Published online 28 November 2012 *Nucleic Acids Research, 2013, Vol. 41, Database issue* ***D579–D583***
doi:10.1093/nar/gks1220

ViralZone: recent updates to the virus knowledge resource

Patrick Masson[1], Chantal Hulo[1], Edouard De Castro[1], Hans Bitter[2], Lore Gruenbaum[2], Laurent Essioux[3], Lydie Bougueleret[1], Ioannis Xenarios[1,4] and Philippe Le Mercier[1,*]

[1]SIB Swiss Institute of Bioinformatics, Swiss-Prot Group, Centre Médical Universitaire, CH-1211 Geneva 4, Switzerland, [2]Translational Research Sciences, Hoffmann-La Roche Drug Discovery & Early Development, 340 Kingsland Street, Nutley, NJ 07110NJ, USA, [3]Hoffmann La Roche, Bioinformatics and Exploratory Statistics Department, pRED, Grenzacherstrasse 124, 4070 Basel, Switzerland and [4]SIB Swiss Institute of Bioinformatics, Vital-IT Group, Quartier Sorge—Bâtiment Génopode, 1015, Lausanne, Switzerland

Received September 15, 2012; Revised October 26, 2012; Accepted October 31, 2012

ABSTRACT

ViralZone (http://viralzone.expasy.org) is a knowledge repository that allows users to learn about viruses including their virion structure, replication cycle and host–virus interactions. The information is divided into viral fact sheets that describe virion shape, molecular biology and epidemiology for each viral genus, with links to the corresponding annotated proteomes of UniProtKB. Each viral genus page contains detailed illustrations, text and PubMed references. This new update provides a linked view of viral molecular biology through 133 new viral ontology pages that describe common steps of viral replication cycles shared by several viral genera. This viral cell-cycle ontology is also represented in UniProtKB in the form of annotated keywords. In this way, users can navigate from the description of a replication-cycle event, to the viral genus concerned, and the associated UniProtKB protein records.

INTRODUCTION

The ViralZone database (http://viralzone.expasy.org) is an online resource that brings together viral molecular biology knowledge with viral genomic and protein sequences (1). ViralZone was created in 2009 and is updated regularly on a bi-monthly basis. The resource contains two main types of information: virus description pages and lists of relevant UniProtKB proteins (which are generated automatically for each virus). The core data in ViralZone are the virus description pages, which provide information on all viral genera referenced by the International Committee for Taxonomy of Viruses (2). Curators combine data from recent publications and textbook knowledge to create the tables, pictures, textual annotations and links to original publications that are found in each virus page. These provide an accessible summary of the available information on a viral genus, including illustrations of the virion and genome schematics, descriptions of the replication cycle, links to many databases (3–8), epidemiology data and lists of manually annotated proteins in UniProtKB (4). Viral description pages are virus-centric and describe the processes and biology that are relevant to each viral genus. To complement these descriptions we have now added another layer of information to ViralZone in the form of a viral ontology. This describes common replication steps or characteristics that are shared between multiple viral genera and is organized in the form of 133 ontology pages. The ontology is used to link common processes in the viral description pages—each of these linking back to the ontology pages.

NEW VIRAL ONTOLOGY COVERING VIRUS-SPECIFIC MOLECULAR PROCESSES

Viruses use a variety of unique molecular mechanisms during replication in hosts (9). These often circumvent or exploit cellular processes, and their study affords a greater understanding of the cellular functions concerned. Viral mechanisms are also widely exploited as tools for biological research and biotechnology; examples include the reverse transcriptase (10) and T7 RNA polymerase (11) enzymes, internal ribosome entry site (12) and lentiviral vectors (13). Most of these replication mechanisms are described in ViralZone fact sheets for the viral genus that uses them. However, these are designed to

*To whom correspondence should be addressed. Tel: +41 22 379 58 70; Fax: +41 22 379 58 58; Email: philippe.lemercier@isb-sib.ch

provide a short overview of the biology of a virus and do not contain detailed explanations of the molecular events that occur. Moreover, information disseminated in fact sheets is not easily extracted and does not offer a means to group viruses sharing a common process. For example, all viruses using ribosomal read-through (14) are annotated as such, but there is no way to list them all in ViralZone.

