# Old Herborn University Seminar Monograph

## 31. EVOLUTIONARY BIOLOGY OF THE VIROME, AND IMPACTS IN HUMAN HEALTH AND DISEASE

**EDITORS:**

PETER J. HEIDT
PEARAY L. OGRA
MARK S. RIDDLE
VOLKER RUSCH

Old Herborn University

# Old Herborn University Seminar Monograph 31

ISBN 3-923022-43-3
ISSN 1431-6579



EDITORS:

Prof. Dr. Peter J. Heidt
Foundation for Microbiology and Gnotobiology
Vlietweg 60A
2323 LE - Leiden
The Netherlands

Prof. em. Pearay L. Ogra, M.D.
Jacobs School of Medicine and Biomedical Sciences
University at Buffalo, State University of New York
219 Bryant Street
Buffalo, NY 14222
USA

Mark S. Riddle, M.D., Dr.P.H.
Uniformed Services University of Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20814
USA

Dr. rer. nat. Volker Rusch
Old Herborn University Foundation
Postfach 1765
D-35727 Herborn-Dill
Germany

Old Herborn University

Old Herborn University Foundation
Postfach 1765
D-35727 Herborn-Dill
Germany
Telephone: +49 - 2772 - 921100
Telefax: +49 - 2772 - 921101

# Contents

# Contents (continued)

# Contents (continued)

# Contents (continued)

# Participating Authors

**Prof. Patrick Forterre, PhD,** Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France.
(E-mail: patrick.forterre@pasteur.fr).

**Scott Allyn Handley, PhD,** Department of Pathology and Immunology, Washington University, 660 South Euclid Avenue, Campus Box 9119, St. Louis, MO 63110, USA.
(E-mail: shandley@wustl.edu).

**Dr. Tim Lachnit,** Zoological Institute, Christian-Albrechts-University Kiel, Am Botanischen Garten 1-9, 24118 Kiel, Germany.
(E-mail: tlachnit@zoologie.uni-kiel.de).

**Uri Laserson, PhD,** Department of Genetics and Genomic Sciences and Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, 1425 Madison Ave, Room L14-52A New York, NY 10029, USA.
(E-mail: uri@lasersonlab.org).

**João Trindade Marques, PhD,** Department of Biochemistry and Immunology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627 - Pampulha, Belo Horizonte, Minas Gerais, CEP 30270-901, Brazil.
(E-mail: jtm@ufmg.br).

**Prof. em. Pearay L. Ogra, MD,** Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, 219 Bryant Street, Buffalo, NY 14222, USA.
(E-mail: pogra@upa.chob.edu).

**Steven D. Quistad, PhD,** Laboratory of Genetic Evolution, Superior School of Industrial Physics and Chemistry of the City of Paris (ESPCI), 10 Rue Vauquelin, 75005 Paris, France.
(E-mail: steven.quistad@gmail.com).

**Mark S. Riddle, MD, DrPH,** Uniformed Services University of Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814, USA.
(E-mail: mark.riddle@usuhs.edu).

**Kimberley Seed, PhD,** Department of Plant & Microbial Biology, 271 Koshland Hall, University of California, Berkeley, Berkeley, CA 94720-3102, USA.
(E-mail: kseed@berkeley.edu).

**Robert H. Yolken, MD,** Johns Hopkins School of Medicine, Blalock 1104, 600 N Wolfe Street, Baltimore, MD 21287, USA.
(E-mail: rhyolken@gmail.com).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 1-3 (2017).

# INTRODUCTION TO THE 31ST OLD HERBORN UNIVERSITY SEMINAR ON EVOLUTIONARY BIOLOGY OF THE VIROME, AND IMPACTS IN HUMAN HEALTH AND DISEASE

**HERBORN-DILL, GERMANY**
**JUNE 22, 2017**

MARK S. RIDDLE[1] and PEARAY L. OGRA[2]

[1]Uniformed Services University of Health Sciences, Bethesda, MD, USA;
[2]Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, NY, USA

***"A MATHEMATICAL PROBLEM SHOULD BE DIFFICULT IN ORDER TO ENTICE US, YET NOT COMPLETELY INACCESSIBLE, LEST IT MOCK AT OUR EFFORTS. IT SHOULD BE TO US A GUIDE POST ON THE MAZY PATHS TO HIDDEN TRUTHS, AND ULTIMATELY A REMINDER OF OUR PLEASURE IN THE SUCCESSFUL SOLUTION."***
(David Hilbert, Paris, 1900)

Dr. Hilbert describes an important notion in science and that it is important to select an area where scientific endeavour is both rewarding and achievable. For the individual, selecting the right frontier in science is almost a matter of taste as it must consider selecting a topic that that truly advance a field that are considered in the context of emerging theories, new techniques, and new ideas (*Greenwood*, 1992). Frontiers in science are everywhere.

Since 1987, the Old Herborn University Seminar has been exploring and celebrating the frontiers of science as they relate to the microbiome of the environment and living things. As with the individual embarking upon a career in a frontier in a particular science, the Old Herborn University Executive and Scientific Program Committees identified a topic which was ripe for consideration as the topic of a seminar, and that was the, "Evolutionary Biology of the Virome and Impacts on Human Health and Disease." At first glance, one might question the consideration of this topic given the relative infancy of our understanding of the virome relative to the vast amount of research that has been conducted on the bacteriome (Figure 1). Indeed, the field is considered in its infancy in understanding the structure, composition, and function of the bacteriome and interactions with human health and our ability to modify it to cure disease (*Marchesi*, 2016). Despite this fundamental challenge, the selection of the topic, which had not previously been considered by the OHUS, emerged from a number of important reasons. Firstly, the evidence to date recognizes the vast success of viruses in abundance (though not biomass) throughout the world. For example, viruses account for 94% of all nucleic acid containing entities in the oceans (*Suttle*, 2007), and the number of viral particles in the human gut roughly equates to the number of bacteria (*Kim*, 2011). To further imagine the power and potential of viruses in the world, Landenmark and colleagues estimated the quantity of DNA (of which viruses play a substantial contributor) in the biosphere as an estimate of the vast storage and

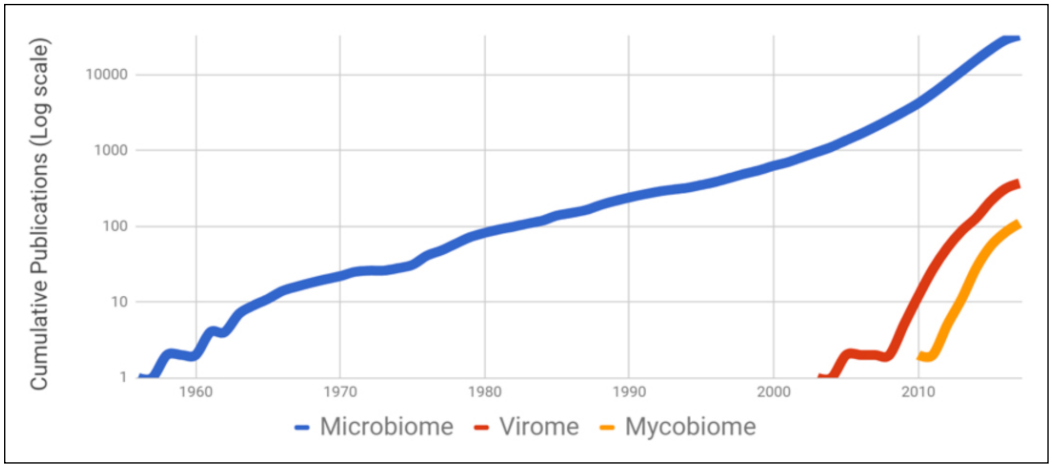


**Figure 1**: Relative force of publications related to phylum-specific research of the microbiome.

biological power that exists (*Landenmark*, 2015). The estimate that was generated was $5.3\times10^{31}$ megabase pairs (Mb) which equates to $10^{21}$ computers with the mean storage capacity of the world's four most powerful supercomputers. To put this into further perspective, $10^{21}$ is the order of magnitude that is estimated to quantify the amount of grains of sand on the beaches of the world. To simply state, with viruses, big things come in small packages. With such vast "biological computational power", the virome is integral to our understanding of the world with applications to evolutionary biology, engineering, botany, synthetic biology, ecology, emerging infectious diseases, agriculture, aquaculture, and animal health (including humans) (*Parker*, 2016).

A second reason to consider the virome is the recognition that with viruses: *We are them, and they are us*. While commonly used as the context underlying many a scientific fiction novel and movie, it is estimated that at least 8% of human DNA is composed of retroviral genome, and over 100,000 known viral particles exist in human genome (*Horie*, 2010). And the impact of viruses on human evolution and success is impressively highlighted by the discovery about two decades ago (*Mi*, 2000), where a gene in the human genome which encoded for a unique protein 'syncytin" made only by placental cells was discovered. Surprisingly, the gene appeared to be viral in origin and in its molecular characterization. A similar gene has also been identified in other primates, including chimpanzee placental cells. Subsequently, Heidmann and colleagues (*Dupressoir*, 2005) discovered a similar gene in other mammals including dogs, cats, pandas, hyenas and other carnivores. The syncytin gene appears to be strikingly similar in all species tested to date. It is noteworthy that syncytin is synthesized in the placental cells which are in direct contact with the uterine mucosal surface. Syncytin is responsible for the cell fusion and development of Syncytiotrophoblast. This cellular layer is essential for foetal survival and the transport of all soluble and cellular products from the maternal to the foetal circulation. It has been proposed that syncytin may have initially evolved to allow the virus to fuse host cells in order to facilitate cell-to-cell virus spread. Such a viral-induced process also allowed developing mam-

malian organisms to induce foetal cell fusion locally in the uterus to maternal cellular genome to facilitate maternal transport of nutrients and gasses essential for foetal survival. Thus, it has been said, “If not for a virus, none of us would ever be born.” (*Zimmer*, 2014)

While the individual scientist (and rather science teams in today’s world) approaches advancement through use of knowledge, honing of laboratory methods, and experimentation, the advancement that is intended from the OHUS is through what may be best termed consilience, or the 'jumping together' of knowledge by the linking of facts and fact-based theory across disciplines (*Wilson*, 2000). Therefore, to address this formidable challenge, a diverse group of leading scientists across multiple disciplines of evolutionary biology, virology, bacteriology, computational biology, immunology and other domains were assembled to review emerging research across the spectrum of fundamentals of virome in single cellular organisms, the virome of multicellular organisms, and the interaction of virome on human health and disease. And a fruitful and productive endeavour it was.

## LITERATURE


Dupressoir, A., Marceau, G., Vernochet, C., Bénit, L., Kanellopoulos, C., Sapin, V., and Heidmann, T.: Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proc. Natl. Acad. Sci. USA. 102, 725-730 (2005).

Greenwood, A.: Science at the frontier. Volume 1. The National Academies Press, Washington, DC (1992).

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J.M., and Tomonaga, K.: Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature 463, 84-87 (2010).

Kim, M.-S., Park, E.-J., Roh, S.W., and Bae, J.-W.: Diversity and abundance of single-stranded DNA viruses in human feces. Appl. Environ. Microbiol. 77, 8062–8070 (2011).

Landenmark, H.K., Forgan, D.H., and Cockell, C.S.: An Estimate of the Total DNA in the Biosphere. PLoS Biol. 13, e1002168 (2015).

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., Keith, J.C. Jr., and McCoy, J.M.: Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403, 785-789 (2000).

Marchesi, J.R., Adams, D.H., Fava, F., Hermes, G.D., Hirschfield, G.M., Hold, G., Quraishi, M.N., Kinross, J., Smidt, H., Tuohy, K.M., Thomas, L.V., Zoetendal, E.G., and Hart, A.: The gut microbiota and host health: a new clinical frontier. Gut 65, 330-339 (2016).

Parker, M.T.: An Ecological Framework of the Human Virome Provides Classification of Current Knowledge and Identifies Areas of Forthcoming Discovery. Yale J. Biol. Med. 89, 339-351 (2016).

Suttle, C.A.: Marine viruses--major players in the global ecosystem. Nat. Rev. Microbiol. 5, 801-812 (2007).

Wilson, E.O.: Consilience: The Unity of Knowledge. Abacus, London (2010).

Zimmer, C.: A Planet of Viruses. The University of Chicago Press, Chicago (2014).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 5-13 (2017).

# EVOLUTIONARY BIOLOGY OF THE VIROME AND IMPACTS ON HUMAN HEALTH AND DISEASE: AN HISTORICAL PERSPECTIVE

PEARAY L. OGRA[1] and MARK S. RIDDLE[2]

[1]Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, NY, USA;
[2]Uniformed Services University of Health Sciences, Bethesda, MD, USA

***"THE SINGLE BIGGEST THREAT TO MAN'S CONTINUED DOMINANCE ON THE PLANET IS A VIRUS"***
(Joshua Lederberg)

This quote opens the Hollywood movie "Outbreak" to introduce the epidemics of Haemorrhagic Virus fever in Zaire in 1967 and again in the mid 1990's. The movie is a fascinating commentary on the contemporary perceptions of serious or fatal virus infections, and subsequent impact on human rights and the societal good versus evil. The perception of viruses as evil life forms has been part of human society for thousands of years, and the word "virus" has been used in Latin, Greek and Sanskrit languages for centuries to describe the venom of snake, a dangerous slimy liquid, a fatal poison, or a substance produced in the body as a result of disease, especially one that is capable of infecting others.

Although numerous studies have attempted to explain the evolution of viruses and their biologic ancestry, the precise origins of viruses continue to remain a mystery. Recent studies have proposed that modern viruses are possibly derived from multiple ancient (now extinct) cell types that harboured segmented RNA genomes and co-existed with the ancestors of modern living cells (*Nasir* and *Caetano-Anolles*, 2015). Despite the mystery surrounding their origins, viruses have played a key role in the exploration of modern cellular and subcellular biologic systems, especially in defining the structure and function of the genes, and in the exploration of many other aspects of molecular biology. Of particular importance has been the role of viruses in human diseases.

Viruses represent the single most successful (numerically) biologic entity to date. It is estimated that $>10^{31}$ virus particles inhabit the earth and over $10^{7}$ viruses exist per ml of water in the oceans. It has been suggested that viruses outnumber their hosts globally by tenfold or more (*Proctor*, 1997). Virus as an infectious entity was initially introduced to describe a process of non-bacterial pathology by Ivanovsky in 1892. In 1898, Beijerinck independently applied the term to identify the causative agent of tobacco mosaic disease (*Johnson*, 1942). Three decades later, Stanley succeeded in crystalizing the tobacco mosaic virus (*Stanley*, 1935). The crystalized preparations retained all the biologic properties, including the ability to infect live cells. As a result, Stanley concluded that the viruses could not be truly alive.

Although the existence of viruses as a distinct morphologic and biologic entity has now been established for over a century, most modern classifications of existing life forms have continued to exclude viruses from the tree of life, in part, because they did not

exhibit the ribosomal genes which are routinely employed in microbial phylogenetic analysis. Recent identification of a giant virus species with dimensions and size of the genomes as big as many other microbes, and the discovery of mimivirus, a parasite of amoeba with a 1.2 megabase-pair DNA genome, and of other giant viruses whose genomes contain over 1.9 to 2.5 million bases have challenged these perceptions of the viruses (*Raoult* et al., 2006; *LaScola* et al., 2008; *Desnues* et al., 2012).

Based on ribosomal gene analysis, the living organisms have now been classified in the following domains:

1) Subcellular organisms such as viruses.
2) Bacteria and prokaryotes, as unicellular organisms without a nucleus or any membrane bound organelle in the cell. The classical examples include bacteria such as *E. coli, Streptococcus pyogenes, and cyanobacteria*. However, *Proteobacteria* are phylogenetically and physiologically related to eukaryotes.
3) Archaea. These resemble bacteria, are prokaryotes and contain no nucleus. They make up their own domain, live in a wide range of extreme environmental conditions.
4) Eukaryotes. Include virtually all nucleated unicellular or multicellular organisms including, fungi, plants, animal (including man), and
5) Protista. These include some multicellular and single cellular organisms which don't fit in other life categories. They are characterized in two major taxons; Animal like protists include amoeba, trypanosomes, plasmodium (malarial parasites), crithidia. Plant like protists include red or green algae and possibly other still to be classified life forms (*Woese* and *Fox*, 1977).

Based on the extensive use of whole genome sequencing, it now appears that most living organisms may in fact be chimeras containing genes from many different sources, including earlier eukaryotic life forms, prokaryotic unicellular non-nucleated organisms; and subcellular life forms including viruses. Recently, based on a series of elegant studies, Forterre and his colleagues have suggested that all life forms on earth could be more appropriately divided into two main groups:

1) ribosome encoding organisms (modern cells) and,
2) capsid encoding organisms (viruses).

Forterre has proposed that all modern cells descend from "the last universal cellular ancestor or the last universal CenAncester (LUCA). The ribosomes of LUCA contain 33 ribosomal proteins and 3 rRNA molecules. He and his colleagues have proposed that the modern universal optimized genetic code was possibly operational in the LUCA. Therefore, it may be possible to draw a new tree of life connecting together all ribosome-encoding organisms. However, there is not a single informational molecule identified to date that is common to all viruses (*Forterre* and *Prangishvilli*, 2009). Although, these authors concluded that "understanding how modern viruses originated, thus appears to be a more complex problem from the start than understanding the evolutionary history of modern cells", they have also argued strongly in favour of the possibility that viruses are essentially parasitic organisms which infect living cells and produce virions in order to spread their genes. Viral genes possibly originated in the "virosphere" during replication of the virus genome, or were recruited from other cellular lineages which are now extinct (*Forterre*, 1992). Certain specific viral proteins are present in the existing cellular domains of life forms

**Table 1**: Relative distribution of different viruses in different life forms*.

| Virus types | No. of virus genera in the life forms | |
|---|---|---|
| | Prokaryotic | Eukaryotic |
| RNA Viruses | Few | Many |
| (+) RNA | <10 | 150-160 |
| (-) RNA | <2 | 50-60 |
| dsRNA | <2 | 30-35 |
| Retroviral | 5-6 | 15-20 |
| ssDNA | 8-10 | >100 |
| dsDNA (All) | 40-50 | 70-80 |

*Adapted from King, et al., 2011.

infected by different viruses, raising the distinct possibility that viruses are very ancient, and at least two types of virions may have originated independently before the evolution of LUCA (*Forterre*, 1992). More recently, these investigators have proposed that DNA was "invented" by viruses, helping to convert a world of RNA based organisms to one of DNA-based hereditary ancestry, and giant viruses represent the origin of the nucleus of eukaryotic life forms (*Forterre*, 2010).

## ENVIRONMENT AND HUMAN VIROME

### Environmental Virome

Viruses inhabit virtually the entire spectrum of unicellular or multicellular life forms described above and which exist on earth today. However, limitations in our ability to detect, isolate and classify many known and still to be discovered viruses have limited our ability to explore in detail the comprehensive evolution of the global environmental virome. It has been proposed that of the over $10^{31}$ viral particles present in the current global microbial population, only about 2200 genomes from double stranded DNA viruses (dsDNA) and retroviruses are well established, as opposed to over 45,000 bacterial genomes identified to date (*Koonin*, 2015).

The viromes of the major cellular life domains are strikingly different in different life forms. Several forms of dsDNA viruses are present in both bacteria and archaea; but no viruses are known to be shared by eukaryotes with other life forms. Based on the classification of viruses introduced by Baltimore nearly 40 years ago, seven classes of viruses have now been identified in the environmental virome (*Baltimore*, 1971; *Koonin* et al., 2015). These include:

1) positive stranded RNA viruses, characterized by virions containing RNA of the same polarity as messenger RNA (mRNA),
2) negative stranded RNA viruses,
3) double stranded (ds)RNA viruses,
4) reverse transcribing viruses with positive stand RNA genome,
5) reverse transcribing virus with dsDNA genome,
6) single stranded DNA (ssDNA) viruses, and
7) dsDNA viruses.

**Table 2**: Relative distribution of DNA or RNA viral families infecting different life forms.

| Life forms | Percent Viral Families | |
|---|---|---|
| | RNA viruses | DNA viruses |
| Eukaryotic: | | |
| Fungi | 80 | 20 |
| Plants | 86 | 14 |
| Animals (invertebrates and vertebrates) | 60 | 40 |
| Protista-like and algae | 57 | 43 |
| Bacteria: | | |
| Proteobacteria | 28 | 79 |
| Other bacteria | 0 | 100 |
| Archaea | 0 | 100 |

In prokaryotic organisms most viruses possess dsDNA genomes, with a significant minority of ssDNA containing viruses, and only a very small number of RNA viruses. On the other hand, RNA viruses represent the majority of virome diversity in Eukaryotic life forms, with a significant, although a smaller number of ssDNA, and dsDNA viruses (*King* et al., 2011), as shown in Table 1. It should however be noted that genome sizes of RNA viruses are much more restricted compared to the DNA viruses which exhibit significantly more genome size diversity, often over 4 orders of magnitude higher genome size ranges (*Campillo-Balderas* et al., 2015). Of all the known viruses, about 9% have 2-3 segments and about 27% exhibit more than 4 segments. RNA viruses appear to be segmented more frequently (50%), than DNA viruses. Of all the segmented DNA and RNA viruses 60% infect plants, 21% infect other phylogenetically distant plants and other animals including 18% vertebrates and, about 1% infects bacteria (*Campillo-Balderas* et al., 2015). The relative distribution of RNA or DNA viruses in other eukaryotic life forms, (including, fungi, plants, animals), bacteria, archaea and Protista-like organisms (including algae) are shown in Table 2 (*Letunic* and *Bork*, 2007).

It is clear that the environmental virome is immense and virtually all life forms are inhabited by viruses. It appears that the genome size of dsDNA viruses have the highest diversity and surpass the far more restricted genome sizes of RNA and ssDNA viruses. However the RNA viruses exhibit a broader range of hosts in eukaryotic life forms, and infect a lot fewer bacterial species when compared to the DNA viruses.

## Human virome

The human virome represents a diverse spectrum of viruses found in other prokaryotic and eukaryotic life forms which inhabit various systemic sites, skin or the mucosal surfaces in humans. The viral component of the human microbiome is also referred to as the human virome.

The repertoire of viruses in human virome includes traditional human viruses, bacterial viruses (bacteriophage) associated with bacterial species which constitute the human microbiome, and viruses infecting fungi and plants associated with the human microbiome.

**Table 3**: Listing of the viral families in the human virome identified to date.

| RNA viruses: | Viral family: |
|---|---|
| dsRNA | Reoviridae |
| ssRNA(+) | Coronaviridae; Astroviridae; Caliciviridae; Flaviviridae; Picornaviridae; Togaviridae; Herpesviridae |
| ssRNA(-) | Rhabdoviridae; Feloviridae; Paramyxoviridae; Arenaviridae; Bunyaviridae; Orthumyxoviridae; Deltavirus. |
| DNA viruses: | Viral family: |
| dsDNA | Herpesviridae; Adenoviridae; Papillomaviridae; Polyomaviridae; Poxviridae |
| ssDNA | Anelloviridae; Parvoviridae |
| Retro viruses: | Viral family: |
| ssRNA (RT) | Retroviridae |
| dsDNA (RT) | Hepadnaviridae |

Human viral infections can manifest as:

a) asymptomatic and acute self-limiting, or as fulminant and progressive infections,
b) chronic symptomatic or asymptomatic infections,
c) endogenous retroviral infections (retroviruses comprise over 8% of the human genome), and
d) still to be discovered viruses related to many diseases currently considered to being of unknown aetiology.

The virome of existing bacterial cell populations is immense, and bacteriophages affect human health, because they have major influence on bacterial cell population, structure, toxin production and virulence. For example, bacteriophage communities in the respiratory tract of healthy individuals have been found to be unique in each individual, representing a random and often transient sampling of the external environment. On the other hand, the bacteriophage communities in patients with pathologic states, such as cystic fibrosis, were found to be similar to other patients with cystic fibrosis. Such colonization is presumably facilitated by the similar underlying airway pathology (*Willner* et al., 2009).

Recent technologic advances to study viruses now include metagenomics analyses, employing comparison of genetic information from next generation sequencing of clinical samples to genomes of all known viruses, and techniques designed to translate viral genes into proteins and computationally search for similar protein sequences in newly discovered agents (*Wylie* et al., 2012). These advances have made it possible to characterize rapidly the existing viral genomes and identify new viral agents, as listed in Table 3.

Viruses have been recovered from all human mucosal surfaces, skin and many systemic sites. In recent studies of the DNA virome in several thousand human samples of blood, faecal samples, respiratory and gastrointestinal secretions, saliva, milk, urine, and other body fluids have yielded a wealth of information about the distribution, genetic diversity and pathogenicity of human viruses identified to date. They are listed in Table 4 (*Pride* et al., 2012; *Delwart*, 2016; *Moustafa* et al., 2017). In one study, sequences for 94 different viruses were identified in different human blood samples, including for 19 human DNA viruses, proviruses and RNA viruses (*Moustafa* et al., 2017).

**Table 4**: Tissue distribution of different viral families in different human body surfaces or secretions.

| Virus | Genetic diversity | Tissue distribution | Pathogenic or commensal |
|---|---|---|---|
| Adenoviridae | High | GI, Resp., Urine, Blood | Both |
| Anelloviridae | High | Blood | Commensal/Pathogen? |
| Astrovirus | Medium | Blood, GI* | Both |
| Flaviviridae | Low | Blood | Both |
| Herpesviridae | Low | Blood, Skin | Both |
| Papillomaviridae | High | Skin | Both |
| Parvoviridae | High | Blood, GI* | Both |
| Picobirnaviridae | High | GI* | Commensal/Pathogen? |
| Picornaviridae | High | Blood, GI*, Resp.**, Skin | Both |
| Polyomaviridae | Medium | Blood, Resp.** Skin | Both |

GI*: Gastrointestinal tract. Resp.**: respiratory tract.
Adapted from: Delwart, 2016; Moustafa et al., 2017; and Pride et al., 2012.

Introduction of metagenomics has indeed revolutionized the study of human viruses. For example until recently, only two polyomaviruses were known as human pathogens. However, about 13 other human polyomaviruses have now been identified and some of these have been implicated in several neurologic or renal disorders and other pathological states in immunocompromised hosts. Some papillomaviruses are also found in asymptomatic healthy skin, while a few can induce anal or cervical cancers (*Foulongne* et al., 2012; *DeCaprio* and *Garcea*, 2013). Another important recent observation relates to the identification of a rarely studied ssDNA virus family, the anelloviruses. It appears that they may be the most common human viral infections identified to date. These viruses have been detected in almost 100% of blood samples from human adults. These viruses are acquired shortly after birth and multiple strains have been identified in the same individual. The anelloviruses exhibit the highest level of genetic diversity known in any viral family to date. Most anelloviruses function as commensals and the infections are generally asymptomatic. Increasing amount of anelloviruses in immunosuppressed subjects have been associated with possible immunologically mediated chronic inflammation and expression of clinical symptoms. However, their role in the pathogenesis of the disease still remains to be clearly defined (*Spandole* et al., 2015).

The spectrum and the load of different viruses in the human body are truly immense. Over $10^{10}$-$10^{11}$ viral particles/gram of faeces representing bacteriophage predators of bacteria and archaea, as well as many other viruses have been identified to date in the humans. Furthermore, as pointed out earlier, endogenous retroviruses constitute at least 8% of human DNA (*Minot* et al., 2011, 2012). In spite of their overwhelming number (possibly billions),

as residents of the human body, only few hundred human viruses have been clearly associated with serious disease in man (*Delwart*, 2016).

Available epidemiologic evidence suggests that the acquisition of the environmental virome by mankind is an evolutionary (and possibly a requisite) adaptation, and the organisms are neither consistently commensal nor pathogenic. The outcome of the microbiome-host interaction at a cellular level is determined by a complex balance between the pathogenic potential of the virus and the immunologic status of the host at the time of their initial and subsequent interactions.

