software consists of two main modules, input module and the compute module. The input module consists of three components: (i) input for identification of a virus family, (ii) input of character values to create identification matrix and (iii) direct input of identification matrix developed by the user. In each of these cases, a predesigned screen allows an user to feed in the data, by clicking one of the possible alternatives. These pop-up screens are user-friendly. The programs are written in HTML and Unix Shell scripts. The compute module is divided into three parts: (i) assignment of the virus family, (ii) creation of the identification matrix, and (iii) identification of the virus species and display of results. The programs for the compute module are written in Unix shell scripts and C.

The assignment of the family to the unknown virus is carried out deterministically using the key given in figure 3. The character-based keys using neg-entropy as a measure were developed. Use of such characters showed that only 12 characters are necessary and sufficient to identify one of the 25 animal virus families [6] (Table 2). The International Committee on Taxonomy of Viruses (ICTV) has also developed the keys giving higher importance to (i) nature of viral genome, (ii) the strandedness of the viral genome and (iii) the replication strategy of the virus [7]. These ICTV keys can also be used to assign the family. The software developed allows choice of any one of the above methods.

Only after virus family is assigned, the process of virus identification is continued. To carry out virus identification, the first step is to create identification matrices for each family using the characters those provide high resolution and thus have different values for different viruses. Development of such matrices is described in detail in our earlier communication [8], but the brief outline of the method used with an example is given in figure 4.

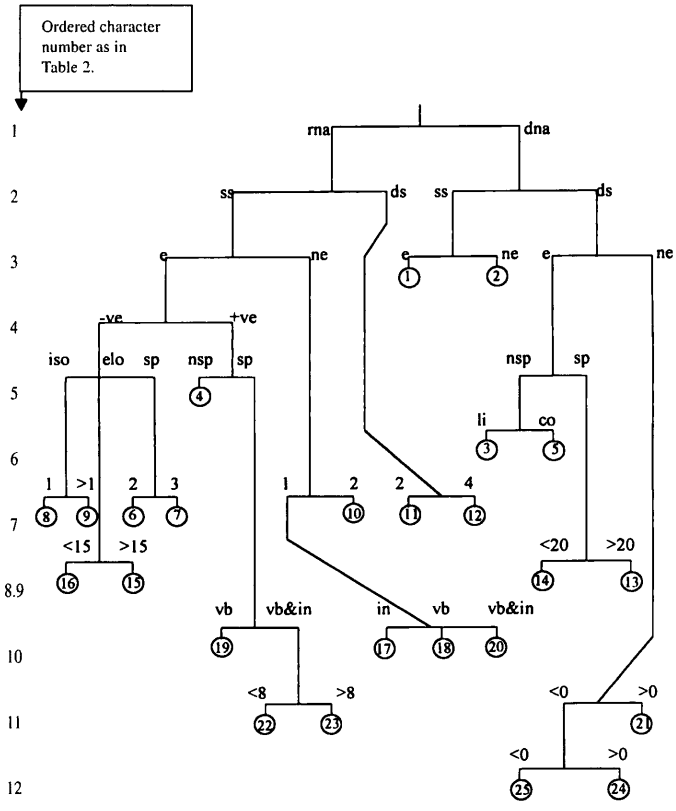

Fig. 3. A tree drawn using 12 ordered characters as given in table 2.The short-form given here are explained in table 2. The Families identified are represented as encircled numbers and correspond to the superscripts in table 2. Note that the tree has 22 nodes and a maximum of 7 steps to identify any family.

Assignment of virus family using sequence data

For viruses belonging to each family, a study has been carried out to find out which protein is sequenced for most of the members of the family. For each family, multiple alignment was carried out for available protein sequences using CLUSTALW [9]. The bias and penalty