To address this need we have created a new section describing viral molecular biology. The information is structured with a vocabulary that is used both in virus fact sheets and molecular mechanisms pages, and represents a basis to develop virus ontology. The long-term goal is to link ViralZone page, UniProt Keywords and Gene Ontology terms. The concept of a central ontology was chosen because it has proven to be efficient for managing large data sets and analysis generated by transcriptomic and proteomic studies (15). In ViralZone, 133 new pages describe the viral ontology. The ontology is divided in five parts that describe the main steps in the viral life cycle: 18 pages linked to viral entry (Figure 1), 29 pages linked to viral replication, 13 pages linked to viral exit,

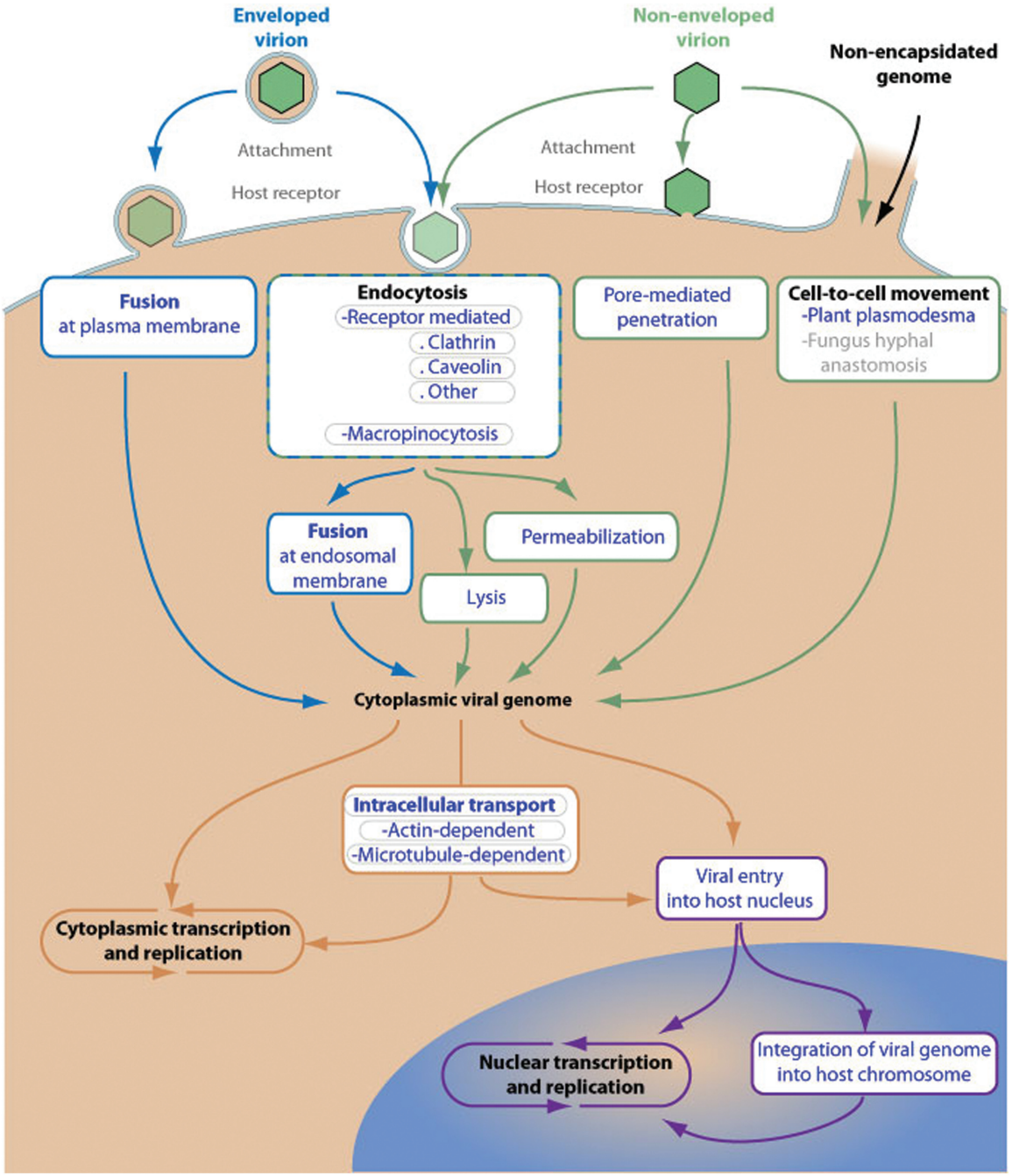

Figure 1. Virus entry ontology. This picture is effectively a graphical menu that provides links to 18 pages describing possible mechanisms of viral entry into host cells and events occurring subsequent to host-cell entry.