Maternally derived antibodies acquired transplacentally or via breast milk induce significant protection against disease against a variety of bacterial as well as viral infections in the neonatal period and early infancy. Acquisition of such viral infections during early years of human growth and development appear to play a very important role in the maturation and functional development of the host immune system at mucosal and systemic sites. Such early infections may also protect against more pathogenic infections with other viruses later in life. For example, there is evidence to suggest that pegivirus or GBV-C virus belonging to the larger family of viruses formerly known as of hepatitis G or HPgV, dengue and zika viruses can significantly improve the outcome of HIV infection. Pegivirus-c infection appears to be quite common and asymptomatic. It is estimated that to date over 700 million humans are infected with the virus. HIV infected subjects with pegivirus-c infection tend to live longer and experience healthier lives (*Linen* et al., 1996) Similarly, experimental animal studies have demonstrated that mouse norovirus, a common human commensal, restored immunologic functions that were previously altered by an induced germ-free status or the use of antibiotics (*Kernbauer* et al., 2014). In another recent study, it has been demonstrated that early enteric infections can influence the development of the gut and murine immune system in a very similar manner as observed with the bacterial microbiome on development of the human gut and maturation of the immune system (*Caldwell*, 2015).

Finally, the retroviral components of the human genome have also been adapted to enhance functional development as well as the survival of the mammalian host. It has been proposed that proteins expressed by such endogenous retroviruses can effectively bind or block receptors for other exogenous retroviruses (*Griffiths*, 2001; *Malfavon-Borja* and *Feschotte*, 2015). In an elegant study, endogenous retroviral envelope proteins were found to be responsible for the fusion of foetal trophoblast cells with the mammalian placental cells during the onset of pregnancy. Such fusion is essential for nutritional exchange between maternal and foetal system during gestation (*Mi* et al., 2000).

Based on the information summarized above, it is clear that we are just beginning to explore and define the complexity of human virome relative to the immense load of environmental virome. Our understanding of viruses has evolved as a result of specific infections associated with many eukaryotic life forms, beginning with the discovery of tobacco mosaic disease virus. However, the human virome, as we understand it today, is immensely larger than the few hundred disease producing viruses. While some viruses are prone to produce acute, chronic, fulminant or fatal infections in man (*Hulo* et al., 2011), the human virome predominantly represents a

large reservoir of symbiotic interactions between the viruses and the human and other life forms on earth. In fact, the human virome may provide significantly far more beneficial effects on human health, immunologic homeostasis, and maternal-neonatal interactions, than the development of serious or fatal disease.

## LITERATURE


Baltimore, D.: Expression of animal virus genomes. Bacteriol. Rev. 35, 235-241 (1971).

Caldwell, K.: Expanding the role of the virome: Commensalism in the gut. J. Virol. 89, 1951-53 (2015).

Campillo-Balderas, J.A., Lazcano, A., and Becerra, A.: Viral genome size distribution does not correlate with the antiquity of the host lineages. Front. Ecol. Evol. 3, doi: 10.3389/fevo.2015.00143 (2015).

DeCaprio, J.A. and Garcea, R.L.: A cornucopia of human polyomaviruses. Nat. Rev. Microbiol. 11, 264-276 (2013).

Delwart, E.: The good viruses. The Scientist 115, 41-46 (2016)

Desnues, C., La Scola, B., Yutin, N., Fourhous, G., Robert, C., Azza, S., Jardot, P., Monteil, S., Campocasso, A., Koonin, E.V., and Raoult, D.: Provirophage and transpovirions as the diverse mobilome of giant viruses. Proc. Nat. Acad. Sci. USA 109, 18078-18083 (2012).

Forterre, P.: New hypotheses about the origins of viruses, prokaryotes and eukaryotes. In.: Frontiers of life (Eds.: Trân Thanh Vân, J. & K., Mounolou, J.C., Shneider, J., and McKay, C.). Edition Frontieres, Gif-sur-Yvette, France, 221-234 (1992).

Forterre, P.: Giant viruses: conflicts in revisiting the virus concept. Intervirology 53, 362-378 (2010).

Forterre, P. and Prangishvili, D.: The origin of viruses. Res. Microbiol. 160, 466-472 (2009).

Foulongne, V., Sauvage, V., Dereure, H.C., and Cheval, O.: Human skin microbiota: High diversity of DNA viruses identified on the human skin by high throughput sequencing. PLoS One 7, e38499 (2012).

Griffiths, D.J.: Endogenous retroviruses in the human genome sequence. Genome Biol. 2, reviews1017.1-reviews1017.5 (2001).

Hulo, C., deCastro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., and Le Mercier, P.: Viral zone: a knowledge resource to understand virus diversity. Nucleic Acids Res. D576-582 (2011).

Johnson. J. (Ed.): Phytopathologic classics, Vol. 7. American Phytopathologic Society, St. Paul, MN, USA, 27-30 (1942).

Kernbauer, E., Ding, Y., and Caldwell, K.: An enteric virus can replace the beneficial function of commensal bacteria. Nature 516, 94-98 (2014).

King, A.M.Q., Lefkowitz, E., Adams, M.J., and Carstens, B. (Eds.): Virus taxonomy; Ninth report of the international Committee on Taxonomy of Viruses. Elsevier Academic Press, Amsterdam (2011).

Koonin, E.V., Dolja, V.V., and Krupovic, M.: Origins and evolution of viruses of eukaryotes: the modularity. Virology 479-480, 2-25 (2015).

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P, Koonin, E., and Raoult, D.: The virophage as a unique parasite of the giant mimivirus. Nature 455, 100-104 (2008).

Letunic, L. and Bork, P.: Interactive tree of life (ITOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23, 127-128 (2007).

Linnen, J., Wages, J. Jr., Zhang-Keck, Z.Y., Fry, K.E., Krawczynski, K.Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., Hadziyannis, S., Karayiannis, P., Fung, K., Nakatsuji, Y., Shih, J.W., Young, L., Piatak, M. Jr., Hoover, C., Fernandez, J.,

Chen, S., Zou, J.C., Morris, T., Hyams, K.C., Ismay, S., Lifson, J.D., Hess, G., Foung, S.K., Thomas, H., Bradley, D., Margolis, H., and Kim J.P.: Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Science 271,505-508 (1996).

Malfavon-Borja, R. and Feschotte, C.: Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. J. Virol. 89, 4047-4050 (2015).

Mi S, Lee X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.-Y., Edouard, P., Howes, S., Keith, J.C. Jr., and McCoy, J.M.: Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403, 785-789 (2000).

Minot, S., Grunberg, S., Wu, G., Lewis, J., and Bushman, F.D.: Hypervariable loci in the human gut virome. Proc. Natl. Acad. Sci. USA 109, 3962-3966 (2012).

Minot, S., Sinha, R., Chen, J, Li, H., Keilbaugh, S.A., Wu, G., Lewis, J., and Bushman, F.D.: Dynamic response of the human virome to dietary intervention. Genome Res. 21, 1616-1625 (2011).

Moustafa, A., Xie, C., Kirkness, E., Biggs, W., Wong, E., Turpas, Y., Bloom, K., Delwart, E., Nelson, K.E., Venter, J.C., and Telenti, A.: The blood DNA virome in 8,000 humans. PLOS Pathogens 13, e1006292 (2017).

Nasir, A. and Caetano-Anolles, G.: A phylogenomic data-driven exploration of viral origins and evolution. Sci. Adv. 1, e1500527 (2015).

Pride, D.T., Salzman, J., Haynes, M., Rohwer, F., Davis-Long, C., White, R.A. III, Loomer, P., Armitage, G.C., and Relman, D.A.: Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. ISME J. 6, 915-926 (2012)

Proctor, L.M.: Advances in the study of marine viruses. Microsc. Res. Tech. 37, 136-161(1997).

Raoult, D., Renesto, P., and Brouqui, P.: Laboratory infection of a technician by mimivirus. Antern. Med. 144, 702-703 (2006).

Spandole, S., Cimponeriu, D., Berca, L.M., and Mihăescu, G.: Human anelloviruses: an update of molecular, epidemiological and clinical aspects. Arch. Virol. 160, 893-908 (2105).

Stanley, W.M.: Isolation of a crystalline protein processing the properties of tobacco mosaic virus. Science 81, 644-655 (1935).

Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F.E., Silva, J., Tammadoni, S., Nosrat, B., Conrad, D., and Rohwer, F.: metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. PLoS One 4, e7370 (2009).

Woese, C. and Fox, G.: Phylogenetic structure of the bacteria domain in the primary kingdom. Proc. Natl. Acad. Sci. USA 74, 5088-5090 (1977).

Wylie, K.M., Weinstock, G.M., and Storch, G.A.: Emerging view of the human virome. Transl. Res. 160, 283-290 (2012).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 15-26 (2017).

# THE ORIGIN, NATURE AND DEFINITION OF VIRUSES (AND LIFE): NEW CONCEPTS AND CONTROVERSIES

PATRICK FORTERRE

Institute Pasteur, Department of Microbiology, and
Institute for Integrated Cell Biology, Department of Microbiology,
University of Paris-Saclay, Paris, France

## SUMMARY

During the last two decades, our vision of the viral world has been profoundly modified by several discoveries in different fields of biology. Many of these discoveries conflict with the traditional view of viruses as inert biological objects that played a minor role in evolution and mainly evolve by picking genes from their hosts. It has been realized that viruses are very ancient, predating the Last Universal Cellular Ancestor (LUCA), extremely diverse, and have played a major role in life evolution. New definitions and concepts of viruses have been proposed to take these new discoveries into account. In particular, the virocell concept states that viruses are cellular organisms and emphasizes their ability to produce their own genes. Although the virocell concept remove arguments against the non-living status of viruses, the definition of life and living organisms remains challenging. Here, viruses are defined as capsid encoding organisms and life as the mode of existence of biological entities (individuals in philosophical term).

## INTRODUCTION

For years, most biologists considered viruses as by-products of biological evolution that could have only play a minor role in the history of the living world. This has gradually changed recently as a result of several advances in different fields of biology. The molecular ecologists have highlighted the extraordinary abundance of viral particles and viral genes in the environment (*Kristensen* et al., 2010, *Suttle*, 2013), the structuralists biologists have shown unexpected kinship between viruses infecting organisms belonging to different cellular domains (archaea, bacteria or eukaryotic) determining the structure of the proteins forming the viral capsid (*Abrescia* et al., 2012). At the same time, the study of archaeal virus revealed a fascinating world of different viruses previously unknown in bacteria and eukaryotes (*Prangishvili*, 2013). To top it all, the discovery of giant virus has caught the imagination of the scientific community by revealing the existence of viruses whose genomes are greater than those of many bacteria and archaea (*Raoult* et al., 2013). All these findings have revived interest in viruses and rested the issue of their nature - living or not - and the definition of life itself. Here, I review the definition of viruses and life that I proposed recently (*Forterre*, 2016, and references therein) and I discuss the virocell concept, that challenges the traditional view that assimilate viruses to their virions.

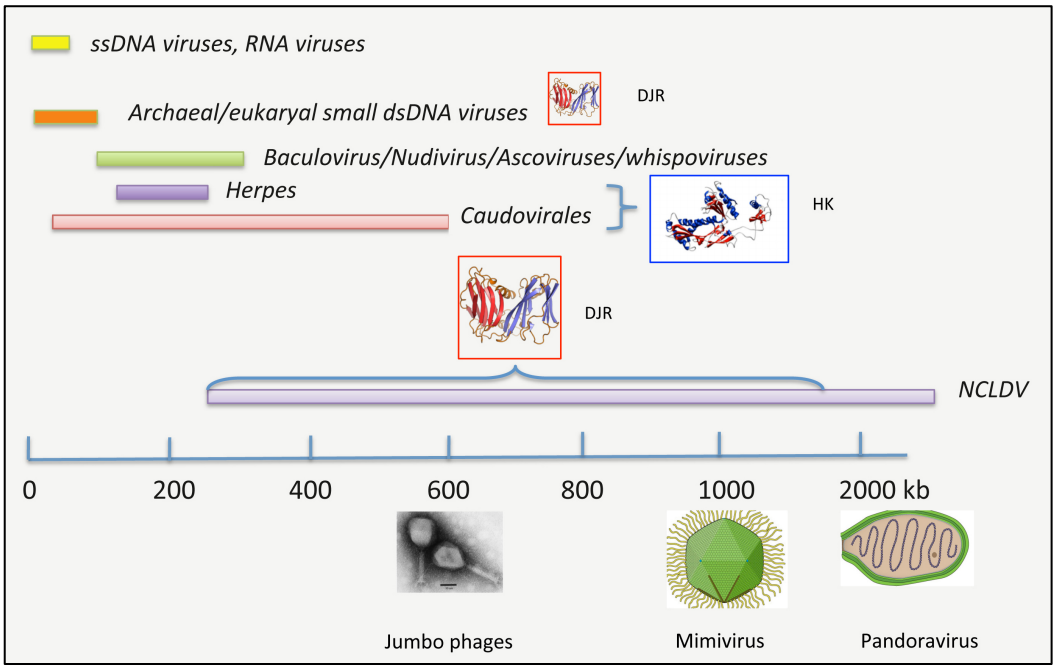


**Figure 1**: Genome size scale of viruses. DJH: double-jelly roll fold, HK: Hong Kong fold. Drawings and pictures are from ViralZone (*Hulo de Castro* et al., 2012).

## WHAT IS A VIRUS?

The discovery of giant viruses, such as Mimivirus and Pandoravirus, has raised new interest in the problem of the nature, origin and role of viruses in life evolution (*Raoult* et al., 2004, *Forterre*, 2010, *Philippe* et al., 2013, *Forterre*, 2017). These viruses produce virions that are visible under the light microscope and have genomes larger than the genomes of some free-living microbes. However, dividing viruses between 'giants' and 'small' viruses is artificial since there is a continuous gradient in genome size between the smallest virus (encoding two genes) and the Pandoravirus, encoding more than 2000 genes (*Forterre* et al., 2014) (Figure 1). The challenge is to find a definition of viruses that takes into account this diversity. Ten years ago, Didier Raoult and myself suggested classifying the living world in two major realms: "*capsid-encoding organisms*" (viruses) and "*ribosome-encoding organisms*" (Archaea, Bacteria and Eukarya) (*Raoult* and *Forterre*, 2008). We proposed the term "*orphan replicons*" for mobile elements such as plasmids, transposons, etc. that are evolutionary related to viruses (capsidless viruses according to *Koonin* and *Dolja*, 2013). Notably, considering the capsid to be the hallmark of the virus allows distinguishing between viruses and orphan replicons. This can be illustrated by comparing the smallest known plasmid encoding one protein (a replication protein) to the smallest known virus that encode two proteins, a replication protein and a capsid protein (*Krupovič* and *Bamford*, 2010) (Figure 2). Importantly capsids should themselves be defined as a set of proteins (at least one) associated to the viral nucleic acid to form a virion.

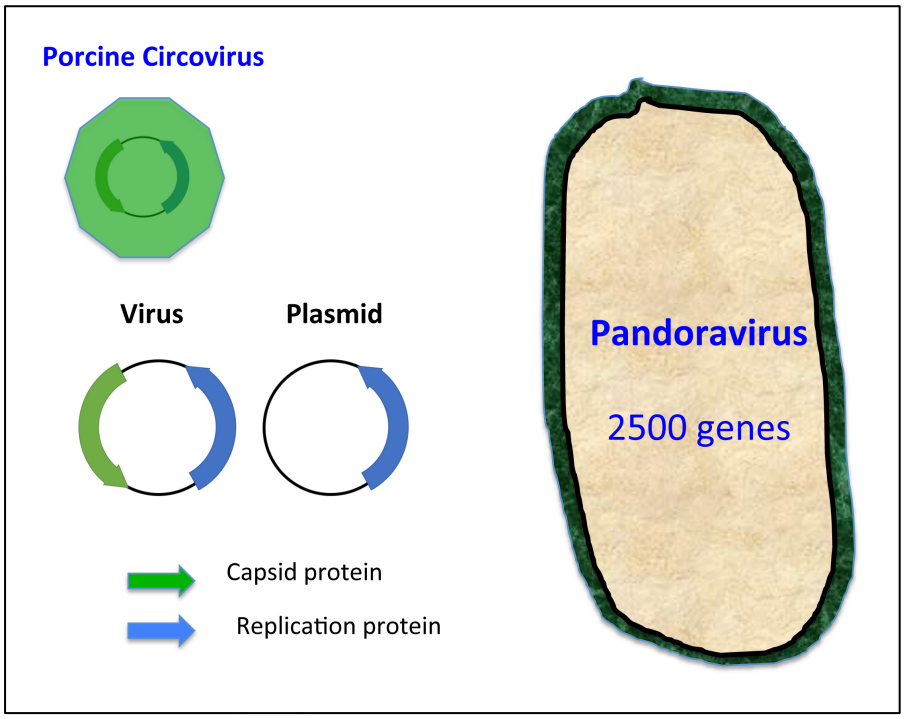


**Figure 2**: Definition of viruses as capsid encoding organism allows distinguishing viruses and plasmids and is valid for both the smallest and the largest viruses (drawing partly inspired by *Krupovič* and *Bamford*, 2010).

## THE ORIGIN OF VIRUSES

The definition of viruses as organisms encoding protein-based capsids implies that viruses originated after the emergence of the ribosome, i.e. after the emergence of rather sophisticated cells (Figure 3). This definition thus clearly refutes all "virus first" theory for the origin of life. On the other hand, comparative analyses of some key viral proteins, have shown that viruses were most probably already present in the biosphere at the time of the last universal common ancestor (LUCA) (*Abrescia* et al., 2012) (Figure 4). Indeed, at least two major lineages of viruses characterized by their specific major capsid proteins (MCP) and packaging ATPases have members in the three ensembles of viruses infecting each of the three domains of life. In addition to the MCP characteristics of these two lineages, many other types of non-homologous MCP have been identified by structural biologists (*Krupovic* and *Koonin*, 2017), confirming that viruses are polyphyletic, and preventing the definition of a viral "LUCA". This indicates that viruses originated several times independently, some of them before LUCA, other possibly later on, by recombination between various replicons and cassettes encoding sets of genes required to make a virion (see for instance *Krupovic*, 2013).

The first viruses were most likely RNA viruses infecting RNA/protein cells that originated from the association of parasitic RNA replicons and simple capsid proteins (*Forterre*, 2006a). Ancient structures present in

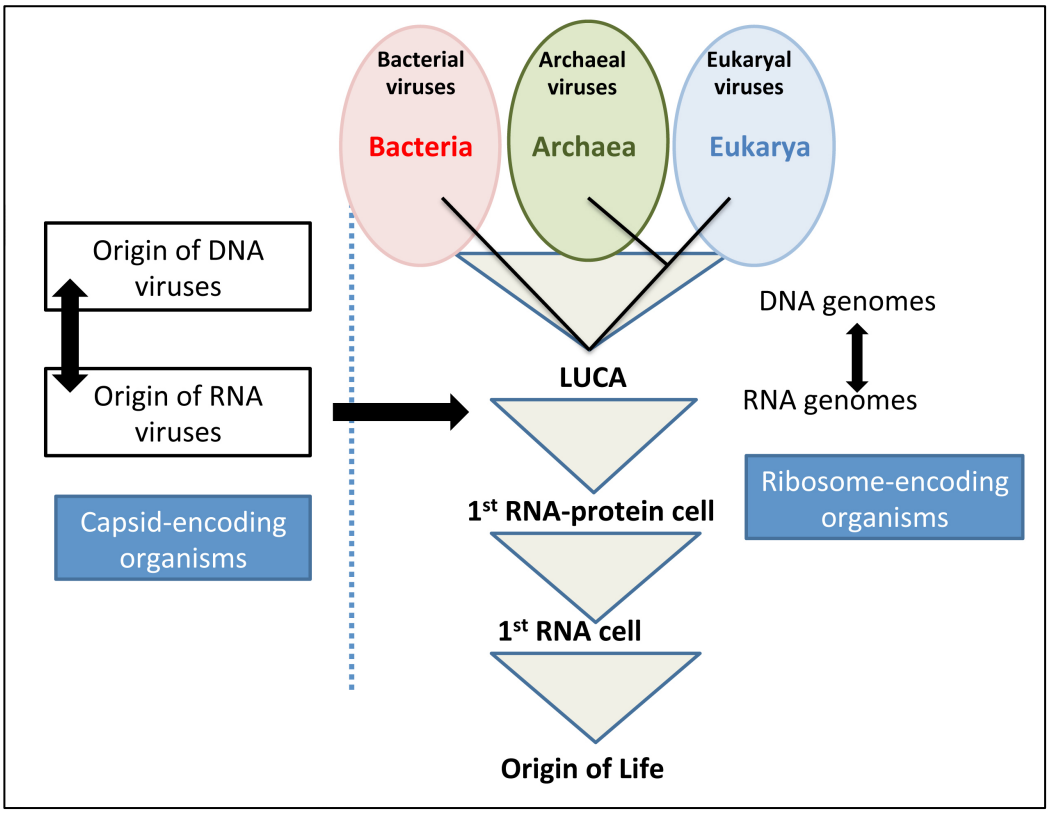


**Figure 3**: Schematic evolutionary pathway from the origin of life to the modern world of ribosome and capsid encoding organisms. LUCA: the Last Universal Common Ancestor.

these RNA/protein cells, such as membrane vesicles, intracellular compartments, or primitive chromosome scaffolds, may have provided the basis for the emergence of different types of simple virions (pleomorphic vesicle-like virions, icosahedral capsids and nucleocapsids) (*Forterre* and *Krupovic*, 2012). Later on, DNA viruses possibly originated from RNA viruses (Figure 3) and/or from the association of DNA replicons with capsids from RNA viruses. I suggested that DNA itself might have appeared in the ancient virosphere, being originally a particular type of modified RNA genome (the out of virus hypothesis for DNA origin) (*Forterre*, 2002, 2006ab). The early emergence of DNA and DNA replication machineries in such ancient virosphere would explain why these mechanisms are much more diverse in the viral world than in the cellular world (*Forterre*, 2013a). Later on, DNA and two non-homologous viral DNA replication mechanisms would have been transferred to cells, one in the bacterial lineage and the other in the "arkaryal lineage" (Arkarya being the name proposed for the clade grouping Archaea and Eukarya) (*Forterre*, 2015).

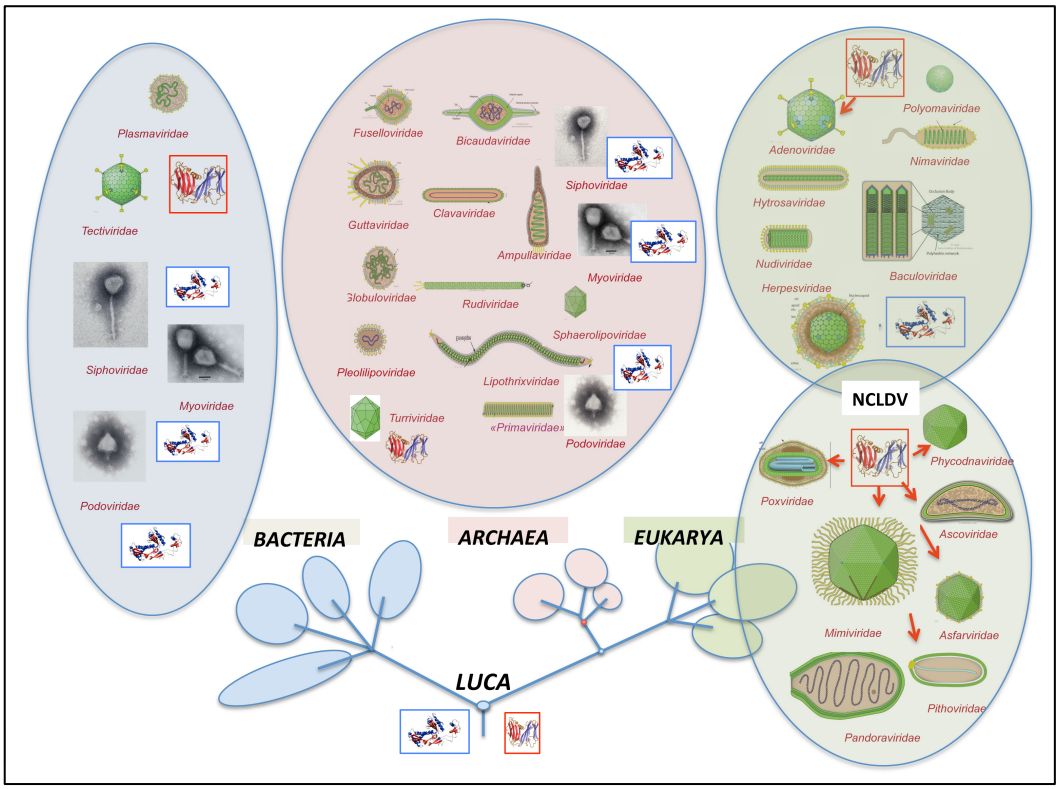


**Figure 4**: The Universal tree of life and the three ensembles of double-stranded DNA viruses corresponding to each domain (some families) are not indicated in eukaryotes. NCLDV: NucleoCytoplasmic Large DNA Viruses. LUCA: The Last Universal Common Ancestor. The depicted folds correspond to those in Figure 1 (see legend). This picture illustrates the diversity of archaeal dsDNA viruses (*Prangishvili*, 2013) compared to bacterial ones. Caudovirales (Myoviridae, Siphoviridae, Podoviridae) are common to Archaea and Bacteria and represent 95% of known bacterial viruses. Eukaryotes are infected by many different RNA virus families not shown here, whereas Bacteria are also infected by RNA viruses, but much more rarely. Drawings and pictures are from ViralZone (*Hulo de Castro* et al., 2012)

## ARE VIRUSES ORGANISMS? NO, ACCORDING TO THE TRADITIONAL VIEW OF VIRUSES ASSIMILATED TO THEIR VIRIONS

Defining virions as "*capsid encoding organisms*" raises the question: are viruses organisms? Usually, viruses are not considered to be organisms because as stated by *Lwoff* (1966) "*an organism is constituted of cells*", whereas viruses are assimilated to macromolecular machines. This is because viruses are usually confused with their virions. The tradition to identify viruses to their virions has both historical and practical reasons **(***Forterre*, 2016). Historically, the assimilation of viruses to their virions can be traced back to the discovery of viruses. The term "virus" was indeed first used to describe the infectious entities able to pass through filters that retain bacteria (*Bos*, 1999) and it turned out that these infectious entities are the viral particles. Later on, images of

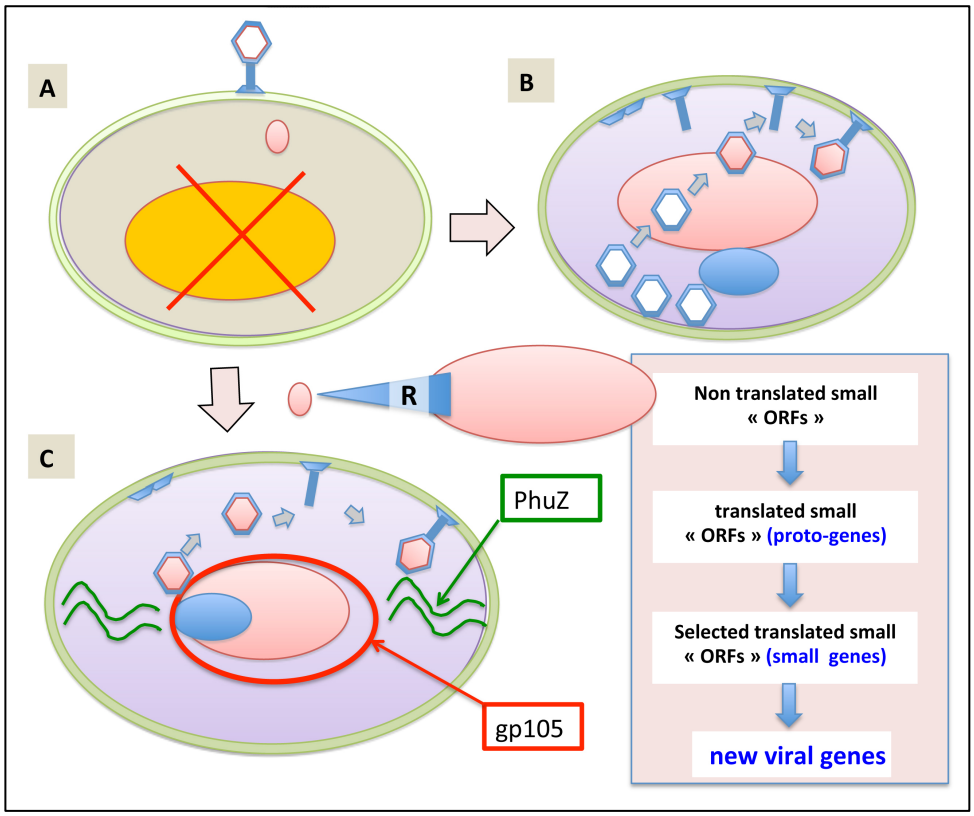


**Figure 5**: Infection by a virulent bacteriovirus transforms a bacterium (A) into a virocell (B) in which the only present nucleic acid is often the viral DNA (pink) after destruction of the bacterial chromosome (orange). The infection transforms the cellular metabolism and membranes (indicated by the differences in colour). Some bacteriovirus transform the bacterium into a virocell with a nucleus (C). A nuclear membrane is formed by a viral encoded protein (gp105) and the nucleus is positioned at the middle of the cell by a tubuline-like protein (PhuZ) (*Chaikeeratisak* et al. 2017). New genes can originate in the virocell (as in ribocells) during the replication (R) of the viral genome by the mechanism summarized on the lower right panel (*Carvunis* et al., 2012, *Zhao* et al., 2014).

virions have been constantly used to illustrate and popularize the virus concept in publications, textbooks and conferences (as it is the case in Figures 1, 3 and 4 of this paper!).