Table 2. Hierarchically ordered characters used to form identification key

SL No.	Ordered characters	Families identified	
		Number	Name
1	RNA/DNA (rna/dna)	0	
2	Double/Single stranded (ds/ss)	0	
3	Enveloped/Nonenveloped (e/ne)	2	*Polydnaviridae*[1], *Parvoviridae*[2]
4	RNA -ve sense/+ve sense (-ve/+ve)	0	
5	Virion shape isometric/ elongated/nonspherical/spherical (iso/elo/nsp/sp)	2	*Herpesviridae*[3], *Coronaviridae*[4]
6	Genome linear/coiled (li/co)	1	*Baculoviridae*[5]
7	No. of genome segments (1/> 1/2/3/4)	7	*Arenaviridae*[6], *Bunyaviridae*[7], *Paramyxoviridae*[8], *Orthomyxoviridae*[9], *Nodaviridae*[10], *Birnaviridae*[11], *Reoviridae*[12]
8	DNA size in kb (< 20/> 20)	2	*Poxviridae*[13], *Hepednaviridae*[14]
9	RNA size in kb (< 15/> 15)	2	*Filoviridae*[15], *Rhabdoviridae*[16]
10	Host type vertebrate/insect/ vertebrate + insect (in/vb/vb & in)	4	*Tetraviridae*[17], *Caliciviridae*[18], *Retroviridae*[19], *Picornaviridae*[20]
11	Percentage Carbohydrate content (< 0/> 0 & < 8/> 8)	3	*Adenoviridae*[21], *Togaviridae*[22], *Flaviviridae*[23]
12	Percentage Lipid content (< 0/> 0)	2	*Iridoviridae*[24], *Papovaviridae*[25]

Note: The animal virus family identified using these ordered keys is also given.

	Viruses (OTUs) / Characters	BHK-21	BS-C1	Chick embryo	PK	Mouse embryo	PS	Vero	KB	SP1K	MA111	PK15
P1	Uukuniemi	1	1	1	1	0	0	0	0	0	0	0
P2	Sabo	1	0	0	0	1	0	0	0	0	0	0
P3	Jamestown Canyon	1	0	0	0	0	1	1	0	0	0	0
P4	California Encephalitis	0	1	1	0	0	1	0	1	1	0	0
P5	Buttonwillow	1	0	0	0	0	1	1	0	1	1	1

The data is represented as follows : The cytopathic effects observed on these cell lines by the five chosen viruses are represented as : 1 for yes, 0 for no and unknown. Number of OTUS = n = 5

The distances dP_k dP_1 is obtained by calculating the Euclidean distances of the viruses by using the Jaccard's coefficent is :

dP_1	0.00				
dP_2	0.89	0.00			
dP_3	0.91	0.86	0.00		
dP_4	0.84	1.00	0.92	0.00	
dP_5	0.94	0.92	0.70	0.88	0.00
	dP_1	dP_2	dP_3	dP_4	dP_5

The average distance for the P_kth row with respect to other OTUs is :

$$W_k = \left\lceil \left[\sum_{\substack{l=1 \\ k \neq l}}^{n} dP_k \; dP_l \right] / (n-1) \right\rceil \; . \; 100$$

Where dP_k dP_1 is the euclidean distance between $OTUP_k$ and $OTUP_1$

The resulting weight matrix obtained is as follows :

P_1	89	89	89	89	01	01	01	01	01	01	01
P_2	91	01	01	01	91	01	01	01	01	01	01
P_3	84	01	01	01	01	84	84	01	01	01	01
P_4	01	91	91	01	01	91	01	91	91	01	01
P_5	86	01	01	01	01	86	86	01	86	86	86

Fig. 4. Procedure used to develop the Virus Identification Matrix in Probabilistic identification. An example of a few viruses of the *Bunyaviridae* family.

values are varied so as to get the best multiple alignment results [10]. Peptides, common in every virus of a given family or genus having a length greater than or equal to six was picked up. Such peptides are called "putative signature peptides" of the virus family or genus under study. Presence of each of these peptides in the whole database was studied through a program called MATCH. If the putative signature peptides were found to be present only in the viruses of the family/genus under study but absent in every other sequence in the whole data base even with three mismatches, then that peptide is called the signature peptide of the family/ genus. Unique peptides are obtained for 16 families by analyzing protein sequence data. These peptides are given in table 3. Such unique peptides thus can be used to assign the family of the unknown virus.

