

Viruses at the source

Arnthor Ævarsson, coordinator of project Virus-X, Matis, Reykjavik, Iceland

The European project Virus-X explores the genomes of viruses contained in extreme environments: geothermal hot springs in Iceland, deep seabed of the North Atlantic... The consortium members are now trying to identify the enzymes that can constitute new or better tools for molecular biology applications.

The estimated number of viruses inhabiting the Earth by far outnumber the estimated number of stars in the observable Universe⁽¹⁾. Every type of organism, unicellular or multicellular, may potentially be infected by viruses. Even in the most extreme environments such as in geothermal hot springs in Iceland, at the boiling point of water, you find viruses that propagate by infecting the heat-loving bacteria that thrive under such conditions. Viruses carry their own genomes harboring genes required for their successful propagation through infection of their host cell. The massive efforts in DNA sequencing of the genomes of various organisms over the past decades still leaves the genomes of viruses as the largest uncharacterized genomic territory still to be explored⁽¹⁾. Although the function of a gene product (a protein) can often be deduced by the sequence

of the corresponding gene, the function of proteins encoded by viral genomes is to a large extent unknown. Given the fact that many proteins with special functions (enzymes) originating from viruses have found important use as tools in molecular biology applications, such as for manipulation and sequencing of DNA (e.g. DNA polymerases), it is clear that further exploration of the genomes of viruses in nature may be of great reward in the discovery of new and improved enzymes of commercial value and benefits for society.

The Virus-X project, initiated in 2015 by Icelandic scientists and funded by the European research and innovation program Horizon 2020, is a scientific expedition to explore and exploit uncharted territory of genetic diversity in nature. Specifically, the project aims to explore the genomes of viruses in geothermal hot springs and ocean ecosystems. The Virus-X Consortium, that was formed to



PH.D. IN MOLECULAR BIOPHYSICS

Arnthor Ævarsson worked in several companies before becoming independent. He is now conducting his research for the Icelandic company Matis.

FURTHER INFORMATION

■ www.virus-x.eu
The Virus-X project website

carry out the ambitious multidisciplinary research plan over four years, includes seven universities, two research institutes and five companies from eight European countries (www.virus-x.eu).

No cultivation

The core of Virus-X is a bio-discovery pipeline that starts with environmental sampling in selected habitats. Scientists at the research company Matis in Iceland lead the sampling of various geothermal sites in Iceland, including the famous Geysir geothermal area harboring some of the extremely rare erupting hot springs (geysers). A research unit at The Institute Pasteur in France, specializing in viruses that infect a group of bacteria-like microorganisms called Archaea that live in extreme environments, further extends the geothermal sampling at sites in the western part of Georgia and in Beppu on Kyushu island in Japan⁽²⁾. A third group at the University of Bergen in Norway, focuses on ocean sampling in the Northern Atlantic and the Arctic Ocean including the Jan Mayen Fracture Zone and unique ecosystems of deep-sea hydrothermal vents.

Context

Viruses are poorly known entities in our environment. Yet, we find them everywhere. New techniques for analyzing environmental DNA offer hope for a better understanding of these ubiquitous entities.

A hallmark of Virus-X is a research approach termed metagenomics that bypasses the conventional isolation and cultivation of organisms in a laboratory and rather involves extracting genetic material directly from environmental samples which is then sequenced by the latest generation of high technology sequencing machines to obtain multiple genomic sequences from the mixture of organisms in the sample (i.e. a metagenome). The processing of the wealth of information obtained in the project this way, in the form of billions of the letters A, T, G and C in the DNA sequences, is a discipline in itself called bioinformatics and is a fundamental expertise in the project provided by several expert groups including the Max-Planck Institute and Bielefeld University in Germany, Université Clermont Auvergne in France and Bio-Product, a bioinformatics company in The Netherlands. The computational bioinformatics algorithms have identified over 20 millions genes in the genetic data obtained so far and to some extent predicted the function of the corresponding gene products.

Enzymes that can become new and improved tools for molecular biology applications, such as enzymes that can cut, ligate or amplify DNA (i.e. nucleases, DNA ligases and DNA polymerases, respectively), are among enzymes of special interest in the project. Thousands of genes coding for enzymes in this category have been identified from sequence information and some of these genes selected for cloning into gene vehicles (vectors) and subsequent introduction into specific microbial production hosts that allow the desired enzymes to be produced for further studies. The cloning efforts are led by a group at the University of Stuttgart in Germany but due to the great scale of the project require the participation of several partners in the project as well as the protein

production efforts led by Lund University in Sweden.

Enzyme functions

The biochemical characterization, led by University of Gdansk in Poland, involves measuring some functional characteristics of the enzymes such as activity at different temperatures and pH and substrate specificity. This information gives indications of the potential of using these enzymes for new or improved applications and the potential of commercial value for the participating enzyme developers (Arctic-Zymes in Norway, Matis in Iceland and A&A Biotechnology in Poland). The societal impact of Virus-X is imperative.

Although the function of some of the identified genes can be predicted from sequence information through similarity to genes with defined function, an important part of Virus-X is to develop a pipeline to determine the function of the discovered, novel proteins, in particular for those whose function cannot be predicted reliably from sequence. These genes are abundant in viral genomes and often constitute the majority of the genes in a given viral genome.



▲ Viruses survive in hot springs up to 100° C. Here in Seltún, Iceland.



▲ The EuroScience Open Forum (Esof), in which Arnthor Ævarsson participated, took place from 9 to 14 July, 2018, in Toulouse, France. More info on www.esof.eu/en/home.html



GRANT AWARDED
Virus-X receives funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Project Virus-X, grant agreement n° 685778).

The 3-dimensional structure of a gene product (e.g. an enzyme) can give better indications of its function by structural similarity than sequence information may give. The application of a special technique, called X-ray crystallography, performed at University of Durham in the UK and SARomics Biostructures in Sweden, is used to determine the 3-dimensional structures of selected gene products. This may reveal structural similarity and an evolutionary relationship to proteins with related and defined functions, a kindredness that may be apparent by conserved structure but unrecognizable by sequence comparisons.

Last but not least, Virus-X will provide invaluable information on ecological aspects of microbial ecosystems to better understand certain aspects, like host specificity and various defense and anti-defense mechanisms, in the interplay between viruses and their hosts in nature. An interplay and battle that undoubtedly will continue as long as life exists on Earth. ■

- (1) G. Ofir et R. Sorek, *Cell*, 172, 1260, 2018.
- (2) D. Prangishvili et al., *Nat. Rev. Microbiol.*, 15, 724, 2017.