The assimilation of viruses to their virions had important consequences on the previous definitions of viruses. For instance, *Lwoff* (1957) stated that, in contrast to cells, viruses have only one type of nucleic acid (either RNA or DNA). However, DNA viruses actually possess both DNA and RNA (messenger RNA). Assuming that viruses have only one type of nucleic acid thus clearly means that one identifies the virus and the virion. Another example of this confusion is provided by environmental virologists who always determine the number of viruses in the environment by counting the number of viral-like particles and equal this number to the number of viruses (*Forterre*, 2013b). This is in fact the equivalent to count fish eggs to estimate the number of fishes in the ocean! A striking example can also be found in a seminal review by Jacob and

Wollman in which these authors first described the three possible forms of viruses (including intracellular one) to conclude by defining a virus as "*a genetic element enclosed in a protein coat"* (*Jacob* and *Wollman*, 1961).

## VIRUS DEFINITION AND THE ORIGIN OF VIRAL GENES

A damaging consequence of the assimilation of viruses to their virions is that many biologists, especially evolutionists interested in the history of life, underestimate the capacity of viruses to "produce" new genes *de novo*. This is because, once assimilated to their virions, viruses are considered to be passive and inert objects, entirely dependent of their cellular hosts (*Forterre*, 2011). As a consequence, most biologists wrongly assumed that all (or almost all) viral genes are derived from their cellular hosts/victim (*Moreira* and *Lopez-Garcia*, 2009). However, this is probably not correct. One can safely assume that most viral genes originated *de novo* in viral genomes during the intracellular cycle of virus reproduction by the same mechanisms that produce novel genes in cellular genomes (*Forterre*, 2011) (Figure 5). The mechanisms of *de novo* gene emergence have now been revealed by comparative analyses of multiple closely related genomes of *Saccharomyces cerevisiae* and *Drosophila melanogaster* (*Carvunis* et al., 2012, *Zhao* et al., 2014). Most new genes did not originate from gene duplication, as often assumed, but by the selection of short open reading frames (protogenes) arising randomly in intergenic regions. In the case of RNA viruses, such new genes can also originate on the non-coding strand of ancient genes, producing overlapping genes that can be unrelated from one viral strain to another (*Rancurel* et al., 2009). The massive creation of viral genes *de novo* in viral genomes explains well why most of them have no cellular homologues.

The continuous creation of new viral genes *de novo* provides an unlimited resource of new genes for cellular organism, since viral genes can become cellular following the integration of viral genome into cellular chromosomes. Notably, viral genomes can integrate into cellular genomes with few constraints in term of size, whereas the amount of cellular DNA or RNA that a virus can take up is limited by the size of the virion. This explains why the gene flux is overwhelming from viruses to cells than the other way around (*Forterre* and *Prangishvili*, 2013). The integration of viral DNA is a critical mechanism speeding the rate of evolution. There are now many examples of viral proteins whose exaptation has been at the origin of major evolutionary transition in life evolution. Well-known examples are Peg10 and Syncitins which derived from the Gag and Env proteins of an endogenous retrovirus, respectively, and are used by mammals to build the placenta and protect the embryo against the immune system of the mother (*Chen* et al.2015, *Villarreal*, 2016). A less known but also dramatic example is provided by the Arc protein, a master gene of memory acquisition, which is derived from the Gag gene of endogenous retroviruses (*Day* and *Schepherd*, 2015). Many authors now recognize the important role that viruses have played at several critical transitions in life evolution, especially in the emergence of new cellular lineages (*Forterre* and *Prangishvili*, 2009, *Koonin* and *Dolja*, 2013). This is not surprising, considering that viruses are major actors of both variation and selection, the two pillars of Darwinism (*Forterre*, 2012).

## ARE VIRUSES ORGANISMS? YES, IF ONE FOCUSES ON THE VIROCELL

The assimilation of viruses to their virions led to an underestimation of the intracellular phase of the virus reproduction cycle, often called the "eclipse phase' (since virions are no more visible) or the vegetative phase. These are unfortunate choices since the intracellular phase is precisely the active phase of the virus reproduction during which the virus indeed behaves as an organism. During this phase, the viral genome is transcribed and replicated and the cellular metabolism is partly or completely reorganized in favour of the virus. The transformation of the host/victim metabolism into a viral metabolism is especially critical for the whole process. For that purpose, the viral genome encodes proteins that redirect the metabolism of its host/victim and/or encodes viral metabolic enzymes complementary to or replacing those of the host cell (*Rosenwasser* et al., 2016).

A few years ago, I proposed the "virocell concept", to focus attention on the intracellular phase of the virus reproduction cycle (*Forterre*, 2011, 2013). In particular, the virocell concept should help taking into account the possible *de novo* origin of most viral genes. However, the virocell concept should not be confused with the definition of viruses because it does not cover the whole process corresponding to the viral organism. The term organism in the definition of viruses describes a biological process and integrates all aspects of the viral reproduction cycle: the virion, the virocells, and the viral genome.

The virocell concept was also proposed to fit with the definition of viruses as capsid encoding ORGANISMS, since the virocell corresponds to a new type of cellular organism (Figure 5B). At the beginning of the infection, two organisms thus co-exist in the same cell, the virus and the ribocell (a bacterium, an archaeon or an eukaryote), fighting each other (in particular via the CRISPR and anti-CRISPR systems). Later on, the two organisms can manage to co-exist pacifically in a form of symbiosis sometimes called carrier state or persistence forming a ribovirocell. However, very often, the virus predominates and the ribocell disappeares, leaving a transient virocell that commit suicide while liberating a wealth of virions. To paraphrase the metaphor from François Jacob: "*the dream of a cell is to produce two cells*", one can say: "*the dream of a virocell is to produce as much virions as possible*". In the ribovirocell, both organisms manage to fulfil their own dream but making a compromise, dividing more slowly for the cell and producing less virions for the virus. The co-existence of different organisms in the same cell is a frequent situation in biology, indicating that one should not confuse the notions of cell and organisms. Many eukaryotic cells harbour a multitude of intracellular bacteria and/or archaea. An amazing example is provided by amoeba infected by a bacterium and a giant virus; the giant virus itself being infected by a virophage (a virus of a virus) (*Moliner* et al., 2010). In that case, four organisms are present in the same cell as in a Russian doll (*Forterre*, 2010, *Mart Krupovic*, personal communication).

By analogy with the eukaryotic nucleus, the viral factories produced by many eukaryotic viruses replicating in the cytoplasm can be considered as the nuclei of these virus virocells. In that case, it is possible that this analogy reflects homology since several authors

have suggested the existence of an evolutionary link between the nucleus of eukaryotic cells and the viral factories (nucleus) of giant DNA viruses infecting eukaryotic cells (*Forterre* and *Gaia*, 2016). Strikingly, it turned out recently that some viruses infecting bacteria can also produce a nucleus in which viral transcription and replication takes place, as well as a simple mitotic-like apparatus to localize this nucleus at the centre of the infected bacterium (*Chaikeeratisak* et al., 2017) (Figure 5C). This again emphasizes the power of viral creativity and increases the appeal of the so-called viral eukaryogenesis hypothesis for the origin of eukaryotes (*Forterre* and *Raoult*, 2017).

## ARE VIRUSES ALIVE?

The answer to this question changed frequently depending of the period and the authors. Originally, viruses were often considered to be living because the "infectious fluid" detected by the pioneers of virology displayed all the classical properties of life: reproduction, multiplication and evolution (*Bos*, 2000). Later on, most biologists conclude that viruses are not living when they realized that virions are "simply" inert nucleoprotein particles devoid of metabolism (*Bos*, 2000, *Van Regenmortel*, 2003; *Moreira* and *Lopez-Garcia*, 2009). This was of course another consequence of the assimilation of viruses to their virions. For instance, *Van Regenmortel* (2003), former president of the ICTV, wrote that: "*viruses do not possess many of the essential attributes of living organisms, such as the ability to capture and store free energy and they lack the characteristic autonomy arising from the presence of integrated, metabolic activities*". However, this is not true for the virocell, since the latter is characterized by a specific metabolism working for the benefit of the virus (*Rosenwasser* et al., 2016). It is thus tempting to conclude that viruses are actually living after all. This conclusion then raises further questions. If viruses are living, one can go one step further and ask: are plasmids living? As previously mentioned, the only difference between the smallest plasmid and the smallest virus is the presence in the latter of a gene encoding a capsid protein (Figure 2). Does this mean that addition of a single gene is sufficient to transform a non-living biological object (the plasmid) into a living organism (the virus)?

I recently noticed another example that dramatically illustrates the difficulty to define a living organism, the transition between an intracellular bacterium and an organelle (*Forterre*, 2016). All biologists would agree that intracellular bacteria are living and most of them would also assume that mitochondria are not because "*they lack autonomy and a life cycle*" (*Van Regenmortel*, 2010). However, it is impossible to determine when the transition from living to non-living occurred in the evolution leading from the alphaproteobacterium ancestor of mitochondria to *bona fide* mitochondria. This is because the autonomy of an endosymbiont towards its host decreases in a continuous manner in the course of reductive evolution **(***Forterre*, 2016**).** It is thus hopeless to search for the gene or the gene set that would define life and/or determine the degree of autonomy of an organism. One cannot define a clear-cut border between living and non-living organisms/organelles based on quantitative or qualitative features.

This conclusion raises a challenging question: should we exclude the terms "life" and "living" from the biological literature, since they cannot be rigorously defined scientifically? This seems difficult considering that biology is the science of "life". To solve this conundrum, I suggested considering as living all biological entities as long as they are operational in the process of "life" (*Forterre*, 2016). To discriminate between biological entities that can be living (i.e. a protein or a chromosome) and biological entities that cannot, such as a protein domain or a gene, I have proposed using the philosophical distinction between "individual" and "particular" (*Chauvier*, 2008, *Pradeu*, 2010). Individuals should be "*separable, countable and have acceptable clear-cut spatial boundaries*" (a protein, a chromosome) whereas a particular is "*everything that can be designed through a demonstrative reference*" (a protein domain or a gene) (*Chauvier*, 2008, *Pradeu*, 2010). In these proposals, life can be defined as *"the mode of existence of biological individuals*" (*Forterre*, 2016) to paraphrase the definition proposed by Friedrich Engels in the 19th century "*life is the mode of existence of an albuminoid body"* (*Engels*, 2006 [1883]).

## CONCLUSION

Viruses and evolutionary related mobile elements are a major component of the biosphere beside the descendants of LUCA (archaea, bacteria and eukarya). Deciphering the history of the co-evolution between viruses, mobile elements and cells will be a major task of this century. Recently, most efforts in the fields have been made in analysing viromes from various environments. The limitation of this approach is that most sequences in viromes correspond to unknown viruses and data analyses end up focusing on the small set of sequences retrieved from already known viruses. A major effort should be now to isolate new virus-host systems, especially for viruses infecting some understudied organisms that represent most of the biodiversity on earth, such as the various phyla of protists in the eukaryotic domain, or the recently described new archaeal lineages that seem widespread in all types of environments.

## ACKNOWLEDGMENT

Patrick Forterre research groups are supported by an ERC grant from the European Union's Seventh Framework Program [340440 Project EVOMOBIL].

## LITERATURE

Abrescia, N.G., Bamford, D.H., Grimes, J.M., and Stuart, D.I.: Structure unifies the viral universe. Annu. Rev. Biochem. 81, 795-822 (2012).

Bos, L.: Beijerinck's work on tobacco mosaic virus: historical context and legacy. Philos. Trans. R. Soc. Lond., B Biol. Sci. 354, 675-685 (1999).

Bos L.: 100 years of virology: from vitalism via molecular biology to genetic engineering. Trends Microbiol. 8, 82-87 (2000).

Carvunis A-R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., Brar, G.A., Weissman, J.S., Regev, A., Thierry-Mieg,

N., Cusick, M.E., and Vidal, M.: Protogenes and de novo gene birth. Nature 487, 370-374 (2012).

Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A.F., Erb, M.L., Coker, J.K., Vavilina, A., Newton, G.L., Buschauer, R., Pogliano, K., Villa, E., Agard, D.A., and Pogliano, J.: Assembly of a nucleus-like structure during viral replication in bacteria. Science 355, 194-197 (2017).

Chauvier, S.: Particuliers, individus et individuation. In: L'individu: perspectives contemporaines (Eds.: Ludwig P. and Pradeu T.). Librairie Philosophique J. Vrin, Paris: 11-35 (2008).

Chen, H., Sun, M., Liu, J., Tong, C., and Meng, T.: Silencing of Paternally Expressed Gene 10 Inhibits Trophoblast Proliferation and Invasion. PLoS One.10, e0144845 (2015).

Cortez, D., Forterre, P., and Gribaldo, S.: Hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. Genome Biol. 10, R65 (2009).

Daubin, V. and Ochman, H.: Start-up entities in the origin of new genes. Curr. Opin. Genet. Dev. 14, 616-619 (2004).

Day, C. and Shepherd, J.D.: Arc: building a bridge from viruses to memory. Biochem. J. 469, e1-3 (2015).

Engels, F.: Dialectics of Nature. (Translation of Dialektik der Nature, 1883). Wellred publications, London (2006).

Forterre, P.: The origin of DNA genomes and DNA replication proteins. Curr. Opin. Microbiol. 5, 525-532 (2002).

Forterre, P.: The origin of viruses and their possible roles in major evolutionary transitions. Virus Res. 117, 5-16 (2006a).

Forterre, P.: Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc. Natl. Acad. Sci. USA 103, 3669-3374 (2006b).

Forterre, P.: Giant viruses: conflicts in revisiting the virus concept. Intervirology 53, 362-378 (2010).

Forterre, P.: Manipulation of cellular syntheses and the nature of viruses: the virocell concept. C.R. Chim. 14, 392-399 (2011).

Forterre, P.: Darwin's goldmine is still open: variation and selection run the world. Front. Cell. Infect. Microbiol. 2, 106 (2012).

Forterre P.: Why are there so many diverse replication machineries? J. Mol. Biol. 425, 4714-4726 (2013a)

Forterre, P.: The virocell concept and environmental microbiology. ISME J. 7, 233-236 (2013b).

Forterre, P.: The universal tree of life: an update. Front. Microbiol. 6, 717 (2015).

Forterre, P.: To be or not to be alive: How recent discoveries challenge the traditional definitions of viruses and life. Stud. Hist. Philos. Biol. Biomed. Sci. 59, 100-108 (2016).

Forterre, P. and Gaïa, M.: Giant viruses and the origin of modern eukaryotes. Curr. Opin. Microbiol. 31, 44-49 (2016).

Forterre, P. and Krupovic, M.: The origin of virions and virocells: the escape hypothesis revisited. In: Viruses: Essential Agents of Life (Ed.: Witzany, G.). Springer Science + Business Media, Dordrecht: 43-60 (2012).

Forterre, P. and Prangishvili, D.: The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. Ann. N Y Acad. Sci. 1178, 65-77 (2009).

Forterre, P. and Prangishvili, D.: The major role of viruses in cellular evolution: facts and hypotheses. Curr. Opin. Virol. 3, 558-565 (2013).

Forterre, P. and Raoult, D.: eLetter: (http://science.sciencemag.org/content/355/6321/194/tab-e-letters) (2017).

Forterre, P., Krupovic, M., and Prangishvili, D.: Cellular domains and viral lineages. Trends Microbiol. 22, 554-558 (2014).

Hulo de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., and Le Mercier, P.: ViralZone: a knowledge resource to understand virus diversity. Nucleic Acids Res. 39, D576-D582 (2011).

Jacob, F. and Wollman, E.L.: Viruses and genes. Sci. Am. 204, 93-107 (1961).

Koonin, E.V. and Dolja, V.V.: A virocentric perspective on the evolution of life. Curr. Opin. Virol. 3, 546-557 (2013).

Koonin, E.V. and Dolja, V.V.: Virus world as an evolutionary network of viruses and capsidless selfish elements. Microbiol. Mol. Biol. Rev. 78, 278-303 (2014).

Kristensen, D.M., Mushegian, A.R., Dolja, V.V., and Koonin, E.V.: New dimensions of the virus world discovered through metagenomics. Trends Microbiol. 18, 11-19 (2010).

Krupovic, M: Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Curr. Opin. Virol. 3, 578-586 (2013).

Krupovič, M. and Bamford, D.H.: Order to the viral universe. J. Virol. 84, 12476-12479 (2010).

Krupovic M, and Koonin E.V.: Multiple origins of viral capsid proteins from cellular ancestors. Proc. Natl. Acad. Sci. USA 114, E2401-E2410 (2017).

Lwoff, A.: The concept of virus. J. Gen. Microbiol.17, 239-253 (1957).

Lwoff A.: Interaction among virus, cell, and organism. Science 152, 1216-1220 (1966

Moliner, C., Fournier, P.E., and Raoult, D.: Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS Microbiol. Rev. 34, 281-294 (2010).

Moreira, D. and López-García, P.: Ten reasons to exclude viruses from the tree of life. Nat. Rev. Microbiol. 7, 306-311 (2009).

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., Garin, J., Claverie, J.M., and Abergel, C.: Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341, 281-286 (2013).

Prangishvili, D.: The wonderful world of archaeal viruses. Annu. Rev. Microbiol. 67, 565-585 (2013).

Pradeu, T.: What is an organism? An immunological answer. Hist. Phil. Life Sci. 32, 247-268 (2010).

Rancurel, C., Khosravi, M., Dunker, A.K., Romero, P.R., and Karlin, D.: Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. J. Virol. 83, 10719-10736 (2009).

Raoult, D. and Forterre, P.: Redefining viruses: lessons from Mimivirus. Nat. Rev. Microbiol. 6, 315-319 (2008).

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola B, Suzan, M., and Claverie, J.M.: The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344-1350 (2004).

Rosenwasser, S., Ziv, C., Creveld, S.G., and Vardi, A.: Virocell Metabolism: Metabolic Innovations During Host-Virus Interactions in the Ocean. Trends Microbiol. 24, 821-832 (2016).

Suttle, C.A.: Viruses: unlocking the greatest biodiversity on Earth. Genome 56, 542-544 (2013).

Van Regenmortel, M.H.: Viruses are real, virus species are man-made, taxonomic constructions. Arch. Virol. 148, 2481-2488 (2003).

Van Regenmortel, M.H.: Logical puzzles and scientific controversies: the nature of species, viruses and living organisms. Syst. Appl. Microbiol. 33, 1-6 (2010).

Villarreal, L.P.: Viruses and the placenta: the essential virus first view. APMIS 124, 20-30 (2016).

Zhao, L., Saelao, P., Jones, C.D., and Begun, D.J.: Origin and spread of de novo genes in Drosophila melanogaster populations. Science 343, 769-772 (2014).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 27-32 (2017).

# CO-EVOLUTIONARY DYNAMICS BETWEEN EPIDEMIC *VIBRIO CHOLERAE* AND PREDATORY PHAGE

KIMBERLEY SEED

Department of Plant & Microbial Biology, University of California, Berkeley, Berkeley, CA, USA

## SUMMARY

Waterborne pathogens like *Vibrio cholerae* pose significant threats to global health. *V. cholerae* can persist in the aquatic environment, and it can emerge to cause devastating cholera outbreaks in endemic regions and in vulnerable areas where infrastructure has been compromised and populations have been displaced. The host-pathogen interactions that dictate disease outcome and cholera transmission dynamics occur in the context of a complex microbial ecosystem that includes predatory bacterial viruses (phages). Such phages are found in the aquatic environment and are co-ingested with *V. cholerae*, permitting continued phage predation of *V. cholerae* within the human intestinal tract. In our efforts to understand the interactions between phage and *V. cholerae* during human cholera infection, we interrogate the molecular mechanisms underpinning phage-bacterial co-evolution in naturally evolved microbial populations within and between cholera patients. We found that phage predation in the human intestinal tract places mutational constraints on *V. cholerae*, ensuring that phages have access to a susceptible bacterial population. However, phage resistant variants do emerge, displacing sensitive variants. Our data indicate that the key to phage resistance in successful epidemic strains of *V. cholerae* resides in mobile genetic elements. Additionally, we have identified novel genomic signatures associated with the co-evolution of phage that allow phage to overcome resistance barriers and thrive in otherwise resistant cells. Understanding *V. cholerae*-phage interactions serves as a useful platform to understand how predatory phages structure microbial communities, including those that flourish in disease states and those that comprise our microbiome.

## INTRODUCTION

Illness and death caused by infectious diarrheal disease agents, like *Vibrio cholerae*, are major threats to public health and significant barriers to socio-economic development worldwide (*Havelaar* et al., 2015). Upon ingestion of food or water contaminated with pathogenic *V. cholerae* individuals can succumb to cholera, an acute diarrheal disease that leads to severe dehydration and death. The incidence of cholera worldwide is steadily increasing, the global disease burden is estimated to be 3-5 million cases resulting in more than 100,000 deaths annually (*Harris* et al., 2012). Following ingestion, *V. cholerae* must survive the gastric acid in the stomach and those bacteria that do go on to colonize the mucosal surface of the small intestine. It is here

that the organism elaborates cholera toxin, the major virulence factor for pathogenic strains (*Waldor* and *Mekalanos*, 1996). Cholera toxin binds to a receptor on enterocytes and activates adenylate cyclase leading to chloride secretion and secretory diarrhoea. Our understanding of *V. cholerae* pathogenesis has been built almost exclusively on studies to understand bacterial virulence factors and virulence gene regulation (for example: (*Finkelstein* and *LoSpalluto*, 1969; *Miller* et al., 1987; *Taylor* et al., 1987; *Merrell* et al., 2002a,b; *Mandlik* et al., 2011; *Fu* et al., 2013; *Kamp* et al., 2013). However, the role that other constituents of the microbial community, including predatory phages, play in disease outcome and the evolution and epidemiology of *V. cholerae* are not well understood.

Phages are bacterial viruses that act with exquisite specificity to kill their perpetually evolving bacterial targets. Many studies have demonstrated the presence of predatory phage co-existing with *V. cholerae* in stool when cholera victims present to the clinic with severe disease (*d'Herelle* and *Malone*, 1927; *Pasricha* et al., 1931; *Nelson* et al., 2007; *Seed* et al., 2010; 2014; *David* et al., 2015). These observations, coupled with fluctuations in environmental phage levels, have implicated predatory phages in shaping cholera outbreaks (*d'Herelle* and *Malone*, 1927; *Pasricha* et al., 1931; *Faruque* et al., 2005a,b). An interesting feature of predatory phages is that, similar to environmental factors like rainfall amounts, phage may modulate the inter-epidemic persistence of *V. cholerae* in the environment, thus impacting the occurrence of outbreaks; however, uniquely, these phages are co-ingested with *V. cholerae* into the human host and have the potential to continue to prey on their bacterial host during the course of infection. Predatory phages are also shed in appreciable amounts by infected patients where, like *V. cholerae,* they can be spread to others via faecal-oral transmission. Phages, therefore, have the potential to impact all aspects of the *V. cholerae* life cycle (including environmental persistence, transmission, infection and dissemination), on both a short and long-term evolutionary scale.

## THE DOMINANT PREDATORY PHAGES

Previous work done at the International Centre for Diarrheal Disease Research, Bangladesh (ICDDR,B) in Dhaka, Bangladesh documented an inverse correlation between cholera disease burden and the levels of phages that specifically infect and kill *V. cholerae* in the aquatic environment (*Faruque* et al., 2005b), suggesting these viral predators have a role in shaping cholera outbreaks (*Faruque* et al., 2005a). We analysed phages isolated from cholera patient stool samples that had been collected over a decade long period at the ICDDR,B. Surprisingly, we found that the genetic diversity of predatory phages associated with *V. cholerae* in patient samples is strikingly low (*Seed* et al., 2010). We hypothesize that these particular phages have evolved unique strategies to maintain a long-term association with their bacterial host. In total, three unique and unrelated predatory phages have been identified: ICP1, ICP2 and ICP3. The continued prevalence of these phages (at least in Bangladesh) is supported by recent metagenomic data collected to understand microbial succession after cholera infection (*David* et al., 2015). Additionally, subsequent analysis of phage recovered from a Haitian cholera

patient sample revealed the presence of an ICP2 isolate that was strikingly similar, although clearly distinct from phage recovered two years earlier from a Bangladeshi patient (*Seed* et al., 2014). The limited diversity of phages associated with *V. cholerae* in human cholera patient samples allows us to simultaneously assess phage and *V. cholerae* evolution.

## MECHANISTIC CHARACTERIZATION OF VIRULENCE REDUCTION IN PHAGE RESISTANT *V. CHOLERAE*

We identified the surface structure used by ICP1, the most prominent phage, to initiate infection and found that in both simulated aquatic environments and within experimental infection models phage predation can reduce bacterial loads and select for receptor mutants (*Seed* et al., 2012). In order to conclusively determine if phages influence the abundance of *V. cholerae* in the human intestinal tract, we analysed intra-patient microbial populations (*Seed* et al., 2014). We demonstrated that the predatory phage ICP2 can exert selective pressure resulting in a reduction of wild-type bacteria and the proliferation of heterogeneous resistant populations during infection in humans. We found that resistant populations had altered receptors, validating our previous observations in experimental models. We found that regardless of mechanism, receptor alterations were accompanied by decreased virulence and/or transmission potential. This observation, in part, explains how these phages maintain their long-term association: phage predation in the human intestinal tract places mutational constraints on *V. cholerae*, ensuring that phages have access to a susceptible population. Collectively, this work demonstrates that adaptations to phage predation involve trade-offs in evolutionary fitness and provides a molecular mechanism for how phage predation impacts *V. cholerae* transmission and seeding of environmental reservoirs.

## IDENTIFICATION OF NOVEL GENOMIC SIGNATURES ASSOCIATED WITH CO-EVOLUTION OF PHAGE AND *V. CHOLERAE*

The pervasiveness of specific predatory phages in Bangladesh with continued cholera epidemics suggests that *V. cholerae* has strategies to limit phage predation, and that phages can evolve to overcome such defences. Our work has focused on ICP1, since it is the most prevalent phage found associated with *V. cholerae* in cholera patient samples in this region (*Seed* et al., 2010). Similar to what we discovered for ICP2 (discussed above), ICP1 uses an essential virulence factor on the *V. cholerae* surface to initiate infection. Interestingly, we have not observed surface modifications allowing escape from phage for ICP1-*V. cholerae* in the context of human infection (*Seed* et al., 2012). Instead, we discovered that novel anti-phage mobile genetic elements called PLEs are responsible for ICP1 resistance in epidemic *V. cholerae* (*O'Hara* et al., 2017). PLEs are present in epidemic *V. cholerae* isolates recovered between 1949-2011 (spanning the entire collection period for which strains were available), and from different locations including Egypt, Mozambique and Bangladesh. PLEs have no sequence similarity to other known anti-phage systems, highlighting the genetic novelty found in

studying phage-host co-evolution and, most importantly, highlighting the need to study naturally evolved bacterial and viral populations. Although the mechanistic basis for how PLE protects *V. cholerae* from ICP1 infection is not fully understood, our studies thus far have revealed a multi-faceted mode of phage interference. Perhaps the most striking finding to come from characterization of these genetic elements is that all PLEs, regardless of geographic or temporal origin, respond uniquely to ICP1. These results indicate that the molecular battle between ICP1 and *V. cholerae* has been going on for at least 60 years, and that PLEs are a key bacterial weapon in this battle.

