



Bioinformatics Applications in Advancing Animal Virus Research

23

Ablesh Gautam, Ashish Tiwari, and Yashpal Singh Malik

Abstract

Viruses serve as infectious agents for all living entities. There have been various research groups that focus on understanding the viruses in terms of their host-viral relationships, pathogenesis and immune evasion. However, with the current advances in the field of science, now the research field has widened up at the 'omics' level. Apparently, generation of viral sequence data has been increasing. There are numerous bioinformatics tools available that not only aid in analysing such sequence data but also aid in deducing useful information that can be exploited in developing preventive and therapeutic measures. This chapter elaborates on bioinformatics tools that are specifically designed for animal viruses as well as other generic tools that can be exploited to study animal viruses. The chapter further provides information on the tools that can be used to study viral epidemiology, phylogenetic analysis, structural modelling of proteins, epitope recognition and open reading frame (ORF) recognition and tools that enable to analyse host-viral interactions, gene prediction in the viral genome, etc. Various databases that organize information on animal and human viruses have also been described. The chapter will converse on overview of the current advances, online and downloadable tools and databases in the field of bioinformatics that will enable the researchers to study animal viruses at gene level.

A. Gautam (✉)

Central Research Institute, Kasauli, Himachal Pradesh, India

A. Tiwari

University of Kentucky, Louisville, KY, USA

Y. S. Malik

ICAR-Indian Veterinary Research Institute (ICAR-IVRI), Izatnagar, Uttar Pradesh, India

© Springer Nature Singapore Pte Ltd. 2019

Y. S. Malik et al. (eds.), *Recent Advances in Animal Virology*,
https://doi.org/10.1007/978-981-13-9073-9_23

447

Keywords

Animal viruses · Animal diseases · Bioinformatics tools · Online databases · Gene prediction · ORF finding · Host-virus relationship

Preamble

Viruses are notorious to infect all forms of life ranging from bacteria to chordates. In humans, viruses are known to cause infectious diseases such as influenza, hepatitis, AIDS, diarrhoea, encephalitis, dengue fever and, more recently, severe acute respiratory syndrome (SARS), Ebola (Singh et al. 2017a), Zika (Singh et al. 2017b), etc. Despite the vaccines and treatments for such diseases, morbidity and mortality both occur as a result of the viral infections. Viral disease of animals not only affects the production but also is a threat to humans (Saminathan et al. 2016). A rapid growth in the availability of sequencing methods and a vast amount of viral sequence data have been generated during recent times. Thus, it is imperative to decipher this data using more advanced tools such as bioinformatics resources. A large number of bioinformatics tools that can aid in the analysis of viral genomes and develop preventive and therapeutic strategies have been developed for human as well as animal viruses. This chapter will introduce virologists to some of the common as well virus-specific bioinformatics tools that the researches can use to analyse viral sequence data to elucidate the viral dynamics, evolution and preventive therapeutics.

23.1 Applications of Bioinformatics in Virology

Analysis of viral sequence involves use of certain tools that are employable on any novel sequence, for example, gene identification, ORF identification, functional annotation and phylogeny. However, due to small genome size, viruses have complex methods to maximize the coding potential of genomes and evolution. Many viruses utilize overlapping reading frames or translational frameshifts to code for multiple proteins from limited genome sequences. Also, higher rates of mutations and recombination between related viruses pose a challenge in accurate phylogenetic and evolutionary analysis of viruses using general-purpose softwares. Lately, enormous growth in the volume and diversity of viral sequences in the databases has been seen. Now, it has become imperative to organize data of these viral sequences in virus family-specific resources tailored for accurate analysis of a specific virus.

23.1.1 Phylogeny and Molecular Epidemiology

One of the most common applications of bioinformatics in virology was to use phylogenetic analysis of the viral isolates to aid in the epidemiological analysis of viral outbreaks. General-purpose phylogeny programs such as PHYLIP (Felsenstein 1989) have been used extensively for the phylogeny and molecular epidemiology of viruses. A comprehensive list of these packages and web servers is maintained by Joe Felsenstein at <http://evolution.genetics.washington.edu/phylip/software.html>.

23.1.2 ORF/Gene Discovery

An open reading frame (ORF) is the part of genome that translates into a protein. Finding ORF is one of the key steps in viral genome analysis. It forms the basis for further analysis such as homologous search, predicting proteins, functional analysis and viral vaccine and antiviral target discovery. If an ORF translates a surface protein that is unique to that virus, it may elicit immune responses and could potentially be a vaccine candidate. ORF Finder by NCBI is a ORF prediction program (Rombel et al. 2002). The program outputs a range of each ORFs along with its protein translation in six possible reading frames from the input DNA sequence. It can be used to search newly sequenced DNA for potential protein encoding sequences and to verify predicted proteins using SMART BLAST or BLASTP (Altschul et al. 1990). However, the web version of the program is limited to a query sequence length of 50 kb only. A standalone system has no limitation on length but is available only for the Linux 64 operating system. NEG8, a 167-codon novel ORF in segment 8 of influenza virus, was visualized using ORF Finder (Clifford et al. 2009). Using the ORF Finder in association with the basic local alignment search tool BLAST, 154 ORFs were found in the Hz-1 virus genome (Cheng et al. 2002). Due to small genome size, viruses employ multiple strategies to maximize the coding potential including frameshifts and alternative codon usage. Thus, virus-specific programs have been developed to overcome these challenges. GeneMark (<http://opal.biology.gatech.edu/GeneMark/genemarks.cgi>) provides gene prediction tools for viruses (Besemer and Borodovsky 2005). Viral genome organizer (VGO) – a Java-based web tool – offers identification of gene and ORF identification in viral sequences (Upton et al. 2000).

23.1.3 Epitope Recognition

Identification of immune epitopes is important in designing new vaccine candidates and in diagnostics. An epitope is the part of an antigen that is recognized by the receptors of immune system components such as antibodies, B cells or T cells.

Epitopes have been generally classified as either linear or conformational epitopes. T cells recognize linear epitopes, short continuous strings of amino acids derived from protein antigen, presented with MHC class I molecules. B cells and antibodies, on the other hand, recognize conformational epitopes which are formed by interactions of amino acids with multiple discontinuous segments forming a three-dimensional antigen (Barlow et al. 1986). Owing to the simple linear structure of T cell epitopes, their interaction with receptors can be modelled with high accuracy (DeLisi and Berzofsky 1985). A large number of prediction databases and servers thus are available for linear epitope prediction. MHCPEP (Brusic et al. 1998), SYFPEITHI (Rammensee et al. 1999), FIMM (Schonbach et al. 2005), MHCBN (Bhasin et al. 2003) and EPIMHC (Reche et al. 2005) are some of the commonly used T cell epitope prediction programs. Immune epitope database and analysis resource (<https://www.iedb.org>) (Vita et al. 2015) offers the most comprehensive set of tools for epitope analysis for epitope prediction covering HLA-A and HLA-B for humans as well as chimpanzee, macaque, gorilla, cow, pig and mouse and is one of the few databases that cover such a variety of organisms. Since 2011, IEDB uses NetMHCpan as prediction method. NetMHC server uses the artificial neural network method to predict binding of peptides to different alleles from human as well as 41 animals including cattle and pig (38 from core). The database also contains curated data for many viruses including influenza and herpesviruses. B cell receptors and epitope interactions are more complex in nature than the linear epitopes for T cells; thus, accuracy of B cell epitopes is relatively low. Furthermore, most of the current databases are centred on linear rather than conformational epitopes. Bcipep is a tool developed for predicting the linear epitope of B cells (Saha et al. 2005). Epitome is a database of structure-inferred antigenic residues in proteins (Schlessinger et al. 2006). Epitome is especially useful in the prediction of antibody-antigen complex interaction. The database is available at <http://www.rostlab.org/services/epitome/>. AntiJen is an intricate database with entries on both T cell and B cell epitopes. It emphasizes on integration of kinetic, thermodynamic, functional and cellular data within the context of immunology and vaccinology (Toseland et al. 2005) (Fig. 23.1a).

