JVI Journals.ASM.org

Unraveling the Web of Viroinformatics: Computational Tools and Databases in Virus Research

Deepak Sharma, Pragya Priyadarshini, Sudhanshu Vrati

Vaccine and Infectious Disease Research Centre, Translational Health Science and Technology Institute, Gurgaon, India

The beginning of the second century of research in the field of virology (the first virus was discovered in 1898) was marked by its amalgamation with bioinformatics, resulting in the birth of a new domain—viroinformatics. The availability of more than 100 Web servers and databases embracing all or specific viruses (for example, dengue virus, influenza virus, hepatitis virus, human immunodeficiency virus [HIV], hemorrhagic fever virus [HFV], human papillomavirus [HPV], West Nile virus, etc.) as well as distinct applications (comparative/diversity analysis, viral recombination, small interfering RNA [siRNA]/short hairpin RNA [shRNA]/microRNA [miRNA] studies, RNA folding, protein-protein interaction, structural analysis, and phylotyping and genotyping) will definitely aid the development of effective drugs and vaccines. However, information about their access and utility is not available at any single source or on any single platform. Therefore, a compendium of various computational tools and resources dedicated specifically to virology is presented in this article.

Viruses, known to have infected humankind since approximately the 15th century BC, are found in all cellular forms of life, from bacteria to chordates. Pathogenic human viruses are causative agents of many morbid diseases such as influenza, AIDS, dengue fever, encephalitis, hepatitis, diarrhea, and severe acute respiratory syndrome (SARS). However, despite the availability of effective vaccines and treatments for several diseases, ~6 million deaths occur every year due to viruses. Thus, it is imperative to develop remedies against these viral invaders. Advances in molecular biology and bioinformatics have resulted in a deluge of genomic and experimental data. To store, examine, and disseminate all this information, 112 viroinformatics resources have been developed (as of May 2014; Table 1). The International Committee on Taxonomy of Viruses (ICTV) that performs the task of naming and classifying viruses lists 2,618 species. Strikingly, the genomes of ~40,000 viral strains, across more than 900 species, have been sequenced (NCBI Viral Genome Resource, NCBI Influenza Virus Resource [NCBI-IVR], and http://www.viprbrc.org/).

The aims of this article are manyfold. The first aim is to provide an extensive list of viroinformatics resources. At present, the information about these resources is so disparate that it is almost impossible to be aware of all the tools (many of which have been developed recently). The second aim is to categorize the resources based on their specificity for a particular virus, application, or task. The third aim is to compare tools performing similar tasks on the basis of (i) importance, popularity, and reliability (reflected by the citation index of a tool), (ii) the number of genomes and sequences included, (iii) the unique features of a given resource, and (iv) ease of use (whether a Web interface is available). If available, references to publications comparing similar tools have also been provided. The fourth aim is to provide information about tools that have been succeeded by a newer resource, have not been updated for more than past 2 years, or are currently inaccessible. The fifth (and most important) aim is to enable virologists to get an overview of tools capable of carrying out a desired task and select the most suitable one(s). For compiling the comprehensive index of resources presented here, an initial list was first prepared by extensively searching the literature (PubMed [http://www.ncbi.nlm.nih.gov/pubmed] and ScienceDirect [http://www.sciencedirect.com]) using all possible combinations of various keywords. Subsequently, all the publications describing these tools were thoroughly explored to find other resources mentioned in them. This greatly expanded the initial list and highlights the fact that it is almost impossible to learn about several resources by directly searching through the use of keywords. The inventory of databases provided by *Nucleic Acids Research* (http://www.oxfordjournals.org/nar/database/a/) was also examined for those dedicated specifically to virology.

VIRUS-CENTERED RESOURCES AND THE VIRUS PATHOGEN DATABASE AND ANALYSIS RESOURCE (ViPR)

Influenza virus. Influenza virus, one of the major human health public threats, was brought to the forefront of the world's attention due to the recent emergence of 2009 pandemic influenza A (H1N1, swine flu) virus. To date, 11 distinct Web portals and tools have been developed exclusively for influenza virus (Table 2). The first one to be developed was the Influenza Virus Database (IVDB), and the NCBI Influenza Virus Resource (NCBI-IVR) and Influenza Research Database (IRD) were subsequently concurrently established (1–3). Although all three websites provide access to sequence information as well as on-site access to various generic tools (BLAST, multiple-sequence alignment, and phylogenetic tree construction), each has a distinct focus. IVDB is exclusively armed with (i) the Sequence Distribution Tool, to display the worldwide geographic distribution of the chosen viral genotypes and to couple genomic data with epidemiological data, and (ii) the Quality Filter System, to classify the nucleotide se-

Accepted manuscript posted online 26 November 2014

Citation Sharma D, Priyadarshini P, Vrati S. 2015. Unraveling the web of viroinformatics: computational tools and databases in virus research. J Virol 89:1489–1501. doi:10.1128/JVI.02027-14.

Editor: S. P. Goff

Address correspondence to Deepak Sharma, deepak@thsti.res.in.

doi:10.1128/JVI.02027-14

TABLE 1 Databases, Web servers, and tools for virus research

Website	Key feature(s)	CI[a]	Access[b]	URL (reference)
Alvira[c]	For multiple-sequence alignment, especially of large no. of genomes (has graphical interface)	0	W	http://bioinfo.genopole-toulouse.prd.fr/Alvira (59)
ATIVS	Analysis tools for influenza virus surveillance	1.2	W	http://influenza.nhri.org.tw/ATIVS/ (11)
AVPdb	Resource of experimentally validated antiviral peptides targeting medically important viruses	4.8	W	http://crdd.osdd.net/servers/avpdb (111)
AVPpred	Antiviral peptide prediction algorithm	4.7	W	http://crdd.osdd.net/servers/avppred (112)
Base-By-Base	Software package to identify differences between genomes	4.7	D	http://athena.bioc.uvic.ca/virology-ca-tools/base-by-base/ (56)
bNAber	Database of broadly neutralizing HIV antibodies	4.8	W	http://bnaber.org (26)
CAPIH[d]	Comparative host–HIV-1 protein interaction network between humans and model animals	1.1	W	http://bioinfo-dbb.nhri.org.tw/capih (93)
CoVDB[d]	Repository of annotated coronavirus genes/genomes	3.7	W	http://covdb.microbiology.hku.hk (60)
DengueNet[d]	For global surveillance of dengue fever and dengue hemorrhagic fever	0.2	W	http://www.who.int/denguenet (31)
DPVweb	Information about viruses, viroids, and satellites of plants, fungi, and protozoa	3.1	W	http://www.dpvweb.net/ (49)
Dr.VIS	Catalogs human disease-related viral integration sites	2.6	W	http://bminfor.tongji.edu.cn/drvis (79)
EpiFlu	Comprehensive collection of influenza virus sequences containing associated clinical/epidemiological metadata	4.3	W	http://platform.gisaid.org (7)
euHCVdb	Resource of computer-annotated sequences, protein sequences/structures, and functional analysis tools (now succeeded by ViPR)	12.4[e]	W	http://euhcvdb.ibcp.fr (38)
EuResist	For predicting response to anti-HIV therapy	2.7	W, D	http://www.euresist.org (19)
FLAVIdB[d]	Web portal combining antigenic data of flaviviruses and analysis tools	5.0	W	http://cvc.dfci.harvard.edu/flavi/ (46)
Flavitrack[c]	Manually annotated database of flavivirus sequences	1.8	W	http://carnot.utmb.edu/flavitrack (45)
FluGenome	For genotyping influenza A virus and analyzing reassortment events	8.1	W	http://www.flugenome.org/ (8)
FluTE	Influenza epidemic simulation model	23.0[e]	D	http://www.cs.unm.edu/~dlchao/flute/ (14)
GATU	Annotation of viral genomes	5.6	W	http://athena.bioc.uvic.ca/virology-ca-tools/gatu/ (98)
Geno2pheno	For predicting drug resistance in HIV-1, HBV, and HCV	17.9[e]	W	http://www.geno2pheno.org/ (106)
GIB-V[c]	Platform for comparative analysis of viral genomes	1.4	W	http://gib-v.genes.nig.ac.jp/ (57)
GiRaF	Identification of influenza virus reassortments	3.6	D	http://www.cbcb.umd.edu/software/giraf/ (9)
HBVdb	Provides access to computer-annotated sequences as well as generic (BLAST, FASTA, ClustalW) and specialized tools for annotation, genotyping, and drug resistance profiling	2.1	W	http://hbvdb.ibcp.fr (33)
HBVPathDB[c]	Pathway information of HBV infection-related reactions	0.4	W	http://www.bio-inf.net/HBVPathDB/HBV/ (89)
HBVRegDB	Database for detection of regulatory elements	1.9	W	http://lancelot.otago.ac.nz (35)
HepSEQ	Hosts molecular, clinical, and epidemiological information as well as Sequence Matcher, Genotyper, and Polymerase Annotator tools for HBV	2.0	W	http://www.hepseq.org/ (32)
HERVd[d]	Repository of human endogenous retroviruses	4.2	W	http://herv.img.cas.cz (54)
HESAS[d]	Database to understand the role of HERVs in the human genome	1.4	W	http://www.primate.or.kr/HESAS (53)
HIPdb	Repository of experimentally validated HIV-inhibiting peptides	4.8	W	http://crdd.osdd.net/servers/hipdb (110)
HIVCD	Tool for contamination screening in HIV sequence laboratory	1.0	D	http://sourceforge.net/projects/hivcd/ (25)
HIV Positive Selection Mutation DB[d]	Selection pressure maps of HIV protease and reverse transcriptase	2.4	W	http://fold.doe-mbi.ucla.edu/HIV/ (23)
HIVSIM	For comparing levels of effectiveness of novel HIV therapy regimens	0.4	D	https://sites.google.com/site/hivsimulator/ (21)
HIVsirDB[d]	Repository of HIV-inhibiting siRNAs	3.2	W	http://crdd.osdd.net/raghava/hivsir/ (82)
HIV Systems Biology	Houses Gene Overlapper, Replication Cycle site, GPS-Prot, and AIDSVu tools	0.9	W	http://hivsystemsbiology.org (28)
HPV-QUEST	Tool for genotyping HPV	0	W	http://www.ijbcb.org/HPV/ (102)
HTLV-1 Molecular Epidemiology DB	For HTLV-1 sequence management and data mining	1.3	W	http://htlv1db.bahia.fiocruz.br (29)
HVDB	Database of hepatitis A, B, C, D, and E virus sequences; also provides a phylogenetic analysis tool	5.3	W	http://s2as02.genes.nig.ac.jp (47)