In response to the diverse strategies that bacteria use to defend against the threat of predatory phages (*Dy* et al., 2014), phages can co-evolve to circumvent any resistance barrier that they face (*Samson* et al., 2013). In an unexpected twist, we discovered that ICP1 has co-evolved to overcome PLEs using a CRISPR-Cas adaptive immune system (*Seed* et al., 2013). We discovered that half of all ICP1 isolates recovered from cholera patients in Bangladesh encode a functional CRISPR-Cas system. CRISPR-Cas is a sequence specific adaptive immune system that is typically encoded by bacteria to defend against phage ( *Barrangou* et al., 2007; *Marraffini*, 2015). The ICP1-encoded CRISPR-Cas system targets and degrades PLE to block its anti-phage activity (*Seed* et al., 2013). Collectively, we have established that the long-term interactions between *V. cholerae* and ICP1 serve as a useful platform to understand the evolution of phage-resistance and counter-resistance in the context of human disease.

## CONCLUSIONS

Host-pathogen interactions are strongly affected by the complex microbial community that surrounds them. Predator-prey dynamics are notably absent between phages and their bacterial hosts in the intestinal microbiome in healthy humans (*Reyes* et al., 2010). However, disease states, which are often accompanied by significant bacterial proliferation, provide optimal conditions for rampant phage predation. Phage-mediated perturbations of microbial communities have recently been implicated in inflammatory bowel disease (*Norman* et al., 2015) and cystic fibrosis (*James* et al., 2014). Understanding the molecular consequences of phage predation on the long-term evolution of *V. cholerae* serves as a relevant platform to understand similar dynamics that are likely applicable to many bacterial diseases. In addition, there is significant interest in developing phage as biocontrol agents in human infections as well as in agriculture and food safety (*Doss* et al., 2017). Future therapies directed at using phages in therapeutic regimens will rely on a deeper understanding of the molecular mechanisms underpinning phage-host interactions in the context of human disease.

## ACKNOWLEDGEMENTS

The work related to this review was supported by the National Institute of Allergy and Infectious Diseases (USA). K.S. is a Chan Zuckerberg Biohub investigator.

# LITERATURE


Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P.: CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-1712 (2007).

d'Herelle, F. and Malone, R.H.: A Preliminary Report oí Work carried out by the Cholera Bacteriophage Enquiry. Ind. Med. Gaz. 62, 614-616 (1927).

David, L.A., Weil, A., Ryan, E.T., Calderwood, S.B., Harris, J.B., Chowdhury, F., Begum, Y., Qadri, F., LaRocque, R.C., and Turnbaugh, P.J.: Gut microbial succession follows acute secretory diarrhea in humans. mBio 6, e00381-15 (2015).

Doss, J., Culbertson, K., Hahn, D., Camacho, J., and Barekzi, N.: A review of phage therapy against bacterial pathogens of aquatic and terrestrial organisms. Viruses 9, 50 (2017).

Dy, R.L., Richter, C., Salmond, G.P.C., and Fineran, P.C.: Remarkable mechanisms in microbes to resist phage infections. Annu. Rev. Virol. 1, 307-331 (2014).

Faruque, S.M., Islam, M.J., Ahmad, Q.S., Faruque, A.S.G., Sack, D.A., Nair, G.B., and Mekalanos, J.J.: Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. Proc. Natl. Acad. Sci. USA 102, 6119-6124 (2005a).

Faruque S,M., Naser, I.B., Islam, M.J., Faruque, A.S.G., Ghosh, A.N., Nair, G.B., Sack, D.A., and Mekalanos, J.J.: Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. Proc. Natl. Acad. Sci. USA 102, 1702-1707 (2005b).

Finkelstein, R.A. and LoSpalluto, J.J.: Pathogenesis of experimental cholera. Preparation and isolation of choleragen and choleragenoid. J. Exp. Med. 130, 185-202 (1969).

Fu, Y., Waldor, M.K., and Mekalanos, J.J.: Tn-Seq analysis of Vibrio cholerae Intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. Cell Host Microbe 14, 652-663 (2013).

Harris, J.B., LaRocque, R.C., Qadri, F., Ryan, E.T., and Calderwood, S.B.: Cholera. Lancet 379, 2466-2476 (2012).

Havelaar, A.H., Kirk, M.D., Torgerson, P.R., Gibb, H.J., Hald, T., Lake, R.J., Praet, N., Bellinger, D.C., de Silva, N.R., Gargouri, N., Speybroeck, N., Cawthorne, A., Mathers, C., Stein, C., Angulo, F.J., Devleesschauwer, B.; World Health Organization Foodborne Disease Burden Epidemiology Reference Group: World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. PLoS Med. 12: e1001923 (2015).

James, C.E., Davies, E.V., Fothergill, J.L., Walshaw, M.J., Beale, C.M., Brockhurst, M.A., and Winstanley, C.: Lytic activity by temperate phages of Pseudomonas aeruginosa in long-term cystic fibrosis chronic lung infections. ISME J. 9, 1391-1398 (2014).

Kamp, H.D., Patimalla-Dipali, B., Lazinski, D.W., Wallace-Gadsden, F., and Camilli, A.: Gene fitness landscapes of Vibrio cholerae at important stages of its life cycle. PLoS Pathog. 9, e1003800 (2013).

Mandlik, A., Livny, J., Robins, W.P., Ritchie, J.M., Mekalanos, J.J., and Waldor, M.K.: RNA-Seq-based monitoring of infection-linked changes in Vibrio cholerae gene expression. Cell Host Microbe 10, 165-174 (2011).

Marraffini, L.A.: CRISPR-Cas immunity in prokaryotes. Nature 526, 55-61 (2015).

Merrell, D.S., Butler, S.M., Qadri, F., Dolganov, N.A., Alam, A., Cohen, M.B., Calderwood, S.B., Schoolnik, G.K., and Camilli, A.: Host-induced epidemic spread of the cholera bacterium. Nature 417, 642-645 (2002a).

Merrell, D.S., Hava, D.L., and Camilli, A.: Identification of novel factors involved in colonization and acid tolerance of Vibrio cholerae. Mol. Microbiol. 43, 1471-1491 (2002b).

Miller, V.L., Taylor, R.K., and Mekalanos, J.J.: Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. Cell 48, 271-279 (1987).

Nelson, E.J., Chowdhury, A., Harris, J.B., Begum, Y.A., Chowdhury, F., Khan, A.I., Larocque, R.C., Bishop, A.L., Ryan, E.T., Camilli, A., Qadri, F., and Calderwood, S.B.: Complexity of rice-water stool from patients with Vibrio cholerae plays a role in the transmission of infectious diarrhea. Proc. Natl. Acad. Sci. USA 104, 19091-19096 (2007).

Norman, J.M., Handley, S.A., Baldridge, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P., Stappenbeck, T.S., McGovern, D.P., Keshavarzian, A., Mutlu, E.A., Sauk, J., Gevers, D., Xavier, R.J., Wang, D., Parkes, M., and Virgin, H.W.: Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell 160, 447-460 (2015).

O'Hara, B.J., Barth, Z.K., McKitterick, A.C., and Seed, K.D.: A highly specific phage defense system is a conserved feature of the Vibrio cholerae mobilome. PLoS Genet. 13, e1006838 (2017).

Pasricha, C.L., de Monte, A., and Gupta, S.K.: Seasonal variations of cholera bacteriophage in natural waters and in man, in Calcutta during the year 1930. Ind. Med. Gaz. 66, 543-546 (1931).

Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I.: Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466, 334-338 (2010).

Samson, J.E., Magadán, A.H., Sabri, M., and Moineau, S.: Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. 11, 675-687 (2013).

Seed, K.D., Bodi, K.L., Kropinski, A.M., Ackermann, H.-W., Calderwood, S.B., Qadri, F., and Camilli, A.: Evidence of a dominant lineage of Vibrio cholerae-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. mBio 2, e00334-10 (2010).

Seed, K.D., Faruque, S.M., Mekalanos, J.J., Calderwood, S.B., Qadri, F., and Camilli, A.: Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in Vibrio cholerae O1. PLoS Pathog. 8, e1002917 (2012).

Seed, K.D., Lazinski, D.W., Calderwood, S.B., and Camilli, A.: A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature 494, 489-491 (2013).

Seed, K.D., Yen, M., Shapiro, B.J., Hilaire, I.J., Charles, R.C., Teng, J.E., Ivers, L.C., Boncy, J., Harris, J.B., and Camilli, A.: Evolutionary consequences of intra-patient phage predation on microbial populations. Elife. 3, e03497 (2014).

Taylor, R.K., Miller, V.L., Furlong, D.B., and Mekalanos, J.J.: Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. Proc. Natl. Acad. Sci. USA 84, 2833-2837 (1987).

Waldor, M.K. and Mekalanos, J.J.: Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272, 1910-1914 (1996).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 33-38 (2017).

# THE IMPORTANCE OF INSECT VIROMES: HUMAN HEALTH AND BEYOND

ERIC ROBERTO GUIMARÃES ROCHA AGUIAR,
ROENICK PROVETI OLMO, and JOÃO TRINDADE MARQUES

Department of Biochemistry and Immunology, Institute of Biological Sciences,
Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

## SUMMARY

Insects are the most abundant group of multicellular organisms and carry an incredible diversity of viruses. As such, insects can be seen as major reservoirs of viral biodiversity that can also be transmitted to other hosts. Therefore, the collection of viruses (e.g. virome) circulating in insects is of special interest. For example, we have seen many recent worldwide outbreaks by viruses transmitted by mosquitoes, such as Dengue and Zika. In addition, other components of the insect virome might not directly infect other hosts but could affect transmission of arboviruses by modulating vector competence. Finally, characterizing the incredible diversity of insect viruses could lead to the discovery of novel genes that can be tapped for biotechnological applications. Considering the importance of insect viromes, our group has used small RNA-based metagenomic approaches to identify and characterize viruses in different insects. This strategy has important advantages such as the fact that viral sequences are naturally enriched in the small RNA fraction of insects. In addition, the size distribution of small RNAs allows classification of viral sequences independent of homology searches against reference databases. Small RNA based metagenomics has greatly improved our ability to detect and identify novel viruses in insects thus allowing very detailed surveillance of insect viromes.

## INTRODUCTION

Viruses are obligatory intracellular pathogens that are fully dependent on the host machinery to replicate (*Marsh* and *Helenius*, 2006). These pathogens are found in all branches of the tree of life, from bacteria to higher eukaryotes (*Lauring* et al., 2013). Viruses are remarkably diverse, varying in nucleic acid content, genome structure and organization. Viral genomes can be DNA or RNA, single or double stranded, linear or circular containing one or multiple segments (*Edwards* and *Rohwer*, 2005).

Insects are reservoirs of an incredible diversity of viruses that is just beginning to be uncovered (*Aguiar* et al., 2015; *Li* et al., 2015; *Shi* et al., 2016). From a basic biology point of view, insect viruses help to understand the evolution of animal viruses. From an applied perspective, this immense viral biodiversity could also lead to the discovery of novel genes that could be tapped for biotechnological purposes.

Viruses that circulate in insects also include an important group of arthropod borne viruses (arboviruses) that

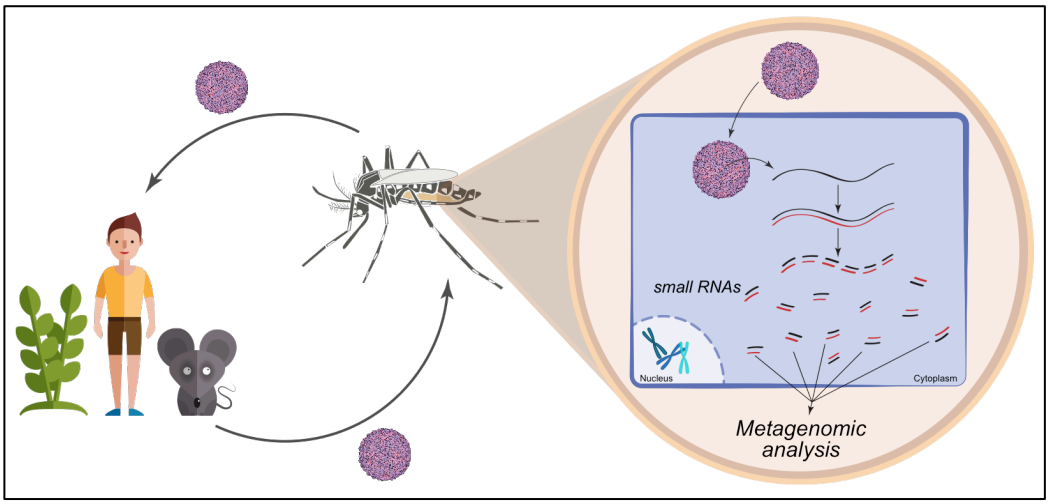


**Figure 1**: Insects are major reservoirs of viruses.
Insects are the most abundant multicellular organisms. Viruses circulating in insects represent the majority of the viral biodiversity in multicellular organisms at any given point. In addition, insects are able to directly transmit viruses that can infect plants, humans, and other vertebrate animals. Therefore, characterizing viruses in insects is of great importance for basic biology as well as for human health. Metagenomic strategies allow a general overview of the collection of organisms found in any sample. In insects specifically, small RNA based metagenomics is a sensitive approach to identify viruses since virus-derived small RNAs are commonly found in insects. These virus-derived small RNAs are generated by processing of viral RNAs by the host antiviral response and can be used to reconstruct virus sequences.

can be transmitted to other hosts (Figure 1) (*Liang* et al., 2015). Arboviral infections still represent a major threat to humans, livestock and agriculture (*Shepard* et al., 2013; *Cargnelutti* et al., 2014; *Diez-Domingo* et al., 2014; *Redinbaugh* and *Zambrano*, 2014). We highlight recent increases in human infections by arboviruses, such as Zika, Dengue and Chikungunya virus (*Vijayakumar* et al., 2013; *Carrington* and *Simmons*, 2014; *Diallo* et al., 2014). Despite their importance, we still lack effective vaccines or treatments for diseases caused by most of arboviruses. Monitoring of viruses circulating in vector insects is still an important strategy to prevent outbreaks. Constant surveillance allows early detection of viruses that could eventually cause outbreaks (*Gubler*, 2001; *Parrish* et al., 2008).

For reasons highlighted above, characterization of the collection of viruses in insects (e.g. the virome) is extremely important. In this scenario, metagenomics has been an important tool to study viral diversity.

## SURVEILLANCE OF INSECT VIROMES UTILIZING METAGENOMIC STRATEGIES

Metagenomics allow broad analysis of genetic material within a sample without any previous knowledge. Different metagenomic strategies have been successfully applied to virus discovery based on both DNA and RNA sequencing with their own advantages and limitations (*Kreuze* et al., 2009; *Yamao* et

al., 2009; *Webster* et al., 2015; *Hang* et al., 2016). DNA sequencing tends to bias the identification to microorganism that have DNA at some stage of their life cycle, which is not the case of all viruses (*Venter* et al., 2004; *Edwards* and *Rohwer*, 2005). In contrast, RNA sequencing is a more unbiased approach since all viruses generate RNA during their replication cycle. Many studies have taken advantage of RNA sequencing as proxy to infer presence of viruses in diverse organisms (*Kreuze* et al., 2009; *Zhuang* et al., 2014; *Li* et al., 2015; *Webster* et al., 2015). These studies have relied on sequencing of long or small RNA fractions derived from infected hosts. Interestingly, while long RNAs represent direct products of the viral replication cycle, small RNAs are a result of the antiviral response. Indeed, antiviral mechanisms commonly target exposed viral RNAs that are further degraded to generate virus-derived small RNAs.

Previous work from our group indicated that small RNAs are enriched for viral sequences in comparison to long RNAs (Figure 1) (*Aguiar* et al., 2015). We showed that small RNAs also allow identification and classification of viral sequences through pattern-based analyses independent of homology to known references (*Aguiar* et al., 2016). The requirement for sequence similarity comparisons is a major limitation of metagenomics studies especially considering that viruses have high mutation rates. Identification through sequence homology searches requires considerable similarity to previously identified viruses. In order to overcome this limitation, sequence-independent strategies may increase chances of identifying highly divergent viruses. Since small RNAs are products of specific host antiviral pathways, they show unique molecular signatures that can be used to infer their origin (*Aguiar* et al., 2016). Unfortunately, small RNA pattern-based analyses are mainly limited to organisms and conditions where there is a functional antiviral RNAi pathway. Absence or inhibition of the RNAi pathway may make it difficult to perform pattern-based analyses.

## THE IMPORTANCE OF STUDYING THE DIVERSITY OF VIRUSES CIRCULATING IN INSECTS

Insects are important hosts for viral infection since they are the most abundant group of multicellular organisms. It is becoming apparent that the collection of viruses found in insects is highly diverse (*Li* et al., 2015; *Shi* et al., 2016). The collection of viruses circulating in insects includes most viral families circulating in animals (*Shi* et al., 2016). Therefore, studies on insect viromes can provide important information about viral biodiversity and evolution. The characterization of this incredible biodiversity in insects could also lead to the discovery of novel genes with potential biotechnological applications.

In addition, insects are major reservoirs of viral biodiversity that can be transmitted to plants, humans and other vertebrate animals thus threatening agriculture, livestock and public health (*Shepard* et al., 2013; *Cargnelutti* et al., 2014; *Diez-Domingo* et al., 2014; *Redinbaugh* and *Zambrano*, 2014). More than 100 different human arboviruses have been described although this is likely not to be the final number (*Gubler*, 2001). Of note, we have also seen recent worldwide outbreaks by

different mosquito-borne viruses including *Dengue virus* (DENV), *Chikungunya virus* (CHIKV) and *Zika virus* (ZIKV) (*Vijayakumar* et al. 2013; *Carrington* and *Simmons*, 2014; *Diallo* et al., 2014). In addition to the human and social impact, the economical burden of insect borne viruses is enormous. Only in the Americas, between 2000 and 2007, the estimated economic impact of Dengue was of approximately US$ 2.1 billion annually (*Shepard* et al., 2011).

The study of insect viromes can directly impact public health by providing insights into the circulation of arboviruses and help prevent outbreaks. Nevertheless, it is important to note that insects also carry a large diversity of viruses that are unlikely to directly infect vertebrate hosts (*Bolling* et al., 2015). These insect specific-viruses are not infectious to other organisms but may indirectly modulate the transmission of arboviruses by affecting vector competence (e.g. the ability of the insect to function as a vector).

## CONCLUSION

In summary, we have highlighted here the importance of characterizing insect viromes and how it can impact human health, biotechnology and evolutionary biology. New strategies such as small RNA based metagenomics have impacted the characterization of viruses circulating in insects. Studies on insect viromes are central to the understanding of viral evolution since insect viruses are the largest and most diverse group in multicellular organisms. In addition, knowledge on the circulation of viruses in insects would help prevent outbreaks. Finally, these diverse insect viruses could lead to the discovery of novel genes with potential use in the biotechnology industry.

## ACKNOWLEDGEMENTS

We would like to thank members of our laboratory for invaluable suggestions and criticism. This work was a collaboration with the laboratory of Dr. Jean-Luc Imler at the University of Strasbourg and was supported by grants from CAPES, CNPq, and FAPEMIG.

## LITERATURE

Aguiar, E.R., Olmo, R.P., Paro, S., Ferreira, F.V., de Faria, I.J., Todjro, Y.M., Lobo, F.P., Kroon, E.G., Meignin, C., Gatherer, D., Imler, J.L., and Marques, J.T.: Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. Nucleic Acids Res. 43, 6191-6206 (2015).

Aguiar, E.R., Olmo, R.P., and Marques, J.T.: Virus-derived small RNAs: molecular footprints of host-pathogen interactions. Wiley Interdiscip. Rev. RNA 7, 824-837 (2016).

Bolling, B.G., Weaver, S.C., Tesh, R.B., and Vasilakis, N.: Insect-Specific Virus Discovery: Significance for the Arbovirus Community. Viruses 7, 4911-4928 (2015).

Cargnelutti, J.F., Olinda, R.G., Maia, L.A., de Aguiar, G.M., Neto, E.G., Simoes, S.V., de Lima, T.G., Dantas, A.F., Weiblen, R.,

Flores, E.F., and Riet-Correa, F.: Outbreaks of Vesicular stomatitis Alagoas virus in horses and cattle in northeastern Brazil. J. Vet. Diagn. Invest. 26, 788-794 (2014).

Carrington, L.B. and Simmons, C.P.: Human to mosquito transmission of dengue viruses. Front. Immunol. 5, 290 (2014).

Diallo, D., Sall, A.A., Diagne, C.T., Faye, O., Faye, O., Ba, Y., Hanley, K.A., Buenemann, M., Weaver, S.C., and Diallo, M.: Zika virus emergence in mosquitoes in southeastern Senegal, 2011. PLoS One 9, e109442 (2014).

Diez-Domingo, J., Perez-Yarza, E.G., Melero, J.A., Sanchez-Luna, M., Aguilar, M.D., Blasco, A.J., Alfaro, N., and Lazaro, P.: Social, economic, and health impact of the respiratory syncytial virus: a systematic search. BMC Infect. Dis. 14, 544 (2014).

Edwards, R.A. and Rohwer, F.: Viral metagenomics. Nat. Rev. Microbiol. 3, 504-510 (2005).

Gubler, D.J.: Human arbovirus infections worldwide. Ann. NY Acad. Sci. 951, 13-24 (2001).

Hang, J., Klein, T.A., Kim, H.C., Yang, Y., Jima, D.D., JRichardson, J.H., and Jarman, R.G.: Genome Sequences of Five Arboviruses in Field-Captured Mosquitoes in a Unique Rural Environment of South Korea. Genome Announc. 4, e01644-15 (2016).

Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., and Simon, R.: Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. Virology 388, 1-7 (2009).

Lauring, A.S., Frydman, J., and Andino, R.: The role of mutational robustness in RNA virus evolution. Nat. Rev. Microbiol. 11, 327-336 (2013).

Li, C.X., Shi, M., Tian, J.H., Lin, S.D., Kang, Y.J., Chen, L.J., X. Qin, X.C., Xu, J., Holmes, E.C., and Zhang, Y.Z.: Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife 4, e05378 (2015).

Liang, G., Gao, X., and Gould, E.A.: Factors responsible for the emergence of arboviruses; strategies, challenges and limitations for their control. Emerg. Microbes Infect. 4, e18 (2015).

Marsh, M. and Helenius, A.: Virus entry: open sesame. Cell 124, 729-740 (2006).

Parrish, C.R., Holmes, E.C., Morens, D.M., Park, E.C., Burke, D.S., Calisher, C.H., Laughlin, C.A., Saif, L.J., and Daszak, P.: Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol. Mol. Biol. Rev. 72, 457-470 (2008).

Redinbaugh, M.G. and Zambrano, J.L.: Control of virus diseases in maize. Adv. Virus Res. 90, 391-429 (2014).

Shepard, D.S., Coudeville, L., Halasa, Y.A., Zambrano, B., and Dayan, G.H.: Economic impact of dengue illness in the Americas. Am. J. Trop. Med. Hyg. 84, 200-207 (2011).

Shepard, D.S., Undurraga, E.A., and Halasa, Y.A.: Economic and disease burden of dengue in Southeast Asia. PLoS Negl. Trop. Dis. 7, e2055 (2013)..

Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X.C., Li, J., Cao, J.P., Eden, J.S., Buchmann, J., Wang, W., Xu, J., Holmes, E.C., and Zhang, Y.Z.: (2016). Redefining the invertebrate RNA virosphere. Nature 540, 539-543 (2016).

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., and Smith, H.O.: Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66-74 (2004).

Vijayakumar, K., George, B., Anish, T.S., Rajasi, R.S., Teena, M.J., and Sujina, C.M.: Economic impact of chikungunya epidemic: out-of-pocket health expenditures during the 2007 outbreak in Kerala, India.

Southeast Asian J. Trop. Med. Public Health 44, 54-61 (2013).

Webster, C.L., Waldron, F.M., Robertson, S., Crowson, D., Ferrari, G., Quintana, J.F., Brouqui, J.M., Bayne, E.H., Longdon, B., Buck, A.H., Lazzaro, B.P., Akorli, J., Haddrill, P.R., and Obbard, D.J.: The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster. PLoS Biol. 13, e1002210 (2015).

Yamao, T., Eshita, Y., Kihara, Y., Satho, T., Kuroda, M., Sekizuka, T., Nishimura, M., Sakai, K., Watanabe, S., Akashi, H., Rongsriyam, Y., Komalamisra, N., Srisawat, R., Miyata, T., Sakata, A., Hosokawa, M., Nakashima, M., Kashige, N., Miake, F., Fukushi, S., Nakauchi, M., Saijo, M., Kurane, I., Morikawa, S., and Mizutani, T.: Novel virus discovery in field-collected mosquito larvae using an improved system for rapid determination of viral RNA sequences (RDV ver4.0). Arch. Virol. 154, 153-158 (2009).

Zhuang, L., Zhang, Z., An, X., Fan, H., Ma, M., Anderson, B.D., Jiang, J., Liu, W., Cao, W., and Tong, Y.: An efficient strategy of screening for pathogens in wild-caught ticks and mosquitoes by reusing small RNA deep sequencing data. PLoS One 9, e90831 (2014).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 39-52 (2017).

# CNIDARIA - AN EMERGING MODEL PHYLUM TO INVESTIGATE THE EVOLUTION OF METAZOAN IMMUNITY

STEVEN D. QUISTAD

Laboratory of Genetic Evolution, Superior School of Industrial Physics and Chemistry of the City of Paris (ESPCI), Paris, France

## SUMMARY

The Cnidarian phylum (sea anemones, jellyfish, *Hydra*, corals) is considered to be phylogenetically basal to Bilateria (files, worms and humans). Due to their relative position within the tree of life, Cnidarians provide insight into the genetic toolkit of the last common metazoan ancestor. Their simple body plan consists of only two epithelial cell layers connected by a jelly-like mesoglea. Despite their morphological simplicity, the cnidarian immune system is surprisingly complex, with multiple immune components conserved from Cnidarians to humans. In addition, the cnidarian immune system appears to be more similar to the human immune system compared to canonical model organisms such as *Drosophila melanogastor* and *Caenorhabditis elegans*. Here, I begin with a brief overview of cnidarian immunology and demonstrate that while the field has advanced rapidly in the past decade, large gaps exist that warrant additional investigations. Next, I outline a novel approach to identify predicted immune proteins in Cnidarians that are missed using traditional protein annotation methods. This approach is based on the identification of similarities between viral gene products and host proteins and can be applied to virtually any animal system.

## INTRODUCTION

### Cnidarians: Aggressors of oceans and protectors of continents

The phylum Cnidaria (sea anemones, jellyfish, *Hydra*, and corals) contains over 10,000 species ranging in size from a few millimetres to over 75 meters and inhabits both fresh and saltwater environments across the globe (*Daly* et al., 2007). The lifestyle of a Cnidarian is either completely sessile or slow moving. Therefore, it would be reasonable to assume that they are relatively docile organisms. However, at the cellular level they are amongst the most aggressive hunters of the aquatic environment. Using phylum-specific molecular harpoons, Cnidarians inject host tissue with toxic venom capable of paralyzing prey and even killing humans (*Beckmann* and *Özbek*, 2012). At the macro-level a specific group of Cnidarians, reef-building corals, are responsible for the only biologically generated structure that can be viewed from space: the coral reef. Despite covering less than 0.5% of the ocean surface, coral reefs support almost one third of all marine fish species, providing both food and coastal-protection against storms (*Moberg* and *Folke*, 1999. Whether attending to a jellyfish sting or consuming one of the many

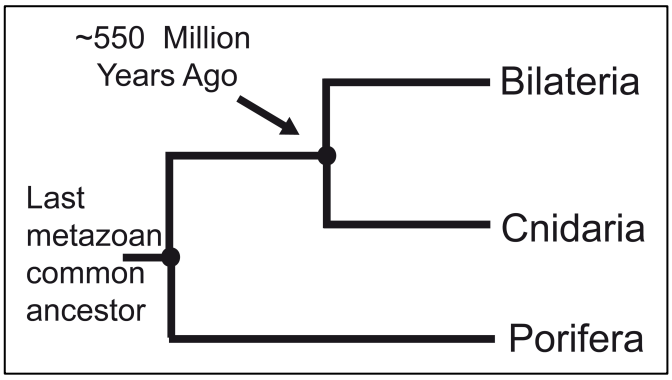


**Figure 1**: Cnidarians are phylogenetically basal to Bilaterians. The Bilaterian phylum diverged from Cnidarians approximately ~550 million years ago predating the Cambrian Explosion.

species of fish supported by the coral reef ecosystem, millions of people are directly impacted by Cnidarians.