23.1.4 Structural Modelling

Three-dimensional prediction of viral proteins can be used to predict the correlation between actual protein structure and antigenic sites, folding surfaces and functional motifs. Such structural modelling tools may be implicated to identify and design novel candidates for antiviral inhibitors and vaccine targets. Secondary structures may be predicted using the tool PredictProtein (<http://www.predictprotein.org/>) (Rost et al. 2004). Using this online tool, along with secondary structures, solvent accessibility and possible transmembrane helices can be predicted. Further, it also provides expected accuracy of prediction methods. SWISS-MODEL (<http://>

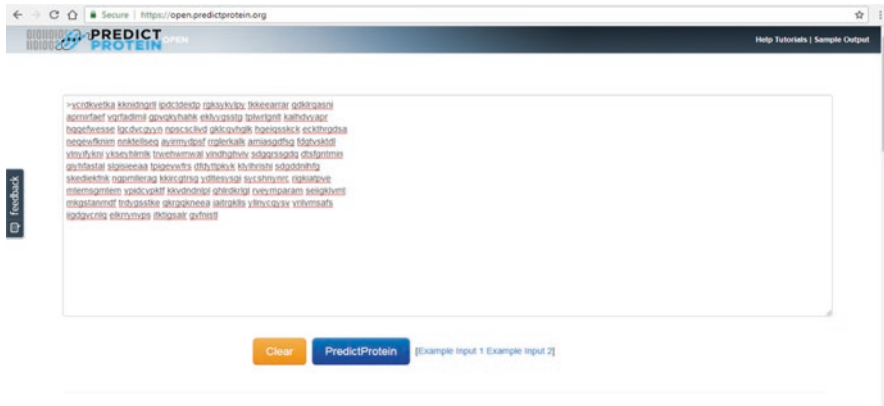


Fig. 23.1a The online tool PredictProtein predicts various secondary structures in a given viral protein. The amino acid sequences of viral protein are required to be fed in Fasta format

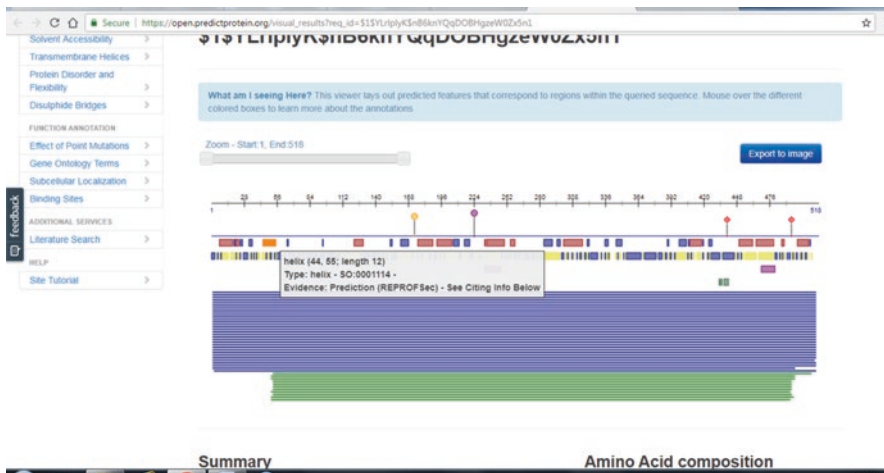


Fig. 23.1b Prediction of various secondary structures (Helical) in a given viral protein using the online tool Predictprotein

swissmodel.expasy.org/) is a popular tool for the prediction of a 3-D structure of a protein. 3-D structure prediction programs usually employ homology searching using similar and known protein structures as templates. One of the most commonly used database for such templates is Protein Data Bank (PDB) (Reddy et al. 2001). Output from the SWISS-MODEL program includes the template selected, alignment between the query sequence and the template, and the predicted 3-D model. Results of SWISS-MODEL are, however, only sent by email (Figs. 23.1b, 23.1c, 23.1d and 23.1e).

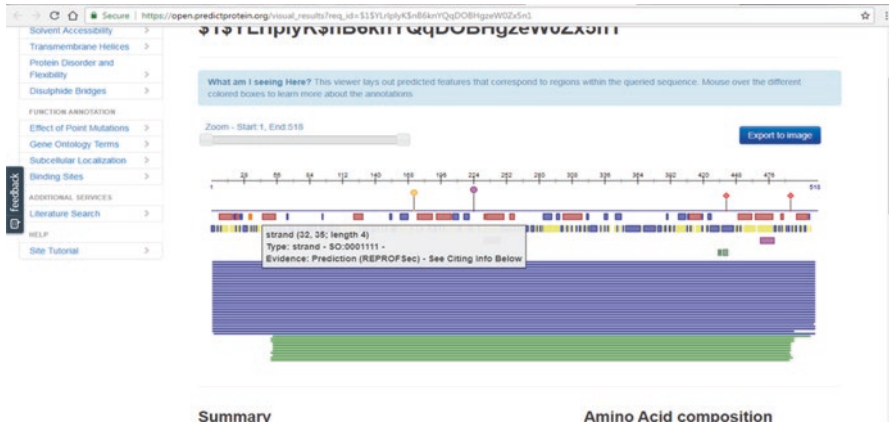


Fig. 23.1c Prediction of various secondary structures (Strand) in a given viral protein using the online tool Predictprotein

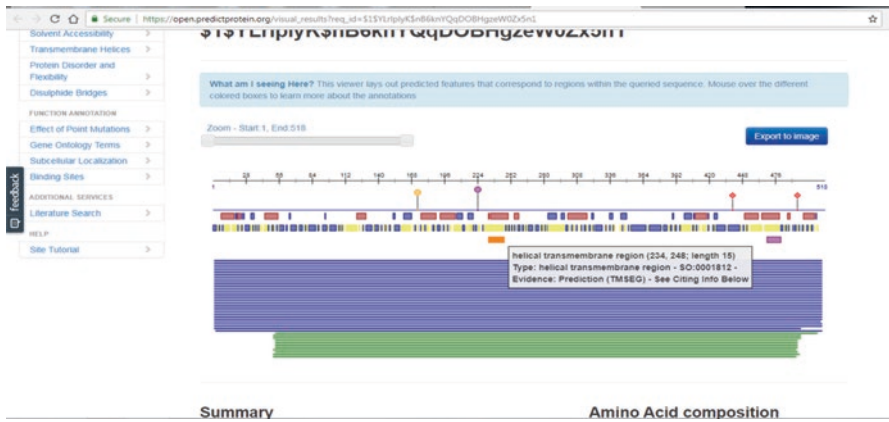


Fig. 23.1d Prediction of various secondary structures (Helical transmembrane region) in a given viral protein using the online tool Predictprotein

23.2 Virus-Centred Bioinformatics Tools

For long, bioinformatic analysis of viruses utilized common bioinformatics tools developed for other organisms. However, analysing viral genomes using general bioinformatics tools could compromise the accuracy and sensitivity of analysis. Virus genomes are too small (e.g. < 10 kb) to compute statistics with their codon usage. To maximize the coding potential, viruses work with unusual codon usage patterns comprising of overlapping coding and non-coding functional elements. Additionally, viruses also rely on other translational mechanisms such as stop codon read-through, frameshifting, leaky scanning and internal ribosome entry sites.

