

RESEARCH MATTERS

Virologists—Heroes need weapons

Franziska Hufsky^{1,2}, Bashar Ibrahim^{1,2}, Martin Beer^{1,3}, Li Deng^{1,4}, Philippe Le Mercier^{1,5}, Dino P. McMahon^{1,6,7}, Massimo Palmarini^{1,8}, Volker Thiel^{1,9,10}, Manja Marz^{1,2*}

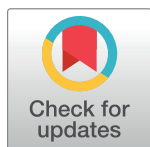
1 European Virus Bioinformatics Center, Jena, Germany, **2** RNA Bioinformatics and High-Throughput Analysis Jena, Friedrich Schiller University Jena, Jena, Germany, **3** Institute of Diagnostic Virology, Friedrich-Loeffler-Institute, Greifswald, Germany, **4** Institute of Virology, Helmholtz Zentrum Munich, Munich, Germany, **5** Swiss-Prot group, SIB, CMU, University of Geneva Medical School, Geneva, Switzerland, **6** Host parasite evolution and ecology, Institute of Biology, Free University of Berlin, Berlin, Germany, **7** Department for Materials and Environment, BAM, Federal Institute for Materials Research and Testing, Berlin, Germany, **8** MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom, **9** Federal Department of Home Affairs, Institute of Virology and Immunology, Bern and Mittelhäusern, Switzerland, **10** Department of Infectious Diseases and Pathobiology, University of Bern, Bern, Switzerland

* evbc@uni-jena.de

Virologists. You might know a couple of them, but unless you are a virologist yourself, the probability that you have collaborated with one in the past is low. The community is relatively small, but they pack a heavy punch and are expected to play a leading role in the research into pathogens that lies ahead. You may ask why we think virologists are our future. Suffice it to say that it is not just because they have invented technologies that belong to the space age, including use of viruses as vehicles to shuttle genes into cells[1], organic nanoparticles with specific tools attached to their surfaces to get inside target cells[2], and using genetically modified viruses as therapies to fight against cancer[3]. Did you know that virologists currently only know of about 3,200 viral species but that more than 320,000 mammal-associated viruses[4] are thought to await discovery? Just think about the viruses hidden in the Arctic ice[5] or in the insects and other animals from once cut-off regions in the world, which now face ever-increasing human exposure[6]. But a heroic (as well as an apocalyptic) role for virologists may also be on the horizon, as the adoption of phage therapy may, in the future, be used to control harmful bacteria when antibiotics fail[7].

Bioinformaticians. You may know a couple of them, but usually you “just” need one to do proper math for your -omics experiments. Nowadays, nearly everyone in the life sciences has used BLAST [8] at least once, or made an alignment, or asked a bioinformatician to analyze high-throughput sequencing data. Of course, bioinformaticians do more than straightforward tasks. Software is based on algorithms and on data structures. Biological data are intrinsically complicated, and it is demanding to find appropriate data structures to process them efficiently[9]. Bioinformaticians routinely have to develop tailored, study-specific algorithms and tools used by a wide variety of scientists, including biochemists, biologists, geneticists, and molecular life scientists; but we rarely find virus-specific tools used by virologists. Why is this?

It is possibly not the problem of virologists but of bioinformaticians themselves. Although it was a virus that was the first organism to have its genome completely sequenced[10], bioinformaticians quickly focused attention on larger organisms, including humans, mice, plants, fungi, and bacteria. Perhaps there was no time left to take care of viruses? It also may be that, to a bioinformatician, a virus may appear uninteresting, at least at first. For example, the most dangerous human pathogenic (mostly RNA) viruses are short, usually contain only a single-digit number of genes that lack introns, contain only mononucleotide or dinucleotide repeats, and very few regulatory elements[11]. But astonishingly, we now know that the human



OPEN ACCESS

Citation: Hufsky F, Ibrahim B, Beer M, Deng L, Mercier PL, McMahon DP, et al. (2018) Virologists—Heroes need weapons. *PLoS Pathog* 14(2): e1006771. <https://doi.org/10.1371/journal.ppat.1006771>

Editor: Glenn F Rall, The Fox Chase Cancer Center, UNITED STATES

Published: February 8, 2018

Copyright: © 2018 Hufsky et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors received no specific funding for this study.

Competing interests: The authors have declared that no competing interests exist.

genome consists of 8%–60% virus-derived sequences (depending on how this is measured: 8% can be directly traced back to viruses, whereas a figure of 60% includes LINEs and SINEs that are thought to be of viral origin[12]). Viruses have therefore played a very large part in shaping the evolution of the human genome as well as the genomes of organisms from all domains of life. Furthermore, we can observe evolution in viruses in realtime thanks to their extraordinary mutation rates, rapid replication cycles, large population sizes, and immense recombination potential[13]; this means that we are able to observe molecular evolution in a matter of days! Can you imagine that nearly all existing bioinformatical tools are not specifically designed for the context of viruses? We argue that without bioinformaticians, virologists are fighting with one arm tied behind their backs.

Clearly virologists are well positioned to tackle some of the major disease threats that will inevitably face humans in the decades to come—but they are currently not the best equipped. In order to give virologists weapons, i.e., bioinformatic tools, we recently founded the European Virus Bioinformatics Center (EVBC).