**Cnidarians: A basal metazoan phylum**

In addition to their modern-day impact, Cnidarians also occupy a basal position within the animal tree of life. The first cnidarian fossils were discovered in Cambrian-era rock and can be traced back ~550 million years ago during a period referred to as the Cambrian Explosion (*Van Iten* et al., 2016). Often referred to as the "Big Bang" of life on Earth, the Cambrian Explosion describes a dramatic period in the fossil record during which the ancestors of all extant animal phyla can be found (*Shu*, 2008). Molecular phylogenetic data combined with fossil evidence suggest the first Cnidarians existed approximately 542-720 Myr ago (*Davidson* and *Erwin*, 2009) and it was during this period that their sister group Bilateria, which includes flies, worms, and humans, diverged from the last common metazoan ancestor (Figure 1) (*Kortschak* et al., 2003; *Hemmrich* et al., 2012). Therefore, Cnidarians are considered to be amongst the most phylogenetically basal to all metazoan life and provide important insight into the ancestral state of the first metazoans.

The relatively simple Cnidarian body plan consists of two cell layers, an endoderm and ectoderm, held together by the jelly-like mesoglea. Due to the morphological simplicity and basal status, the Cnidarian phylum has traditionally been viewed as "primitive" and less complex than "higher" organisms such as flies and worms. However with the genomic sequencing of *Nematostella vectensis* (*Putnam* et al., 2007), *Hydra magnipapillata* (*Chapman* et al., 2010), and *Acropora digitifera* (*Shinzato* et al., 2011), as well as numerous functional studies, it has become clear the "primitive" hypothesis is consistently unsupported with regards to development (*Kusserow* et al., 2005; *Desvignes* et al., 2010) and immunity (*Miller* et al., 2007; *Zmasek* et al., 2007; *Franzenburg* et al., 2012; *Quistad* et al., 2014). In addition, many gene families that have been lost in canonical model organisms such as *Caenorhabditis elegans* and *Drosophila melanogaster* are present in Cnidarians (*Miller* et al., 2005; *Technau* et al., 2005).

Model organisms have provided invaluable insight into the origin of immunity and how the human immune system functions. For example, studies in chickens led to the identification of

T and B cells, while self-versus non-self was first investigated in starfish (*Cooper* et al., 1965; *Litman* and *Cooper*, 2007). However, our current perspective of the evolution of immunity is largely based on only three of the thirty extant phyla (Bilateria, Nematoda, Arthropoda), therefore, a broader representation of phyla is required (*Litman* and *Cooper*, 2007)**.** Here, I argue that the conservation and complexity of the Cnidarian immune system, combined with well-developed molecular tools, strongly supports Cnidarians as a valuable and emerging model phylum to investigate the ancestral state of the animal immune system. First, I briefly review the status of cnidarian immune research focusing on studies that involve direct experimentation. Then, I present a novel approach to investigate the immune systems of Cnidarians based on predicted viral-host interactions that can be applied to virtually any animal system.

## THE CNIDARIAN IMMUNE SYSTEM - ANCIENT AND CONSERVED

### Mucosal surfaces: Dynamic regions of host-microbe interactions

Mucosal surfaces likely appeared with the first Cnidarians and can now be found throughout the metazoan tree of life (*Lang* et al., 2007). The Surface-Mucus Layer (SML) is composed of a multi-layered matrix of glycoproteins (mucins), located at the apical region of specific epithelial cell types (*Bäckhed* et al., 2005). Glycosylation of mucins is tightly controlled by the host repertoire of glycosyltransferases resulting in a diverse array of mucin macromolecules that either remain tethered to the cell surface or is secreted into the surrounding environment (*Ferez-Vilar* and *Hill*, 1999; *Hang* and *Bertozzi*, 2005). These energy-rich mucins not only provide food to commensal bacteria (*Derrien* et al., 2010) but also serve as a particle trap for their major predators: the bacteriophage (phage). Despite high mucin turnover, metazoans maintain specific phage and bacteria species within the SML (*Lozupone* et al., 2012; *Grasis* et al., 2014) and alterations of the SML microbiome have been associated with various disease states from Cnidarians (*Closek* et al., 2014) to humans (*Johansson* et al., 2013). Resident bacteria protect their hosts via niche exclusion of pathogenic species or strains while phage can directly lyse invading pathogens that traverse the SML (*Barr* et al., 2013). In the Bacteriophage Adherence to Mucus (BAM) model, mucus-adherent phages bind to host mucins and provide the host with immune protection (*Barr* et al., 2013, 2015). For an in-depth discussion of the origin and evolution of microbiome selection within the first metazoan SML the readers are pointed to the article of *Quistad* et al. (2016a). Taken together the physical properties of the SML combined with the host-specific microbiome provide the first line of immune protection against invading pathogens.

### Detection and response to microbes: The Toll-Like Receptors (TLRs)

While the SML provides a dynamic layer of immune defence, pathogens have evolved various mechanisms to traverse the SML and invade host cells. To detect and respond to extracellular and intracellular microbes, metazoans utilize a class of Pattern Recognition Receptors (PRRs) that rely on Microbe-Associated Molecular Patterns (MAMPs). One of the primary classes of PRRs are the Toll-Like Receptors

(TLRs) (*Muzio* et al., 2000). Upon activation by MAMPs, TLRs recruit the adaptor protein MyD88 resulting in a range of host responses promoting either cell-survival (*Iwasaki* and *Medzhitov*, 2004; *Redfern* et al., 2011) or elimination of the pathogen via programmed cell death (apoptosis) (*Aliprantis* et al., 2000).

The Toll-pathway was first described in *Drosophila* (*Anderson* et al., 1985) and has been shown to be involved with both development (*Wang* et al., 2005) and pathogen defence (*Rosetto* et al., 1995). Additional work in *C. elegans* demonstrated components of the Toll-signalling pathway are present, however, the central proteins involved with TLR-signalling are lacking. Based on these data, the TLR-pathway and its role in immunity was proposed to have evolved within Bilaterians (*Kim* and *Ausubel*, 2005). However, the *Nematostella*, *Hydra*, and *Acropora* genomes revealed that the major components of TLR-signalling are present. To investigate whether the TLR-signalling pathway is functional in *Hydra*, MyD88-deficient and germ-free *Hydra* were generated and the transcriptional profiles were determined. Multiple components central to TLR-signalling were down-regulated in MyD88-deficient and germ-free *Hydra* including members of the TRAF family, MAP-kinase p38, and the kinase TAK1. To determine the role of TLR-signalling in the establishment of the resident microbiome, germ-free MyD88 and wild-type *Hydra* were generated and reinfected with complex microbial communities. MyD88-deficient *Hydra* were found to exhibit a delayed response in bacterial recolonization suggesting TLR-signalling plays a role in the establishment of *Hydra*-associated microbiome (*Franzenburg* et al., 2012). While functional studies have yet to be performed in other cnidarians, the *Acropora digitifera* genome suggests the TLR repertoire is significantly more complex than *Nematostella* or *Hydra* in terms of total TLR proteins present and associated protein domains (*Shinzato* et al., 2011). Future work should focus on determining the binding partners of other cnidarian TLRs, the associated signalling cascades, and the resulting cellular response. Taken together these data support the hypothesis that TLR-signalling and its role in immunity predates the evolutionary split between Bilaterians and Cnidarians. For a more detailed discussion of host-microbe interactions in *Hydra* the reader is pointed to the review by *Schröder* and *Bosch* (2016).

**Detection of intracellular microbes: The Nod-Like Receptors (NLRs)**

If a microorganism is able to successfully traverse the SML and enter a host cell, Nod-Like Receptors (NLRs) are involved with the detection of MAMPs and activation of the associated cellular response. In humans, NLRs consist of a central NACHT domain, a N-terminal effector domain, and a C-terminal Leucine-Rich Repeat (LRR) that directly binds to bacterial MAMPs (*Hansen* et al., 2011). Activation of NLRs results in the formation of a specialized structure called the inflammasome which is involved with caspase activation and subsequent processing of proinflammatory cytokines (*Franchi* et al., 2009).

While cnidarian intracellular NLRs are relatively understudied compared to the mucosal immune system, *in vitro* experimentation in *Hydra* suggests NLRs may play a conserved role in inflammasome formation. Specifically, the HyNLR Type I protein was found to co-immunoprecipitate with the *Hydra* Caspase 1, a central component of the mature inflammasome. In addition, RT-PCR analysis detected an up-regulation of the HyNLR Type I gene

in response to bacterial stimulation (*Lange* et al., 2011). Bioinformatic analysis of the *Acropora* NLR repertoire identified more NACHT-domain encoding genes (~500 predicted proteins) than any metazoan investigated thus far, including humans, and includes many novel domain combinations (*Hamada* et al., 2013).

**The Cnidarian complement system: Symbiosis or cell lysis?**

In addition to being directly targeted by antimicrobial peptides, pathogenic microorganisms can also be targeted for destruction by the complement system. Upon activation, the complement cascade labels foreign cells for direct lysis by host proteins. While multiple pathways can lead to complement activation, all converge on the central effector molecule C3, which is cleaved into C3a and C3b by the C3 convertase complex (*Gros* et al., 2008). C3a subsequently activates an inflammatory response while C3b remains attached to the microbe leading to phagocytosis by the host cell and ultimately cell lysis by the membrane attack complex (MAC) (*Noris* and *Remuzzi*, 2013). Three major pathways lead to complement activation: the classical pathway involving antigen-antibody complexes, the lectin pathway involving Mannose-Binding Lectins (MBL) and the alternative pathway. In Cnidarians C3 and components of both the MBL and alternative pathways are present however, many of the proteins involved with the formation of the MAC are absent, suggesting that the ancestral role of complement may have been the promotion of phagocytosis rather than lysing of invading cells (*Cerenius* et al., 2010). This hypothesis is supported by investigations into the role of the complement pathway in the establishment and maintenance of cnidarian-algal symbiosis. In the reef-building coral *Acropora millepora*, the MBL Millectin was found to bind directly to symbiotic *Symbodinium* species *in vivo* (*Kvennefors* et al., 2010). In addition, members of the complement pathway were activated in response to bacterial challenge in corals though direct opsonisation of cells was not observed (*Kvennefors* et al., 2008, 2010). To determine whether the ancestral role of the metazoan complement system was lysis, symbiosis, or both, future work should directly test the impact of complement binding on cellular viability across a range of symbionts, pathogens, and cnidarian species.

**Apoptosis: The final defence against invading microorganisms**

In response to an invading pathogen, the host cell can also elect to undergo apoptosis or programmed cell death in an effort to prevent further dissemination of the pathogenic entity (*Barber*, 2001). While many types of apoptosis exist, metazoans appear to be unique through their use of Tumour Necrosis Factor Receptors (TNFRs) (*Quistad* and *Traylor-Knowles*, 2016). The TNF Receptor-Ligand superfamily (TNFRSF) is a central mediator of apoptosis and misregulation of the TNFRSF is involved in a variety of inflammatory disorders including multiple sclerosis, type 2 diabetes, and rheumatoid arthritis (*Bahia* and *Silakari*, 2010). Upon ligand binding, TNFR activation can lead to activation of the NF-κB transcription factor (among others), promoting cell survival, or associate with the Fas-Associated Death domain Protein (FADD) resulting in caspase activation and apoptosis (*Lin* et al., 1999; *Micheau* and *Tschopp*, 2003).

Prior to the sequencing of the *Acropora* genome, expansion of the TNF ligand-receptor superfamily was predicted to have occurred with the emergence of adaptive immunity in

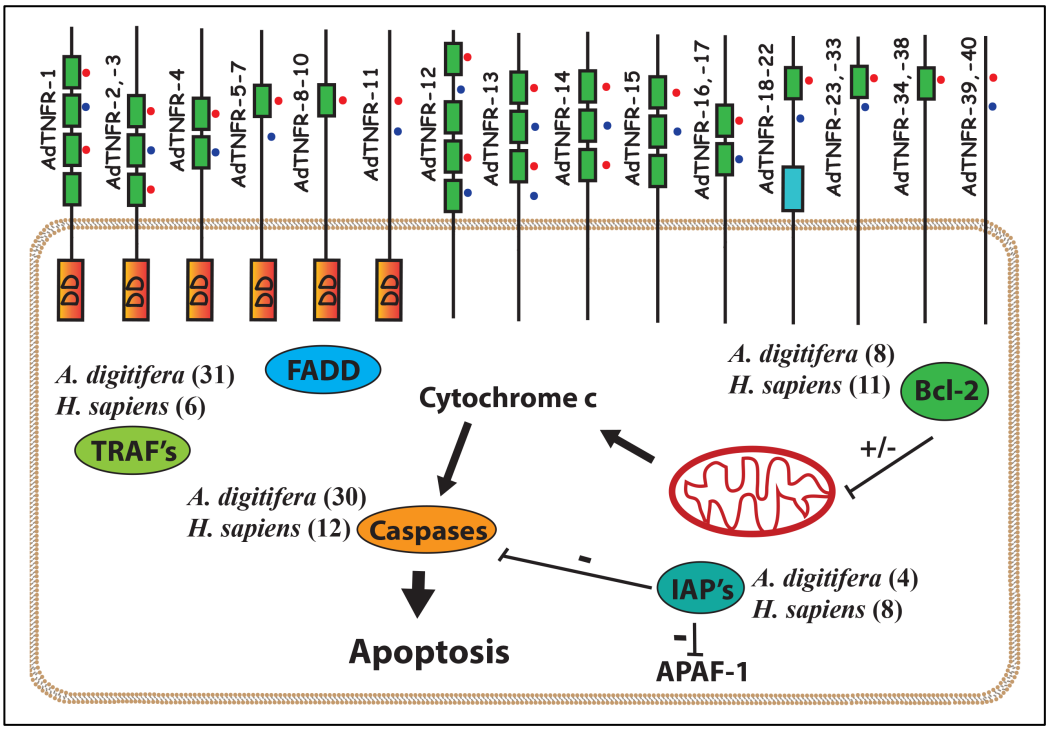


**Figure 2**: The coral apoptotic repertoire. The putative TNFR repertoire of *Acropora digitifera* (top) with Death Domain (DD), Cysteine Rich Domain (green boxes), Immunoglobin Domain (blue box), 50s loop-TNF binding site (red dot), and 90s loop TNF binding site (blue dot) indicated. Members of the Death Receptor Signalling Pathway (bottom) found in the *Acropora digitifera* genome with number of proteins within a specific protein family indicated for both *Acropora digitifera* and *Homo sapiens* including TNF-Receptor Associated Factors (TRAFs), B-Cell Lymphoma family members (Bcl-2), Inhibitor of Apoptosis proteins (IAP's), FADD, APAF-1 and Caspases. (From: *Quistad* et al., 2014).

vertebrates (*Wiens* and *Glenney*, 2011). Corals and other Cnidarians were predicted to have one, if any, members of the TNFRSF. However, genomic analysis of the *Acropora digitifera* genome led to the discovery of 40 predicted TNFRs compared to 29 found in humans (*Quistad* et al., 2014). In addition to TNFRs, the *Acropora* genome was found to have all of the central components of the canonical apoptotic cascade including three TRAF proteins, thirty caspases, and eight members of the Bcl-2 family (Figure 2) (*Shinzato* et al., 2011). This conservation led us to wonder what might happen if coral cells were exposed to a human TNF. Exposure of coral cells to Human TNFα was found to cause apoptotic blebbing, caspase activation, cell death, and finally coral bleaching. Next, human T-cell lymphocytes were exposed to a member of the coral TNFSF and it was found that coral TNF also caused apoptosis in humans, demonstrating remarkable conservation of TNF-induced apoptosis across 550 million years of evolution (*Quistad* et al., 2014). Further investigations into the origin of TNFRs suggests their role in apoptosis evolved before the Cnidarian-Bilateria split, though the individual domains of the TNFR protein are even more ancient (*Quistad*

and *Traylor-Knowles*, 2016).

Additional work involving other cnidarian apoptotic proteins has further supported a Precambrian origin for the canonical apoptotic cascade (*David* et al., 2005; *Moya* et al., 2016). For example, expression of a coral caspase induced cell death in mammalian cells and coral FADD protein was found to directly associate with a zebrafish caspase (*Sakamaki* et al., 2014). TNFRs have also been implicated in the ability of coral to resist heat stress which is expected to increase in frequency with future climate change (*Hoegh-Guldberg* et al., 2007). Specific TNFR genes were found to be "front-loaded" in corals that were naturally heat-resistant compared to corals of the same species that were heat-sensitive revealing a genomic basis of coral resilience (*Barshis* et al., 2013). While functional investigations into cnidarian apoptosis are still in their infancy they have already revealed novel and exciting insights into the origins and evolution of apoptosis.

**The Cnidarian immune system: Complex, conserved, minimally explored**

Many publications to date investigating Cnidarians use the term "unexpected complexity" however it is now clear this phrase is no longer appropriate. At the molecular level Cnidarians have proven to be extremely complex and often more similar to humans when compared to established model organisms such as *Drosophila* and *C. elegans*. The basic properties and processes of the SML and mucosal immunity in general began in Cnidarians and continue to operate in the human gut (*Schröder* and *Bosch*, 2016). Immune receptors such at NLRs are more diverse in Cnidarians than any animal investigated thus far (*Lange* et al., 2011). Investigations into the origin and evolution of fundamental immune processes such as TLR-signalling and apoptosis have forced us to rethink multiple decade-long hypotheses (*Kim* and *Ausubel*, 2005; *Wiens* and *Glenney*, 2011). In summation, the Cnidarian phylum has already provided us with unexpected and novel insights into the ancestral state of the first animals, however, all investigations thus far have relied on previously characterized immune proteins described in other systems. How can we begin to understand immune processes that may be phylum-specific or completely novel?

## USING VIRUSES TO PREDICT NOVEL IMMUNE PROTEINS

Viruses are the master manipulators of host immunity (*Tortorella* et al., 2000). As obligate intracellular parasites, viruses must be able to complete their life cycle while avoiding immune detection. To successfully invade and replicate within a host cell, viruses express proteins that mimic host immune proteins (*Liang* et al., 2008; *Hagai* et al., 2014). For example, if a virally encoded cytokine was compared to the human proteome *in silico*, then the corresponding human cytokine could be identified without any *a priori* knowledge of the human immune system. An expansion of this concept led to the hypothesis that an *in silico* comparison of all viral gene products against the host proteome would identify both known and potentially novel immune-associated proteins. To test this hypothesis we created a mock viral metagenome using 16 human viruses and compared them against the human proteome using tBLASTn revealing multiple proteins involved with

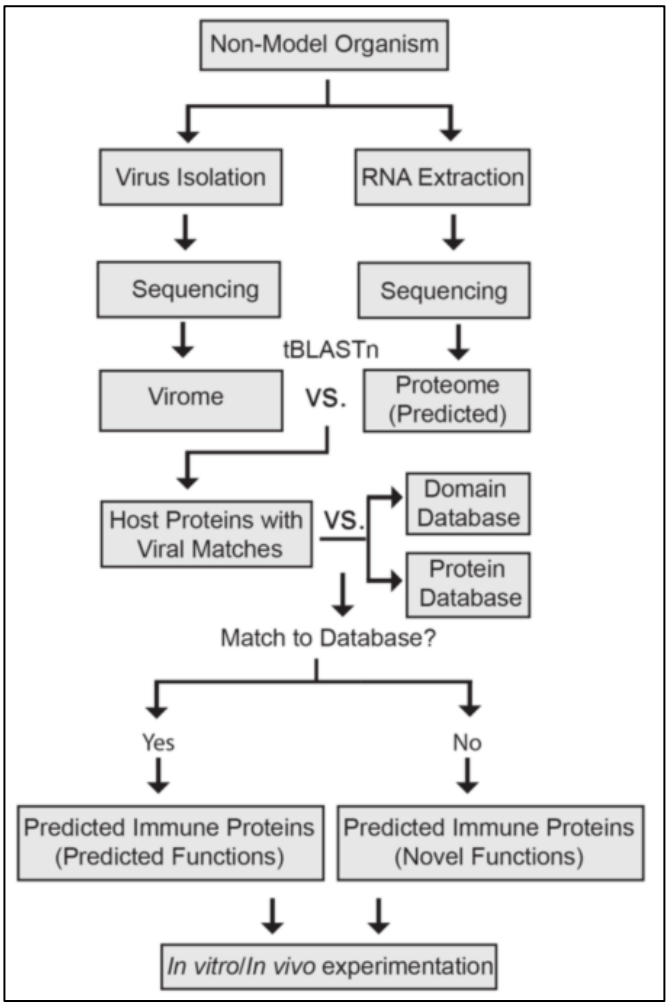


**Figure 3**: Proposed pipeline to predict immune proteins in uncharacterized systems. Viruses are extracted from the non-model organism of interest using a Virus-Like Particle extraction protocol of choice and nucleic acid is sequenced. If a gene model is not available total RNA is also extracted from host tissue and an assembled transcriptome is prepared. Viral gene segments are compared to the translated transcriptome through tBLASTn to identify matches to host proteins. These proteins are further analysed through comparison to a well-characterized immune system using BLASTp (e.g., human or mouse) and domain prediction database (e.g., Conserved Domain CDD, Pfam). Proteins that lack hits to either database represent predicted immune genes with unknown function and are selected for further biochemical investigations using *in vitro* and *in vivo* experimentation. (From: *Quistad* et al., 2016b).

complement activation, apoptosis, and cytokine signalling (*Quistad* et al., 2016b). Next the same analysis was performed using a coral viral metagenome. As expected, the group of coral proteins identified through the virome analysis were significantly enriched for pathogen-sensing and apoptotic proteins. In total 159 coral proteins were identified that contain no known protein domains, suggesting that they may be involved in novel immune processes. Based on these results, a general pipeline to predict immune proteins was created that can be applied to virtually any animal

system. First, viruses and total RNA are extracted from the organism of interest and nucleic acid is sequenced. Next the virome is compared to an assembled host transcriptome and host proteins matching viral gene segments are identified. The group of proteins with viral matches are then compared to well-characterized host proteomes (i.e. model organisms) and protein domain databases to determine if their function can be predicted. Proteins that fail to match previously characterized proteins or contain previously annotated domains are candidate genes predicted to be involved with phylum-specific immunity or novel immune processes. These candidate proteins can be investigated further using *in vivo* and *in vitro* experimentation (Figure 3). The proposed pipeline combines the power of existing databases with new predictions generated by viral communities providing a more comprehensive prediction of the host immune repertoire (*Quistad* et al., 2016b). While the conserved portion of the Cnidarian immune system has already been identified based on protein annotations from other systems, this domain-independent approach provides novel protein targets to potentially discover cnidarian-specific immune processes.

## FUTURE DIRECTIONS

The majority of published work focusing on cnidarian immunity has been performed in the *Hydra* system due to its ease of culture, fully sequenced genome, and well-developed tools for genetic manipulation (*Bosch*, 2013). Cnidarians are divided into two subphyla: Medusozoa (*Hydra* and jellyfish) and Anthozoa (corals and sea anemones) therefore our current understanding of cnidarian immunity is currently biased towards the Medusozoans. While there have been many important discoveries in cnidarian immunology using the coral system, determining the underlying mechanisms is challenging due to lack of molecular tools. To broaden our understanding of cnidarian immunity Anthozoans with developed molecular tools such as *Nematostella vectensis* should also be utilized (*Darling* et al., 2005). Currently, the *Nematostella* system provides a sequenced genome (*Putnam* et al., 2007), gene-knockout techniques (*Ikmi* et al., 2014), and the ability to create transgenic animal lines (*Renfer* et al., 2010) yet, essentially the entire field uses these tools to address development-based questions, leaving the *Nematostella* immune system uncharacterized. By continuing to expand research into *Hydra* and coral immunity as well as developing additional species to investigate mechanistic questions such as *Nematostella*, we will obtain a more comprehensive understanding of the cnidarian immunity and thereby better understand the ancestral state of the animal immune system.

## ACKNOWLEDGEMENTS

I would like to thank Dr. Juris Grasis for his helpful comments with this manuscript.

For full funding details, see: *Quistad* et al., 2016a,b; *Quistad* and *Traylor-Knowles*, 2016.

# LITERATURE


Aliprantis, A.O., Yang, R.B., Weiss, D.S., Godowski, P., and Zychlinsky, A.: The apoptotic signaling pathway activated by Toll-like receptor-2. EMBO J. 19, 3325-3336 (2000).

Anderson, K.V., Jürgens, G., and Nüsslein-Volhard, C.: Establishment of dorsal-ventral polarity in the Drosophila embryo: Genetic studies on the role of the Toll gene product. Cell 42, 779-789 (1985).

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I.: Host-bacterial mutualism in the human intestine. Science 307, 1915-1920 (2005).

Bahia, M.S. and Silakari, O.: Tumor necrosis factor alpha converting enzyme: An encouraging target for various inflammatory disorders. Chem. Biol. Drug Des. 75, 415-443 (2010).

Barber, G.N.: Host defense, viruses and apoptosis. Cell Death Differ. 8, 113-126 (2001).

Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., Doran, K.S., Salamon, P., Youle, M., and Rohwer, F.: Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc. Natl. Acad. Sci. USA 110, 10771-10776 (2013).

Barr, J.J. Auro, R., Sam-Soon, N., Kassegne, S., Peters, G., Bonilla, N., Hatay, M., Mourtada, S., Bailey, B., Youle, M., Felts, B., Baljon, A., Nulton, J., Salamon, P., and Rohwer, F.: Subdiffusive motion of bacteriophage in mucosal surfaces increases the frequency of bacterial encounters. Proc. Natl. Acad. Sci. USA 112, 13675-13680 (2015).

Barshis, D.J., Ladner, J.T., Oliver, T.A., Seneca, F.O., Traylor-Knowles, N., and Palumbi, S.R.: Genomic basis for coral resilience to climate change. Proc. Natl. Acad. Sci. USA 110, 1387-1392 (2013).

Beckmann, A. and Özbek, S.: The nematocyst: A molecular map of the cnidarian stinging organelle. Int. J. Dev. Biol. 56, 577-582 (2012).

Bosch, T.C.G.: Cnidarian-microbe interactions and the origin of innate immunity in metazoans. Annu. Rev. Microbiol 67, 499-518 (2013).

Cerenius, L., Kawabata, S., Lee, B.L., Nonaka, M., and Söderhäll, K.: Proteolytic cascades and their involvement in invertebrate immunity. Trends Biochem. Sci. 35, 575-583 (2010).

Chapman, J.A., Kirkness, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., Disbennett, K., Pfannkoch, C., Sumin, N., Sutton, G.G., Viswanathan, L.D., Walenz, B., Goodstein, D.M., Hellsten, U., Kawashima, T., Prochnik, S.E., Putnam, N.H., Shu, S., Blumberg, B., Dana, C.E., Gee, L., Kibler, D.F., Law, L., Lindgens, D., Martinez, D.E., Peng, J., Wigge, P.A., Bertulat, B., Guder, C., Nakamura, Y., Ozbek, S., Watanabe, H., Khalturin, K., Hemmrich, G., Franke, A., Augustin, R., Fraune, S., Hayakawa, E., Hayakawa, S., Hirose, M., Hwang, J.S., Ikeo, K., Nishimiya-Fujisawa, C., Ogura, A., Takahashi, T., Steinmetz, P.R., Zhang, X., Aufschnaiter, R., Eder, M.K., Gorny, A.K., Salvenmoser, W., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Böttger, A., Tischler, P., Wolf, A., Gojobori, T., Remington, K.A., Strausberg, R.L., Venter, J.C., Technau, U., Hobmayer, B., Bosch, T.C., Holstein, T.W., Fujisawa, T., Bode, H.R., David, C.N., Rokhsar, D.S., and Steele, R.E.: The dynamic genome of Hydra. Nature 464, 592-596 (2010).

Closek, C.J., Sunagawa, S., DeSalvo, M.K., Piceno, Y.M., DeSantis, T.Z., Brodie, E.L., Weber, M.X., Voolstra, C.R., Andersen, G.L., and Medina, M.: Coral transcriptome and bacterial community profiles reveal distinct Yellow Band Disease states in Orbicella faveolata. ISME J. 8, 1-12 (2014).