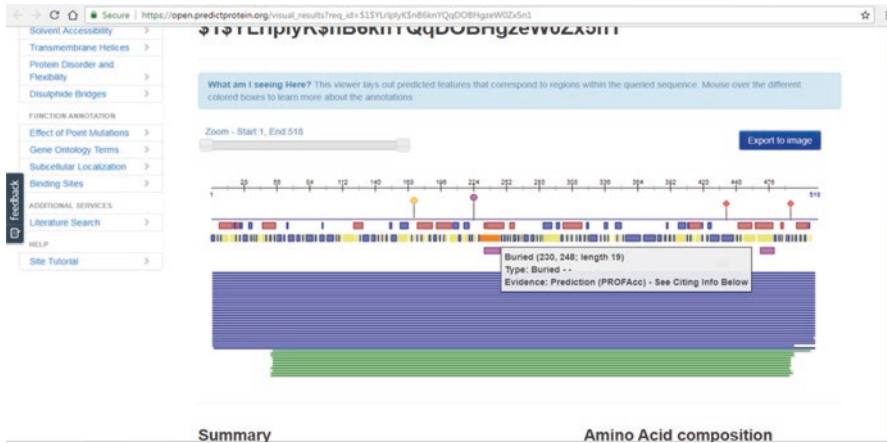


Fig. 23.1e Prediction of various secondary structures (buried sequence motifs) in a given viral protein using the online tool Predictprotein

Comparative genomic analysis of viruses is complicated by the fact that highly conservative sequences may not be coding for anything. Presence of overlapping pairs may be indicated by conservation for the sequences where there is overlapping of CDSs and/or non-coding functional elements. Novel virus types comprise of new CDSs that are different than previously known CDSs. There are multiple databases and tools available for analysis of human viruses; however, there are still only a limited number of resources designed specifically for veterinary viruses. In this section, some of the databases and resources useful for the analysis of veterinary viruses are discussed (Table 23.1).

23.2.1 Comparative and Diversity Analysis of Viral Sequences

Viruses are one of the most diversified and dynamic microorganisms. With increasing viral genome sequencing, there was a need to develop bioinformatics tools to compare and analyse the voluminous data. To meet this requirement, one such downloadable software package is Base-By-Base, which aids in analysis of whole viral genome alignments at single nucleotide level (Brodie et al. 2004). Moreover, with the online resource Genome Information Broker for Viruses (GIB-V), comparative studies can be made using the generic tools such as ClustalW, BLAST and Keyword Search algorithms (Hirahata et al. 2007). Another downloadable web server tool, ViroBLAST, is an exclusive BLAST tool that can be used for queries against multiple databases (Deng et al. 2007). Sequences from a variety of viral strains can be analysed simultaneously using the Alvira software, which is a multiple sequence alignment tool that provides graphical representation as well (Enault et al. 2007). Furthermore, comparative analysis of genes and genomes of coronavirus can be carried out by using the CoVDB (coronavirus database) (Huang et al. 2008).

Table 23.1 Virus-specific bioinformatics tools

S. no.	Applications	Name of the tools
1.	ORF gene finding	ORF Finder by NCBI SMART BLAST GeneMark VGO VIDA VIGOR
2.	Epitope recognition	MHCPEP SYFPEITHI FIMM MHCBN NetMHC Epitome
3.	Comparative and diversity analysis of viral sequences	GIB-V ViroBLAST ViralZone ViralORFeome PriSM PHACCS VIROME VMGAP Metavir Wommack
4.	Viral recombination and integration-specific resources/ tools	jpHMM ViReMa VIPR HMM SeLOX VIRAPOPS SeqMap VirusSeq ViralFusionSeq VirusFinder
5.	Small-RNA analysis tools	ViTa VIRsiRNadb siVirus Paparazzi Visitor VIROME
6.	Virus-host interaction tools	PhEVER Virus-PLoc iLoc-virus pLoC-mVirus
7.	Genome annotation tools	NCBI genotyping tool VIGOR VGO GATU ZCURVE_V STAR

(continued)

Table 23.1 (continued)

S. no.	Applications	Name of the tools
8.	Primer design tools	PrimerHunter PhyloType RotaC VirOligo
9.	Virus structural modelling tools	PredictProtein SWISS-MODEL LearnCoil-VMF VIPERdb Jmol STRAP
10	Phylogeny and molecular epidemiology	PHYLIP

The digital resource ViralZone is designed specifically to comprehend viral diversity and acquire information on viral molecular biology, hosts, taxonomy, epidemiology and structures (Hulo et al. 2011). The Simmonds program was upgraded to the simple sequence editor (SSE) software package, wherein the user-given sequences can be aligned and annotated and further can be analysed for diversity and phylogeny (Simmonds 2012). Evolutionary changes in viral genome lead to polymorphisms in their proteins, which in turn result into changes in viral phenotype such as viral virulence, viral-host interactions, etc. The digital database, ViralORFeome, not only stores all variants and mutants of viral ORFs, but also provides tools to design ORF-specific cloning primers (Pellet et al. 2010). Further, degenerate primer pairs can be selected and matched to amplify user-defined viral genomes using the online tool PriSM (Yu et al. 2011). The recent advances in next-generation sequencing and technologies have facilitated to study viral population at an advanced level. The viral population biodiversity and dynamics can be studied using the first such tool developed, PHACCS (Phage Communities from Contig Spectrum), that can analyse the shotgun sequence data to estimate the structure and diversity of phages (Angly et al. 2005). Later on, more tools/resources were developed to analyse viral metagenomics sequences, such as Viral Informatics Resource for Metagenomic Exploration (VIROME), Viral MetaGenome Annotation Pipeline (VMGAP) and Metavir (Lorenzi et al. 2011, Roux et al. 2011, Wommack et al. 2012). Novel viruses can be identified from a pool of specimen types using a specific computational pipeline, VirusHunter (Zhao et al. 2013).

23.2.2 Viral Recombination and Integration-Specific Resources

The phenomenon of genetic recombination in viruses is responsible for the emergence of new viruses, increased virulence and host range, immune evasion and development of antiviral resistance. This distinct process of viral recombination can be detected by two bioinformatics tools, viz. jpHMM (Jumping Profile Hidden

Markov Model) and ViReMa (Virus Recombination Mapper) genomes (Schultz et al. 2009; Routh and Johnson 2014). The jpHMM, a web server, can be used for predicting recombination in HIV-1 and HBV, whereas ViReMa, a downloadable software, can be used to analyse next-generation sequencing data. Additionally, another software called VIPR HMM (Viral Identification with a PRobabilistic algorithm incorporating hidden Markov model) can detect recombinant and non-recombinant viruses using microbial detection microarrays (Allred et al. 2012). Further, viral genome sequences can be searched for degenerate locus of recombination (lox)-like sites by a web server called SeLOX (Surendranath et al. 2010). A downloadable software, VIRAPOPS, is a forward simulator that allows simulation of RNA virus population (Petitjean and Vanet 2014). With this software, the drastic changes in rapidly evolving RNA viruses such as mutability, recombination, variation, covariation, etc. can be simulated to predict their effects on viral populations. SeqMap is a tool capable of identifying viral integration sites (VIS) from ligation-mediated PCR (LM-PCR), linear amplification-mediated PCR (LAM-PCR) and nonrestrictive LAM-PCR (nrLAM-PCR) reactions and mapping short sequences to the genome (Hawkins et al. 2011). Further, VIS can also be detected by three more distinct tools, VirusSeq, ViralFusionSeq, and VirusFinder (Chen et al. 2013, Li et al. 2013, Wang et al. 2013). For more precise VIS prediction, all four tools can be employed by virologists.