Identification using signature peptides

Identification of a member virus of a particular family is possible with the use of signature sequences. This procedure can be used to confirm identification. Signature peptides were obtained for the members of the *Flavivirus* genus. The sequences of the envelope glycoprotein (Egp) of 19 *Flaviviruses* extracted from sequence databases were studied. In the multiple alignment program, a single linkage cluster is developed which brings two sequentially most similar proteins close together. Amino acid sequences of such most homologous protein pairs were aligned. By examining the alignment, an oligopeptide having at least 10 amino

Table 3. Unique peptide sequences for animal virus families obtained from analysis of virus protein sequence data

Family	Genus	Protein	Unique peptide
Togaviridae	*Alphavirus*	Structural polyprotein	AYEHXXV/TXPN
Filoviridae	*Filovirus*	Nucleocapsid protein	PQLSAIALGVAT AHGSTLAGVNV GEQYQQLREAA
Iridoviridae	*Lymphocystivirus* *Iridovirus*	Capsid protein	TSXFIDXAT IEKXXYGG
Papovaviridae	*Papillomavirus*	L1 protein	CKYPDF/Y GHPLF/YNKV/L
	Polyomavirus	Coat protein VP1	PDPXXNEN GVGPLCK QVEEVR
		Coat protein VP2	WXLPLXLGLYG
Arenaviridae	*Arenavirus*	Surface glycoprotein	MLXKEYXXRQXXTP PTHXHIXGXXCPXPHR LXLXGRSC
Flaviviridae	*Flavivirus*	Nonstructural protein 1 Envelope glycoprotein	CWYXMEIRP DRGWGNXCGXFGKG
Adenoviridae		Hexon protein	FKPYSGTA GVLAGQ PNYCFPL NPFNHHRN
Caliciviridae	*Calicivirus*	Coat protein	LXPXXNPYLXH SGSGVFGGKLAA MLQYPHVLFDARQ HGSIPSDLIP VFQXNRHFDF TXGWSTP PDGWPDTTI
Paramyxoviridae	*Paramyxovirus* *Morbillivirus*	Hemagglutinin neuraminidase Hemagglutinin	NRKSCS DVLTPLFKIIGDE
Picornaviridae	*Enterovirus* *Rhinovirus* *Hepatovirus* *Cardiovirus* *Apthovirus*	Genome polyprotein	DGYXXQXXXXXDD
Nodaviridae	*Nodavirus*	Coat protein precursor	FLKCAFA DPGKGIPD FRYASM
Parvoviridae	*Parvovirus* *Dependovirus*	Coat protein VP1	TPWXXXXXNXXXXXFXP
Orthomyxoviridae	Influenza-A *Influenza-B* *Dhori*	Nucleoprotein	GQIXXXPXFXXXR
	Influenza-C	Nucleoprotein	TQIXXXAXFXXXR
Hepadnaviridae	*Orthohepadnavirus*	Core antigen	EYLVSFGVWI
Herpesviridae	*Rhadinovirus* Major *Lymphocryptovirus*	Capsid protein	PXXYXXXXXXXXNXVTA GNXPXXLXPXXF
	Simplexvirus *Varicellavirus* *Betaherpesvirinae* *Cytomegalovirus*		HPGXXXTXVRXD
Birnaviridae	*Birnavirus*	Structural protein VP2	GPASIPD

Note: Single letter amino acid code is used x represent any amino acid.

Table 4. Identification matrices created for different virus families

Family	Type of identification matrices	No. of OTUs/ viruses in the matrix	No. of characters considered
Bunyaviridae	Cytopathic effects observed on cell lines	63	43
	Natural host range — mosquitoes	95	120
Rhabdoviridae	Natural host range	46	150
Togaviridae	Natural host	28	218
	Experimental viremia	28	21
	Cytopathic effects observed on cell lines	22	18
Reoviridae	Natural host range	62	124
Flaviviridae	Natural host range	62	278
	Experimental viremia	58	53
	Amino acid composition of envelope protein	14	20
	Amino acid composition of nucleocapsid protein	13	20
	Antigenic relationship	48	48
	Cytopathic effects observed on cell lines	52	18
	Human disease	28	27
Coronaviridae	Amino acid composition of nucleocapsid protein	10	20
Parvoviridae	Amino acid composition of coat protein VP1	8	20
Poxviridae	Amino acid composition of fusion protein	8	20

acid residues out of which at least four amino acids are different in two proteins were picked up. These unique peptides in the Egp of the virus are termed as candidate signature peptides. Using the program such as MATCH, a search was carried out to confirm its uniqueness, by allowing three mismatches. The initial search was carried out on related proteins, namely Egp of *Flaviviridae* family. Only after the uniqueness with three mismatches was established at this level, a search on the full database of protein sequences was carried out again by allowing three mismatches. If no other matching peptide was observed, then the peptide was termed as signature peptide.