Cooper, M.A.X.D., Peterson, R.D.A., and Good, R.A.: Delineation of the thymic and bursal lymphoid systems in the chicken. Nature 205, 143-146 (1965).

Daly, M., Brugler, M.R., Cartwright, P.,

Collins, A.G., Dawson, M.N., Fautin, D.G., France, S.C., McFadden, C.S., Opresko, D.M., Rodriguez, E., Roman, S.L., and Stake, J.L.: The phylum Cnidaria: a review of phylogenetic patterns and diversity 300 years after Linnaeus. Zootaxa 1668, 127-182 (2007).

Darling, J.A., Reitzel, A.R., Burton, P.M., Mazza, M.E., Ryan, J.F., Sullivan, J.C. and Finnerty, J.R.: Rising starlet: the starlet sea anemone, Nematostella vectensis. Bioessays 27, 211-221 (2005).

David, C.N., Schmidt, N., Schade, M., Pauly, B., Alexandrova, O., and Böttger, A.: Hydra and the evolution of apoptosis. Integr. Comp. Biol. 45, 631-638 (2005).

Davidson, E.H. and Erwin, D.H.: An integrated view of precambrian eumetazoan evolution. Cold Spring Harb. Symp. Quant. Biol. 74, 65-80 (2009).

Derrien, M., van Passel, M.W., van de Bovenkamp, J.H., Schipper, R.G., de Vos, W.M., and Dekker, J.: Mucin-bacterial interactions in the human oral cavity and digestive tract. Gut Microbes 1, 254-268 (2010).

Desvignes, T., Pontarotti, P., and Bobe, J.: Nme gene family evolutionary history reveals pre-metazoan origins and high conservation between humans and the sea anemone, nematostella vectensis. PLoS One 5, e15506 (2010).

Ferez-Vilar, J. and Hill, R.L.: The structure and assembly of secreted mucins. J. Biol. Chem. 274, 31751-31754 (1999).

Franchi, L., Warner, N., Viani, K., and Nuñez, G.: Function of Nod-like receptors in microbial recognition and host defense. Immunol. Rev. 227, 106-128 (2009).

Franzenburg, S., Fraune, S., Künzel, S., Baines, J.F., Domazet-Lošo, T., and Bosch, T.C.G.: MyD88-deficient Hydra reveal an ancient function of TLR signaling in sensing bacterial colonizers. Proc. Natl. Acad. Sci. USA 109, 19374-19379 (2012).

Grasis, J.A., Lachnit, T., Anton-Erxleben, F., Lim, Y.W., Schmieder, R., Fraune, S., Franzenburg, S., Insua, S., Machado, G., Haynes, M., Little, M., Kimble, R., Rosenstiel, P., Rohwer, F.L., and Bosch, T.C.: Species-specific viromes in the ancestral holobiont hydra. PLoS One 9, e109952 (2014).

Gros, P., Milder, F.J., and Janssen, B.J.C.: Complement driven by conformational changes. Nat. Rev. Immunol. 8, 48-58 (2008).

Hagai, T., Azia, A., Babu, M.M., and Andino, R.: Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. Cell Rep. 7, 1729-1739 (2014).

Hamada, M., Shoguchi, E., Shinzato, C., Kawashima, T., Miller, D.J., and Satoh, N.: The complex NOD-like receptor repertoire of the coral acropora digitifera includes novel domain combinations. Mol. Biol. Evol. 30, 167-176 (2013).

Hang, H.C. and Bertozzi, C.R.: The chemistry and biology of mucin-type O-linked glycosylation. Bioorganic Med. Chem. 13, 5021-5034 (2005).

Hansen, J.D., Vojtech, L.N. and Laing, K.J.: Sensing disease and danger: A survey of vertebrate PRRs and their origins. Dev. Comp. Immunol. 35, 886-897 (2011).

Hemmrich, G., Khalturin, K., Boehm, A.M., Puchert, M., Anton-Erxleben, F., Wittlieb, J., Klostermeier, U.C., Rosenstiel, P., Oberg, H.H., Domazet-Loso, T., Sugimoto, T., Niwa, H., and Bosch, T.C.: Molecular signatures of the three stem cell lineages in hydra and the emergence of stem cell function at the base of multicellularity. Mol. Biol. Evol. 29, 3267-3280 (2012).

Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E., Harvell, C.D., Sale, P.F., Edwards, A.J., Caldeira, K., Knowlton, N., Eakin, C.M., Iglesias-Prieto, R., Muthiga, N., Bradbury, R.H., Dubi, A., and Hatziolos, M.E.: Coral reefs under rapid climate change and ocean acidification. Science 318, 1737–17-2 (2007).

Ikmi, A., McKinney, S.A., Delventhal, K.M., and Gibson, M.C.: TALEN and CRISPR/Cas9-mediated genome editing in the early-branching metazoan Nematostella vectensis. Nat. Commun. 5, 5486 (2014).

Iwasaki, A. and Medzhitov, R.: Toll-like receptor control of the adaptive immune responses. Nat. Immunol. 5, 987-995 (2004).

Johansson, M.E.V, Sjövall, H., and Hansson, G.C.: The gastrointestinal mucus system in health and disease. Nat. Rev. Gastroenterol. Hepatol. 10, 352–361 (2013).

Kim, D.H. and Ausubel, F.M.: Evolutionary perspectives on innate immunity from the study of Caenorhabditis elegans. Curr. Opin. Immunol. 17, 4-10 (2005).

Kortschak, R.D., Samuel, G., Saint, R., and Miller, D.J.: EST analysis of the cnidarian Acropora millepora reveals extensive gene loss and rapid sequence divergence in the model invertebrates. Curr. Biol. 13, 2190-2195 (2003).

Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H.A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M.Q., and Holstein, T.W.: Unexpected complexity of the Wnt gene family in a sea anemone. Nature 433, 156-160 (2005).

Kvennefors, E.C.E., Leggat, W., Hoegh-Guldberg, O., Degnan, B.M., and Barnes, A.C.: An ancient and variable mannose-binding lectin from the coral Acropora millepora binds both pathogens and symbionts. Dev. Comp. Immunol. 32, 1582-1592 (2008).

Kvennefors, E.C.E., Leggat, W., Kerr, C.C., Ainsworth, T.D., Hoegh-Guldberg, O., and Barnes, A.C.: Analysis of evolutionarily conserved innate immune components in coral links immunity and symbiosis. Dev. Comp. Immunol. 34, 1219-1229 (2010).

Lang, T., Hansson, G.C., and Samuelsson, T.: Gel-forming mucins appeared early in metazoan evolution. Proc. Natl. Acad. Sci. USA 104, 16209-16214 (2007).

Lange, C., Hemmrich, G., Klostermeier, U.C., López-Quintero, J.A., Miller, D.J., Rahn, T., Weiss, Y., Bosch, T.C.G., and Rosenstiel, P.: Defining the origins of the NOD-like receptor system at the base of animal evolution. Mol. Biol. Evol. 28, 1687-1702 (2011).

Liang, C., Lee, J.-S., and Jung, J.U.: Immune evasion in Kaposi's sarcoma-associated herpes virus associated oncogenesis. Semin. Cancer Biol. 18, 423-436 (2008).

Lin, Y., Devin, A., Rodriguez, Y., and Liu, Z.G.: Cleavage of the death domain kinase RIP by Caspase-8 prompts TNF-induced apoptosis. Genes Dev. 13, 2514-2526 (1999).

Litman, G.W. and Cooper, M.D.: Why study the evolution of immunity? Nat. Immunol. 8, 547-548 (2007).

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R.: Diversity, stability and resilience of the human gut microbiota. Nature 489, 220-230 (2012).

Micheau, O. and Tschopp, J.: Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. Cell 114, 181-190 (2003).

Miller, D.J., Ball, E.E., and Technau, U.: Cnidarians and ancestral genetic complexity in the animal kingdom. Trends Genet. 21, 536-539 (2005).

Miller, D.J., Hemmrich, G., Ball, E.E., Hayward, D.C., Khalturin, K., Funayama, N., Agata, K., and Bosch, T.C.G.: The innate immune repertoire in cnidaria--ancestral complexity and stochastic gene loss. Genome Biol. 8, R59 (2007).

Moberg, F.F. and Folke, C.: Ecological goods and services of coral reef ecosystems. Ecol. Econ. 29, 215-233 (1999).

Moya, A. Sakamaki, K., Mason, B.M., Huisman, L., Forêt, S., Weiss, Y., Bull, T.E., Tomii, K., Imai, K., Hayward, D.C., Ball, E.E., and Miller, D.J.: Functional conservation of the apoptotic machinery from coral to man: the diverse and complex Bcl-2 and caspase repertoires of Acropora millepora. BMC Genomics 17, 62 (2016).

Muzio, M., Polentarutti, N., Bosisio, D., Manoj Kumar, P.P., and Mantovani, A.: Toll-like receptor family and signalling pathway. Biochem.Soc.Trans. 28, 563-566 (2000).

Noris, M. and Remuzzi, G.: Overview of complement activation and regulation. Semin. Nephrol. 33, 479-492 (2013).

Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N.,

and Wörheide, G.: Genomic data do not support comb jellies as the sister group to all other animals. Proc. Natl. Acad. Sci. USA 112, 15402-15407 (2015).

Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V., Jurka, J., Genikhovich, G., Grigoriev, I.V., Lucas, S.M., Steele, R.E., Finnerty, J.R., Technau, U., Martindale, M.Q., Rokhsar, D.S.: Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317, 86-94 (2007).

Quistad, S.D., Stotland, A., Barott, K.L., Smurthwaite, C.A., Hilton, B.J., Grasis, J. A., Wolkowicz, R., and Rohwer, F.L.: Evolution of TNF-induced apoptosis reveals 550 My of functional conservation. Proc. Natl. Acad. Sci. USA 111, 9567-9572 (2014).

Quistad, S.D., Grasis, J.A., Barr, J.J., and Rohwer, F.L.: Viruses and the origin of microbiome selection and immunity. ISME J. 11, 835-840 (2016a).

Quistad, S.D., Lim, Y.W., Silva, G.G.Z., Nelson, C.E., Haas, A.F., Kelly, L.W., Edwards, R.A., and Rohwer, F.L.: Using viromes to predict novel immune proteins in non-model organisms. Proc. R. Soc. B. 283, 20161200 (2016b).

Quistad, S.D. and Traylor-Knowles, N.: Precambrian origins of the TNFR superfamily. Cell Death Discov. 2, 16058 (2016).

Redfern, R.L., Reins, R.Y., and McDermott, A.M.: Toll-like receptor activation modulates antimicrobial peptide expression by ocular surface cells. Exp. Eye Res. 92, 209-220 (2011).

Renfer, E., Amon-Hassenzahl, A., Steinmetz, P.R.H., and Technau, U.: A muscle-specific transgenic reporter line of the sea anemone, Nematostella vectensis. Proc. Natl. Acad. Sci. USA 107, 104-108 (2010).

Rosetto, M., Engström, Y., Baldari, C.T., Telford, J.L., and Hultmark, D.: Signals from the IL-1 receptor homolog, Toll, can activate an immune response in a Drosophila hemocyte cell line. Biochem. Biophys. Res. Commun. 209, 111-116 (1995).

Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P.; NISC Comparative Sequencing Program, Smith, S.A., Putnam, N.H., Haddock, S.H., Dunn, C.W., Wolfsberg, T.G., Mullikin, J.C., Martindale, M.Q., and Baxevanis, A.D.: The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. Science 342, 1242592 (2013).

Sakamaki, K., Shimizu, K., Iwata, H., Imai, K., Satou, Y., Funayama, N., Nozaki, M., Yajima, M., Nishimura, O., Higuchi, M., Chiba, K., Yoshimoto, M., Kimura, H., Gracey, A.Y., Shimizu, T., Tomii, K., Gotoh, O., Akasaka, K., Sawasaki, T., and Miller, D.J.: The apoptotic initiator caspase-8: its functional ubiquity and genetic diversity during animal evolution. Mol. Biol. Evol. 31, 3282-3301 (2014).

Schröder, K. and Bosch, T.C.G.: The Origin of Mucosal Immunity: Lessons from the Holobiont Hydra. MBio 7, e01184-16 (2016).

Shinzato, C., Shoguchi, E., Kawashima, T., Hamada, M., Hisata, K., Tanaka, M., Fujie, M., Fujiwara, M., Koyanagi, R., Ikuta, T., Fujiyama, A., Miller, D.J., and Satoh, N.: Using the Acropora digitifera genome to understand coral responses to environmental change. Nature 476, 320-323 (2011).

Shu, D.: Cambrian explosion: Birth of tree of animals. Gondwana Res. 14, 219-240 (2008).

Silveira, C.B. and Rohwer, F.L.: Piggyback-the-Winner in host-associated microbial communities. NPJ Biofilms Microbiomes 2, 16010 (2016).

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G., and Manuel, M.: A Large and Consistent Phylogenomic Dataset

Supports Sponges as the Sister Group to All Other Animals. Curr. Biol. 27, 958-967 (2017).

Technau, U., Rudd, S., Maxwell, P., Gordon, P.M., Saina, M., Grasso, L.C., Hayward, D.C., Sensen, C.W., Saint, R., Holstein, T.W., Ball, E.E., and Miller, D.J.: Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. Trends Genet. 21, 633-639 (2005).

Tortorella, D., Gewurz, B.E., Furman, M.H., Schust, D.J., and Ploegh, H.L.: Viral subversion of the immune system. Annu. Rev. Immunol. 18, 861-926 (2000).

Van Iten, H., Leme, J.M., Pacheco, M.L.A.F., Simões, M.G., Fairchild, T.R., Rodrigues, F., Galante, D., Boggiani, P.C., and Marques, A.C.: Origin and Early Diversification of Phylum Cnidaria: Key Macrofossils from the Ediacaran System of North and South America. In: The Cnidaria, Past, Present and Future: The world of Medusa and her sisters (Eds.: Goffredo, S. and Dubinsky, Z.). Springer International Publishing AG, Cham (ZG) Switzerland, 31-40 (2016).

Wang, J., Tao, Y., Reim, I., Gajewski, K., Frasch, M., and Schulz, R.A.: Expression, regulation, and requirement of the toll transmembrane protein during dorsal vessel formation in Drosophila melanogaster. Mol. Cell. Biol. 25, 4200-4210 (2005).

Wiens, G.D. and Glenney, G.W.: Origin and evolution of TNF and TNF receptor superfamilies. Dev. Comp. Immunol. 35, 1324-1335 (2011).

Zmasek, C., Zhang, Q., Ye, Y., and Godzik, A.: Surprising complexity of the ancestral apoptosis network. Genome Biol. 8, R226 (2007).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 53-59 (2017).

# HOST-ASSOCIATED VIROMES IN HEALTH AND DISEASE

SCOTT ALLYN HANDLEY

Department of Pathology and Immunology, Washington University,
St. Louis, MO 63110, USA

## SUMMARY

The microbiome (defined here as all microorganisms, commensal, symbiotic or pathogenic that share our body space) plays a significant role in human health and disease (*Virgin*, 2014; *Norman* et al., 2015). Many studies have focused on the bacterial component of the microbiome. For example, Crohn's disease (CD) and ulcerative colitis (UC), two forms of inflammatory bowel disease (IBD), are characterized by changes in the gut bacterial microbiome in adults and children (*Gevers* et al., 2014; *Kostic* et al., 2014). The virome is composed of both eukaryote infecting viruses and bacteriophages (phages, hereafter). Recently, associations of the virome with disease have been made in HIV/AIDS (*Handley* et al., 2012, 2016), malnutrition (*Reyes* et al., 2015), and IBD (*Norman* et al., 2015). AIDS in both humans and non-human primate models is characterized by an expansion of enteric eukaryotic viruses, in particular picornaviruses and adenoviruses (*Handley* et al., 2012, 2016). In IBD, we showed that in a cross-sectional comparison of adult IBD cases and household controls there was a significant expansion of the *Caudovirales*, an order of phages, and anelloviruses, a family of eukaryotic DNA viruses of unknown pathogenicity (*Norman* et al., 2015). Importantly, we also observed statistically significant associations between phages and bacteria that might play a protective or causative role in IBD.

## INTRODUCTION

Recent associations of the virome with disease have been observed in a diverse set of diseases including in HIV/AIDS (*Handley* et al., 2012, 2016), malnutrition (*Reyes* et al., 2015), and IBD (*Norman* et al., 2015). AIDS in both humans and non-human primate models is characterized by an expansion of enteric eukaryotic viruses, in particular picornaviruses and adenoviruses (*Handley* et al., 2012, 2016). In IBD, there is a significant expansion of the *Caudovirales*, an order of phages and anelloviruses, a family of eukaryotic DNA viruses of unknown pathogenicity (*Norman* et al., 2015). The pathological consequences of viral community changes in the enteric microbiome are unclear. As eukaryotic viruses infect host cells, it is safe to speculate that expansion of these viruses may impact host cell physiology. Our understanding of the impact for many of these viruses is relatively well-established as these viruses regularly cause infectious disease. The role of bacteriophage expansion is less clear due to the fact that they can infect, modulate and lyse bacterial cells, as well as influence host-cell physiology via their free-virion state. This later concept has an anaemic literature and

would benefit from further scientific investigation. This review focuses on summarizing what is known about the largely ignored associations between phage and human disease, and highlights an even greater challenge for the virome research community, that of the currently unclassifiable components of the virome frequently referred to as 'dark matter'. The impact of viral dark matter is that we are blind to a significant component of host-associated viromes and may be missing important disease associated viruses. A concerted research effort is required to define the mechanisms behind virome changes in causing disease in concert with defining the full virome by shedding light onto the void of our own viromes.

## REVIEW

The human intestinal microbiome consists of an estimated $10^{14}$ bacterial cells that play a vital role in gut metabolism, nutrient uptake, immune system development, intestinal physiology, and competition with pathogenic bacteria (*Van Praet* et al., 2015). Bacterial communities are parasitized by complex array of bacteriophages. Next generation sequencing (NGS)-based studies have revealed that enteric bacteriophage communities are highly variable among individuals, incredibly diverse, and affected by environmental stimuli (*Reyes* et al., 2010; *Minot* et al., 2013). Since bacteriophage influence bacterial fitness by killing bacteria with specific receptors on their surfaces and by transferring mobile genetic elements and antibiotic resistance genes among bacteria, an altered intestinal virome might well trigger shifts in bacterial flora relevant to diseases such as IBD (*Duerkop* et al., 2012; *Sougakoff* et al., 1988; *Doucet-Populaire* et al., 1991).

Bacteriophages have been proposed to be associated with inflammatory disease (*Riley*, 2004), however their role in disease remains largely undefined. Increased tailed bacteriophage (order *Caudovirales*) have been observed by electron microscopy in intestinal biopsy washes from adults with CD compared to healthy controls (*Colombet* et al., 2008). These findings were further supported by a sequence-based study of paediatric CD, where the analysis of mucosal washes and intestinal biopsies revealed an increase in bacteriophage reads when compared to non-IBD tissue biopsies (*Wagner* et al., 2013). Increased bacteriophages may be explained by the activation of bacterial-integrated prophage by environmental stimuli, such as oxidative stress or antibiotic use (*Fortier* and *Sekulovic*, 2013), which are commonly associated with IBD, but this specific association has yet to be proven.

Free bacteriophage particles in the intestine could come into contact with epithelial cells as well as the lamina propria and antigen presenting cells, via breaks in the intestinal mucosa, the lamina propria and host myeloid cells resulting in host immune surveillance and systemic spread (*Górski* et al., 2006). It has long been established that antibody responses are made against bacteriophage particles, indicating that they are immunogenic (*Uhr* et al., 1962). Additionally, even though a known eukaryotic cell receptor is unknown, endotoxin-free bacteriophage particles stimulate inflammatory cytokine production (including IL-1β and TNF-α) by macrophages in a MyD88-dependent manner (*Eriksson* et al., 2009). This may be due in part to CpG motifs present in bacteriophage

genomes, which stimulate interferon production and protect against vaccinia virus challenge *in vivo* (*Mori* et al., 1996). Alternately, the cytokine production may be related to the immunogenicity of bacteriophage coat proteins that have been reported to enhance DNA vaccine potency by stimulating adaptive immune responses (*Cuesta* et al., 2006). This adjuvant effect of bacteriophages has not been explored in mechanistic detail. Therefore, changes in bacteriophage populations may have effects on intestinal inflammation indirectly by pruning the structure of the enteric bacterial microbiome or directly by stimulating inflammation.

A recent study defined the faecal virome in IBD patients and controls (*Norman* et al., 2015). This study found that the enteric virome differs between CD and UC patients and controls with a significant, and disease-specific, expansion of bacteriophages that was not secondary to changes in bacterial populations. These data support a 'virus-predator-bacterial prey' model in which the virome may contribute to IBD associated intestinal inflammation through altering bacterial community structure or through direct interactions with the host.

In animal models, viruses that infected eukaryotic cells (murine norovirus) have been shown to interact with IBD risk genes to alter intestinal disease (*Cadwell* et al., 2010; *Basic* et al., 2014). In the same study that identified enteric bacteriophage expansion associations with IBD, an increased abundance of anellovirus was observed in IBD patients when compared to controls. Taken together with the bacteriophage data, these data suggest that both bacterial and eukaryotic viruses may contribute to the IBD pathologies.

The virome expansion observed in IBD was extraordinarily patient specific. Bacteriophage populations were different between individuals, but remained relatively constant within a patient over time. This parallels what is observed for bacterial community structure in IBD patients, suggesting that a full understanding of a patient's personal microbiome may be required to properly realize safe and effective personal treatment strategies. It is well recognized, for example, that the course of disease (or prognosis) varies substantially between patients with IBD. Recent studies have begun to investigate the biology that determines prognosis, and have demonstrated that pathways associated with CD8+ T-cell activation are up-regulated in a subset of IBD patients who subsequently experience a significantly more aggressive disease course (*Lee* et al., 2011). Indeed, the balance between T-cell activation and exhaustion seems to determine prognosis in a range of autoimmune diseases, including IBD (*McKinney* et al., 2015). The causes of T-cell exhaustion in IBD, however, remain unknown, but it is striking that this phenomenon is typically associated with chronic viral infection. Nevertheless, to date, no studies have investigated the associations between an individual's personal enteric viral content virome and their disease status.

## THE DARK MATTER CHALLENGE

Typically a majority of the sequences present in purified virus preparations cannot be classified due to the lack of statistically significant sequence similarity to reference virus sequence (*Reyes* et al., 2010, 2015; *Minot* et al.,

2013; *Norman* et al., 2015). Unlike bacteria, viruses lack a universally conserved marker sequence (*Rohwer* and *Edwards*, 2002). Thus, viromes require analysis via metagenomic (sequencing random fragments) methods. These fragments are classified using sequence alignment to known viral sequences. In all previous virome studies, fewer than 50% of viral sequences are classifiable with the rest remaining as viral "dark matter" with unknown taxonomic and functional assignment, compromising our ability to detect important associations. Recent efforts have attempted to address the dark matter in human enteric viromes. One striking example was the discovery of crAssphage, a previously unrecognized phage that is present in ~50% of the population and is the most abundant known phage in the human gut (*Dutilh* et al., 2014). The majority of crAssphage proteins have no sequence similarity to known viral proteins and its extraction from the viral dark matter relied on novel computational approaches. Thus, there is a fundamental need to apply novel computational tools to analyse viral dark matter.

In addition to more robust computational analysis of virome sequence data, concerted wet laboratory techniques need to be employed for functional characterization of unclassifiable sequence data. The advent of NGS methods has dramatically increased our ability to detect nucleic acid sequences derived from novel eukaryotic viruses and phages, forming the basis of most virome studies (*Minot* et al., 2013; *Ogilvie* et al., 2013; *Lim* et al., 2015; *Krishnamurthy* et al., 2016; *Manrique* et al., 2016), but the potential to bring this knowledge to clinical application has not yet been realized. Over the last decade, a tremendous number of previously unrecognized eukaryotic viruses from human stool samples have been discovered (*Finkbeiner* et al., 2008a, 2008b, 2009; *Holtz et al.,* 2008; *Kapoor* et al., 2008, 2009; *Phan* et al., 2016). However, of these a cell culture system has only been described for human Theiler's-like cardiovirus (*Chiu* et al., 2010) and astrovirus VA1 (*Janowski* et al., 2017). Likewise, many novel phage sequences in the human enteric tract have been discovered, but culture has not been attempted (*Reyes* et al., 2010; *Minot* et al., 2011, 2013; *Reyes* et al., 2015). This paucity of culturable eukaryotic viruses and phages precludes any functional assessment of the role of these agents in disease. Therefore, it is essential that culture systems be developed for eukaryotic viruses and phages identified in virome studies in order to study their impact on disease. These efforts will aid in functional characterization of dark matter as classifiable viruses, making these efforts extraordinarily valuable for characterizing the viromes influence of health and disease.

The increased literature emerging on enteric virome demonstrate that both eukaryotic viruses and phage populations take on different forms in disease states. In particular, the emergence of both eukaryotic and bacterial viral populations in people with IBD is a prominent display of the extent and complexity of the virome in context of a dysbiotic bacterial microbiome and diseased intestine. How these viruses are contributing to disease, or if they are just innocuous passengers or markers of the diseased state has yet to be answered. As highlighted here, the path to establishing causal connections between viromes and disease require advancements in total understanding of the virome through advances in both computational and laboratory biology. Only then will we fully appreciate the impact our smallest enteric passengers on our own health.

# LITERATURE


Basic, M., Keubler, L.M., Buettner, M., Achard, M., Breves, G., Schröder, B., Smoczek, A., Jörns, A., Wedekind, D., Zschemisch, N.H., Günther, C., Neumann, D., Lienenklaus, S., Weiss, S., Hornef, M.W., Mähler, M., and Bleich, A.: Norovirus triggered microbiota-driven mucosal inflammation in interleukin 10-deficient mice. Inflamm. Bowel Dis. 20, 431-443 (2014).

Cadwell, K., Patel, K.K., Maloney, N.S., Liu, T.C., Ng, A.C.Y., Storer, C.E., Head, R.D., Xavier, R., Stappenbeck, T.S., and Virgin, H.W.: Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. Cell 141, 1135-1145 (2010).

Chiu, C.Y., Greninger, A.L., Chen, E.C., Haggerty, T.D., Parsonnet, J., Delwart, E., DeRisi, J.L., and Ganem, D.: Cultivation and serological characterization of a human Theiler's-like cardiovirus associated with diarrheal disease. J. Virol. 84, 4407-4414 (2010).

Colombet, J., Marteau, P., Sime-Ngando, T., and Doré, J., and Leclerc, M.: Dysbiosis in inflammatory bowel disease: a role for bacteriophages? Gut. 57, 424-425(2008).

Cuesta, A.M., Suárez, E., Larsen, M., Jensen, K.B., Sanz, L., Compte, M., Kristensen, P., and Alvarez-Vallina, L.: Enhancement of DNA vaccine potency through linkage of antigen to filamentous bacteriophage coat protein III domain I. Immunology 117, 502-506 (2006).

Doucet-Populaire, F., Trieu-Cuot, P., Dosbaa, I., Andremont, A., and Courvalin, P.: Inducible transfer of conjugative transposon Tn1545 from Enterococcus faecalis to Listeria monocytogenes in the digestive tracts of gnotobiotic mice. Antimicrob. Agents Chemother. 35, 185-187 (1991).

Duerkop, B.A., Clements, C.V., Rollins, D., Rodrigues, J.L., and Hooper, L.V.: A composite bacteriophage alters colonization by an intestinal commensal bacterium. Proc. Natl. Acad. Sci. USA 109, 17621-17626 (2012).

Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., Felts, B., Dinsdale, E.A., Mokili, J.L., and Edwards, R.A.: A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat. Commun. 5, 4498 (2014).

Eriksson, F., Tsagozis, P., Lundberg, K., Parsa, R., Mangsbo, S.M., Persson, M. A.A., Harris, R.A., and Pisa, P.: Tumor-specific bacteriophages induce tumor destruction through activation of tumor-associated macrophages. J. Immunol. 182, 3105-3111 (2009).

Finkbeiner, S.R., Allred, A.F., Tarr, P.I., Klein, E.J., Kirkwood, C.D., and Wang, D.: Metagenomic analysis of human diarrhea: viral detection and discovery. PLoS Pathog. 4, e1000011 (2008a).

Finkbeiner, S.R., Kirkwood, C.D. and Wang, D.: Complete genome sequence of a highly divergent astrovirus isolated from a child with acute diarrhea. Virol J. 5, 117 (2008b).