23.2.3 Small-RNA Analysis Tools

miRNAs: A microRNA (miRNA) is a small, regulatory, non-coding RNA molecule that regulates the translation or stability of viral and host target mRNAs, thereby affecting viral pathogenesis. This host-viral regulatory relationship can be investigated by a database called ViTa, capable of curating known viral miRNA genes and known/putative target sites of host miRNA (Hsu et al. 2007). ViTa exploits miRanda and TargetScan to scan viral genomes and determine miRNA targets. ViTa is also capable of annotating the viruses, virus-infected tissues and tissue specificity of host miRNAs. Subtypes of viruses, for example, influenza viruses, and the conserved regions in various viruses can also be compared using the ViTa database. Viral miRNA candidate hairpins can be predicted using the database Vir-Mir. It serves as a platform to query the predicted viral miRNA hairpins (based on taxonomic classification) and host target genes (based on the use of the RNAhybrid program) in human, mouse, rat, zebrafish, rice and *Arabidopsis* (Li et al. 2008).

siRNA: A siRNA is similar to miRNA that operates within the RNA interference (RNAi) pathway. It interferes in expression of specific genes and, therefore, is used in post-transcriptional gene silencing. VIRsiRNAdb is an online curated repository that stores experimentally validated research data of siRNA and short hairpin RNA (shRNA) targeting diverse genes of 42 important human viruses, including influenza virus (Tyagi et al. 2011, Thakur et al. 2012). The current database includes experimental information on siRNA sequence, virus subtype, target gene, GenBank accession, design algorithm, cell type, test object, method, efficacy, etc. A

web-based software, siVirus, is an antiviral sRNA design software that allows analysis of influenza virus, HIV-1, HCV and SARS coronavirus (Naito et al. 2006). Further, viral siRNA sequence data sets can be analysed using the softwares Visitor and VIROME (Antoniewski 2011; Watson et al. 2013). A Perl script, called Papparazzi, enables reconstitution of viral genome using a viral siRNA in a given sample (Vodovar et al. 2011).

23.2.4 Virus-Host Interaction and Miscellaneous Softwares

Host-pathogenic interactions play an important role in determining the pathogenicity of a pathogen or immune evasion mechanism of a host. To comprehend such interactions between viral and host cellular proteins, various databases and softwares are available. One such database is PhEVER that enables to explore virus-virus and virus-host lateral gene transfers by providing evolutionary and phylogenetic information (Palmeira et al. 2011). This distinct database catalogues homologous families between different viral sequences and between viral and host sequences. It compiles the extensive data from completely sequenced genomes (2426 non-redundant viral genomes, 1007 non-redundant prokaryotic genomes, 43 eukaryotic genomes ranging from plants to vertebrates). Thus, it enables compiling of various proteins into homologous families by selecting at least one viral sequence, related alignments and phylogenies for each of these families.

With increasing availability of viral genome sequences, data mining, curation and genome annotation have become essential components to better comprehend the structure and function of genome components. This information can further be exploited to develop diagnostics, vaccines and therapeutics.

There are a number of tools available capable of annotation and classification of viral sequences, such as NCBI genotyping tool (Rožanov et al. 2004), VIGOR (Viral Genome ORF Reader) (Wang et al. 2010), Viral Genome Organizer (VGO) (Upton et al. 2000), Genome Annotation Transfer Utility (GATU) (Tcherepanov et al. 2006), Virus Genotyping Tools (Alcantara et al. 2009), ZCURVE_V (Guo and Zhang 2006) and STAR (Subtype Analyser) (Myers et al. 2005).

VGO is a web-based genome browser that allows viewing and predicting genes and ORFs in one or more viral genomes. It also allows performing searches within viral genomes and acquiring information about a genome such as locating genes, ORFs, start/stop codons, etc. Within genome, the sequences can be searched for regular expression, fuzzy motif pattern, genes with highest AT composition, etc. Using VGO, comparative analyses can be made between different viral genomes. VGO uses the graphical user interface (GUI) for constructing alignments and display orthologues in a set of genomes. It also allows searching the translated genome for matches to mass spec peptides.

VIGOR is a gene prediction online tool that was developed by J. Craig Venter Institute in 2010. It started with gene prediction in small viral genomes such as coronavirus, influenza, rhinovirus and rotavirus. With the updated version in 2012 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3394299/>), VIGOR is now

capable of gene prediction in 12 more viruses: measles virus, mumps virus, rubella virus, respiratory syncytial virus, alphavirus and Venezuelan equine encephalitis virus, norovirus, metapneumovirus, yellow fever virus, Japanese encephalitis virus, parainfluenza virus and Sendai virus. With VIGOR, based on sequence similarity searches, users are able to predict protein coding regions, start and stop codons and other complex gene features such as RNA editing, stop codon leakage and ribosomal shunting. Further, various features such as frameshifts, overlapping genes, embedded genes, etc. can be predicted in the virus genome. Additionally, a mature peptide can be predicted in a given polypeptide open reading frame. VIGOR is also capable of genotyping influenza virus and rotavirus. Four output files – a gene prediction file, a complementary DNA file, an alignment file, and a gene feature table file – are produced by VIGOR. GenBank submission can be directly done using the gene feature table.

Genome Annotation Transfer Utility (GATU) facilitates quick and efficient annotation of similar target genome using the reference genomes that have already been annotated. Later, the users can manually curate the annotated genome. The newly annotated genomes can be saved as GenBank, EMBL or XML file format. Although it doesn't provide a complete annotation system, GATU serves as a very useful tool for the preliminary work in genome annotation. GATU utilizes tBLASTn and BLASTn algorithms to map genes onto the new target genome by using an annotated reference genome. As a result, majority of the new genome's genes are annotated in a single step. With GATU, users can also identify open reading frames present in the target genome and absent from the reference genome. These ORFs can further be scrutinized by using other bioinformatics tools such as BLAST and VGO, which can determine if the ORFs should be included in the annotation. Multiple-exon genes and mature peptides can also be analysed using GATU.

A primer design tool, PrimerHunter, allows to design highly sensitive and specific primers for virus subtyping by PCR (Duitama et al. 2009). PrimerHunter allows predicting specific forward and reverse primers with respect to a given set of DNA sequences. PhyloType is a web-based as well as downloadable software that uses parsimony to reconstruct ancestral traits and to select phylotypes (Chevenet et al. 2013). RotaC is an automated genotyping tool for group A rotaviruses (Maes et al. 2009). It works by comparing a complete ORF of interest to other complete ORFs of cognate genes available in the GenBank database by performing BLAST searches.