(Continued on following page)

TABLE 1 (Continued)

Website	Key feature(s)	CI[a]	Access[b]	URL (reference)
ICTV	Taxonomic classification and nomenclature of viruses		W	http://ictvonline.org
IPDR	Designing primers/probes for influenza diagnostics	1.3	W	http://www.ipdr.mcw.edu (13)
IRD	Resource of influenza virus host-pathogen interactions as well as BLAST, alignment, and phylogenetic tools	22.6[e]	W	http://www.fludb.org (6)
ISED	Catalogs influenza virus sequence and epitope information of viruses from Asia	1.7	W	http://influenza.cdc.go.kr (10)
IVDB[d]	Hosts influenza virus sequence data and Q-filter system as well as BLAST, alignment, phylogenetic, and sequence distribution tools	5.6	W	http://influenza.big.ac.cn (1)
jpHMM	For analysis of recombinations in viruses	7.7	W, D	http://jphmm.gobics.de (70)
LANL HCV Database	Contains sequence data and immunological epitopes in HCV (now succeeded by ViPR)	24.3[e]	W	http://hcv.lanl.gov (36, 37)
LANL HFV Database	Houses annotated HFV sequences and several analysis tools	0.8	W	http://hfv.lanl.gov (44)
LANL HIV Database	Comprehensive resource on HIV sequences, immunological epitopes, drug resistance, and vaccine trials	7.0	W	http://hiv.lanl.gov (15, 16)
LearnCoil-VMF[c]	Identifying coiled-coil-like motifs in viral membrane-fusion proteins	10.3[e]	W	http://web.wi.mit.edu/kim (115)
Metavir	Web server for analysis of viral metagenomes (viromes)	10.1[e]	W	http://metavir-meb.univ-bpclermont.fr (66)
NCBI Genotyping Tool	Resource for genotyping of viral sequences	14.4[e]	W	http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi (95)
NCBI-HHPID	Catalogs all interactions between HIV-1 and human proteins	28.4[e]	W	http://www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions (92)
NCBI-IVR	Provides access to influenza virus sequence data as well as BLAST, alignment, phylogenetic, and genome annotation tools	78.5[e]	W	http://www.ncbi.nlm.nih.gov/genomes/FLU/ (2)
NCBI Viral Genome Resource	Contains sequences for ~3,600 viral genomes		W	http://www.ncbi.nlm.nih.gov/genomes/VIRUSES/viruses.html
NCBI-VVR	Virus-specific databases that provide Web retrieval interfaces and analysis and visualization tools (currently for dengue and West Nile viruses)	3.4	W	http://www.ncbi.nlm.nih.gov/genomes/VirusVariation/ (30)
OpenFluDB	Isolate-centered database for influenza virus (contains information about virus type, host, and collection date/place)	2.4	W	http://openflu.vital-it.ch (12)
Paparazzi	Perl script to reconstruct entire viral genome from virus-derived siRNAs	6.8	D	http://carla.saleh.free.fr/softwares.php (88)
PaVE	Hosts annotated papillomavirus genomes and protein sequences/structures	6.7	W	http://pave.niaid.nih.gov (43)
PBRC	Comprehensive *Poxviridae* informational and analytical resource (now merged into ViPR)	4.6	W	http://www.poxvirus.org (40)
PHACCS[c]	For assessment of the diversity of uncultured/environmental viral communities	13.1[e]	W	http://phage.sdsu.edu/phaccs (65)
PhEVER[d]	Houses evolutionary and phylogenetic information for analysis of the nature of virus-virus and virus-host lateral gene transfers	0.3	W	http://pbil.univ-lyon1.fr/databases/phever (94)
phiSITE	Database of gene regulation in bacteriophages	0.7	W	http://www.phisite.org/ (52)
PhyloType	For searching for viral phylotypes	2.4	W, D	http://www.phylotype.org (105)
PIRSpred	For HIV-1 protein-inhibitor resistance/susceptibility prediction	1.3	W	http://protinfo.compbio.washington.edu/pirspred/ (20)
PrimerHunter	Primer-designing tool for virus subtyping	2.6	W, D	http://dna.engr.uconn.edu/software/PrimerHunter/ (104)
PriSM	Tool to select optimal primers for amplification of viral genomes	0.9	W	http://www.broadinstitute.org/perl/seq/specialprojects/primerDesign.cgi (64)
RetroTector	Detection of retroviral sequences in vertebrate genomes	7.9	W	http://retrotector.neuro.uu.se/ (55)
RNA Virus Database[d]	Devoted to RNA viruses	1.5	W	http://tree.bio.ed.ac.uk/rnavirusdb (50)
RotaC	Genotyping tool for group A rotaviruses	19.5[e]	W	http://rotac.regatools.be (107)
SCUEAL	Prediction of HIV-1 subtypes	0.2	W	http://www.datamonkey.org/dataupload_scueal.php (22)
SeLOX	For identifying degenerate lox-like sites in genomic sequences	1.6	W	http://selox.mpi-cbg.de/cgi-bin/selox/index (73)
SeqHepB	Genome sequence analysis program and relational database for HBV	7.4		http://www.seqhepb.com (34)

(Continued on following page)

TABLE 1 (Continued)

Website	Key feature(s)	CI[a]	Access[b]	URL (reference)
SeqMap 2.0	Identifying integration sites from LM-PCR, LAM-PCR, and nrLAM-PCR analysis	3.0	W	http://seqmap.compbio.iupui.edu/ (75)
siVirus	Antiviral siRNA design software	3.4	W	http://sivirus.rnai.jp/ (84)
SQUAT	Examining the quality of HIV-1 PR/RT sequences	3.9	D	http://www.stat.brown.edu/CFAR/SQUAT (24)
SSE	Nucleotide and amino acid sequence analysis platform	13.3[e]	D	www.virus-evolution.org/Downloads/Software (62)
Stanford HIV Drug Resistance DB	Contains genotype-treatment, genotype-phenotype, and genotype-outcome correlation data	14.6[e]	W	http://hivdb.stanford.edu (18)
STAR	Subtyping tool for HIV-1 and HBV	4.9	W	http://www.vgb.ucl.ac.uk/starn.shtml (101)
Subviral RNA Database	For research/analysis of viroids, satellite RNAs, satellite viruses, and human hepatitis *delta* virus	6.8	W	http://subviral.med.uottawa.ca (51)
VaZyMolO	Tool for defining/classifying viral protein modules	3.7	W	http://www.vazymolo.org (114)
VBRC	Viral resource of informational and analytical tools (now merged into ViPR)		W	http://www.vbrc.org/
vFitness	For estimations of HIV-1 fitness (replication rate)	1.5	W	http://bis.urmc.rochester.edu/vFitness/ (27)
VGDB[c]	Storing/analyzing genes and proteins from complete genomes (now succeeded by ViPR)	1.9	W	http://athena.bioc.uvic.ca/genomes/index.html (41)
VGO	Annotation of complete viral genomes, especially those of large poxviruses	2.4	W	http://athena.bioc.uvic.ca/virology-ca-tools/vgo/ (97)
VIDA[d]	Organizes ORFs from animal virus genomes	5.9	W	http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA.html (48)
VIGOR	Gene prediction program	3.1	W	http://jcvi.org/vigor (96)
VIPERdb	Database for icosahedral virus capsid structures	23.8[e]	W	http://viperdb.scripps.edu (116)
ViPR	Integrated repository of data (sequences, protein structures, epitopes, clinical/surveillance metadata) and analysis tools (BLAST, alignment, phylogeny) for multiple virus families	7.6	W	http://www.viprbrc.org (39)
VIPR HMM	For detecting recombinant and nonrecombinant viruses hybridized to diagnostic microarray	0	D	http://ibridgenetwork.org/wustl/vipr (72)
VIPS	Prediction of viral IRESs	2.4	W	http://140.135.61.250/vips/ (113)
ViralFusionSeq	For discovery of integration events by combining soft-clipping information, read-pair analysis, and targeted *de novo* assembly	4.8	D	http://www.hkbic.cuhk.edu.hk/software/viralfusionseq (76)
ViralORFeome[d]	Database to generate a versatile collection of viral ORFs and design ORF-specific primers	6.9	W	http://www.viralorfeome.com (63)
ViralZone	Fact sheets on all known virus families/genera with sequence data	6.0	W	http://www.expasy.org/viralzone/ (61)
VIRAPOPS	Forward simulator to model specific RNA virus functions	0	D	http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html (74)
ViReMa	Algorithm for detection of recombination junctions in viral genomes	2.0	D	http://sourceforge.net/projects/virema (71)
VirGen	Organizes the "sequence space" of viral genomes (now succeeded by ViPR)	1.5	W	http://117.239.43.117/virgen/virgen.asp (42)
VirHostNet[d]	Resource of virus-host interaction networks coupled to their functional annotation	11.3[e]	W	http://pbildb1.univ-lyon1.fr/virhostnet (91)
Vir-Mir db	Catalogs potential viral miRNA hairpins and allows prediction of viral miRNA target genes	5.6	W	http://alk.ibms.sinica.edu.tw (81)
ViroBLAST	Extends utility of BLAST to query against multiple databases and user-customized datasets	6.7	W, D	http://indra.mullins.microbiol.washington.edu/viroblast/viroblast.php (58)
VirOligo[d]	Repository of virus-specific oligonucleotides	1.7	W	http://viroligo.okstate.edu/ (108)
VIROME	Pipeline for classification of viral metagenomic sequences	6.9	W	http://virome.dbi.udel.edu (68)
viRome	R package for analysis of viral small RNA sequence datasets	0	D	http://virome.sf.net (87)
VIRsiRNAdb	Curated database of siRNA/shRNA targeting 42 important human viruses	3.2	W	http://crdd.osdd.net/servers/virsirnadb (83)
VIRsiRNApred	Viral siRNA efficacy prediction tool	0	W	http://crdd.osdd.net/servers/virsirnapred/ (85)
VirusFinder	Software for detecting integration sites by blending SVdetect with CREST	4.4	D	http://bioinfo.mc.vanderbilt.edu/VirusFinder/ (78)
Virus-Genotyping Tools	High-throughput genotyping of recombinant and nonrecombinant viruses	8.3	W	http://bioafrica.mrc.ac.za/rega-genotype/html/ (99)
VirusHunter	Pipeline for detection of novel/known viruses from various specimens	8.0	D	http://www.ibridgenetwork.org/wustl/virushunter (69)

(Continued on following page)

TABLE 1 (Continued)