Finkbeiner, S.R., Holtz, L.R., Jiang, Y., Rajendran, P., Franz, C.J., Zhao, G., Kang, G., and Wang, D.: Human stool contains a previously unrecognized diversity of novel astroviruses. Virol J. 6, 161 (2009).

Fortier, L.C. and Sekulovic, O.: Importance of prophages to evolution and virulence of bacterial pathogens. Virulence 4, 354-365 (2013).

Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., Morgan, X.C., Kostic, A.D., Luo, C., González, A., McDonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., Stephens, M., Heyman, M., Markowitz, J., Baldassano, R., Griffiths, A., Sylvester, F., Mack, D., Kim, S., Crandall, W., Hyams, J., Huttenhower, C., Knight, R., and Xavier, R.J.: The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 15, 382-392 (2014).

Górski, A., Ważna, E., Dąbrowska, B.W., Switała-Jeleń, K., and Miedzybrodzki, R.:

Bacteriophage translocation. FEMS Immunol. Med. Microbiol. 46, 313-319(2006).

Handley, S.A., Desai, C., Zhao, G., Droit, L., Monaco, C.L., Schroeder, A.C., Nkolola, J.P., Norman, M.E., Miller, A.D., Wang, D., Barouch, D.H., and Virgin, H.W.: SIV Infection-Mediated Changes in Gastrointestinal Bacterial Microbiome and Virome Are Associated with Immunodeficiency and Prevented by Vaccination. Cell Host Microbe 19, 323-335 (2016).

Handley, S.A., Thackray, L.B., Zhao, G., Presti, R., Miller, A.D., Droit, L., Abbink, P., Maxfield, L.F., Kambal, A., Duan, E., Stanley, K., Kramer, J., Macri, S.C., Permar, S.R., Schmitz, J.E., Mansfield, K., Brenchley, J.M., Veazey, R.S., Stappenbeck, T.S., Wang, D., Barouch, D.H., and Virgin, H.W.: Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. Cell 151, 253-266 (2012).

Holtz, L.R., Finkbeiner, S.R., Kirkwood, C.D., and Wang, D.: Identification of a novel picornavirus related to cosaviruses in a child with acute diarrhea. Virol. J. 5, 159 (2008).

Janowski, A.B., Bauer, I.K., Holtz, L.R., and Wang, D.: Propagation of astrovirus VA1, a neurotropic human astrovirus, in cell culture. J. Virol. 91, e00740–17 (2017).

Kapoor, A., Li, L., Victoria, J., Oderinde, B., Mason, C., Pandey, P., Zaidi, S.Z., and Delwart, E.: Multiple novel astrovirus species in human stool. J. Gen. Virol. 90, 2965-2972 (2009).

Kapoor, A., Victoria, J., Simmonds, P., Slikas, E., Chieochansin, T., Naeem, A., Shaukat, S., Sharif, S., Alam, M.M., Angez, M., Wang, C., Shafer, R.W., Zaidi, S., and Delwart, E.: A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. Proc. Natl. Acad. Sci. USA 105, 20482-20487 (2008).

Kostic, A.D., Xavier, R.J., and Gevers, D.: The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology 146, 1489-1499 (2014).

Krishnamurthy, S.R., Janowski, A.B., Zhao, G., Barouch, D., and Wang, D.: Hyperexpansion of RNA Bacteriophage Diversity. PloS Biol. 14, e1002409 (2016).

Lee, J.C., Lyons, P.A., McKinney, E.F., Sowerby, J.M., Carr, E.J., Bredin, F., Rickman, H.M., Ratlamwala, H., Hatton, A., Rayner, T.F., Parkes, M., and Smith, K.G.C.: Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. J. Clin. Invest. 121, 4170-4179 (2011).

Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., and Holtz, L.R.: Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat. Med. 21, 1228-1234 (2015).

Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J.: Healthy human gut phageome. Proc. Natl. Acad. Sci. USA 113, 10400-10405 (2016).

McKinney, E.F., Lee, J.C., Jayne, D.R.W., Lyons, P.A., and Smith, K.G.C.: T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. Nature 523, 612-616 (2015).

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D.: The human gut virome: inter-individual variation and dynamic response to diet. Genome Res. 21, 1616-1625 (2011).

Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D.: Rapid evolution of the human gut virome. Proc. Natl. Acad. Sci. USA 110, 12450-12455 (2013).

Mori, K., Kubo, T., Kibayashi, Y., Ohkuma, T., and Kaji, A.: Anti-vaccinia virus effect of M13 bacteriophage DNA. Antiviral Res. 31, 79-86 (1996).

Norman, J.M., Handley, S.A., Baldridge, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P., Stappenbeck, T.S., McGovern, D.P.B., Keshavarzian, A., Mutlu, E.A., Sauk, J., Gevers, D., Xavier, R.J., Wang, D., Parkes, M., and Virgin, H.W.: Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell 160, 447-460

(2015).
Ogilvie, L.A., Bowler, L.D., Caplin, J., Dedi, C., Diston, D., Cheek, E., Taylor, H., Ebdon, J.E., and Jones, B.V.: Genome signature-based dissection of human gut metagenomes to extract subliminal viral sequences. Nat. Commun. 4, 2420 (2013).
Phan, M.V.T., Anh, P.H., Cuong, N.V., Munnink, B.B.O., van der Hoek, L., My, P.T., Tri, T.N., Bryant, J.E., Baker, S., Thwaites, G., Woolhouse, M., Kellam, P., Rabaa, M.A., Cotten, M., VIZIONS Consortium: Unbiased whole-genome deep sequencing of human and porcine stool samples reveals circulation of multiple groups of rotaviruses and a putative zoonotic infection. Virus Evol. 2, vew027 (2016).
Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I.: Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466, 334-338 (2010).
Reyes, A., Blanton, L.V., Cao, S., Zhao, G., Manary, M., Trehan, I., Smith, M.I., Wang, D., Virgin, H.W., Rohwer, F., and Gordon, J.I.: Gut DNA viromes of Malawian twins discordant for severe acute malnutrition.Proc. Natl. Acad. Sci. USA 112, 11941-11946 (2015).
Riley, P.A.: Bacteriophages in autoimmune disease and other inflammatory conditions. Med. Hypotheses 62, 493-498(2004).
Rohwer, F. and Edwards, R.: The Phage Proteomic Tree: a genome-based taxonomy for phage. J. Bacteriol. 184, 4529-4535 (2002).
Sougakoff, W., Goussard, S., Gerbaud, G., and Courvalin, P.: Plasmid-mediated resistance to third-generation cephalosporins caused by point mutations in TEM-type penicillinase genes. Rev. Infect. Dis. 10, 879-884 (1988).
Uhr, J.W., Finkelstein, S., and Franklin, E.C.: Antibody response to bacteriophage phi-X-174 in non-mammalian vertebrates. Proc. Soc. Exp. Biol. Med. 111, 13-15 (1962).
Van Praet, J.T., Donovan, E., Vanassche, I., Drennan, M.B., Windels, F., Dendooven, A., Allais, L., Cuvelier, C.A., van de Loo, F., Norris, P.S., Kruglov, A.A., Nedospasov, S.A., Rabot, S., Tito, R., Raes, J., Gaboriau-Routhiau, V., Cerf-Bensussan, N., Van de Wiele, T., Eberl, G., Ware, C.F., and Elewaut, D.: Commensal microbiota influence systemic autoimmune responses. EMBO J. 34, 466-474 (2015).
Virgin, H.W.: The virome in mammalian physiology and disease. Cell 157, 142-150 (2014).
Wagner, J., Maksimovic, J., Farries, G., Sim, W.H., Bishop, R.F., Cameron, D.J., Catto-Smith, A.G., and Kirkwood, C.D.: Bacteriophages in gut samples from pediatric Crohn's disease patients: metagenomic analysis using 454 pyrosequencing. Inflamm. Bowel Dis. 19, 1598-1608 (2013).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 61-66 (2017).

# VIRUSES - IMPORTANT REGULATORS IN THE METAORGANISM *HYDRA*

TIM LACHNIT and THOMAS C.G. BOSCH

Zoological Institute, Christian-Albrechts-University Kiel, Kiel, Germany

## SUMMARY

*Hydra* is not only associated with host specific bacteria, but also with a host specific viral community. The viral community is composed of eukaryotic and prokaryotic viruses (phages) infecting both *Hydra* and its associated bacteria. The observation of persistent viral infections in *Hydra* implies a cross-talk between eukaryotic viruses and their host. The virome of *Hydra* encodes viral structural and replication associated proteins, but also genes that interact with genetic processing and cellular regulation of their host suggesting that viruses interfere with the innate immune system and modulate *Hydra*'s functions. The bacteriophage population identified by virome sequencing is dominated by phages infecting Gammaproteobacteria and Betaproteobacteria. However, bacteriophages are not only present in their lytic lifecycle, we could also identify prophages in the genomes of *Hydra* associated bacteria. Reactivation of this hidden phage community may serve as internal regulator of host associated bacterial community or function as weapon to control bacterial colonization from the bacterioplankton community. Regarding the diversity, abundance and genetic repertoire of *Hydra* associated viruses we expect viruses to be important regulators of metaorganisms.

## INTRODUCTION

All multicellular organisms are associated with a host specific bacterial community. This close association between bacteria and its host is beneficial for both partners and forms a functional unit termed "metaorganism" (*Deines* et al., 2017). Selection of host specific microbiota and the control of bacterial community composition are essential for the stability of the metaorganism. The freshwater polyp *Hydra* is an ideal model organism to study general principles of host-microbe interactions regarding selective colonization and the role of the innate immune system in establishing and maintaining the holobiont composition (*Schröder* and *Bosch*, 2016). *Hydra* features as basal metazoan only an innate immune system to detect and interact with microbes (*Bosch*, 2013). Surprisingly these basal control mechanisms enable *Hydra* to maintain its host specific bacterial community. Even under laboratory culture conditions for over 20 years *Hydra* maintains its specific bacteria and features a similar bacterial community to individuals living in the wild (*Fraune* and *Bosch*, 2007; *Franzenburg* et al., 2013). Bacterial communities are intensively studied and regarded as important part of holobionts. Compared to bacteria very little is known about viruses although they

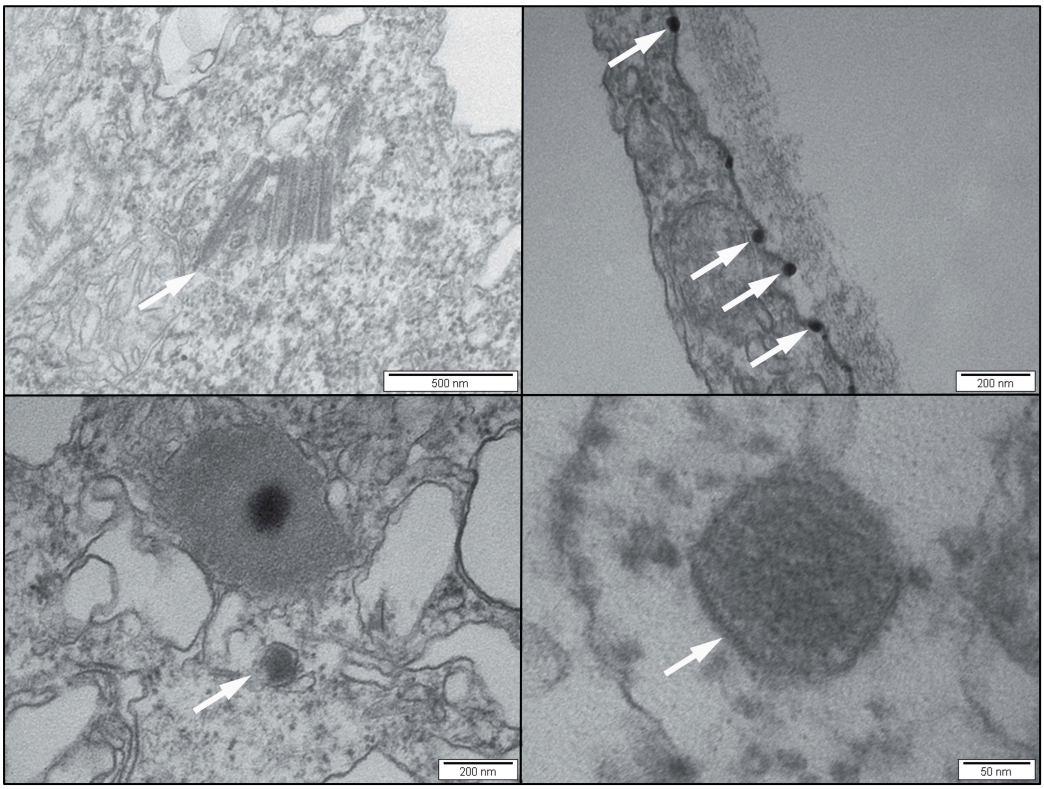


**Figure 1**: Virus-like particles within *Hydra* tissue. Transmission electron micrographs of ultrathin sections of *Hydra* stained with uranyl acetate. Baculovirus replication (top-left), accumulation of virus-like particles at the ectodermal epithelial layer (top-right), virus-like particles featuring morphological similarity to Phycodnaviridae (bottom).

are most abundant and highly diverse (*Suttle*, 2007). It is well documented that viruses have a huge impact on population dynamics and are important regulators within ecosystems (*Brussaard*, 2004). However, since viruses are genetically diverse lacking genes that could be targeted by universal primers and the majority cannot be propagated by culturing, both composition and function of the viral community within metaorganisms resisted analysis (*Reyes* et al., 2012). In this article we summarize recent knowledge about the composition and function of viral communities in the basal metazoan *Hydra*.

## VIRUSES OF *HYDRA*

*Hydra* is not only associated with a specific bacterial community they are also associated with a host-specific viral community (*Grasis* et al., 2014). The viral community is composed of eukaryotic viruses that can be expected to infect *Hydra* and a diverse bacteriophage community that might interact with the associated bacterial community. Surprisingly, *Hydra* as basal metazoan is already associated with viruses that are known as causative

agents of infective disease in vertebrates, like Herpesviridae and Poxoviridae, but features also viruses that infect invertebrates, like insect viruses of the family Baculoviridae and even plant viruses of the family Phycodnaviridae. However, these viruses are not causing noticeable disease symptoms in *Hydra*, so that a persistent or chronic viral infection can be proposed. Observed shifts within the viral community composition under environmental stress conditions suggest changes in host-viral interactions or the presence of latent viral infection of *Hydra*. During latency viral genome persists as nucleic acid either integrated or as episome in the nucleus of host cells. Stress exposure, such as temperature, toxins or UV-radiation switch this viral life cycle to a lytic stage and new virons are produced (*Traylen* et al., 2011; *Kenney* and *Mertz*, 2014).

The presence of virus-like particles (VLPs) within the tissue of *Hydra* could be confirmed by transmission electron microscopy (TEM). Microscopic investigation of *Hydra* tissue supports virome sequencing data and the view that viral replication in *Hydra* is balanced or in a controlled mode of viral replication. Baculovirus replicate in *Hydra* cells (Figure 1) and are released via budding from the ectodermal epithelial cell layer (*Deines* et al., 2017). Apart from Baculoviruses VLPs with different morphology could be observed in close proximity to the ectodermal membrane (Figure 1). The observation that VLPs are released to the environment via controlled budding together with an apparent sporadic appearance of VLPs within *Hydra* tissue, e.g. VLPs with morphological similarity to Phycodnaviridae (Figure 1), indicate that several viruses have established a chronic viral infection. In a metaorganismic view this points to a homeostatic relation between virus and *Hydra*. In this balanced state viruses are still able to replicate but in a controlled manner to ensure that host expenses for viral replication do not become detrimental. This suggests a cross-talk between both partners.

## FUNCTION OF VIRUSES WITHIN *HYDRA*

*Hydra* is permanently associated with a host specific viral community. To establish persistent viral infections, viruses have to evade the host immune defence so that viral infections are not entirely cleared. In *Hydra* viral infections can be chronic with a continuous proliferation of virons or lysogenic phages (*Deines* et al., 2017). However, in both cases some viral genes are active and interact with their host and modulate *Hydra's* functions. Viruses have developed a variety of different mechanisms to escape host immune response (*Christiaansen* et al., 2015; *Kang* and *Kieff*, 2015), while the host has coevolved to control viral infections (*Klotman* and *Chang*, 2006; *Iwasaki* and *Medzhitov*, 2015). Virus-host interactions influence cellular pathways and host metabolism (*Goodwin* et al., 2015; *Powdrill* et al., 2016). This functional remodelling of the host *Hydra* is not limited to active viral replication; also viruses in latency remain active and cross-talk with their host. Baculoviridae interfere with cellular pathways of their host during latency (*Davis* et al., 2015). Several genes are transcribed and manipulate immune system, metabolism and cell-cycle (*Monteiro* et al., 2012). Viral infection and functional manipulation may only affect some cells and the im-

pact on the entire individual is limited. Environmental stress conditions can change virus-host interactions and switch latent to lytic viral replication (*Traylen* et al., 2011) leading to an imbalance, uncontrolled viral replication and finally to the onset of disease.

## BACTERIOPHAGES ASSOCIATED WITH *HYDRA*

Bacteriophage community accounts for approximately 60% of the total viral community of *Hydra* and is composed of Myoviridae, Podoviridae, Inoviridae and Siphoviridae (*Grasis* et al., 2014). Predicted host range of bacteriophages based on sequence analysis suggests that the largest proportion of *Hydra* associated phages infect Betaproteobacteria and Gammaproteobacteria (*Grasis* et al., 2014). It has to be pointed out that the phage population does not simply mirror the host specific bacterial community. The bacterial community is dominated by Betaproteobacteria accounting for more than 90% of the bacterial community while Gammaproteobacteria are underrepresented in the bacterial population of *Hydra*. In contrast bacteriophages infecting Beta- and Gammaproteobacteria were equally abundant. It can be expected that the observed difference between phage and bacterial population is caused by:

i) transient phages that originate from the surrounding water, adhere to *Hydra's* surface (*Barr* et al., 2014) and/or infect *Hydra's* associated microbiota or,
ii) resident phages originating from *Hydra's* associated bacterial community act as internal regulators of host specific bacterial community and downregulate Gammaproteobacteria by phage infection.

It is well known that phages are important regulators within bacterial populations (*Proctor* and *Fuhrman*, 1990; *Brussaars*, 2004; *Shapiro* et al., 2010) and we expect phages to play an important role in stabilizing and maintaining host specific bacterial community of the metaorganisms *Hydra*. First evidence that phages could be involved in controlling *Hydra* associated bacteria we got from a simple bacteria-bacteria interaction experiment (*Li* et al., 2015). In this experiment Li and colleagues analysed the interaction of the two main colonizers of *Hydra Curvibacter sp.* and *Duganella sp*. Surprisingly they observed a frequency-dependent, non-linear growth rate of *Duganella sp.*, which could not simply be explained by the presence of *Curvibacter sp.* (*Li* et al., 2015). For this reason we hypothesized that a phage could be hidden in form of a prophage within the genome of *Curvibacter sp.* and serves as third player in this interaction experiment. Screening the genome of both bacteria revealed the presence of a prophage signature in the genome of *Curvibacter sp.* Finally, we could prove that this phage is inducible and able to cross infect *Duganella sp*. This observation emphasizes that phages have a regulatory function within the host associated bacterial community and the occurrence of a hidden phage population, which is present in form of prophages in the genomes of associated bacterial community. Analogue to latent eukaryotic viral infections prophages are transcriptional active and able to modulate their bacterial host (*Mann* et al., 2015). This lysogenic conversion of bacteria increase their genetic repertoire by horizontal gene transfer and may change host bacterial interactions (*Madera* et

al., 2009). Being associated with a prophage can protect bacteria from phage infections by superinfection exclusion. Switching from a lysogenic to a lytic lifecycle can be advantages for the bacterium but also for the eukaryotic host. On one hand reactivated phages can serve as weapon against other bacteria and eliminate competitors on the other hand induction of prophages can function as internal regulation of host associated bacterial community.

## CONCLUSION

Viruses are key components of metaorganisms, nevertheless their composition and function has been neglected. Regarding their diversity, abundance and genetic repertoire viruses have a huge impact on a cellular, organismic and population level. Under normal environmental conditions a homeostatic relation between viral community and their hosts can be expected. Due to their fast evolution and dependency on the host cell replication machinery cross-talk between eukaryotic viruses and their host has coevolved and fine-tuned the innate immune system, while host specific bacteriophages inherit regulatory function within the associated bacterial population.

## ACKNOWLEDGEMENTS

This work was supported by the Volkswagen Foundation (Funding program "Experiment! – *In search of bold research ideas*") and by the Collaborative Research Centre 1182 "Origin and Function of Metaorganisms" granted by the Deutsche Forschungsgemeinschaft DFG.

## LITERATURE

Barr, J.J., Youle, M., and Rohwer, F.: Innate and acquired bacteriophage-mediated immunity. Bacteriophage 3, e25857 (2014).

Bosch, T.C.G.: Cnidarian-microbe interactions and the origin of innate immunity in metazoans. Annu. Rev. Microbiol. 67, 499-518 (2013).

Brussaard, C.P.D.: Viral control of phytoplankton populations - A review. J. Eukaryot. Microbiol. 51, 125-138 (2004).

Christiaansen, A., Varga, S.M., and Spencer, J.V.: Viral manipulation of the host immune response. Curr. Opin. Immunol. 36, 54-60 (2015).

Davis, Z.H., Verschueren, E., Jang, G.M., Kleffman, K., Johnson, J.R., Park, J., Von Dollen, J., Maher, M.C., Johnson, T., Newton, W, Jäger, S., Shales, M., Horner, J., Hernandez, R.D., Krogan, N.J., and Glaunsinger, B.A.: Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes. Mol. Cell 57, 349-360 (2015).

Deines, P., Lachnit, T., and Bosch, T.C.G.: Competing forces maintain the Hydra metaorganism. Immunol. Rev. 279, 123-136 (2017).

Franzenburg, S., Walter, J., Künzel, S., Wang, J., Baines, J.F., Bosch, T.C.G., and Fraune, S.: Distinct antimicrobial peptide expression determines host species-specific bacterial associations. Proc. Natl. Acad. Sci. USA 110, E3730-3738 (2013).

Fraune, S. and Bosch, T.C.G.: Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra.

Proc. Natl. Acad. Sci. USA 104, 13146-13151 (2007).

Goodwin, C.M., Xu, S., and Munger, J.: Stealing the Keys to the Kitchen: Viral Manipulation of the Host Cell Metabolic Network. Trends Microbiol. 23 789-798 (2015).

Grasis, J.A., Lachnit, T., Anton-Erxleben, F., Lim, Y.W., Schmieder, R., Fraune, S., Franzenburg, S., Insua, S., Machado, G., Haynes, M., Little, M., Kimble, R., Rosenstiel, P., Rohwer, F.L., and Bosch, T.C.G.: Species-specific viromes in the ancestral holobiont hydra. PLoS One 9, e109952 (2014).

Iwasaki, A. and Medzhitov, R.: Control of adaptive immunity by the innate immune system. Nat. Immunol. 16, 343-353 (2015).

Kang, M.-S. and Kieff, E.: Epstein-Barr virus latent genes. Exp. Mol. Med. 47, e131 (2015).

Kenney, S.C. and Mertz J.E.: Regulation of the latent-lytic switch in Epstein-Barr virus. Semin. Cancer Biol. 26, 60-68 (2014)

Klotman, M.E. and Chang, T.L.: Defensins in innate antiviral immunity. Nat. Rev. Immunol. 6, 447-456 (2006).

Li, X.-Y., Pietschke, C., Fraune, S., Altrock, P.M., Bosch, T.C.G., and Traulsen, A.: Which games are growing bacterial populations playing? J. R. Soc. Interface 12, 20150121 (2015).

Madera, C., García, P., Rodríguez, A., Suárez, J.E., and Martínez, B.: Prophage induction in Lactococcus lactis by the bacteriocin Lactococcin 972. Int. J. Food Microbiol. 129, 99-102 (2009).

Mann, E., Ovchinnikova, O.G., King, J.D., and Whitfield, C.: Bacteriophage-mediated glucosylation can modify lipopolysaccharide O-antigens synthesized by an ATP-binding Cassette (ABC) Transporter-dependent assembly mechanism. J. Biol. Chem. 290, 25561-25570 (2015).

Monteiro, F., Carinhas, N., Carrondo, M.J.T., Bernal, V., and Alves, P.M.: Toward system-level understanding of baculovirus-host cell interactions: From molecular fundamental studies to large-scale proteomics approaches. Front. Microbiol. 3, 391 (2012).

Powdrill, M.H., Desrochers, G.F., Singaravelu, R., and Pezacki, J.P.: The role of microRNAs in metabolic interactions between viruses and their hosts. Curr. Opin. Virol. 19, 71-76 (2016).

Proctor, L.M. and Fuhrman, J.A.: Viral mortality of marine bacteria and cyanobacteria. Nature 343, 60-62 (1990).

Reyes, A., Semenkovich, N.P., Whiteson, K., Rohwer, F., and Gordon, J.I.: Going viral: next-generation sequencing applied to phage populations in the human gut. Nat. Rev. Microbiol. 10, 607-617 (2012).

Schröder, K. and Bosch, T.C.G.: The Origin of Mucosal Immunity: Lessons from the Holobiont Hydra. MBio 7, e01184-16 (2016).

Shapiro, O.H., Kushmaro, A., and Brenner, A.: Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. ISME J. 4, 327-336 (2010).

Suttle, C.A.: Marine viruses--major players in the global ecosystem. Nat. Rev. Microbiol. 5, 801-812 (2007).

Traylen, C.M, Patel, H.R., Fondaw, W., Mahatme, S., Williams, J.F., Walker, L.R., Dyson, O.F., Arce, S., and Akula, S.M. Virus reactivation: a panoramic view in human infections. Future Virol. 6, 451-463 (2011).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 67-87 (2017).

# THE VIROME, COGNITION, AND HUMAN PSYCHIATRIC DISORDERS: A POSSIBLE ROLE FOR VIRUSES WITH HUMAN, BACTERIAL AND ALGAL HOSTS

R.H. YOLKEN[1], E.G. SEVERANCE[1], R. McFARLAND[1], L. JONES-BRANDO[1], D.D. DUNIGAN[2], K.L. GRESSITT[1], M.V. PLETNIKOV[1], T.M. PETRO[3], V.L. NIMGAONKAR[4], J.L. VAN ETTEN[2], and F.B. DICKERSON[5]

[1]Johns Hopkins School of Medicine, Baltimore, MD, USA;
[2]University of Nebraska-Lincoln, Lincoln, NE, USA;
[3]University of Nebraska Medical Center, Lincoln, NE, USA;
[4]University of Pittsburgh, Pittsburgh, PA, USA;
[5]Sheppard Pratt Hospital, Baltimore MD; USA

## SUMMARY

Human psychiatric diseases such as schizophrenia, bipolar disorder, and major depression are important causes of mortality, morbidity, economic hardship, and social disruption worldwide. Cognitive impairments are hallmarks of these disorders and are major contributors to social dysfunction in affected individuals. Although genetic factors have been found which increase the risk of acquiring these disorders, environmental factors are likely to play an important role in disease progression and severity.

We have previously found that serological evidence of exposure to common herpesviruses such as herpes simplex virus type 1 is associated with decreased functioning in memory and other cognitive domains in individuals diagnosed with schizophrenia or bipolar disorder as well as in individuals without a diagnosed psychiatric disorder. Due to the limitations of serological methods, we applied metagenomic sequencing methods to characterize the nasopharyngeal virome of individuals with and without psychiatric disorders. We found surprisingly that DNA mapping to several bacteriophages, including the lactobacillus phage phi-adh was more prevalent in individuals with schizophrenia as compared to controls and was associated with different clinical patterns and responses to medications. We also found that the chlorovirus Acanthocystis turfacea chlorella virus 1 (ATCV-1) was associated with decreased cognitive functioning in individuals without a psychiatric disorder. A possible role for ATCV-1 in animal biology was supported by animal models and the measurement of an immune response to ATCV-1 proteins in humans.

There is ample evidence that exposure to viruses can alter human cognition and behaviour. The human virome is also likely to contain viruses, which can replicate in non-animal hosts and nonetheless have effects on human health and disease.