VirOligo is a database of virus-specific oligonucleotides. The VirOligo database acts as a repository for virus-specific oligonucleotides for virus detection (Onodera and Melcher 2002). The database comprises of Oligo data and Common data tables. The Oligo data table enlists PCR primers and hybridization probes that are used for viral nucleic acid detection, while Common data table contains PCR and hybridization experimental conditions used in their detection. Each Oligo data entry provides information on the name of the oligonucleotide, oligonucleotide sequence, target region, type of usage (PCR primer, PCR probe, hybridization or other), note and direction of the PCR oligonucleotide (forward or reverse). Each oligonucleotide entry also contains direct links to PubMed, GenBank, NCBI Taxonomy databases and

BLAST. On the updated version of VirOligo as of September 2015, the database contains complete listing of oligonucleotides specific to various animal viruses. The viruses are vaccinia virus; canine parvovirus; porcine parvovirus; rodent parvovirus; tobamovirus; potyvirus; borna virus; bovine herpesvirus types 1, 3, 4 and 5; bovine viral diarrhoea virus; bovine parainfluenza 3 virus; bovine respiratory syncytial virus; bovine adenovirus; bovine rhinovirus; bovine coronavirus; bovine reovirus; bovine enterovirus; foot-and-mouth disease (FMD) virus; and alcelaphine herpesvirus.

Virus-PLOC is a web server for prediction of subcellular localization of viral proteins within host and virus-infected cells (Shen and Chou 2007). Another web server developed a little later, iLoc-Virus, is a multi-label learning classifier that predicts the subcellular locations of viral proteins with single and multiple sites (Xiao et al. 2011). Similarly, a most recent web server, pLoC-mVirus (Cheng et al. 2017), is a new predictor that identifies subcellular localization of viral proteins with both single and multiple location sites. It works by extracting information from the Gene Ontology (GO) database and is claimed to be more successful than the state-of-the-art method, iLoc-Virus, in predicting subcellular localization of viral proteins. AVPPred is an antiviral peptide prediction algorithm that contains the peptides with experimentally proven antiviral activity (Thakur et al. 2012). The prediction is based on peptide sequence features, peptide motifs, sequence alignment, amino acid composition and physicochemical properties. VIPS is a viral internal ribosomal entry site (IRES) prediction system that can predict IRES secondary structures (Hong et al. 2013). VIPS uses the RNA fold program that predicts local RNA secondary structures, RNA align program that compares predicted structures and pknobsRG program (Reeder et al. 2007) that calculates the pseudoknot structures. VaZyMoLo, a database that deals with viral sequences at protein level, defines and classifies viral protein modularity (Ferron et al. 2005). It extracts information of complete genome sequences of various viruses from GenBank and RefSeq and organizes the acquired information about modularity on viral ORFs (Fig. 23.1f).

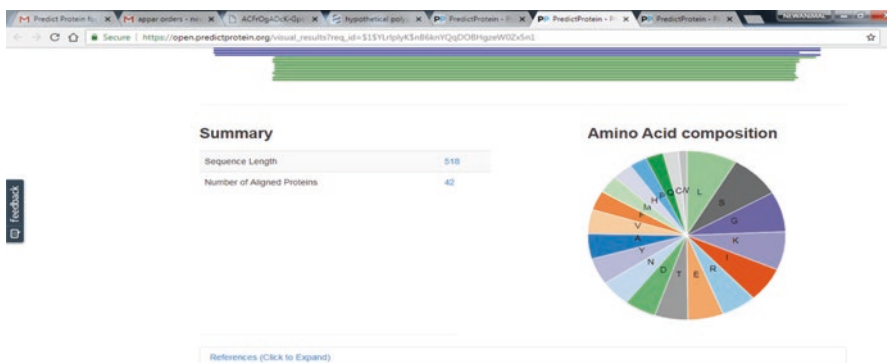


Fig. 23.1f Representation of amino acid composition in a given viral protein using the online tool PredictProtein

There are web-based tools available to predict and analyse structural aspects of viruses. The LearnCoil-VMF is a computational tool that allows to predict coiled-coil-like regions in viral membrane fusion proteins (Singh et al. 1999). The membrane fusion proteins are known to be diverse and share no sequence similarity between most pairs of viruses in the same or different families. The LearnCoil-VMF is also capable of characterizing the core structure of these membrane fusion proteins.

VIPERdb (Virus Particle Explorer database) is a web-based database that enables manual curation of icosahedral virus capsid structures (Carrillo-Tripp et al. 2009). This database serves as a comprehensive resource for specific needs of structural virology and comparatives of data derived from structural and computational analyses of capsids. With the updated version, VIPERdb (2), capsid protein residues in the icosahedral asymmetric unit (IAU) can be deduced using Phi-Psi (Phi-Psi) diagrams (azimuthal polar orthographic projections) (Ref: <https://www.ncbi.nlm.nih.gov/pubmed/18981051>). These diagrams can be depicted as dynamic interface and surface residues and interface and core residues and can be mapped to the database using a new application programming interface (API). This aids in identifying family-wide conserved residues at the interfaces. Additionally, Jmol and STRAP are built in the system to visualize an interactive model of viral molecular structures.

VIDA is a database that organizes animal virus genome open reading frames from partial and complete genomic sequences (Alba et al. 2001). Presently, VIDA includes a complete collection of homologous protein families from GenBank for *Herpesviridae*, *Papillomaviridae*, *Poxviridae*, *Coronaviridae* and *Arteriviridae*. The homologous proteins in VIDA include both orthologous and paralogous sequences. VIDA retrieves virus sequences from GenBank and the files are parsed into subfields. The parsed fields contain all the information such as GenBank accession number, GenBank identifier (GI numbers), protein sequence source, sequence length, gene name and gene product. In order to eliminate 100% redundancy, the virus protein sequences thus retrieved are filtered and a list of synonymous GIs is created for reference. The ORFs from complete and partial virus genomes are further organized into homologous protein families, on the basis of sequence similarity. Furthermore, the structure of known viral proteins or homologous to viral proteins is also mapped onto homologous protein families. VIDA also provides functional classification of virus proteins into broad functional classes based on typical virus processes such as DNA and RNA replication, virus structural proteins, nucleotide and nucleic acid metabolism, transcription, glycoproteins and others. This database also provides alignment of the conserved regions based on potential functional importance. Apart from functional classification, VIDA also provides a taxonomical classification of the proteins and protein families. The protein families serve as a tool for functional and evolutionary studies, whereas alignments of conserved sequences provide crucial information on conserved amino acids or construction of sequence profiles.

23.3 Virus Bioinformatics Databases

23.3.1 Viral Bioinformatics Resource Center (VBRC)

The Viral Bioinformatics Resource Center (VBRC) is one of eight NIH-sponsored Bioinformatics Resource Centers (<http://www.oxfordjournals.org/nar/database/summary/798>). It is an online platform that provides informational and analytical tools and resources to scientific community. The VBRC is oriented to conduct basic and applied research to better comprehend the viruses included on the NIH/NIAID list of priority pathogens. These viruses are selected based on their possibility of bioterrorism threats or as emerging or re-emerging infectious diseases. The VBRC focuses specifically on large DNA viruses. It includes the viruses that belong to the *Arenaviridae*, *Bunyaviridae*, *Filoviridae*, *Flaviviridae*, *Paramyxoviridae*, *Poxviridae* and *Togaviridae* families. It serves as a relational database and web application tool that allows data storage, annotation, analysis and information exchange of the data. The current version (V 4.2) consists of 369 complete genomic sequences.