Website	Key feature(s)	CI[a]	Access[b]	URL (reference)
VirusMINT	Hosts interactions between viral and human proteins	16.5[e]	W	http://mint.bio.uniroma2.it/virusmint/ (90)
Virus-PLoc	Prediction of subcellular location of viral proteins	14.6[e]	W	http://www.csbio.sjtu.edu.cn/bioinf/virus (109)
VirusSeq	For finding integration events using discordant read-pair information	6.4	D	http://odin.mdacc.tmc.edu/~xsu1/VirusSeq.html (77)
Visitor[c]	Perl script for analysis of viral siRNA datasets from Illumina sequencing platform	0.4	D	http://drosophile.org/GEDlab/?page_id = 254 (86)
ViTa[d]	Houses known host/viral miRNAs and known/putative host miRNA targets	5.9	W	http://vita.mbc.nctu.edu.tw/ (80)
VMGAP	Pipeline for functional annotation of metagenomic data	3.5		Several open source programs and public databases (67)
ZCURVE_V	Gene-finding tool for viral and phage genomes	4.2	W, D	http://tubic.tju.edu.cn/Zcurve_V/ (100)

[a] CI, citation index (number of citations per year).
[b] W, Web based; D, downloadable; W, D, both Web based and downloadable.
[c] Currently inaccessible.
[d] Has not been updated for more than 2 years.
[e] CI value > 10.

quences into the 7 categories of C1 to C4 and P1 to P3, respectively, according to their sequence content (coding sequence [CDS], 5′-untranscribed region [5′-UTR], and 3′-UTR) and integrity (complete [C] or partial [P]). NCBI-IVR (obviously one of the resources most often cited and used) is equipped with a genome annotation tool, FLAN (FLu ANnotation), for user-provided sequences and the ability to visualize large phylogenetic trees in an aggregated form with a special representation of subscale details (4, 5). IRD is a comprehensive platform (with data quantity almost equal to that of NCBI-IVR) that integrates genomic, proteomic, immune epitope, and surveillance information (6). Additionally, it emphasizes influenza virus host-pathogen interactions to gain insights into the nature of virulence and host range and the effect of sequence variation on these phenomena. Like NCBI-IVR and IRD, another extensive repository is the EpiFlu database hosted by the Global Initiative on Sharing Avian Influenza Data (GISAID) consortium (7). On the other hand, FluGenome is the only available tool for genotyping influenza A virus and, in addition, facilitates identification of reassortment events between divergent lines (8). A distinct downloadable program, GiRaF (Graph-incompatibility-based Reassortment Finder), designed specifically for the detection of reassortments in influenza viruses, is also available (9). In order to understand and monitor the evolution and migration of strains, an Influenza Sequence and Epitope Database (ISED) was established that catalogs sequence and epitope information of viruses from Asian countries (10). Another useful Web server is ATIVS (Analytical Tool for Influenza Virus Surveillance), which analyzes serological data of all influenza subtypes and hemagglutinin sequence data of influenza A/H3N2 viruses so as to generate antigenic maps for influenza surveillance and vaccine strain selection (11). Recently, OpenFluDB (an isolate-centered inventory) was developed; it contains general information about an isolate, including virus type, host, and collection date and geographical location as well as computationally predicted antiviral resistance, enhanced pathogenicity, or human adaptation propensity (12). The Influenza Primer Design Resource (IPDR) is designed to aid the development of primers and probes that can be used in diagnostic assays (13). To help prepare for future influenza seasonal epidemics or pandemics, a stochastic model, FluTE, capable of simulating the spread of influenza in the United States has also been constructed (14).

HIV and human T-cell lymphotropic virus (HTLV). In agreement with the fact that tremendous efforts are going on worldwide to control and treat AIDS, numerous resources that target human immunodeficiency virus (HIV) have been developed (Table 2). The LANL HIV database provides access to data on genetic sequences, immunological epitopes, drug resistance-associated mutations, and vaccine trials and to various useful tools (Geography Search, which retrieves sequences based on geographic distribution; Sequence Locator, which shows features and positions of nucleotide and protein sequences in the genome; HIValign, an alignment program; Gap Squeeze, which removes columns with more than a given percentage of gaps; N-Glycosite, which identifies potential N-linked glycosylation sites; Highlighter, which highlights matches, mismatches, and transition and transversion mutations; Entropy, which quantifies positional variation in an alignment; FindModel, which finds which evolutionary model best fits user sequences; PhyloPlace, which reports phylogenetic relatedness with reference sequences; Protein Feature Accent, which maps a desired sequence on the three-dimensional [3D] structure of a protein; and Recombinant Identification Program, which detects HIV-1 subtypes and recombination) (15–17). A distinct but partially overlapping platform, the Stanford HIV Drug Resistance Database, which contains genotype-treatment, genotype-phenotype, and genotype-outcome correlation data, is also available (18). In addition, it not only hosts the HIVdb tool for subtype and drug resistance predictions but also hosts the HIValg tool for comparison of three drug resistance algorithms (HIVdb, ANRS, and REGA). Likewise, the EuResist Network provides access to (i) a database of HIV genotypes and clinical responses to antiretroviral therapy, (ii) a locally installable tool to store clinical data of patients and to predict the activity of antiretroviral drugs, and (iii) a Web tool to predict subtypes as well as the efficacy of antiretroviral drug combinations (19). The Web tool PIRSpred is also available to predict drug resistance from HIV-1 genotypes (20). Additionally, HIV Therapy Simulator (HIVSIM) is a software application that enables users to assess the effectiveness of different therapeutic strategies for treatment of HIV infection (21). The Web server SCUEAL (Subtype Classification Using Evo-

TABLE 2 List and comparison of virus-centered resources

Species	Resource[a]	No. of genomes	No. of sequences	Specific feature[b]	CI[c]	Access[d]
Dengue virus	NCBI-VVR	3,204	13,125	Comprehensive sequence database with analysis tools	3.4	W
	DengueNet			For global surveillance of DF/DHF	0.2	W
HBV	SeqHepB		3,391	For determination of genotype and resistance-associated mutations	7.4	
	HBVdb	~4,000	46,667	Hosts all known sequences and generic/specialized tools	2.1	W
	HepSEQ	40	4,296	Focuses on HBV variants from patients	2.0	W
	HBVRegDB			For detection of regulatory elements	1.9	W
HCV	LANL HCV database	1,656[e]	155,260	Contains sequence data and immunological epitopes	24.3	W
	euHCVdb	62	96,645	Resource of sequences, protein structures, analysis tools	12.4	W
HFV	LANL HFV database	~6,000	63,773	Houses annotated sequences and several analysis tools	0.8	W
HIV	Stanford HIV drug resistance DB		157,681	Drug resistance mutations/predictions in HIV-1	14.6	W
	LANL HIV database	4,631	584,970	Hosts sequences, epitopes, mutations, clinical trials, tools	7.0	W
	bNAber			Database of broadly neutralizing antibodies	4.8	W
	SQUAT			Evaluating quality of HIV-1 PR/RT sequences	3.9	D
	EuResist		140,104	For predicting response to therapy and subtyping	2.7	W, D
	HIV positive selection mutation DB		50,634	Provides selection pressure maps of PR/RT	2.4	W
	vFitness			Assessment of HIV-1 replication rate	1.5	W
	PIRSpred			Prediction of HIV-1 drug resistance	1.3	W
	HIVCD			Tool for contamination detection in HIV laboratory	1.0	D
	HIV Systems Biology			Protein interactions, host-gene data, and epidemic maps	0.9	W
	HIVSIM			Assesses effectiveness of novel therapy regimens	0.4	D
	SCUEAL			For subtyping HIV-1	0.2	W
HTLV-1	HTLV-1 molecular epidemiology DB	16	2,571	Stores annotated proviral sequences	1.3	W
Influenza virus	NCBI-IVR	17,784	297,660	Houses sequence data and analysis tools	78.5	W
	FluTE			Influenza epidemic simulation model	23.0	D
	IRD	17,988	306,164	Host-pathogen interactions and analysis tools	22.6	W
	FluGenome	7,135	118,810	For genotyping influenza A virus	8.1	W
	IVDB	~3,000	43,875	Catalogs sequences and geographic distribution of genotypes	5.6	W
	EpiFlu	16,144	367,672	Sequences and associated clinical/epidemiological data	4.3	W
	GiRaF			For detection of reassortments	3.6	D
	OpenFluDB	11,443	232,688	Isolate-centered inventory	2.4	W
	ISED		181,552	Catalogs influenza virus sequences/epitopes from Asia	1.7	W
	IPDR		305,990	Primer designing tool for diagnostics	1.3	W
	ATIVS			Tool for influenza virus surveillance	1.2	W
Papillomavirus	PaVE	274	2,994	Hosts annotated genomes and proteins sequences/structures	6.7	W
Poxvirus	PBRC (later merged into ViPR)	328	22,289		4.6	W
West Nile virus	NCBI-VVR	627	2,438	Comprehensive sequence database with analysis tools	3.4	W

[a] For each virus, resources have been listed in order of decreasing citation index values (numbers of citations per year).
[b] DF, dengue fever; DHF, dengue hemorrhagic fever; PR, protease; RT, reverse transcriptase.
[c] CI, citation index.
[d] W, Web based; D, downloadable; W, D, both Web based and downloadable.
[e] Many genomes classified as "complete genome" have been listed as "partial genome" by euHCVdb.

lutionary ALgorithms) has been specifically established for subtyping HIV-1 (the source code is also available) (22). The HIV Positive Selection Mutation Database provides detailed selection pressure maps of HIV protease (PR) and reverse transcriptase (RT) (23). SQUAT (Sequence Quality Analysis Tool), a standalone tool that runs in the R statistical environment, was created to evaluate the quality of protease and reverse transcriptase sequences (24). In contrast, the open-source tool HIV Contamination Detection (HIVCD) was developed for identifying potential laboratory errors (25). The bNAber (broadly Neutralizing Antibodies electronic resource) database catalogs data on broadly neutralizing HIV antibodies, including sequences, structures, and

neutralization 50% inhibitory concentration (IC_{50}) values (26). The accurate assessment of HIV replication rates (or fitness) is important for analysis of drug resistance. vFitness is a Web tool based on statistical methods (the least-squares approach and measurement error models) to estimate HIV-1 fitness (27). Recently, the HIV Systems Biology website was developed; it houses (i) Gene Overlapper, which collects data from genome-wide surveys of host-cell genes linked to HIV infection and allows user-configured exploration of overlaps among studies, (ii) the Replication Cycle site, which provides an interactive account of the HIV replication cycle, including downloadable movies of structures and a link to the human genome annotated with HIV integration sites, (iii) GPS-Prot, a browser for visualization of human-HIV protein interactions, and (iv) AIDSVu, which hosts interactive online maps and allows visitors to visually explore the HIV epidemic in the United States (28).