Results

It has been observed that the software developed to identify a virus on the WWW can identify an unknown virus with a high degree of confidence [8]. Character based identification matrices have been prepared using AVIS to identify viruses belonging to different families. The list of identification matrices created for *Bunyaviridae, Rhabdoviridae, Togaviridae, Flaviviridae, Coronaviridae, Parvoviridae* and *Poxviridae* are listed in table 4. It can be seen from table 4 that seven different identification matrices are generated for *Flaviviridae* family. A combined matrix was also developed. Accuracy of identification of virus species in this family was almost 99%. This is because of our interest in development of vaccines against flaviviruses. The usefulness of such matrix approach in identification of viruses was checked. Our results have shown that one can identify the virus with 99% accuracy provided the matrix is sufficiently large and prepared using independent character values. Additional matrices for various other families will soon be added to make the web based identification more useful to virologists. The users are requested to visit the web site http://bioinfo.ernet.in/www/prob.html at regular intervals to find out the new matrices.

Sequence information of viruses is increasing at a rapid rate. Peptides specific to each of the 16 animal virus families are identified from our analysis of protein sequence alignment approach. These virus family specific peptides are listed in the table 3. The proteins used to obtain the virus family specific peptides are also listed in this table. It can be seen from table 3 that wherever possible an attempt has been made to identify the genus specific peptide. For example the peptide NRKSCS from hemagglutinin neuraminidase is specific to genus *Paramyxovirus*. The peptide DVLTPLFKIIGDE from hemagglutinin is specific to genus *Morbillivirus*. Both these genus belong to virus family *Paramyxoviridae*. The peptides listed in the last column of table 3 are present only in the protein from the virus family under consideration and in no other protein present in the protein sequence data bank. Note that MATCH was run with zero mismatches. In most cases, more than one peptide are found to be unique

Table 5. Unique peptides for members of the family *Flaviviridae*

Virus	Unique peptide	Unique up to number of mismatches
St. Louis encephalitis virus	VNPFISTGGAN	3
	EGRPAT	0
Murray valley encephalitis virus	VTANPYVASSTA	3
Japanese encephalitis virus	LDVRMINIEA[S/V]Q	3
West Nile virus	TTKATGWIIQK	3
Kunjin virus	STKATGRTILKE	3
Langat virus	DGAEAWNEAGR	3
	FTCEDKK	0
	VGFSGTRP	0
Yellow fever virus	MRVTKDTN[D/G][N/S]NL	3
Powassan virus	KDNQDWNSVE	3
Dengue type 1 virus	GTVLVQV	0
Dengue type 2 virus	GTIVIRV	0
Dengue type 3 virus	TEATQL	0
	GTILIKV	0
Dengue type 4 virus	TTAKEVA	0
	GTTVVKV	0
Tick borne encephalitis virus	GFLTSVGKA	0
Louping ill virus	NPHWNNVER	0

Note: Peptides if unique up to 3 number of mismatches, they are termed as signature peptides. The peptides are from the envelope glycoproteins from the respective viruses.

and thus the presence of all these peptides in the proteins of the virus mentioned in this table helps to assign the virus family with high degree of confidence. Table 3 shows that one can use these unique peptides to assign the family of the virus for 16 virus families.