Using the VBRC, each of the viral gene and genome can be curated. As a result, a comprehensive and searchable summary is acquired that details about the genotype and phenotype of the genes. The role of the genes in host-pathogen relationships is also being emphasized in these curations. Additionally, the VBRC also houses multiple analytical tools such as tools for genome annotation, comparative analysis, whole genome alignments and phylogenetic analysis. Further, this database also looks forward to include high-throughput data derived from other studies such as microarray gene expression data, proteomic analyses and population genetics data.

23.3.2 Poxvirus Bioinformatics Resource Center (PBRC)

The Poxvirus Bioinformatics Resource Center (PBRC, now merged into VBRC) is an online platform that serves as an informational and analytical resource to better comprehend the *Poxviridae* family of viruses. It allows data storage, annotation, analysis and information exchange of the data.

23.3.3 Influenza Virus Database (IVDB)

Influenza virus is one the major global concern. It gained attention after the emergence of pandemic influenza A virus (H1N1, swine flu) in 2009. There are a total of 11 web portals and tools that focus only on influenza virus. This includes the Influenza Virus Database (IVDB), Influenza Research Database (IRD) and NCBI Influenza Virus Resource (NCBI-IVR) (Chang et al. 2007; Bao et al. 2008; Squires et al. 2008). Researchers can exploit all the three websites mentioned for sequence databases as well as various basic tools such as BLAST, multiple-sequence alignment, phylogenetic tree construction, etc.

IVDB provides access to additional tools such as (i) the Sequence Distribution Tool, which provides global geographical distribution of a given viral genotype as well as correlates its genomic data with epidemiological data, and (ii) the Quality Filter System, which according to their sequence content (coding sequence [CDS], 5'untranslated region [5'UTR], and 3'UTR) and integrity (complete [C] or partial [P]) categorizes a given viral nucleotide sequence into either of the seven categories of C1 to C4 and P1 to P3, respectively. NCBI-IVR is the most widely used and cited online resource. With NCBI-IVR, the given viral genomic sequences can be annotated using a genome annotation tool and Flu ANnotation (FLAN) tool. Additionally, large phylogenetic trees may be constructed and can be visualized in aggregated form with sub-scale details (Bao et al. 2007; Bao et al. 2008; Zaslavsky et al. 2008). IRD provides tools for genomic and proteomic intervention, immune epitope prediction and surveillance data for viral nucleotide sequences (Squires et al. 2012). Furthermore, this resource is also equipped with tools that provide insight into host-pathogen interactions, type of virulence, host range and a correlation of sequence variation and these processes. There are other repositories available: Global Initiative on Sharing Avian Influenza Data (GISAID) consortium that mediated the EpiFlu database and FluGenome database that exclusively provides genotyping of influenza A virus and aids in detecting reassortments taking place in divergent lines (Lu et al. 2007). Furthermore, reassortment events in influenza viruses exclusively can be identified by a program GiRaF (Graph-incompatibility-based Reassortment Finder) that can be downloaded (Nagarajan and Kingsford 2011). Another distinct repository, Influenza Sequence and Epitope Database (ISED), provides viral sequences and epitopes from Asian countries; the information could be exploited to understand and study evolutionary divergence and migration of strains (Yang et al. 2009). The web server ATIVS (Analytical Tool for Influenza Virus Surveillance) provides an antigenic map for conducting surveillance and selection of vaccine strains by scrutinizing the serological data of haemagglutinin sequence data of influenza A/H3N2 viruses and influenza subtypes (Liao et al. 2009). There is another online repository OpenFluDB (an isolate-centred inventory), where information of an isolate such as virus type, host, date of isolation, geographical distribution, predicted antiviral resistance, enhanced pathogenicity or human adaptation propensity may be obtained (Liechti et al. 2010). For influenza viruses, primers and probes can be designed using the Influenza Primer Design Resource (IPDR) (Bose et al. 2008). Further, prospective influenza seasonal epidemics or pandemics can be predicted using a stochastic model, FluTE (Chao et al. 2010) (Table 23.2).

23.3.4 Virus Variation Resource (NCBI-VVR)

The NCBI Virus Variation Resource (NCBI-VVR) is a web-based database of a set of viruses, viz. influenza virus, dengue virus, rotavirus, West Nile virus, Ebola virus, Zika virus and MERS coronavirus (Resch et al. 2009). It enables the user to submit

Table 23.2 Virus-specific online databases/repositories

S. no.	Name of the databases/repositories	Available information
1.	VirOligo	Virus-specific oligonucleotides
2.	VIDA	Animal viruses comprising of the families of <i>Herpesviridae</i> , <i>Papillomaviridae</i> , <i>Poxviridae</i> , <i>Coronaviridae</i> and <i>Arteriviridae</i>
3.	VBRC	Large DNA viruses belonging to the families of <i>Arenaviridae</i> , <i>Bunyaviridae</i> , <i>Filoviridae</i> , <i>Flaviviridae</i> , <i>Paramyxoviridae</i> , <i>Poxviridae</i> , <i>Togaviridae</i> , etc.
4.	PBRC (now merged into VBRC)	<i>Poxviridae</i> family of viruses
5.	IVDB, IRD, NCBI-IVR, FluGenome, ISED, ATIVS, OpenFluDB, GISAIID	Influenza group of viruses
6.	NCBI-VVR	Influenza virus, dengue virus, rotavirus, West Nile virus, Ebola virus, Zika virus and Corona virus

their viral sequences along with relevant metadata such as sample collection time, isolation source, geographic location, host, disease severity, etc. It further allows integrating and analysing the viral sequences using the generic tools such as multiple sequence alignment and phylogenetic tree construction.

23.3.5 Web-Based Genotyping Tools

Rotavirus A (RVA) is the most frequent cause of severe diarrhoea in human and animal infants worldwide and remains as a major global threat for childhood morbidity and mortality (Minakshi et al. 2005; Basera et al. 2010). In recent years, extensive research efforts have been done for the development of live, orally administered vaccines. In India, an orally administered vaccine ROTAVAC was also introduced after successful clinical trials in 2014 which became available to clinicians in 2016, although these vaccines will have to be scrutinized and have to be updated regularly to accommodate the emerging rotavirus genotype variations, following which molecular and genetic characterization of new circulating and emerging genotypes of rotavirus strains in humans and animals becomes necessary. Recently, a classification system for RVAs has been described by the Rotavirus Classification Working Group (RCWG) in which all the 11 genomic RNA segments are assigned a particular alphabet followed by the particular genotype number. The classification system will be helpful in explaining the importance of genetic reassortments among RVAs, host range, transfer of gene segments among two different genotypes and adaptation to different hosts. To differentiate between different gene segments of RVAs, an online web-based tool RotaC was developed by the leading researchers from Rega Institute, KU Leuven, Belgium, in 2009 (Table 23.3). It's an easy-to-use

Table 23.3 A list of virus database and sequence-based typing tools that are used globally