Human T-cell lymphotropic virus type 1 (HTLV-1) is the causal agent of adult T-cell leukemia and inflammatory diseases, including HTLV-1-associated myelopathy/tropical spastic paraparesis, uveitis, and infective dermatitis. The HTLV-1 Molecular Epidemiology Database stores annotated proviral sequences from clinical, epidemiological, and geographical studies (29).

Dengue and West Nile viruses. The NCBI Virus Variation Resources (NCBI-VVR) are a set of virus-specific databases that integrate sequence information with relevant metadata (sample collection time and location, disease severity, and serotype) as well as generic tools (multiple-sequence alignment and phylogenetic tree construction) (30). Currently, the NCBI-VVR covers dengue and West Nile viruses (Table 2). NCBI-IVR is also now a part of NCBI-VVR. In addition, DengueNet is a data management system created by the World Health Organization (WHO) for global epidemiological and virological surveillance of dengue fever and dengue hemorrhagic fever (31).

Hepatitis viruses, papillomavirus, and ViPR. The most common causes of viral hepatitis are the five unrelated hepatotropic viruses: hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV), and hepatitis E virus (HEV). The public health repository for HBV, HepSEQ, provides (i) information on epidemiological, virological, clinical, nucleotide sequence, and mutational aspects of HBV infection and (ii) access to the programs Sequence Matcher (searches the database for similar sequences), Genotyper (for genotyping strains), and Polymerase Annotator (annotates the input sequence for known mutations) (32). In a similar vein, HBVdb hosts (i) a collection of computer-annotated sequences based on reference genomes and (ii) integrated generic tools (BLAST, FASTA, and ClustalW) as well as specialized tools for annotation, genotyping, and drug resistance profiling (33). The HepSEQ database focuses on HBV variants from patients (4,296 sequences), while HBVdb catalogs all known HBV sequences (46,667 sequences). Another sequence analysis program (not freely available, and the user has to pay for its services) to determine HBV genotypes, identify key mutations associated with antiviral resistance, and identify clinically important mutants is SeqHepB (34). A knowledge database, HBVRegDB, for the detection, comparison, and visualization of regulatory elements in HBV sequences has also been established (35). The LANL HCV database contains sequence data as well as a curated inventory of immunological epitopes and their interaction with the immune system (36, 37). In comparison, the European HCV database (euHCVdb), an extension of the French HCV database, is a collection of computer-annotated sequences based on reference genomes (38). Moreover, euHCVdb is oriented toward HCV protein sequences, 3D structures, and functional analyses. However, both the LANL HCV database and euHCVdb have been succeeded by the Virus Pathogen Database and Analysis Resource (ViPR). Nevertheless, the tools hosted on these two HCV Web portals are still available and of use to virologists.

ViPR is an integrated comprehensive repository of data (sequences, gene and protein annotations, protein structures, epitopes, and clinical and surveillance metadata) and analysis tools (BLAST, multiple-sequence alignment, phylogenetic tree construction, and comparative analysis) (39). Currently, it contains information for several human-pathogenic viruses belonging to the *Arenaviridae*, *Bunyaviridae*, *Caliciviridae*, *Coronaviridae*, *Filoviridae*, *Flaviviridae*, *Hepeviridae*, *Herpesviridae*, *Paramyxoviridae*, *Picornaviridae*, *Poxviridae*, *Reoviridae*, *Rhabdoviridae*, and *Togaviridae* families. Due to the concern that variola virus (the causative organism of smallpox) and other related viruses might be used as biological weapons, the Poxvirus Bioinformatics Resource Center (PBRC) was developed (40). However, both PBRC and another useful Web server, VBRC (Viral Bioinformatics Resource Center), were subsequently merged into ViPR. Two of the initially established viral repositories, the Viral Genome Database [VGDB] (41) and VirGen (42), have also been succeeded by ViPR.

The *Papillomaviridae* are a diverse family of nonenveloped, double-stranded DNA viruses that infect several amniotes. The Papillomavirus Episteme (PaVE) provides access to data on computationally annotated genomes (for viral open reading frames [ORFs] and *cis* regulatory elements), genes, and protein sequences and structures (43).

VIRUS GROUP-SPECIFIC TOOLS

A few databases have been designed to target a group of viruses instead of a specific virus. Hemorrhagic fever viruses (HFVs) are a diverse set of over 80 viral species defined by their pathogenicity rather than by taxonomical classification. The LANL HFV database provides access to annotated sequences and several analysis tools (Geography Search, Sequence Locator, HFValign, Gapstreeze, N-Glycosite, Highlighter, FindModel, PhyloPlace, and Protein Feature Accent [functions of these tools are similar to those mentioned above for the LANL HIV database]) (44). Similarly, (i) Flavitrack is an annotated database of flavivirus sequences (45), (ii) FLAVIdB houses antigenic data of flaviviruses and analysis tools with applications in immunology and vaccinology (46), (iii) the hepatitis virus database (HVDB) catalogs all the HAV, HBV, HCV, HDV, and HEV sequences as well as providing a tool for their phylogenetic analysis (47), (iv) the VIDA virus database organizes open reading frames (ORFs) from partial or complete genome sequences of animal viruses (*Herpesviridae*, *Poxviridae*, *Papillomaviridae*, *Coronaviridae*, and *Arteriviridae*) (48), (v) DPVweb contains information about viruses, viroids, and satellites of plants, fungi, and protozoa (49), (vi) the RNA Virus Database hosts genomic organization and analytical tools for RNA viruses (50), (vii) the Subviral RNA database is designed to facilitate research and analysis of viroids, satellite RNAs, satellites viruses, the human hepatitis *delta* virus, and related RNA sequences (51), and (viii) the phiSITE database catalogs regulatory elements from bacteriophages (52). Three distinct resources that focus on endogenous retroviruses have also been developed. HERVd is a database of human endogenous retroviruses (HERVs) that provides com-

plex information on and analysis of retroviral elements found in the human genome, while HESAS (HERVs Expression and Structure Analysis System) was established to understand the effect of HERVs on the expression of human functional genes (53, 54). Both HERVd and HESAS depend on RepeatMasker, a program for identification of repetitive sequences. On the other hand, the RetroTector package was constructed for automated recognition of retroviral sequences in genomic data and is independent of RepeatMasker (55). The robustness of RetroTector is due to incorporation of several novel heuristic algorithms.

COMPARATIVE AND DIVERSITY ANALYSIS OF VIRAL SEQUENCES

With ever-increasing numbers of complete and incomplete viral genomic sequences, the development of tools for comparative analysis was inevitable. To this end, a software package (which has to be downloaded and installed), Base-By-Base, was established for single-nucleotide-level analysis of whole-viral-genome alignments (56). On the other hand, the Web-based platform Genome Information Broker for Viruses (GIB-V) is equipped with BLAST, ClustalW, and Keyword Search algorithms for carrying out comparison studies (57). A stand-alone BLAST Web server, ViroBLAST, for flexible queries against multiple databases and user sequence data sets has also been developed (and is available as a downloadable Web server too) (58). The Alvira software is a multiple-sequence alignment tool with a graphical interface that can be explored at different levels of resolution and has been devised specifically to address the problem of simultaneous analysis of a large number of viral strains (59). In addition, the CoVDB database was constructed exclusively for carrying out comparative analysis of coronavirus (CoV) genes and genomes (60).

Viruses are presumably among the most diverse and dynamic biological entities. To understand virus diversity, ViralZone provides information on viral molecular biology, taxonomy, hosts, epidemiology, and structures (virions) (61). On the other hand, the simple sequence editor (SSE) software package, an upgraded version of the Simmonics program, was developed to create an integrated environment where sequences can be aligned and annotated as well as analyzed by the use of a variety of diversity, phylogeny, and RNA structure algorithms (62). The high rate of evolution in viruses leads to numerous polymorphisms in their proteins, which in turn may affect viral virulence and/or host adaptation. ViralORFeome is a database used to manage all possible variants and mutants of viral ORFs and facilitates the designing of ORF-specific cloning primers (Table 3) (63). In comparison, PriSM is a Web tool to select and match degenerate primer pairs for the amplification of user-defined sets of viral genomes (64).

The recent development of next-generation technologies has enabled us to study the biodiversity and dynamics of viral populations isolated from the environment. The first tool developed to this end, PHACCS (Phage Communities from Contig Spectrum), estimates the structure and diversity of uncultured phages (viruses that infect prokaryotes) from shotgun sequence data (Table 3) (65). Subsequently, Metavir, Viral MetaGenome Annotation Pipeline (VMGAP), and Viral Informatics Resource for Metagenomic Exploration (VIROME) were established for the analysis of viral metagenomic sequences (66–68). Among these, both Metavir and VIROME are interactive Web-based resources. A distinct computational pipeline, VirusHunter, has also been built for identification of novel viruses from a wide array of specimen types (69).

VIRAL RECOMBINATION AND INTEGRATION-SPECIFIC RESOURCES

Viruses are known for their ability to mutate and rapidly adapt to new environments. Viral recombination plays a central role in the expansion of viral host range, increases in virulence, evasion of host immunity, evolution of antiviral resistance, and emergence of new viruses. Two distinct tools, jpHMM (jumping profile hidden Markov model) and ViReMa (Virus Recombination Mapper), have been developed for recombination detection in viral genomes (Table 3) (70, 71). While jpHMM is a Web server (source code is also available to run locally) for recombination prediction in HIV-1 and HBV, ViReMa is a command line downloadable program for analyzing next-generation viral sequencing data. On the other hand, VIPR HMM (Viral Identification with a PRobabilistic algorithm incorporating hidden Markov model) software is capable of identifying recombinant and nonrecombinant viruses hybridized to a microbial detection microarray whereas the SeLOX Web server provides an interface for searching degenerate locus of recombination (lox)-like sites within genomic sequences (72, 73). A forward simulator, VIRAPOPS, to predict variations in rapidly evolving RNA viral populations has also been developed (74). Viral integration into the human genome has been implicated in the development of malignant diseases, and several biomedical applications include identification of cancer genes and malignant transformation in gene therapy clinical trials. SeqMap is a tool for identifying viral integration sites (VIS) from ligation-mediated PCR (LM-PCR), linear-amplification-mediated PCR (LAM-PCR), and nonrestrictive LAM-PCR (nrLAM-PCR) analysis (75). Additionally, three more complementary tools, VirusSeq, ViralFusionSeq, and VirusFinder, each with a distinct algorithm, have been established for the detection of VIS (76–78). It would be prudent for a user to employ all the four tools to increase the accuracy of VIS predictions. A database of human disease-related VIS (Dr.VIS) has also been constructed that catalogs data pertaining to characteristics of the malignant disease, chromosome region, genomic position, and virus-host junction sequence (79).