As alluded to above, this Center was developed based on our belief that we lack bioinformatical tools specifically designed for virology. We urgently need to develop algorithms for multiple genome alignments. It sounds trivial, but we still have no tool that can align, for example, 500,000 full-length coronaviruses (approximately 32,000 nt). We cannot visualize such alignments that are clearly needed to rapidly identify mutational hotspots and compensatory mutations.

Can you believe that there is not even a unified database for viruses? We only have good databases for Influenza (EpiFlu [14]), HIV [15], and human pathogenic viruses (ViPR [16]). There is the possibility to archive viral sequences in NCBI, but virologists use this facility only rarely because of its limited utility. For instance, submitters are asked to commit the name of a chromosome, which clearly is not applicable to viruses.

These examples emphasize that we need to bring virologists and bioinformaticians together. Again, this seems obvious and, perhaps, trivial. But in reality, they speak different languages. We urgently need young scientists that understand both virology and bioinformatics so that they can bridge the gap between these two complementary yet different disciplines.

The EVBC aims to develop bioinformatical tools for nearly all areas: (1) for detection of viruses, e.g., from high-throughput sequencing data; (2) virus assembly; (3) quasispecies reconstruction; (4) intraviral interactions; (5) virus entry, i.e., protein—protein interaction; (6) virus—host interactions; (7) phylogeny/cophylogeny; and (8) therapy. For the latter two areas, we have already established tools for HIV and Influenza A, but a large number of outstanding research questions remain. Finally, the EVBC will initiate and coordinate ring trials, undergraduate courses, graduate summer schools, and courses for principal investigators.

Interested? Take a look here: evbc.uni-jena.de.

References

1. Kay MA, Glorioso JC, Naldini L. Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nature medicine*. 2001; 7:33–40. <https://doi.org/10.1038/83324> PMID: 11135613
2. Wang EC, Wang AZ. Nanoparticles and their applications in cell and molecular biology. *Integrative biology: quantitative biosciences from nano to macro*. 2014; 6:9–26. <https://doi.org/10.1039/c3ib40165k> PMID: 24104563
3. Davis JJ, Fang B. Oncolytic virotherapy for cancer treatment: challenges and solutions. *The journal of gene medicine*. 2005; 7:1380–1389. <https://doi.org/10.1002/jgm.800> PMID: 16025557
4. Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrel CM, Solovyov A, et al. A strategy to estimate unknown viral diversity in mammals. *mBio*. 2013; 4:e00598–e00513. <https://doi.org/10.1128/mBio.00598-13> PMID: 24003179

5. Aguirre de Cárcer D, López-Bueno A, Pearce DA, Alcamí A. Biodiversity and distribution of polar fresh-water DNA viruses. *Science advances*. 2015; 1:e1400127. <https://doi.org/10.1126/sciadv.1400127> PMID: 26601189
6. Tatem AJ, Rogers DJ, Hay SI. Global transport networks and infectious disease spread. *Advances in parasitology*. 2006; 62:293–343. [https://doi.org/10.1016/S0065-308X\(05\)62009-X](https://doi.org/10.1016/S0065-308X(05)62009-X) PMID: 16647974
7. Gallois M, Rothkötter HJ, Bailey M, Stokes CR, Oswald IP. Natural alternatives to in-feed antibiotics in pig production: can immunomodulators play a role? *Animal: an international journal of animal bioscience*. 2009; 3:1644–1661. <https://doi.org/10.1017/S1751731109004236> PMID: 22443549
8. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25:3389–3402. PMID: 9254694
9. Li Y, Chen L. Big biological data: challenges and opportunities. *Genomics, proteomics & bioinformatics*. 2014; 12(5):187–189.
10. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*. 1977; 265:687–695. PMID: 870828
11. Marz M, Beerenwinkel N, Drosten C, Fricke M, Frishman D, Hofacker IL, et al. Challenges in RNA virus bioinformatics. *Bioinformatics*. 2014; <https://doi.org/10.1093/bioinformatics/btu105> PMID: 24590443
12. Griffiths DJ. Endogenous retroviruses in the human genome sequence. *Genome biology*. 2001; 2:REVIEWS1017.
13. Brooks T, Roy-Burman A, Tuholske C, Busch MP, Bakkour S, Stone M, et al. Real-Time Evolution of Zika Virus Disease Outbreak, Roatán, Honduras. *Emerging infectious diseases*. 2017; 23:1360–1363. <https://doi.org/10.3201/eid2308.161944> PMID: 28514227
14. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Euro surveillance: bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*. 2017; 22. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494> PMID: 28382917
15. Druce M, Hulo C, Masson P, Sommer P, Xenarios I, Le Mercier P, et al. Improving HIV proteome annotation: new features of BioAfrica HIV Proteomics Resource. *Database: the journal of biological databases and curation*. 2016;2016. <https://doi.org/10.1093/database/baw045> PMID: 27087306
16. Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, et al. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res*. 2012; 40(Database issue): D593–D598. <https://doi.org/10.1093/nar/gkr859> PMID: 22006842