S. no.	Typing tool	Viruses	References	Web link
1	<i>Rotac</i>	Group A rotaviruses	Maes et al. (2009)	http://rotac.regatools.be
2	LANL HCV database	HCV		http://hcv.lanl.gov
3	PaVE	Papillomavirus		http://pave.niaid.nih.gov
4	REGA HPV subtyping tool	Human papillomavirus (HPV)	Alcantara et al. (2009)	http://bioafrica.mrc.ac.za/rega-genotype/html/
5	REGA HIV subtyping tool	Human immunodeficiency virus 1 and 2		http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool
6	BioAfrica HHV8 automated tool	Human herpesvirus type-8	Alcantara et al. (2009)	http://bioafrica.mrc.ac.za/rega-genotype/html/
7	LASP HTLV-1 subtyping tool	Human T-lymphotropic virus type-1	Alcantara et al. (2009)	http://bioafrica.mrc.ac.za/rega-genotype/html
8	Oxford HBV subtyping tool	Hepatitis B virus	Alcantara et al. (2009)	http://bioafrica.mrc.ac.za/rega-genotype/html
9	Oxford HCV subtyping tool	Hepatitis C virus	Alcantara et al. (2009)	http://bioafrica.mrc.ac.za/rega-genotype/html
10	Dengue, Zika and chikungunya viruses typing tool	Dengue, Zika & Chikungunya Viruses		http://bioafrica.mrc.ac.za/rega-genotype/html
11	CoVDB	Coronavirus genes/genomes		http://covdb.microbiology.hku.hk
12	FLAVIdB Flavitrack	Flaviviruses		http://cvc.dfci.harvard.edu/flavi/ http://carnot.utmb.edu/flavitrack
13	FluGenome	Influenza A virus	Lu et al. (2007)	http://www.flugenome.org/
14	HERVd	Repository of human endogenous retroviruses		http://herv.img.cas.cz
15	HVDB	Hepatitis A, B, C, D and E virus sequences		http://s2as02.genes.nig.ac.jp
16	RetroTector	Retroviral sequences in vertebrate genomes		http://retrotector.neuro.uu.se/
17	SCUEAL	HIV-1 subtypes		http://www.datamonkey.org/dataupload_scueal.php

18	SeqHepB	Relational database for HBV			http://www.seqhepb.com
19	STAR	Subtyping tool for HIV-1 and HBV	Myers et al. (2005)		http://pgv19.virol.ucl.ac.uk/download/star_linux.tar
20	Norovirus typing tool – RIVM	Norovirus			www.rivm.nl/mpf/typingtool/norovirus/
21	Enterovirus typing tool – RIVM	Enterovirus			https://www.rivm.nl/mpf/typingtool/enterovirus/
22	Hepatitis E virus genotyping tool – RIVM	Hepatitis E virus			https://www.rivm.nl/mpf/typingtool/hev/
23	Yellow fever virus phylogenetic typing tool	Yellow fever virus			http://biofrica2.mrc.ac.za/rega-genotype/typingtool/yellowfevervirus/
24	FluTE	Influenza A virus	Chao et al. (2010)		http://www.cs.unm.edu/~dlchao/flute/
25	OpenFluDB	Influenza A virus	Liechti et al. (2010)		http://openflu.vital-it.ch

and reliable classification tool for RVAs and works on the agreement with RCWG. It's a platform-independent tool which works on any web browser by simply going to its URL (<http://rotac.regatools.be/>) and has been released without any restriction of use by academicians or anyone else. As claimed, the RotaC web-based tool will be updated regularly to reflect the established as well as newly emerging genotypes announced by the RCWG from time to time.

23.4 Conclusions and Future Prospects

Various researches in animal viral diseases are being conducted at the genomic level. Often, handling an enormous data obtained from sequencing is daunting to researchers. The chapter categorically provides a list of bioinformatics approaches that are useful in data mining. There are tables that list all such bioinformatics programs as per the applications. The tables also list databases that organize information on human and animal viruses such as genomic data, ORFs, oligonucleotides, etc. An illustration has also been provided in the chapter showing the application of the tool PredictProtein, which is used for prediction of three-dimensional structures of viral proteins. The major goal of the chapter has been to provide a roadmap to bioinformatics approaches in the field of animal viral diseases.

Although the chapter elaborates on viruses-specific bioinformatics programs, most of these programs are designed for human viruses. Nevertheless, there are bioinformatics tools that are animal-virus specific, but these are limited in number. Henceforth, in many cases, researchers have to switch to either human virus-specific tools or other generic tools. Application of such tools for studying animal viruses or animal diseases, in many situations, may not be as accurate as with specialized tools. The users should take precautions while using the settings of such tools. Furthermore, the results, thus obtained, also need to be scrutinized. Therefore, development of new bioinformatics programs/tools that are specifically designed for animal viruses/diseases should be taken up robustly. Specialized tools will provide much accurate results and predictions, thereby accelerating the bioinformatics researches in the field of animal viral diseases.

Acknowledgements All the authors of the manuscript thank and acknowledge their respective universities and institutes.

Conflict of Interest There is no conflict of interest.

References

- Alba 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(1):133–136
- 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(Web Server issue):W634–W642
- 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(22):2922–2929
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410
- 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
- Antoniewski C (2011) Visitor, an informatic pipeline for analysis of viral siRNA sequencing datasets. *Methods Mol Biol* 721:123–142
- Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Tatusova T (2007) FLAN: a web server for influenza virus genome annotation. *Nucleic Acids Res* 35(Web Server):W280–W284
- 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(2):596–601
- Barlow DJ, Edwards MS, Thornton JM (1986) Continuous and discontinuous protein antigenic determinants. *Nature* 322(6081):747–748
- Basera SS, Singh R, Vaid N, Sharma K, Chakravarti S, Malik YS (2010) Detection of rotavirus infection in bovine calves by RNA-PAGE and RT-PCR. *Indian J Virol* 21(2):144–147
- Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* 33(Web Server):W451–W454
- Bhasin M, Singh H, Raghava GP (2003) MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics* 19(5):665–666
- 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(1):23–31
- 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
- Brusic V, Rudy G, Harrison LC (1998) MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res* 26(1):368–371
- Carrillo-Tripp M, Shepherd CM, Borelli IA, Venkataraman S, Lander G, Natarajan P, Johnson JE, Brooks CL 3rd, Reddy VS (2009) VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res* 37(Database):D436–D442
- Chang S, Zhang J, Liao X, Zhu X, Wang D, Zhu J, Feng T, Zhu B, Gao GF, Wang J, Yang H, Yu J, Wang J (2007) Influenza Virus Database (IVDB): an integrated information resource and analysis platform for influenza virus research. *Nucleic Acids Res* 35(Database):D376–D380
- Chao DL, Halloran ME, Obenchain VJ, Longini IM Jr (2010) FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol* 6(1):e1000656
- 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(2):266–267
- Cheng CH, Liu SM, Chow TY, Hsiao YY, Wang DP, Huang JJ, Chen HH (2002) Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the Baculoviridae. *J Virol* 76(18):9024–9034
- Cheng X, Xiao X, Chou KC (2017) pLoc-mVirus: predict subcellular localization of multi-location virus proteins via incorporating the optimal GO information into general PseAAC. *Gene* 628:315–321
- Chevenet F, Jung M, Peeters M, de Oliveira T, Gascuel O (2013) Searching for virus phylotypes. *Bioinformatics* 29(5):561–570
- Clifford M, Twigg J, Upton C (2009) Evidence for a novel gene associated with human influenza A viruses. *Virol J* 6:198