SMALL-RNA ANALYSIS TOOLS

miRNAs. MicroRNAs (miRNAs) are key *trans*-acting factors that posttranscriptionally regulate both viral and host gene expression and thus play an important role in viral pathogenesis. ViTa catalogs known host and viral miRNAs as well as known or putative host miRNA targets on viruses (80). It utilizes miRanda and TargetScan to predict miRNA targets within virus genomes. In comparison, the Vir-Mir database houses data on potential viral miRNA hairpins and allows prediction of viral miRNA target genes (based on the use of the RNAhybrid program) in human, mouse, rat, zebrafish, rice, and Arabidopsis (81).

siRNAs. RNA interference (RNAi) has emerged as a powerful modality to inhibit viruses. To this end, curated repositories of experimentally validated small interfering RNA (siRNA) and short hairpin RNA (shRNA) targeting HIV (HIVsirDB) and 42 other important human viruses (VIRsiRNAdb) have been developed (Table 3) (82, 83). siVirus is Web-based antiviral siRNA design software for analysis of influenza virus, HIV-1, HCV, and SARS coronavirus (84). A viral siRNA efficacy prediction tool,

TABLE 3 Categorization and comparison of resources based on specific application

Application	Resource[a]	Specific feature(s)/organism(s) (if any)[b]	CI[c]	Web[d]
Metagenomics	PHACCS[e]		13.1	W
	Metavir		10.1	W
	VIROME		6.9	W
	VMGAP		3.5	
Primer designing	ViralORFeome	For designing ORF-specific primers	6.9	W
	PrimerHunter	For subtype identification	2.6	W, D
	VirOligo	Repository of virus-specific oligonucleotides	1.7	W
	IPDR	For influenza diagnostics	1.3	W
	PriSM	For amplification of viral genomes	0.9	W
Protein-protein interaction	NCBI-HHPID	Catalogs human–HIV-1 interactions	28.4	W
	VirusMINT		16.5	W
	VirHostNet		11.3	W
	CAPIH	Hosts human–animal–HIV-1 interactions	1.1	W
	HIV systems biology (GPS-Prot)	Resource for human–HIV interactions	0.9	W
siRNA analysis	Paparrazi	Reconstruct viral genome from siRNAs	6.8	D
	siVirus	For designing antiviral RNA (HCV, HIV-1, IV, SARS-CoV)	3.4	W
	VIRsiRNAdb	Database of siRNA/shRNA targeting 42 viruses	3.2	W
	HIVsirDB	Repository of HIV-inhibiting siRNAs	3.2	W
	Visitor	Analysis of viral siRNAs from Illumina platform	0.4	D
	viRome	R package for analysis of small RNA sequences	0	D
	VIRsiRNApred	Viral siRNA efficacy prediction tool	0	W
Structural analysis	VIPERdb	Database of icosahedral virus capsid structures	23.8	W
	LearnCoil-VMF[e]		10.3	W
Subtyping/genotyping	RotaC	Group A rotavirus	19.5	W
	Geno2pheno	HBV, HCV, HIV-1	17.9	W
	Stanford HIV drug resistance DB	HIV-1	14.6	W
	NCBI genotyping tool	HBV, HCV, HIV, HTLV, PV	14.4	W
	Virus genotyping tools	HBV, HCV, HHV8, HIV, HPV, HTLV	8.3	W
	FluGenome	Influenza A virus	8.1	W
	SeqHepB	HBV	7.4	
	STAR	HBV, HIV	4.9	W
	EuResist	HIV-1	2.7	W, D
	PrimerHunter	Primer design for subtype identification	2.6	W, D
	HBVdb	HBV	2.1	W
	HepSEQ	HBV	2.0	W
	SCUEAL	HIV-1	0.2	W
	HPV-QUEST	HPV	0	W
Viral integration	VirusSeq		6.4	D
	ViralFusionSeq		4.8	D
	VirusFinder		4.4	D
	SeqMap 2.0		3.0	W
	Dr.VIS	Database of human disease-related VIS	2.6	W
Viral recombination	jpHMM	HBV, HIV-1	7.7	W, D
	ViReMa		2.0	D
	SeLOX		1.6	W
	VIPR HMM		0	D
Miscellaneous[f]	Virus-PLoc	Prediction of subcellular localization of proteins	14.6	W
	SSE	Nucleotide/protein sequence analysis platform	13.3	D
	VirusHunter	Detection of novel/known viruses from specimens	8.0	D
	RetroTector	Identification of retroviral sequences in vertebrates	7.9	W
	ViPR	Repository of data and analysis tools for viruses	7.6	W
	Subviral RNA database	Toolbox for viroids, satellite RNAs, and HDV	6.8	W
	ViroBLAST	For BLAST searches against multiple databases and user's data set	6.7	W, D

(Continued on following page)

TABLE 3 (Continued)

Application	Resource[a]	Specific feature(s)/organism(s) (if any)[b]	CI[c]	Web[d]
	ViralZone	Fact sheets on all known virus families/genera	6.0	W
	ViTa	Houses host/viral miRNAs and known/putative targets	5.9	W
	VIDA	Organizes ORFs from animal virus genomes	5.9	W
	GATU	Annotation of viral genomes	5.6	W
	Vir-Mir db	Prediction of viral miRNA target genes	5.6	W
	HVDB	Database of hepatitis A, B, C, D, and E virus sequences	5.3	W

[a] For each application, resources have been listed in order of decreasing citation index values (numbers of citations per year).
[b] HDV, hepatitis *delta* virus; HHV, human herpesvirus; HTLV, human T-lymphotropic virus; HPV, human papillomavirus; IV, influenza virus; PV, poliovirus; SARS-CoV, SARS coronavirus; VIS, viral integration sites.
[c] CI, citation index.
[d] W, Web based; D, downloadable; W, D, both Web based and downloadable.
[e] Currently inaccessible.
[f] Resources with CI > 5 (the virus-centered tools with CI > 5 listed in Table 2 have been excluded).

VIRsiRNApred, based on a Support Vector Machine (SVM) is also available (85). Visitor and viRome are software packages developed specifically for the analysis of viral siRNA sequence data sets (86, 87). Paparrazi is a useful Perl script for reconstruction of entire viral genomes from viral siRNAs present in a sample (88).

VIRUS-HOST INTERACTION AND OTHER ECLECTIC SOFTWARES

Another major challenge is to understand the complex interplay between viral and host cellular proteins. An attempt to rationally initiate such analysis was made by constructing HBVPathDB, a database of HBV infection-related molecular interaction pathways and networks (89). Subsequently, two knowledge-based resources, VirusMINT and VirHostNet, were almost simultaneously built for the management and analysis of interactions between viral and human proteins (Table 3) (90, 91). Concurrently, the NCBI HIV-1/Human Protein Interaction Database (NCBI-HHPID) was created to exclusively catalog all interactions between HIV-1 and human proteins (92). A Web interface, CAPIH, for comparative analyses of host-HIV-1 protein-protein interactions in human and three model animals (chimpanzee, rhesus macaque, and mouse) has also been established (93). On the other hand, phEVER provides evolutionary and phylogenetic information for exploring virus-host and virus-virus lateral gene transfers (94).

With the availability of numerous viral genome sequences, annotation of genes is the next critical step for understanding the structure and function of genome components and for efficient vaccine development as well as for treatment of viral diseases. Therefore, several tools have been established for annotation and classification of viral sequences: NCBI Genotyping Tool (95), VIGOR (Viral Genome ORF Reader) (96), Viral Genome Organizer (VGO) (97), Genome Annotation Transfer Utility (GATU) (98), Virus-Genotyping Tools (99), ZCURVE_V (100), STAR (Subtype Analyser) (101), and HPV-QUEST (102). The performance of seven HIV-1 subtyping tools in terms of their reproducibility, sensitivity, and specificity was recently evaluated (103). A primer design tool, PrimerHunter, for PCR-based virus subtype identification has also been constructed (104). In addition, there are various other resources that have been developed to address specific issues: (i) PhyloType, which uses parsimony to reconstruct ancestral traits and to select phylotypes (subsets of taxa and strains that share a history) (105); (ii) Geno2pheno, for estimating drug resistance and subtype prediction in HIV-1, HBV, and HCV (106); (iii) RotaC, an automated classification tool for group A rotaviruses (107); (iv) VirOligo, a database of virus-specific oligonucleotides (PCR primers and hybridization probes) as well as of experimental conditions for their usage (108); (v) Virus-PLoc, for predicting the subcellular localization of viral proteins within host cells and virus-infected cells (109); (vi) HIPdb and AVPdb, resources for antiviral peptides targeting HIV and 60 other medicinally important viruses, respectively (110, 111); (vii) AVPpred, for prediction of highly effective antiviral peptides (112); (viii) VIPS, a viral internal ribosomal entry site (IRES) prediction system (113); and (ix) VaZyMolO, devoted to modular (pertaining to a structural or functional unit containing one or more protein domains) classification of viral proteins (114). Web-based tools to explore and analyze the structural aspects of viruses have also been established. The LearnCoil-VMF program detects coiled-coil-like regions in viral membrane-fusion proteins, while VIPERdb (VIrus Particle ExploreR database) is a Web interface for analysis of manually curated icosahedral virus capsid structures (115, 116).

CONCLUSION

This review will not only enable virologists to get a comprehensive overview of tools available for a desired virus or application but will also enable them to select the best tool(s) to carry out their task. Additionally, it will facilitate the development of new and better resources by highlighting the areas where no (or a few) resources exist.

ACKNOWLEDGMENT

This work was supported by the Department of Biotechnology (DBT), government of India (grant BT/MB/01/VIDRC/08 to S.V.).