From the database Egp sequences of 19 *Flaviviruses* were extracted. In figure 5, conserved amino acids from all 19 Egps of the *Flaviviruses* are given. Figure 5 shows that the peptide ^{98}DRGWGNXCGXFGKG^{111} is conserved (X represents any amino acid). As can be seen in figure 5, at position 104 though X is used to represent occurrence of any amino acid, only two amino acids occur namely glycine frequently and histidine rarely. In a similar fashion at position 107, leucine is highly preferred and in rare cases phenylalanine occurs though it is represented by X (Fig. 5). Occurrence of DRGWGNXCGXFGKG, in any other protein sequences in the PIR and Swiss-Prot databank was studied. Even with the allowance of three mismatches, the oligopeptide DRGWGNXCGXFGKG did not occur in any other protein except in Egps of *Flaviviruses*. This suggests that the above peptide can be used as a signature of the genus *Flavivirus*.

Discussion

The animal virus information system AVIS has been analyzed and character values are grouped to form the identification matrix, an approach used by microbiologists for bacterial identification. It has been showed that this approach is equally useful in virus identification. As the quality and quantity of data on viruses increases and it gets captured in the structured computerized database the automatic generation of identification matrices with high resolution becomes easy. The method described here can be extended. The results of *Flaviviridae* family described support the statement that the probabilistic identification with high accuracy can be achieved. Molecular sequence data, from proteins and nucleic acids, that is rapidly accumulating has the potential to substantially improve the identification and taxonomy of viruses. The attempt described in this manuscript of analysis of protein sequences suggests that the method developed is useful and also sufficiently general. The results suggest that

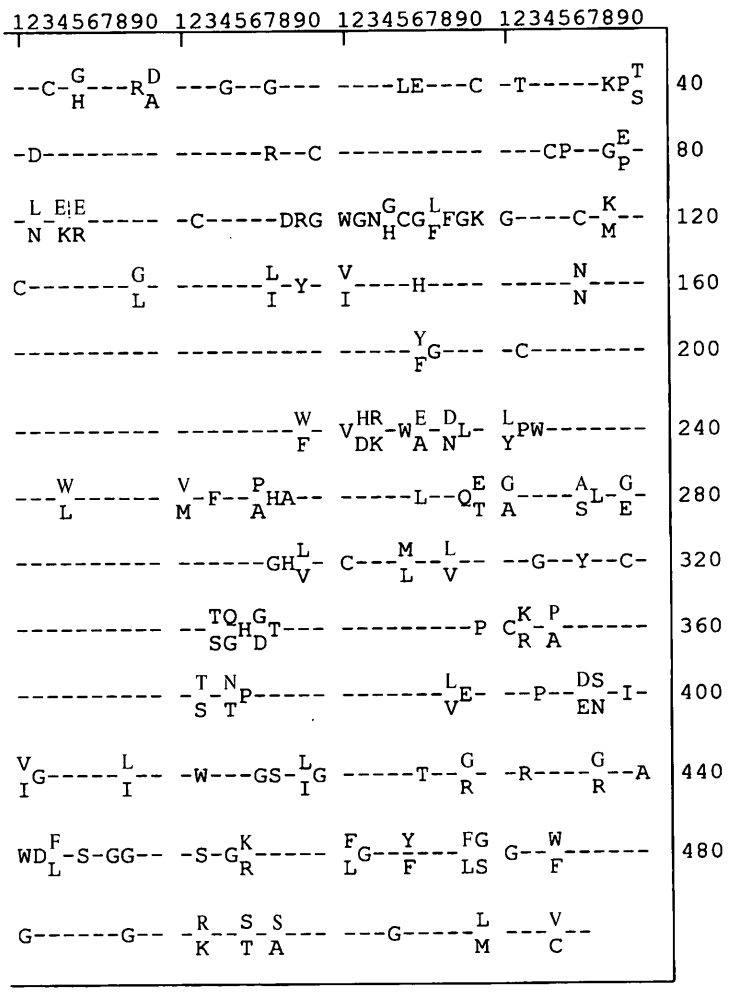

Fig. 5. Conserved amino acids in the envelope glycoproteins of the 19 *Flaviviruses**. "Hyphen mark" indicates nonconserved position with more than 2 types of amino acid residues occurring at each of these positions. The results were obtained after multiple alignment of sequences of envelope glycoproteins. "Highlight" indicates amino acids which occur more than 80% in the alignment.