11 pages linked to the virion structure and 62 pages linked to host–virus interactions. Each of these pages contains a description of the viral process associated with the term, a picture describing the molecular events and pathways, the list of associated viruses and links to original publications. They provide an overview of viruses using a common mechanism and improve the level of detail in virus fact sheets. For instance, the hepatitis B virus (HBV) fact sheet did not specify in what way the viral genome exits the cell nucleus: through nuclear lysis, nuclear egress or nuclear pore. Using the viral ontology, we have now classified the HBV in the nuclear pore export pathway (16). This information and a link have been added to the HBV fact sheet.

ViralZone is a web resource that brings together scientific knowledge and sequence data. This same principle applies to the virus ontology: the same terms described in ViralZone pages are used to annotate UniProtKB virus

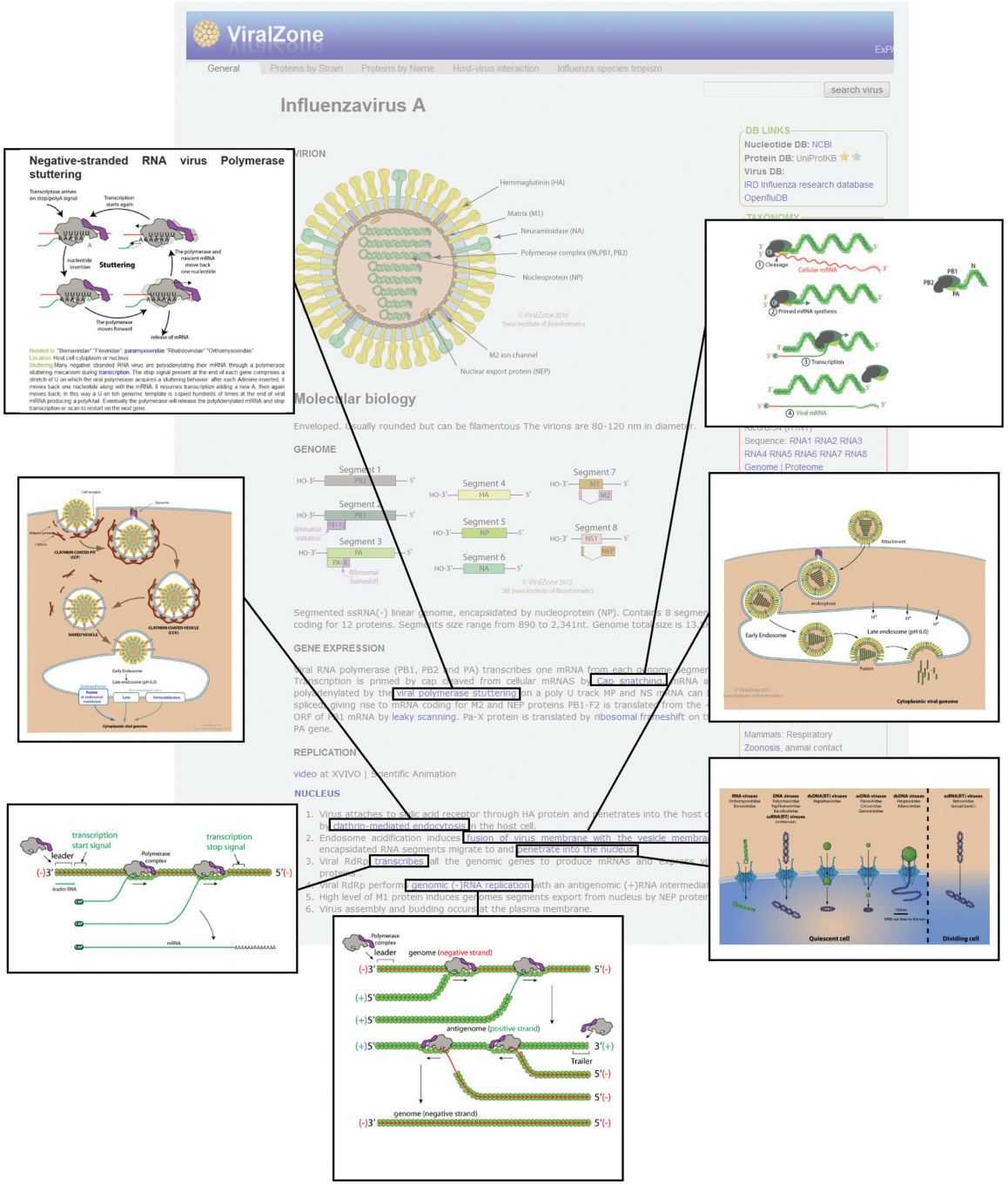

Figure 2. Improved fact sheets. The background image represents the influenza A fact sheet. Links to the ontology section are illustrated by boxes that depict each process. Each step of the viral cycle is explained in detail and users can return to the influenza A page from any of the ontology pages it appears in.

entries. The 120 viral terms created for the ontology correspond to 120 new UniProt keywords that have been assigned to relevant viral entries. For example, the term 'Inhibition of host STAT1 by virus' (17) is linked to 1024 viral protein entries in UniProt release 2012_07.

EXPANDED VIRUS FACT SHEETS

ViralZone virus fact sheets are virus-centric and display specific information on molecular virology, taxonomy, hosts and epidemiology. 'Gene expression' and 'viral replication cycle' sections describe every specific process used by the virus to enter, replicate and exit the host cell. These sections have been updated to link directly to the ontology pages (Figure 2). The advantages of these new links are multiple: first, the amount of available information is increased without overloading the fact sheet page that remains short and concise; second, users can have in-depth insight of the virus molecular biology put into the broader context of all viruses sharing the same feature.

For example, description of the simplex herpes virus replication cycle now contains a total of nine processes that are clickable and lead to specific pages explaining these steps in detail. Among these, the page 'viral penetration into host nucleus' (18) depicts the various mechanisms by which viral genomes enter the host nuclear pores, with a short explanatory text and a schema representing the different strategies employed by the virus. In addition, all viruses penetrating the host nucleus are listed on the right part of the page with a link to original publications.