REFERENCES

1. **Chang S, Zhang J, Liao X, Zhu X, Wang D, Zhu J, Feng T, Zhu B, Gao GF, Wang J, Yang H, Yu J.** 2007. Influenza Virus Database (IVDB): an integrated information resource and analysis platform for influenza virus research. Nucleic Acids Res **35:**D376–D380.
2. **Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D.** 2008. The influenza virus resource at the National Center for Biotechnology Information. J Virol **82:**596–601. http://dx.doi.org/10.1128/JVI.02005-07.
3. **Squires B, Macken C, Garcia-Sastre A, Godbole S, Noronha J, Hunt V, Chang R, Larsen CN, Klem E, Biersack K, Scheuermann RH.** 2008. BioHealthBase: informatics support in the elucidation of influenza virus host pathogen interactions and virulence. Nucleic Acids Res **36:**D497–D503. http://dx.doi.org/10.1093/nar/gkm905.
4. **Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Tatusova T.** 2007. FLAN:

a web server for influenza virus genome annotation. Nucleic Acids Res **35:**W280–W284. http://dx.doi.org/10.1093/nar/gkm354.
5. **Zaslavsky L, Bao Y, Tatusova TA.** 2008. Visualization of large influenza virus sequence datasets using adaptively aggregated trees with sampling-based subscale representation. BMC Bioinformatics **9:**237. http://dx.doi.org/10.1186/1471-2105-9-237.
6. **Squires RB, Noronha J, Hunt V, Garcia-Sastre A, Macken C, Baumgarth N, Suarez D, Pickett BE, Zhang Y, Larsen CN, Ramsey A, Zhou L, Zaremba S, Kumar S, Deitrich J, Klem E, Scheuermann RH.** 2012. Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. Influenza Other Respir Viruses **6:**404–416. http://dx.doi.org/10.1111/j.1750-2659.2011.00331.x.
7. **Bogner P, Capua I, Cox NJ, Lipman DJ.** 2006. A global initiative on sharing avian flu data. Nature **442:**981. http://dx.doi.org/10.1038/442981a.
8. **Lu G, Rowley T, Garten R, Donis RO.** 2007. FluGenome: a web tool for genotyping influenza A virus. Nucleic Acids Res **35:**W275–W279. http://dx.doi.org/10.1093/nar/gkm365.
9. **Nagarajan N, Kingsford C.** 2011. GiRaF: robust, computational identification of influenza reassortments via graph mining. Nucleic Acids Res **39:**e34. http://dx.doi.org/10.1093/nar/gkq1232.
10. **Yang IS, Lee JY, Lee JS, Mitchell WP, Oh HB, Kang C, Kim KH.** 2009. Influenza sequence and epitope database. Nucleic Acids Res **37:**D423–D430. http://dx.doi.org/10.1093/nar/gkn881.
11. **Liao YC, Ko CY, Tsai MH, Lee MS, Hsiung CA.** 2009. ATIVS: analytical tool for influenza virus surveillance. Nucleic Acids Res **37:**W643–W646. http://dx.doi.org/10.1093/nar/gkp321.
12. **Liechti R, Gleizes A, Kuznetsov D, Bougueleret L, Le Mercier P, Bairoch A, Xenarios I.** 2010. OpenFluDB, a database for human and animal influenza virus. Database (Oxford) **2010:**baq004. http://dx.doi.org/10.1093/database/baq004.
13. **Bose ME, Littrell JC, Patzer AD, Kraft AJ, Metallo JA, Fan J, Henrickson KJ.** 2008. The Influenza Primer Design Resource: a new tool for translating influenza sequence data into effective diagnostics. Influenza Other Respir Viruses **2:**23–31. http://dx.doi.org/10.1111/j.1750-2659.2007.00031.x.
14. **Chao DL, Halloran ME, Obenchain VJ, Longini IM, Jr.** 2010. FluTE, a publicly available stochastic influenza epidemic simulation model. PLoS Comput Biol **6:**e1000656. http://dx.doi.org/10.1371/journal.pcbi.1000656.
15. **Kuiken C, Yoon H, Abfalterer W, Gaschen B, Lo C, Korber B.** 2013. Viral genome analysis and knowledge management. Methods Mol Biol **939:**253–261. http://dx.doi.org/10.1007/978-1-62703-107-3_16.
16. **Kuiken C, Korber B, Shafer RW.** 2003. HIV sequence databases. AIDS Rev **5:**52–61.
17. **Shaw TI, Zhang M.** 2013. HIV N-linked glycosylation site analyzer and its further usage in anchored alignment. Nucleic Acids Res **41:**W454–W458. http://dx.doi.org/10.1093/nar/gkt472.
18. **Shafer RW.** 2006. Rationale and uses of a public HIV drug-resistance database. J Infect Dis **194**(Suppl 1):S51–S58. http://dx.doi.org/10.1086/505356.
19. **Zazzi M, Incardona F, Rosen-Zvi M, Prosperi M, Lengauer T, Altmann A, Sonnerborg A, Lavee T, Schulter E, Kaiser R.** 2012. Predicting response to antiretroviral treatment by machine learning: the EuResist project. Intervirology **55:**123–127. http://dx.doi.org/10.1159/000332008.
20. **Jenwitheesuk E, Wang K, Mittler JE, Samudrala R.** 2005. PIRSpred: a web server for reliable HIV-1 protein-inhibitor resistance/susceptibility prediction. Trends Microbiol **13:**150–151. http://dx.doi.org/10.1016/j.tim.2005.02.003.
21. **Lim HC, Curlin ME, Mittler JE.** 2011. HIV Therapy Simulator: a graphical user interface for comparing the effectiveness of novel therapy regimens. Bioinformatics **27:**3065–3066. http://dx.doi.org/10.1093/bioinformatics/btr515.
22. **Kosakovsky Pond SL, Posada D, Stawiski E, Chappey C, Poon AF, Hughes G, Fearnhill E, Gravenor MB, Leigh Brown AJ, Frost SD.** 2009. An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. PLoS Comput Biol **5:**e1000581. http://dx.doi.org/10.1371/journal.pcbi.1000581.
23. **Pan C, Kim J, Chen L, Wang Q, Lee C.** 2007. The HIV positive selection mutation database. Nucleic Acids Res **35:**D371–D375. http://dx.doi.org/10.1093/nar/gkl855.
24. **Delong AK, Wu M, Bennett D, Parkin N, Wu Z, Hogan JW, Kantor R.** 2012. Sequence quality analysis tool for HIV type 1 protease and reverse transcriptase. AIDS Res Hum Retroviruses **28:**894–901. http://dx.doi.org/10.1089/aid.2011.0120.
25. **Ebbert MT, Mallory MA, Wilson AR, Dooley SK, Hillyard DR.** 2013. Application of a new informatics tool for contamination screening in the HIV sequencing laboratory. J Clin Virol **57:**249–253. http://dx.doi.org/10.1016/j.jcv.2013.03.013.
26. **Eroshkin AM, Leblanc A, Weekes D, Post K, Li Z, Rajput A, Butera ST, Burton DR, Godzik A.** 2014. bNAber: database of broadly neutralizing HIV antibodies. Nucleic Acids Res **42:**D1133–D1139. http://dx.doi.org/10.1093/nar/gkt1083.
27. **Ma J, Dykes C, Wu T, Huang Y, Demeter L, Wu H.** 2010. vFitness: a web-based computing tool for improving estimation of *in vitro* HIV-1 fitness experiments. BMC Bioinformatics **11:**261. http://dx.doi.org/10.1186/1471-2105-11-261.
28. **Bushman FD, Barton S, Bailey A, Greig C, Malani N, Bandyopadhyay S, Young J, Chanda S, Krogan N.** 2013. Bringing it all together: big data and HIV research. AIDS **27:**835–838. http://dx.doi.org/10.1097/QAD.0b013e32835cb785.
29. **Araujo TH, Souza-Brito LI, Libin P, Deforche K, Edwards D, de Albuquerque-Junior AE, Vandamme AM, Galvao-Castro B, Alcantara LC.** 2012. A public HTLV-1 molecular epidemiology database for sequence management and data mining. PLoS One **7:**e42123. http://dx.doi.org/10.1371/journal.pone.0042123.
30. **Resch W, Zaslavsky L, Kiryutin B, Rozanov M, Bao Y, Tatusova TA.** 2009. Virus variation resources at the National Center for Biotechnology Information: dengue virus. BMC Microbiol **9:**65. http://dx.doi.org/10.1186/1471-2180-9-65.
31. **World Health Organization.** 2002. DengueNet–WHO's Internet-based system for the global surveillance of dengue fever and dengue haemorrhagic fever (dengue/DHF) http://www.who.int/denguenet. Dengue/DHF–global public health burden. Wkly Epidemiol Rec. **77:**300–304.
32. **Gnaneshan S, Ijaz S, Moran J, Ramsay M, Green J.** 2007. HepSEQ: international public health repository for hepatitis B. Nucleic Acids Res **35:**D367–D370. http://dx.doi.org/10.1093/nar/gkl874.
33. **Hayer J, Jadeau F, Deleage G, Kay A, Zoulim F, Combet C.** 2013. HBVdb: a knowledge database for hepatitis B virus. Nucleic Acids Res **41:**D566–D570. http://dx.doi.org/10.1093/nar/gks1022.
34. **Yuen LK, Ayres A, Littlejohn M, Colledge D, Edgely A, Maskill WJ, Locarnini SA, Bartholomeusz A.** 2007. SeqHepB: a sequence analysis program and relational database system for chronic hepatitis B. Antivir Res **75:**64–74. http://dx.doi.org/10.1016/j.antiviral.2006.11.014.
35. **Panjaworayan N, Roessner SK, Firth AE, Brown CM.** 2007. HBVRegDB: annotation, comparison, detection and visualization of regulatory elements in hepatitis B virus sequences. Virol J **4:**136. http://dx.doi.org/10.1186/1743-422X-4-136.
36. **Kuiken C, Yusim K, Boykin L, Richardson R.** 2005. The Los Alamos hepatitis C sequence database. Bioinformatics **21:**379–384. http://dx.doi.org/10.1093/bioinformatics/bth485.
37. **Yusim K, Richardson R, Tao N, Dalwani A, Agrawal A, Szinger J, Funkhouser R, Korber B, Kuiken C.** 2005. Los Alamos hepatitis C immunology database. Appl Bioinformatics **4:**217–225. http://dx.doi.org/10.2165/00822942-200504040-00002.
38. **Combet C, Garnier N, Charavay C, Grando D, Crisan D, Lopez J, Dehne-Garcia A, Geourjon C, Bettler E, Hulo C, Le Mercier P, Bartenschlager R, Diepolder H, Moradpour D, Pawlotsky JM, Rice CM, Trepo C, Penin F, Deleage G.** 2007. euHCVdb: the European hepatitis C virus database. Nucleic Acids Res **35:**D363–D366. http://dx.doi.org/10.1093/nar/gkl970.
39. **Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z, Zhou L, Larson CN, Dietrich J, Klem EB, Scheuermann RH.** 2012. ViPR: an open bioinformatics database and analysis resource for virology research. Nucleic Acids Res **40:**D593–D598. http://dx.doi.org/10.1093/nar/gkr859.
40. **Lefkowitz EJ, Upton C, Changayil SS, Buck C, Traktman P, Buller RM.** 2005. Poxvirus Bioinformatics Resource Center: a comprehensive *Poxviridae* informational and analytical resource. Nucleic Acids Res **33** (Database issue):D311–D316. http://dx.doi.org/10.1093/nar/gki110.
41. **Hiscock D, Upton C.** 2000. Viral Genome DataBase: storing and analyzing genes and proteins from complete viral genomes. Bioinformatics **16:**484–485. http://dx.doi.org/10.1093/bioinformatics/16.5.484.
42. **Kulkarni-Kale U, Bhosle S, Manjari GS, Kolaskar AS.** 2004. VirGen: a comprehensive viral genome resource. Nucleic Acids Res **32:**D289–D292. http://dx.doi.org/10.1093/nar/gkh098.