- DeLisi C, Berzofsky JA (1985) T-cell antigenic sites tend to be amphipathic structures. *Proc Natl Acad Sci U S A* 82(20):7048–7052
- 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(17):2334–2336
- 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(8):2483–2492
- Enault F, Fremetz R, Baranowski E, Faraut T (2007) Alvira: comparative genomics of viral strains. *Bioinformatics* 23(16):2178–2179
- Felsenstein J (1989) Mathematics vs. evolution: mathematical evolutionary theory. *Science* 246(4932):941–942
- Ferron F, Rancurel C, Longhi S, Cambillau C, Henriessat B, Canard B (2005) VaZyMolO: a tool to define and classify modularity in viral proteins. *J Gen Virol* 86(Pt 3):743–749
- 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
- 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(5):720–722
- 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(Database):D339–D342
- 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(11):e79288
- Hsu PW, Lin LZ, Hsu SD, Hsu JB, Huang HD (2007) ViTa: prediction of host microRNAs targets on viruses. *Nucleic Acids Res* 35(Database):D381–D385
- 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
- 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(Database issue):D576–D582
- Li SC, Shiau CK, Lin WC (2008) Vir-Mir db: prediction of viral microRNA candidate hairpins. *Nucleic Acids Res* 36(Database issue):D184–D189
- 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(5):649–651
- Liao YC, Ko CY, Tsai MH, Lee MS, Hsiung CA (2009) ATIVS: analytical tool for influenza virus surveillance. *Nucleic Acids Res* 37(Web Server issue):W643–W646
- 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
- Lorenzi HA, Hoover J, Inman J, Safford T, Murphy S, Kagan L, Williamson SJ (2011) The viral Meta genome annotation pipeline (VMGAP): an automated tool for the functional annotation of viral metagenomic shotgun sequencing data. *Stand Genomic Sci* 4(3):418–429
- Lu G, Rowley T, Garten R, Donis RO (2007) FluGenome: a web tool for genotyping influenza A virus. *Nucleic Acids Res* 35(Web Server):W275–W279
- Maes P, Matthijssens J, Rahman M, Van Ranst M (2009) Rota C: a web-based tool for the complete genome classification of group A rotaviruses. *BMC Microbiol* 9:238
- Minakshi PG, Malik YS, Pandey R (2005) G and P genotyping of bovine group A rotaviruses in faecal samples of diarrheic calves by DIG-labeled probes. *Indian J Biotechnol* 4:93–99
- 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(17):3535–3540
- Nagarajan N, Kingsford C (2011) GiRaF: robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Res* 39(6):e34
- 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(Web Server):W448–W450

- Onodera K, Melcher U (2002) VirOligo: a database of virus-specific oligonucleotides. *Nucleic Acids Res* 30(1):203–204
- 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(Database issue):D569–D575
- 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(Database issue):D371–D378
- Petitjean M, Vanet A (2014) VIRAPOPS: a forward simulator dedicated to rapidly evolved viral populations. *Bioinformatics* 30(4):578–580
- Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50(3–4):213–219
- Reche PA, Zhang H, Glutting JP, Reinherz EL (2005) EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology. *Bioinformatics* 21(9):2140–2141
- Reddy VS, Natarajan P, Okerberg B, Li K, Damodaran KV, Morton RT, Brooks CL 3rd, Johnson JE (2001) Virus particle explorer (VIPER), a website for virus capsid structures and their computational analyses. *J Virol* 75(24):11943–11947
- Reeder J, Steffen P, Giegerich R (2007) pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic Acids Res* 35(Web Server):W320–W324
- 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
- Rombel IT, Sykes KF, Rayner S, Johnston SA (2002) ORF-FINDER: a vector for high-throughput gene identification. *Gene* 282(1–2):33–41
- Rost B, Yachdav G, Liu J (2004) The PredictProtein server. *Nucleic Acids Res* 32(Web Server):W321–W326
- 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(2):e11
- Roux S, Faubladiet M, Mahul A, Paulhe N, Bernard A, Debros D, Enault F (2011) Metavir: a web server dedicated to virome analysis. *Bioinformatics* 27(21):3074–3075
- Rozanov M, Plikat U, Chappey C, Kochergin A, Tatusova T (2004) A web-based genotyping resource for viral sequences. *Nucleic Acids Res* 32(Web Server):W654–W659
- Saha S, Bhasin M, Raghava GP (2005) Bcipep: a database of B-cell epitopes. *BMC Genomics* 6:79
- Saminathan M, Rana R, Ramakrishnan MA, Karthik K, Malik YS, Dhama K (2016) Prevalence, diagnosis, management and control of important diseases of ruminants with special reference to Indian scenario. *J Exp Biol Agric Sci* 4(3S):3338–3367. [https://doi.org/10.18006/2016.4\(3s\).338.367](https://doi.org/10.18006/2016.4(3s).338.367)
- Schlessinger A, Ofra Y, Yachdav G, Rost B (2006) Epitome: database of structure-inferred antigenic epitopes. *Nucleic Acids Res* 34(Database issue):D777–D780
- Schombach C, Koh JL, Flower DR, Brusci V (2005) An update on the functional molecular immunology (FIMM) database. *Appl Bioinforma* 4(1):25–31
- 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(Web Server):W647–W651
- 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(3):233–240
- Simmonds P (2012) SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes* 5:50
- 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(5):1031–1041

- Singh RK, Dhama K, Karthik K, Tiwari R, Khandia R, Munjal A, Iqbal HM, Malik YS, Bueno-Marí R (2017a) Advances in diagnosis, surveillance, and monitoring of Zika virus: an update. *Front Microbiol* 8:2677. <https://doi.org/10.3389/fmicb.2017.02677>
- Singh RK, Dhama K, Malik YS, Ramakrishnan MA, Karthik K, Tiwari R, Khandia R, Munjal A, Saminathan M, Sachan S, Desingu PA, Kattoor JJ, Iqbal HMN, Joshi SK (2017b) Ebola virus – epidemiology, diagnosis and control: threat to humans, lessons learnt and preparedness plans- an update on its 40 year’s journey. *Vet Q* 37(1):98–135. <https://doi.org/10.1080/01652176.2017.1309474>
- 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(Database issue):D497–D503
- 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(6):404–416
- Surendranath V, Chusainov 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(Web Server issue):W293–W298
- 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
- Thakur N, Qureshi A, Kumar M (2012) VIRsiRNAdb: a curated database of experimentally validated viral siRNA/shRNA. *Nucleic Acids Res* 40(Database issue):D230–D236
- Toseland CP, Clayton DJ, McSparron H, Hemsley SL, Blythe MJ, Paine K, Doytchinova IA, Guan P, Hattotuwigama CK, Flower DR (2005) AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res* 1(1):4
- Tyagi A, Ahmed F, Thakur N, Sharma A, Raghava GP, Kumar M (2011) HIVsirDB: a database of HIV inhibiting siRNAs. *PLoS One* 6(10):e25917
- Upton C, Hogg D, Perrin D, Boone M, Harris NL (2000) Viral genome organizer: a system for analyzing complete viral genomes. *Virus Res* 70(1–2):55–64
- Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B (2015) The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 43(Database issue):D405–D412
- 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(21):11016–11021
- Wang S, Sundaram JP, Spiro D (2010) VIGOR, an annotation program for small viral genomes. *BMC Bioinformatics* 11:451
- 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(5):e64465
- Watson M, Schnettler E, Kohl A (2013) viRome: an R package for the visualization and analysis of viral small RNA sequence datasets. *Bioinformatics* 29(15):1902–1903
- 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(3):427–439
- Xiao X, Wu ZC, Chou KC (2011) iLoc-Virus: a multi-label learning classifier for identifying the subcellular localization of virus proteins with both single and multiple sites. *J Theor Biol* 284(1):42–51
- Yang IS, Lee JY, Lee JS, Mitchell WP, Oh HB, Kang C, Kim KH (2009) Influenza sequence and epitope database. *Nucleic Acids Res* 37(Database):D423–D430

- 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(2):266–267
- 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
- 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(10):e78470