NOVEL HOST–VIRUS INTERACTIONS TAB

Viruses are obligate parasites and have evolved to have many ways of interacting with or hijacking host cell mechanisms. Most of these interactions prevent antiviral cell defense or facilitate viral replication and transmission (19). These interactions are complex because they involve both viral and cellular pathways. Often viruses interact with key cellular regulatory elements, resulting in many complex phenotypes. That is why among the 133 ontology pages in ViralZone, 62 are devoted to host–virus interactions. In virus fact sheets, a new tab termed 'Host–virus interaction' has been created. This new section briefly describes in a few sentences the molecular mechanism by which specific viral proteins interfere with key cellular pathways, including hijacking of the host immune response, perturbation of the cell cycle, apoptosis or autophagy. Key host–virus interactions are clickable and link to the corresponding ontology pages for further details.

Although only around a dozen cellular pathways are commonly targeted by viruses, each of these pathways can be modulated in a variety of ways by different viruses. In ViralZone, the 'Host–virus interaction' tab provides information organized according to the main pathways that are most commonly hijacked. Viral and host proteins involved are described and commented in details with links to publications. As an example, many vertebrate viruses contain information relating to the 'Innate immune response inhibition' in their 'Host–virus interaction' tab. This section is also linked to the viral ontology pages, therefore users can switch between views that are specific to a given virus or mechanism.

NEW HBV INTERACTIVE REPLICATION CYCLE

Hepatitis B is one of the most common infectious diseases in the world. It has been estimated that 350 million people worldwide are chronic HBV carriers (20). The virus genome is only 3.2 kb long and encodes seven proteins (21). Despite the apparent simplicity of HBV, the life cycle of this virus is complex. In partnership with Hoffmann-La Roche, an interactive HBV life cycle resource has been created in ViralZone with 28 new pages linking to 180 publications from PubMed. The entry point to the HBV resource is an illustration depicting the virus replication cycle in a hepatocyte host. The cycle has been divided into 27 steps and molecular events that are each clickable and link to HBV specific description pages. Most of these steps correspond to viral ontology terms described earlier. The replication cycle pages describe the current knowledge on the topic with all major publications and some comments. Many HBV pages are linked to the ViralZone ontology section in such a way that users have access to knowledge that is specific to HBV or that concerns shared viral mechanisms.