43. **Van Doorslaer K, Tan Q, Xirasagar S, Bandaru S, Gopalan V, Mohamoud Y, Huyen Y, McBride AA.** 2013. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. Nucleic Acids Res **41:**D571–D578. http://dx.doi.org/10.1093/nar/gks984.
44. **Kuiken C, Thurmond J, Dimitrijevic M, Yoon H.** 2012. The LANL hemorrhagic fever virus database, a new platform for analyzing biothreat viruses. Nucleic Acids Res **40:**D587–D592. http://dx.doi.org/10.1093/nar/gkr898.
45. **Misra M, Schein CH.** 2007. Flavitrack: an annotated database of flavivirus sequences. Bioinformatics **23:**2645–2647. http://dx.doi.org/10.1093/bioinformatics/btm383.
46. **Olsen LR, Zhang GL, Reinherz EL, Brusic V.** 2011. FLAVIdB: a data mining system for knowledge discovery in flaviviruses with direct applications in immunology and vaccinology. Immunome Res **30:**1–9.
47. **Shin IT, Tanaka Y, Tateno Y, Mizokami M.** 2008. Development and public release of a comprehensive hepatitis virus database. Hepatol Res **38:**234–243. http://dx.doi.org/10.1111/j.1872-034X.2007.00262.x.
48. **Albà MM, Lee D, Pearl FM, Shepherd AJ, Martin N, Orengo CA, Kellam P.** 2001. VIDA: a virus database system for the organization of animal virus genome open reading frames. Nucleic Acids Res **29:**133–136. http://dx.doi.org/10.1093/nar/29.1.133.
49. **Adams MJ, Antoniw JF.** 2006. DPVweb: a comprehensive database of plant and fungal virus genes and genomes. Nucleic Acids Res **34:**D382–D385. http://dx.doi.org/10.1093/nar/gkj023.
50. **Belshaw R, de Oliveira T, Markowitz S, Rambaut A.** 2009. The RNA virus database. Nucleic Acids Res **37:**D431–D435. http://dx.doi.org/10.1093/nar/gkn729.
51. **Rocheleau L, Pelchat M.** 2006. The Subviral RNA Database: a toolbox for viroids, the hepatitis *delta* virus and satellite RNAs research. BMC Microbiol **6:**24. http://dx.doi.org/10.1186/1471-2180-6-24.
52. **Klucar L, Stano M, Hajduk M.** 2010. phiSITE: database of gene regulation in bacteriophages. Nucleic Acids Res **38:**D366–D370. http://dx.doi.org/10.1093/nar/gkp911.
53. **Kim TH, Jeon YJ, Kim WY, Kim HS.** 2005. HESAS: HERVs expression and structure analysis system. Bioinformatics **21:**1699–1700. http://dx.doi.org/10.1093/bioinformatics/bti194.
54. **Paces J, Pavlicek A, Paces V.** 2002. HERVd: database of human endogenous retroviruses. Nucleic Acids Res **30:**205–206. http://dx.doi.org/10.1093/nar/30.1.205.
55. **Sperber GO, Airola T, Jern P, Blomberg J.** 2007. Automated recognition of retroviral sequences in genomic data - RetroTector. Nucleic Acids Res **35:**4964–4976. http://dx.doi.org/10.1093/nar/gkm515.
56. **Brodie R, Smith AJ, Roper RL, Tcherepanov V, Upton C.** 2004. Base-By-Base: single nucleotide-level analysis of whole viral genome alignments. BMC Bioinformatics **5:**96. http://dx.doi.org/10.1186/1471-2105-5-96.
57. **Hirahata M, Abe T, Tanaka N, Kuwana Y, Shigemoto Y, Miyazaki S, Suzuki Y, Sugawara H.** 2007. Genome Information Broker for Viruses (GIB-V): database for comparative analysis of virus genomes. Nucleic Acids Res **35:**D339–D342. http://dx.doi.org/10.1093/nar/gkl1004.
58. **Deng W, Nickle DC, Learn GH, Maust B, Mullins JI.** 2007. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics **23:**2334–2336. http://dx.doi.org/10.1093/bioinformatics/btm331.
59. **Enault F, Fremez R, Baranowski E, Faraut T.** 2007. Alvira: comparative genomics of viral strains. Bioinformatics **23:**2178–2179. http://dx.doi.org/10.1093/bioinformatics/btm293.
60. **Huang Y, Lau SK, Woo PC, Yuen KY.** 2008. CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes. Nucleic Acids Res**36**(Database issue):D504–D511. http://dx.doi.org/10.1093/nar/gkm754.
61. **Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, Le Mercier P.** 2011. ViralZone: a knowledge resource to understand virus diversity. Nucleic Acids Res **39:**D576–D582. http://dx.doi.org/10.1093/nar/gkq901.
62. **Simmonds P.** 2012. SSE: a nucleotide and amino acid sequence analysis platform. BMC Res Notes **5:**50. http://dx.doi.org/10.1186/1756-0500-5-50.
63. **Pellet J, Tafforeau L, Lucas-Hourani M, Navratil V, Meyniel L, Achaz G, Guironnet-Paquet A, Aublin-Gex A, Caignard G, Cassonnet P, Chaboud A, Chantier T, Deloire A, Demeret C, Le Breton M, Neveu G, Jacotot L, Vaglio P, Delmotte S, Gautier C, Combet C, Deleage G, Favre M, Tangy F, Jacob Y, Andre P, Lotteau V, Rabourdin-Combe C, Vidalain PO.** 2010. ViralORFeome: an integrated database to generate a versatile collection of viral ORFs. Nucleic Acids Res **38:**D371–D378. http://dx.doi.org/10.1093/nar/gkp1000.
64. **Yu Q, Ryan EM, Allen TM, Birren BW, Henn MR, Lennon NJ.** 2011. PriSM: a primer selection and matching tool for amplification and sequencing of viral genomes. Bioinformatics **27:**266–267. http://dx.doi.org/10.1093/bioinformatics/btq624.
65. **Angly F, Rodriguez-Brito B, Bangor D, McNairnie P, Breitbart M, Salamon P, Felts B, Nulton J, Mahaffy J, Rohwer F.** 2005. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. BMC Bioinformatics **6:**41. http://dx.doi.org/10.1186/1471-2105-6-41.
66. **Roux S, Faubladier M, Mahul A, Paulhe N, Bernard A, Debroas D, Enault F.** 2011. Metavir: a web server dedicated to virome analysis. Bioinformatics **27:**3074–3075. http://dx.doi.org/10.1093/bioinformatics/btr519.
67. **Lorenzi HA, Hoover J, Inman J, Safford T, Murphy S, Kagan L, Williamson SJ.** 2011. TheViral MetaGenome Annotation Pipeline (VMGAP):an automated tool for the functional annotation of viral Metagenomic shotgun sequencing data. Stand Genomic Sci **4:**418–429. http://dx.doi.org/10.4056/sigs.1694706.
68. **Wommack KE, Bhavsar J, Polson SW, Chen J, Dumas M, Srinivasiah S, Furman M, Jamindar S, Nasko DJ.** 2012. VIROME: a standard operating procedure for analysis of viral metagenome sequences. Stand Genomic Sci **6:**427–439. http://dx.doi.org/10.4056/sigs.2945050.
69. **Zhao G, Krishnamurthy S, Cai Z, Popov VL, Travassos da Rosa AP, Guzman H, Cao S, Virgin HW, Tesh RB, Wang D.** 2013. Identification of novel viruses using VirusHunter–an automated data analysis pipeline. PLoS One **8:**e78470. http://dx.doi.org/10.1371/journal.pone.0078470.
70. **Schultz AK, Zhang M, Bulla I, Leitner T, Korber B, Morgenstern B, Stanke M.** 2009. jpHMM: improving the reliability of recombination prediction in HIV-1. Nucleic Acids Res **37:**W647–W651. http://dx.doi.org/10.1093/nar/gkp371.
71. **Routh A, Johnson JE.** 2014. Discovery of functional genomic motifs in viruses with ViReMa—a virus Recombination Mapper—for analysis of next-generation sequencing data. Nucleic Acids Res **42:**e11. http://dx.doi.org/10.1093/nar/gkt916.
72. **Allred AF, Renshaw H, Weaver S, Tesh RB, Wang D.** 2012. VIPR HMM: a hidden Markov model for detecting recombination with microbial detection microarrays. Bioinformatics **28:**2922–2929. http://dx.doi.org/10.1093/bioinformatics/bts560.
73. **Surendranath V, Chusainow J, Hauber J, Buchholz F, Habermann BH.** 2010. SeLOX–a locus of recombination site search tool for the detection and directed evolution of site-specific recombination systems. Nucleic Acids Res **38:**W293–W298. http://dx.doi.org/10.1093/nar/gkq523.
74. **Petitjean M, Vanet A.** 2014. VIRAPOPS: a forward simulator dedicated to rapidly evolved viral populations. Bioinformatics **30:**578–580. http://dx.doi.org/10.1093/bioinformatics/btt724.
75. **Hawkins TB, Dantzer J, Peters B, Dinauer M, Mockaitis K, Mooney S, Cornetta K.** 2011. Identifying viral integration sites using SeqMap 2.0. Bioinformatics **27:**720–722. http://dx.doi.org/10.1093/bioinformatics/btq722.
76. **Li JW, Wan R, Yu CS, Co NN, Wong N, Chan TF.** 2013. ViralFusionSeq: accurately discover viral integration events and reconstruct fusion transcripts at single-base resolution. Bioinformatics **29:**649–651. http://dx.doi.org/10.1093/bioinformatics/btt011.
77. **Chen Y, Yao H, Thompson EJ, Tannir NM, Weinstein JN, Su X.** 2013. VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. Bioinformatics **29:**266–267. http://dx.doi.org/10.1093/bioinformatics/bts665.
78. **Wang Q, Jia P, Zhao Z.** 2013. VirusFinder: software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. PLoS One **8:**e64465. http://dx.doi.org/10.1371/journal.pone.0064465.
79. **Zhao X, Liu Q, Cai Q, Li Y, Xu C, Li Z, Zhang X.** 2012. Dr.VIS: a database of human disease-related viral integration sites. Nucleic Acids Res **40:**D1041–D1046. http://dx.doi.org/10.1093/nar/gkr1142.
80. **Hsu PW, Lin LZ, Hsu SD, Hsu JB, Huang HD.** 2007. ViTa: prediction of host microRNAs targets on viruses. Nucleic Acids Res **35:**D381–D385. http://dx.doi.org/10.1093/nar/gkl1009.
81. **Li SC, Shiau CK, Lin WC.** 2008. Vir-Mir db: prediction of viral microRNA candidate hairpins. Nucleic Acids Res **36:**D184–D189. http://dx.doi.org/10.1093/nar/gkm610.