*Viruses considered: Dengue 1, Dengue 2, Dengue 3, Dengue 4, Japanese encephalitis, Kunjin, Murray Valley encephalitis, St. Louis encephalitis, Louping ill, Tick borne encephalitis, Langat, West Nile, Yellow fever, Negishi, Omsk hemorrhagic fever, Tyuleniy, Saumarez Reef, Kysanur forest disease and Powassan.

the confidence of assigning virus family using the oligopeptide sequence is high compared to identification of virus species using such an approach. The level of confidence in virus family assignment using family specific peptide goes higher when proteins from number of viruses belonging to a particular family. For example 116 protein sequences of 14 members of *Picornaviridae* family belonging to five genera were studied and the family specific peptide DGYXXQXXXXXDD was obtained. A virus polyprotein having this peptide has more than 95% chance to belong to *Picornaviridae*. However, such a high level of confidence can not be assigned to few other families as protein sequences for few species of those families are available in the sequence databank. As the sequence data on virus increases the above approach becomes powerful and accurate.

The method to identify a virus using the virus species specific peptide is an extension of above approach. In this study one extract a peptide that is present only in that virus and does not exit in any other protein sequence. Unique or virus specific peptides

obtained for members of the *Flaviviridae* family are given in table 5. It may be mentioned that one finds very high identity in the sequences of the Egp for the viruses, which are closely related. For example there is only 6% difference between the Egp of the viruses Tick borne encephalitis and Louping ill and thus it is difficult to pick-up virus specific oligopeptides.

Such oligopeptides could not be obtained for five out of the 19 viruses studied. Virus specific oligopeptides can be used as polymerase chain reaction probes to identify viruses as is described in earlier studies [11]. The procedure used here to identify signature peptides of a particular virus is sufficiently general in nature. False negative results can occur if one has chosen the peptide corresponding to hot spots or if only a few protein sequences from a particular virus family/genus have been selected to obtain unique signature. Available sequence data of strains indicate that the number of mutations in such short regions is smaller than the number of mismatches allowed. Such submolecular markers thus can be used in association with identification matrices to identify unknown virus.

References

1. Mayo MA, Pringle CR. Virus taxonomy — 1997. J Gen Virol 1998;79:649-57.
2. Koonin EV, Dolja VV. Evolution and taxonomy of positive strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit Rev Biochem Mol Biol 1993;28: 375-430.
3. Kolaskar AS, Naik PS. Computerization of virus data and its usefulness in virus classification. Intervirology 1992;34:133-41.
4. Etzold T, Ulyanov A, Argos P. SRS: Information retrieval system for molecular biology data banks. Methods Enzymol 1996;266:114-28.
5. Barker WC, Garavelli JS, Haft DH, Hunt LT, Marzec CR, Orcutt BC, Srinivasarao GY, Yeh LSL, Ledley RS, Mewes HW, Pfeiffer F, Tsugita A. The PIR international protein sequence database. Nucleic Acids Res 1998;26:27-32.
6. Kolaskar AS, Naik PS. Concerted use of multiple database for taxonomic insights. In: Dubois JE, Ghershon N, eds. The Information Revolution: Impact of Science and Technology. Berlin: Springer 1996:236-70.
7. Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD. Virus taxonomy. Classification and nomenclature of viruses, sixth Report of the International Committee on Taxonomy of Viruses. Arch Virol 1995;(Suppl 10):S1-586.
8. Kolaskar AS, Naik PS. Computer-aided virus identification on the world wide web. Arch Virol 1998;143:1513-21.
9. Thompson J D, Higins DG, Gibson TJ. CLUSTALW: improving the sensitivity of progressive multiple alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673-80.
10. Date S, Kulkarni R, Kulkarni B, Kulkarni-Kale U, Kolaskar AS. Multiple alignment of sequences on parallel computers. Comput Appl Biosci 1993;9:397-402.
11. Leary TP, Muerhoff AS, Simons JN, Pilot-Matias TJ, Erker JC, Chalmers ML, Schlauder GG, Dawson GJ, Desai SM, Mushahwar IK. Consensus oligonucleotide primers for the detection of GB virus C in human cryptogenic hepatitis. J Virol Methods 1996;56:119-21.