FUNDING

This activity of the Swiss-Prot group is supported by the Swiss Federal Government through the Federal Office of Education and Science. The HBV replication cycle has been funded by Hoffmann-La Roche. Funding for open access charge: SIB Swiss Institute of Bioinformatics.

Conflict of interest statement. None declared.

REFERENCES

1. Hulo,C., de Castro,E., Masson,P., Bougueleret,L., Bairoch,A., Xenarios,I. and Le Mercier,P. (2011) ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.*, **39**, D576–D582.
2. Pringle,C.R. (1998) The universal system of virus taxonomy of the International Committee on Virus Taxonomy (ICTV), including new proposals ratified since publication of the Sixth ICTV Report in 1995. *Arch. Virol.*, **143**, 203–210.
3. Pruitt,K.D., Tatusova,T., Klimke,W. and Maglott,D.R. (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.*, **37**, D32–D36.
4. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.
5. Adams,M.J. and Antoniw,J.F. (2006) DPVweb: a comprehensive database of plant and fungal virus genes and genomes. *Nucleic Acids Res.*, **34**, D382–D385.
6. Liechti,R., Gleizes,A., Kuznetsov,D., Bougueleret,L., Le Mercier,P., Bairoch,A. and Xenarios,I. (2010) OpenFluDB, a database for human and animal influenza virus. *Database*, **2010**, baq004.
7. Squires,R.B., Noronha,J., Hunt,V., García-Sastre,A., Macken,C., Baumgarth,N., Suarez,D., Pickett,B.E., Zhang,Y., Larsen,C.N. *et al.* (2012) Influenza Research Database: an integrated bioinformatics resource for influenza research and surveillance. *Influenza Other Respi. Viruses*, **6**, 404–416.
8. Carrillo-Tripp,M., Shepherd,C.M., Borelli,I.A., Venkataraman,S., Lander,G., Natarajan,P., Johnson,J.E., Brooks,C.L. 3rd and

Reddy,V.S. (2009) VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res.*, **37**, D436–D442.
9. Forterre,P. (2006) The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.*, **117**, 5–16.
10. Buell,G.N., Wickens,M.P., Payvar,F. and Schimke,R.T. (1978) Synthesis of full length cDNAs from four partially purified oviduct mRNAs. *J. Biol. Chem.*, **253**, 2471–2482.
11. Sousa,R. and Mukherjee,S. (2003) T7 RNA polymerase. *Prog. Nucleic Acid Res. Mol. Biol.*, **73**, 1–41.
12. Kieft,J.S. (2008) Viral IRES RNA structures and ribosome interactions. *Trends Biochem. Sci.*, **33**, 274–283.
13. He,Y. and Falo,L.D. Jr (2007) Lentivirus as a potent and mechanistically distinct vector for genetic immunization. *Curr. Opin. Mol. Ther.*, **9**, 439–446.
14. Weiss,R.B. (1991) Ribosomal frameshifting, jumping and readthrough. *Curr. Opin. Cell Biol.*, **3**, 1051–1055.
15. Bechhofer,S.K., Stevens,R.D. and Lord,P.W. (2005) Ontology driven dynamic linking of biology resources. *Pac. Symp. Biocomput.*, **2005**, 79–90.
16. Li,H.-C., Huang,E.-Y., Su,P.-Y., Wu,S.-Y., Yang,C.-C., Lin,Y.-S., Chang,W.-C. and Shih,C. (2010) Nuclear export and import of human hepatitis B virus capsid protein and particles. *PLoS Pathog.*, **6**, e1001162.
17. Najjar,I. and Fagard,R. (2010) STAT1 and pathogens, not a friendly relationship. *Biochimie*, **92**, 425–444.
18. Peng,L., Ryazantsev,S., Sun,R. and Zhou,Z.H. (2010) Three-dimensional visualization of gammaherpesvirus life cycle in host cells by electron tomography. *Structure*, **18**, 47–58.
19. Faure,M. and Rabourdin-Combe,C. (2011) Innate immunity modulation in virus entry. *Curr. Opin. Virol.*, **1**, 6–12.
20. Dandri,M. and Locarnini,S. (2012) New insight in the pathobiology of hepatitis B virus infection. *Gut*, **61(Suppl. 1)**, i6–17.
21. Urban,S., Schulze,A., Dandri,M. and Petersen,J. (2010) The replication cycle of hepatitis B virus. *J. Hepatol.*, **52**, 282–284.