82. **Tyagi A, Ahmed F, Thakur N, Sharma A, Raghava GP, Kumar M.** 2011. HIVsirDB: a database of HIV inhibiting siRNAs. PLoS One **6:**e25917. http://dx.doi.org/10.1371/journal.pone.0025917.
83. **Thakur N, Qureshi A, Kumar M.** 2012. VIRsiRNAdb: a curated database of experimentally validated viral siRNA/shRNA. Nucleic Acids Res **40:**D230–D236. http://dx.doi.org/10.1093/nar/gkr1147.
84. **Naito Y, Ui-Tei K, Nishikawa T, Takebe Y, Saigo K.** 2006. siVirus: web-based antiviral siRNA design software for highly divergent viral sequences. Nucleic Acids Res **34:**W448–W450. http://dx.doi.org/10.1093/nar/gkl214.
85. **Qureshi A, Thakur N, Kumar M.** 2013. VIRsiRNApred: a web server for predicting inhibition efficacy of siRNAs targeting human viruses. J Transl Med **11:**305. http://dx.doi.org/10.1186/1479-5876-11-305.
86. **Antoniewski C.** 2011. Visitor, an informatic pipeline for analysis of viral siRNA sequencing datasets. Methods Mol Biol **721:**123–142. http://dx.doi.org/10.1007/978-1-61779-037-9_7.
87. **Watson M, Schnettler E, Kohl A.** 2013. viRome: an R package for the visualization and analysis of viral small RNA sequence datasets. Bioinformatics **29:**1902–1903. http://dx.doi.org/10.1093/bioinformatics/btt297.
88. **Vodovar N, Goic B, Blanc H, Saleh MC.** 2011. *In silico* reconstruction of viral genomes from small RNAs improves virus-derived small interfering RNA profiling. J Virol **85:**11016–11021. http://dx.doi.org/10.1128/JVI.05647-11.
89. **Zhang Y, Bo XC, Yang J, Wang SQ.** 2005. HBVPathDB: a database of HBV infection-related molecular interaction network. World J Gastroenterol **11:**1690–1692.
90. **Chatr-aryamontri A, Ceol A, Peluso D, Nardozza A, Panni S, Sacco F, Tinti M, Smolyar A, Castagnoli L, Vidal M, Cusick ME, Cesareni G.** 2009. VirusMINT: a viral protein interaction database. Nucleic Acids Res **37:**D669–D673. http://dx.doi.org/10.1093/nar/gkn739.
91. **Navratil V, de Chassey B, Meyniel L, Delmotte S, Gautier C, Andre P, Lotteau V, Rabourdin-Combe C.** 2009. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. Nucleic Acids Res **37:**D661–D668. http://dx.doi.org/10.1093/nar/gkn794.
92. **Fu W, Sanders-Beer BE, Katz KS, Maglott DR, Pruitt KD, Ptak RG.** 2009. Human immunodeficiency virus type 1, human protein interaction database at NCBI. Nucleic Acids Res **37:**D417–D422. http://dx.doi.org/10.1093/nar/gkn708.
93. **Lin FK, Pan CL, Yang JM, Chuang TJ, Chen FC.** 2009. CAPIH: a Web interface for comparative analyses and visualization of host-HIV protein-protein interactions. BMC Microbiol **9:**164. http://dx.doi.org/10.1186/1471-2180-9-164.
94. **Palmeira L, Penel S, Lotteau V, Rabourdin-Combe C, Gautier C.** 2011. PhEVER: a database for the global exploration of virus-host evolutionary relationships. Nucleic Acids Res **39:**D569–D575. http://dx.doi.org/10.1093/nar/gkq1013.
95. **Rozanov M, Plikat U, Chappey C, Kochergin A, Tatusova T.** 2004. A web-based genotyping resource for viral sequences. Nucleic Acids Res **32:**W654–W659. http://dx.doi.org/10.1093/nar/gkh419.
96. **Wang S, Sundaram JP, Spiro D.** 2010. VIGOR, an annotation program for small viral genomes. BMC Bioinformatics **11:**451. http://dx.doi.org/10.1186/1471-2105-11-451.
97. **Upton C, Hogg D, Perrin D, Boone M, Harris NL.** 2000. Viral genome organizer: a system for analyzing complete viral genomes. Virus Res **70:**55–64. http://dx.doi.org/10.1016/S0168-1702(00)00210-0.
98. **Tcherepanov V, Ehlers A, Upton C.** 2006. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. BMC Genomics **7:**150. http://dx.doi.org/10.1186/1471-2164-7-150.
99. **Alcantara LC, Cassol S, Libin P, Deforche K, Pybus OG, Van Ranst M, Galvao-Castro B, Vandamme AM, de Oliveira T.** 2009. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. Nucleic Acids Res **37:**W634–W642. http://dx.doi.org/10.1093/nar/gkp455.
100. **Guo FB, Zhang CT.** 2006. ZCURVE_V: a new self-training system for recognizing protein-coding genes in viral and phage genomes. BMC Bioinformatics **7:**9. http://dx.doi.org/10.1186/1471-2105-7-9.
101. **Myers RE, Gale CV, Harrison A, Takeuchi Y, Kellam P.** 2005. A statistical model for HIV-1 sequence classification using the subtype analyser (STAR). Bioinformatics **21:**3535–3540. http://dx.doi.org/10.1093/bioinformatics/bti569.
102. **Yin L, Yao J, Gardner BP, Chang K, Yu F, Goodenow MM.** 2012. HPV-QUEST: a highly customized system for automated HPV sequence analysis capable of processing next generation sequencing data set. Bioinformation **8:**388–390. http://dx.doi.org/10.6026/97320630008388.
103. **Pineda-Peña AC, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, Gómez-López A, Camacho RJ, de Oliveira T, Vandamme AM.** 2013. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. Infect Genet Evol **19:**337–348. http://dx.doi.org/10.1016/j.meegid.2013.04.032.
104. **Duitama J, Kumar DM, Hemphill E, Khan M, Mandoiu II, Nelson CE.** 2009. PrimerHunter: a primer design tool for PCR-based virus subtype identification. Nucleic Acids Res **37:**2483–2492. http://dx.doi.org/10.1093/nar/gkp073.
105. **Chevenet F, Jung M, Peeters M, de Oliveira T, Gascuel O.** 2013. Searching for virus phylotypes. Bioinformatics **29:**561–570. http://dx.doi.org/10.1093/bioinformatics/btt010.
106. **Beerenwinkel N, Daumer M, Oette M, Korn K, Hoffmann D, Kaiser R, Lengauer T, Selbig J, Walter H.** 2003. Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Res **31:**3850–3855. http://dx.doi.org/10.1093/nar/gkg575.
107. **Maes P, Matthijnssens J, Rahman M, Van Ranst M.** 2009. RotaC: a web-based tool for the complete genome classification of group A rotaviruses. BMC Microbiol **9:**238. http://dx.doi.org/10.1186/1471-2180-9-238.
108. **Onodera K, Melcher U.** 2002. VirOligo: a database of virus-specific oligonucleotides. Nucleic Acids Res **30:**203–204. http://dx.doi.org/10.1093/nar/30.1.203.
109. **Shen HB, Chou KC.** 2007. Virus-PLoc: a fusion classifier for predicting the subcellular localization of viral proteins within host and virus-infected cells. Biopolymers **85:**233–240. http://dx.doi.org/10.1002/bip.20640.
110. **Qureshi A, Thakur N, Kumar M.** 2013. HIPdb: a database of experimentally validated HIV inhibiting peptides. PLoS One **8:**e54908. http://dx.doi.org/10.1371/journal.pone.0054908.
111. **Qureshi A, Thakur N, Tandon H, Kumar M.** 2014. AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses. Nucleic Acids Res **42:**D1147–D1153. http://dx.doi.org/10.1093/nar/gkt1191.
112. **Thakur N, Qureshi A, Kumar M.** 2012. AVPpred: collection and prediction of highly effective antiviral peptides. Nucleic Acids Res **40:**W199–W204. http://dx.doi.org/10.1093/nar/gks450.
113. **Hong JJ, Wu TY, Chang TY, Chen CY.** 2013. Viral IRES prediction system - a web server for prediction of the IRES secondary structure in silico. PLoS One **8:**e79288. http://dx.doi.org/10.1371/journal.pone.0079288.
114. **Ferron F, Rancurel C, Longhi S, Cambillau C, Henrissat B, Canard B.** 2005. VaZyMolO: a tool to define and classify modularity in viral proteins. J Gen Virol **86:**743–749. http://dx.doi.org/10.1099/vir.0.80590-0.
115. **Singh M, Berger B, Kim PS.** 1999. LearnCoil-VMF: computational evidence for coiled-coil-like motifs in many viral membrane-fusion proteins. J Mol Biol **290:**1031–1041. http://dx.doi.org/10.1006/jmbi.1999.2796.
116. **Carrillo-Tripp M, Shepherd CM, Borelli IA, Venkataraman S, Lander G, Natarajan P, Johnson JE, Brooks CL, III, Reddy VS.** 2009. VIPERdb2: an enhanced and web API enabled relational database for structural virology. Nucleic Acids Res **37:**D436–D442. http://dx.doi.org/10.1093/nar/gkn840.