## INTRODUCTION

Serious psychiatric disorders such as schizophrenia, bipolar disorder and major depression are major causes of mortality and morbidity worldwide

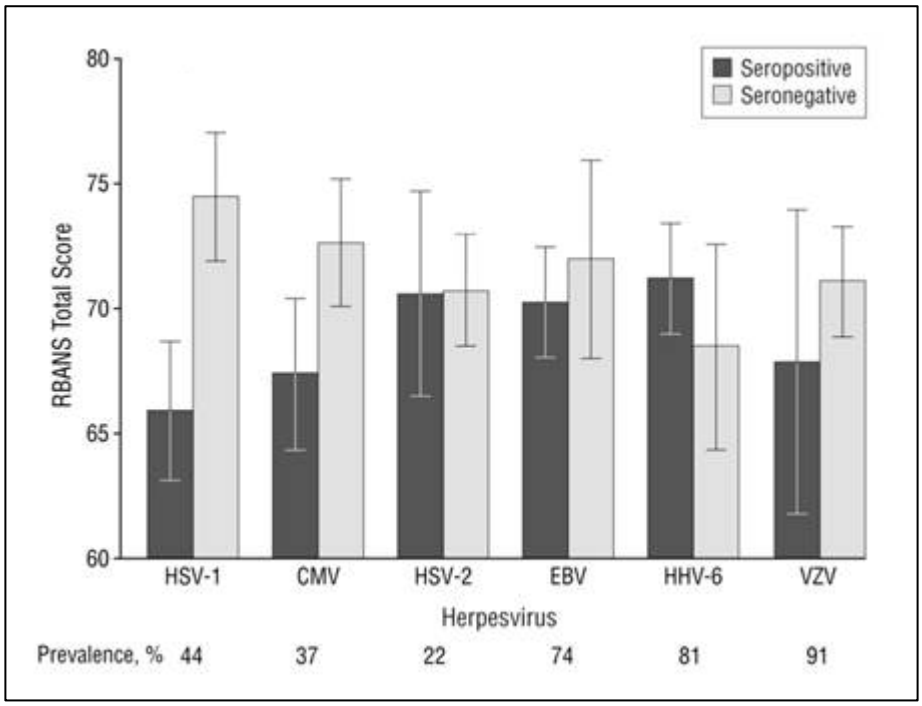


**Figure 1**: Cognitive functioning in individuals with schizophrenia related to serologic evidence of infection with specific herpesviruses. Levels of IgG class antibodies to herpesviruses were measured by means of an enzyme immunoassay, and cognitive functioning was measured by the total score of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) in individuals with schizophrenia (N = 229). Bars indicate the mean scores (±95% confidence intervals [CIs]) of individuals who were seropositive or seronegative for the indicated herpesvirus. The percentage of individuals in the total study population who were seropositive for the indicated herpesvirus is shown in the lower line. The asterisk indicates p<.001 between the RBANS total score for herpes simplex virus 1 (HSV-1) seropositive and seronegative individuals as calculated by 1-way analysis of variance; the differences between the RBANS total scores in terms of seroreactivity to the other viruses were not statistically significant (at the level of α= .008). CMV indicates cytomegalovirus; EBV, Epstein-Barr virus; HSV-2, herpes simplex virus 2; HHV-6, human virus 6; and VZV, varicella-zoster virus. (Figure reproduced from *Dickerson* et al. (2003).

(*Millier* et al., 2014). Numerous studies point to the importance of family history and genetic factors in the aetiology of these disorders (*Bergen* et al., 2014). However, epidemiological studies also point to environmental factors as contributors to disease risk. In particular, studies have documented immune activation in many individuals with psychiatric disorders (*Khandaker* et al., 2015). These studies have intensified interest in understanding the instigating factors for this inflammation as well as for the mechanisms by which immune activation might lead to development of psychiatric disorders in some individuals. Most of this research has focused on exposure to infectious agents and to food antigens in light of the major role of these factors in the generation of immune activation in humans and animal models (*Severance* et al., 2016).

Altered behaviour, mood, and perception are cardinal features of human psychiatric disorders. However, psychiatric disorders are also associated with varying degrees of cognitive

impairment. This is particularly the case in schizophrenia where cognitive impairment is often present and can remain as a significant problem when other symptoms have been alleviated following the administration of antipsychotic medications (*Buchanan* et al., 2005). This residual cognitive impairment is often one of the main barriers to the ability of individuals with schizophrenia to function within work or social environments (*Tas* et al., 2013). Numerous studies in humans and experimental animals have pointed to a role for infectious agents in inducing changes in cognitive functioning particularly in the memory domain (*Williamson* et al., 2011). We thus have explored the relationship between exposure to infectious agents and cognitive functioning in individuals with psychiatric disorders as well as individuals without a psychiatric diagnosis.

## HERPESVIRUSES AND COGNITIVE FUNCTIONING

Our initial studies focused on individuals with schizophrenia because this diagnosis is generally associated with a high rate of cognitive impairment. We also focused on exposure to common infectious agents, including human herpesviruses, because of their high prevalence and their ability to infect the human central nervous systems in some situations (*Dickerson* et al., 2003). As depicted in Figure 1, we found that exposure to herpes simplex virus type 1 (HSV-1), as revealed by the presence of IgG class antibodies, was associated with decreased cognitive functioning as measured by the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). This association correlated with the quantitative level of antibodies, involved several domains including both immediate and delayed memory and was independent of demographic factors such as age, race, gender, and socio-economic status as determined by the level of parental education. The association between decreased cognitive functioning and HSV-1 was also somewhat virus specific in that significant associations were not found with exposures to other herpesviruses such as herpes simplex virus type 2 (HSV-2), Epstein Barr virus (EBV), varicella zoster virus (VZV) or human herpes virus type 6 (HHV-6). There were some associations between cognitive deficits and exposure to cytomegalovirus (CMV), although the strength of these associations was lessened when controlling for demographic factors. Subsequent studies indicated that the association between HSV-1 and cognitive impairment in individuals with schizophrenia is increased in individuals who are cigarette smokers (*Dickerson* et al., 2016) and individuals who have immune activation as evidenced by increased levels of C-reactive protein (*Dickerson* et al., 2012). The association between exposure to HSV-1 and cognitive functioning in individuals with schizophrenia was subsequently found in several other populations in the United States, Europe, and Asia (*Yolken* et al., 2011; *Prasad* et al., 2013; *Thomas* et al., 2013; *Watson* et al., 2013; *Hamdani* et al., 2017)

The possible role of HSV-1 in cognitive functioning in individuals with schizophrenia is supported by additional experimental data. Imaging studies have documented alteration in brain structure and function associated with exposure to HSV-1 in individuals with schizophrenia (*Prasad* et al., 2007; *Schretlen* et al., 2010). It is not clear

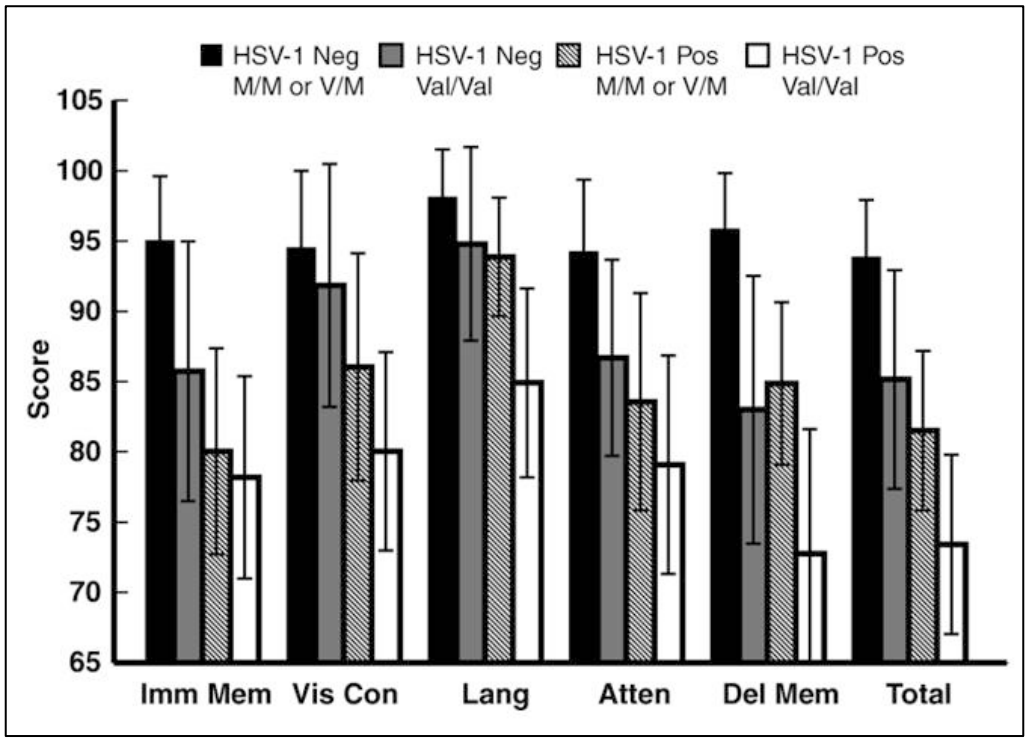


**Figure 2**: Relationship between COMT Val158Met genotype, HSV-1 serological status and RBANS scores. The bars indicate the mean and 95% confidence intervals of the RBANS cognitive scores measured in individuals with bipolar disorder with the indicated HSV-1 and COMT Val158Met polymorphism status. The component scores measure Immediate Memory (ImmMem), Visuospatial Constructional (VisCon), Language (Lang), Attention (Atten), and Delayed Memory (DelMem). All of the components are expressed as age-adjusted scaled scores. The HSV-1 and COMT polymorphism status are depicted as follows: individuals who are HSV-1 seropositive and who have the COMT Met/Met or COMT158 Met/Val genotype (solid bars), individuals who are HSV-1 seronegative and who have the COMT158 Val/Val genotype (shaded bars), individuals who are HSV-1 seropositive and who have the COMT158 Met/Met or COMT158 Met/Val genotype (striped bars), and individuals who are HSV-1 seronegative and who have the COMT158 Val/Val genotype (open bars). (Figure reproduced from *Dickerson* et al. (2006).

whether these effects are related to the replication of HSV-1 in the brain as occurs in encephalitis (*Hahn* et al., 2012), activation of microglia and other immune cells within the central nervous system (*Patterson*, 2015), the development of autoantibodies to brain receptors (*Westman* et al., 2016) or some combination of these and other processes. It is noteworthy in this regard that HSV-1 induces changes in gene expression of neurons derived from human induced pluripotent stem cells (iPSCs), which are consistent with alterations seen in the brains of individuals with schizophrenia (*D'Aiuto* et al., 2012). Furthermore, a pilot study found significant improvement in some cognitive domains but not others or psychiatric symptoms in individuals with schizophrenia treated with the anti-herpes medication valacyclovir (*Prasad* et al., 2013). Follow up studies testing the efficacy of this medication in larger populations of individuals with schizophrenia and serological evidence of infection with HSV-1 are ongoing.

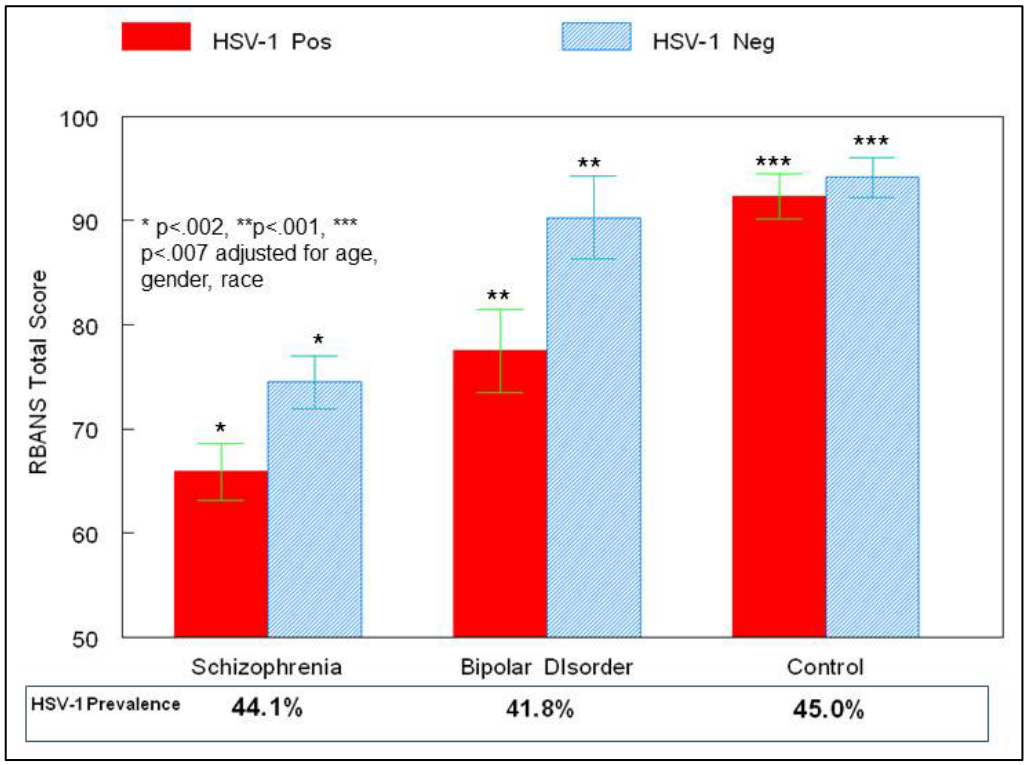


**Figure 3**: The association between serological evidence of exposure to HSV-1 and cognitive functioning as measured by the RBANS Total Score. The numbers in the box indicate similar levels of prevalence to HSV-1 in the 3 populations.

We have also examined the relationship between exposure to common infectious agents and cognitive functioning in individuals with bipolar disorder. We found an association between exposure to HSV-1 and decreased scores on tests of memory in these individuals (*Dickerson* et al., 2004). Overall the level of cognitive impairment in individuals with bipolar disorder is less than that in individuals with schizophrenia. However, the relative effect of exposure to HSV-1 is similar. We also uncovered an example of gene-environmental interaction in the form of polymorphisms in the gene encoding catechol-*O*-methyltransferase (COMT) (*Dickerson* et al., 2006). As shown in Figure 2, there was an additive effect of exposure to HSV-1 and the genotype of COMT. Individuals who had serological evidence indicating exposure to HSV-1 and the high-risk val/val genotype of COMT had the most cognitive impairment. On the other hand, individuals who were not exposed to HSV-1 and who had the low risk Met/Met genotype of COMT had relatively low levels of cognitive impairment. The scores measured in individuals in this group did not differ substantially from the scores of control individuals. HSV-1 has also been reported to be associated with lower levels of cognitive functioning in other populations of individuals with bipolar disorder (*Gerber* et al., 2012; *Hamdani* et al., 2017).

We have also examined the relationship between exposure to HSV-1 and

cognitive functioning in individuals who do not have a psychiatric disorder. Initial studies using small sample sizes did not demonstrate an association. However, when we tested larger sample sizes we found that individuals who had serological evidence of exposure to HSV-1 had small but statistically significant decreases in functioning particularly in the domain of delayed memory (*Dickerson* et al., 2008). We also found an interaction with the COMT gene as described above in the population of individuals with bipolar disorder. These associations were independent of demographic factors such as age, gender, race, and maternal education. Thus as depicted in Figure 3, exposure to HSV-1 is associated with performance on tests of memory in individuals with schizophrenia or bipolar disorder and individuals without a psychiatric disorder in our study population. However the relative levels of cognitive performance differ greatly among these populations. HSV-1 has been associated with decreased cognitive functioning in several other populations including healthy young adults (*Fruchter* et al., 2015; *Hamdani* et al., 2017), middle aged adults (*Gale* et al., 2016), children with a liability to substance use disorder (*Vanyukov* et al., 2017a,b) and elderly individuals (*Itzhaki*, 2014). Recent progress has been made on the development of new medications (*D'Aiuto* et al., 2017) and vaccines (*Chentoufi* et al., 2012) for the prevention of HSV-1 infection and the effective treatment of latent infection. Our studies support the need to develop methods for the prevention of the serious effects of HSV-1 infection on many different populations worldwide.

There are many possible explanations for the HSV-1 related associations with cognitive dysfunction. Foremost is the criticism that many of the reports described above are correlative in nature; moreover, the cognitive changes or the changes in brain imaging variables could be attributed to another variable, such as a coincidental infection or even a demographic variable such as low socio-economic status that correlates well with most infections. Therefore, we have attempted to evaluate whether the aforementioned studies are sufficient to invoke causal relationships between HSV-1 infection and cognitive impairment using the Bradford-Hill criteria (*Hill*, 1965). Dr. Bradford Hill enunciated the eponymous criteria to evaluate causal links in the context of common exposures such as cigarette smoke and common, chronic conditions, such as cancers. We evaluated HSV-1 infection and cognitive dysfunction in an earlier review (*Prasad* et al., 2012) and have updated them in Table 1. We surmise that moderate to strong evidence supporting five of the nine criteria are available, namely the criteria related to strength, consistency, plausibility, temporality and coherence. It should be noted that the criterion for 'biological gradient' cannot be tested in this context because the 'dose' of HSV-1 virions during the initial or the on-going infection cannot be calculated and proxies such as the host antibody response reflects the severity of infection imprecisely, particularly in the chronic stage. Unlike the classic Koch's postulates that require rigid confirmation to all its requirements, Bradford Hill emphasized: 'What I do not believe – and this has been suggested – is that we can usefully lay down some hard-and-fast rules of evidence that *must* be obeyed before we accept cause and effect.' Thus, overall, the HSV-1 related effects described in this monograph cannot be dismissed merely as trivial or chance associations.

**Table 1**: HSV-1 associated cognitive impairments evaluated in relation to Bradford Hill criteria

| Criterion | Explanation | Evidence |
|---|---|---|
| Strength | A strong association is more plausible, but it is not a *sine qua non*. | Association between HSV-1 exposure and cognition has small to medium effect size (*Dickerson* et al., 2003; *Thomas* et al. 2013; *Watson* et al., 2013). |
| Consistency | Causal effects need to be replicable. | The association has been detected in over ten studies (*Dickerson* et al., 2003, 2004, 2008, 2012; *Prasad* et al., 2007, 2012a; *Schretlen* et al., 2010; *Shirts* et al., 2008; *Strandberg* et al., 2003; *Tarter* et al., 2014; *Thomas* et al., 2013; *Watson* et al., 2013; *Yolken* et al., 2011; *Vanyukov* et al., 2017a,b). |
| Specificity | Need to consider alternative explanations, including potential confounding factors. | The association is detectable after accounting for potential confounding factors such as age, gender and socio-economic status. However, other herpes viruses are also associated with cognitive impairments, e.g., cytomegalovirus (*Nimgaonkar* et al., 2016) and herpes simplex virus, type 2 (*Watson* et al., 2013). |
| Plausibility | Plausible biological mechanisms for the associations should be available. | (i) The cognitive impairment could be due to immune reactions to recurrent infection (*Steiner* et al., 2007; *Dantzer* et al., 2008; *Li* et al., 2006) (ii) it could be due to an initial infection in childhood that impairs neurodevelopment (*Meyer* et al., 2009; *Vanyukov* et al., 2017a,b); (iii) it could be due to latent infection in the brain (*Becker*, 1995). |
| Experiment | Another line of evidence supporting the association, such as a treatment trial. | A randomized double blind trial of adjunctive treatment with acyclovir, a specific antiviral drug for HSV-1 infection led to improved cognitive function among HSV-1 seropositive patients with schizophrenia (*Prasad* et al., 2012), a similar trial in patients with chronic schizophrenia seropositive for cytomegalovirus did not show beneficial effects (*Dickerson* et al., 2009). |
| Temporality | HSV-1 exposure should predate cognitive impairments | Evidence from longitudinal follow up studies is mixed, but the majority of studies have been conducted in adults in whom duration of exposure is uncertain (*Strandberg* et al., 2003; *Aiello* et al., 2008; *Prasad* et al., 2012; *Barnes* et al., 2014; *Nimgaonkar* et al., 2015). The only prospective study among children found supportive evidence linking prior HSV-1 exposure with cognitive impairment (*Vanyukov* et al., 2017a,b). |
| Coherence | Congruence between epidemiological and laboratory findings | Latent infection has been modelled in neuron-like cells derived from human induced pluripotent stem cells (hiPSCs) that have features of cortical glutamatergic neurons, suggesting that latent HSV-1 infection can be established in the brain (*D'Aiuto* et al., 2014). This provides a mechanism linking persistent HSV-1 infection directly to cognitive dysfunction. |
| Biological Gradient | Demonstrable link between 'dose' of risk factor such as smoking, and 'effect', such as lung cancer. | It is not possible to test in the context of viral infections and subsequent cognitive impairment, as the 'dose' of initial infective viral load cannot be assayed. |
| Analogy | Other lines of evidence, such as animal studies. | One study in a rodent model of HSV-1 infection showed cognitive impairment in animals following HSV-1 infection *Beers, et al.* (1995). |

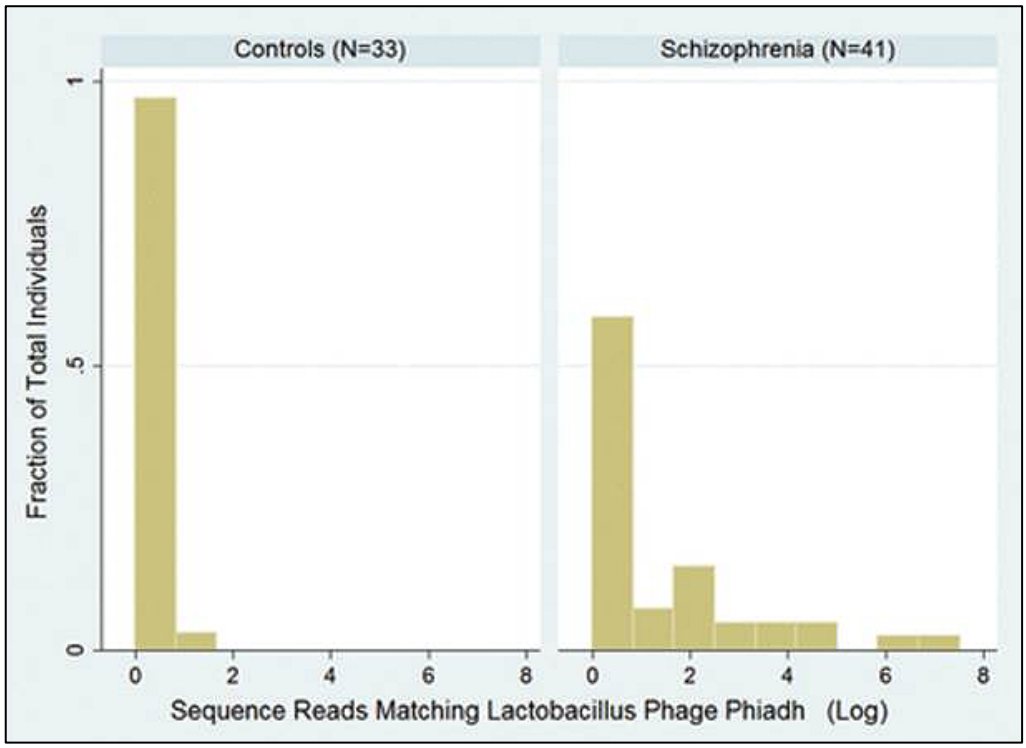


**Figure 4**: Histogram depicting the number of sequence reads matching Lactobacillus phage φ-adh (NC_000896) obtained from 41 individuals with schizophrenia and 33 controls. Sequence matches were originally identified by the use of the CLC homology algorithm and confirmed by Blastn homology searches against the entire non-redundant nucleotide database. (Figure reproduced from *Yolken*, et al. (2015).

## METAGENOMIC SEQUENCING

The above studies indicate that a virus, which replicates at mucosal surfaces, might be associated with an increased risk of psychiatric disorders and cognitive dysfunction. To further explore this interaction we are employing metagenomic sequencing techniques capable of detecting viral sequences at mucosal surfaces. Studies completed to date have been performed using DNA extracted from throat swab samples from individuals with psychiatric disorders and controls. Throat swab samples were initially selected for analysis since they were easy to obtain from the study population in a non-traumatic manner and could be collected on multiple occasions from the same individual. In our initial study, we employed high throughput sequencing to generate more than 100,000,000 sequence reads from samples obtained from 41 individuals with schizophrenia and 33 control individuals without a psychiatric diagnosis (*Yolken* et al., 2015). After matching to available databases, we did not find any known human viruses which distinguished cases from controls or which correlated with cognitive functioning within the groups.

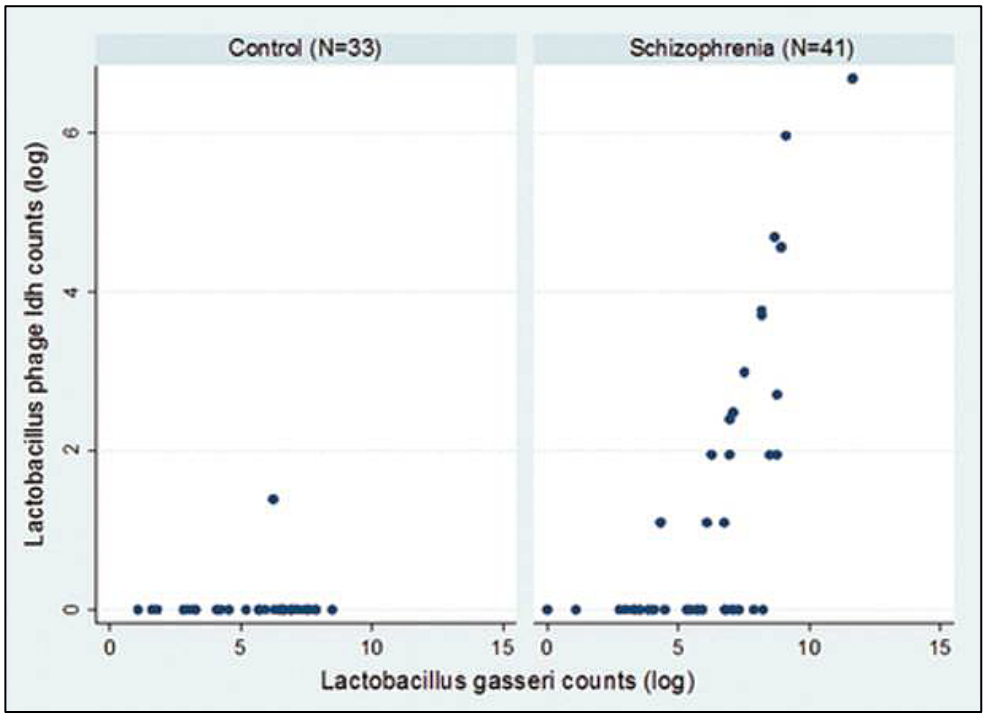


**Figure 5**: Scatterplot indicating the relationship between number of Lactobacillus phage φadh (NC_000896) and *Lactobacillus gasseri* sequence reads in individuals with schizophrenia (right) and controls (left). Note that schizophrenia diagnosis correlates more strongly with the level of φadh as compared with that of the host bacterial species *Lactobacillus gasseri*. (Figure reproduced from *Yolken* et al. (2015).

## A PHAGE ASSOCIATED WITH SCHIZOPHRENIA?

However, we did note some interesting associations with what are usually characterized as "non-human viruses". For example, we identified 79 distinct bacteriophage sequences in the oropharyngeal samples. Of these, one bacteriophage genome, that encoded the lysogenic Lactobacillus phage phi-adh (φadh) **(***Raya* et al., 1989) was significantly different in individuals with schizophrenia ($p<0.00037$, $q<0.03$ adjusted for multiple comparisons) (Figure 4). The different levels of φadh remained significant after controlling for age, gender, race, socioeconomic status, or cigarette smoking ($p<0.006$). Also, the level of φadh correlated better with schizophrenia status than the level of the corresponding host bacteria (Figure 5). Within the group of individuals with schizophrenia, the level of φadh sequences did not correlate with clinical symptoms or demographic variables. However, the level correlated with the prevalence of immunological disorders, particularly type 2 diabetes. The level of sequences homologous to φadh did not correlate with the administration of standard anti-psychotic medications. However, the level did correlate with the administration of the mood stabilizing medication valproic acid. This medication is widely used for the treatment of schizophrenia, particularly in cases of individuals who do not respond to standard medications

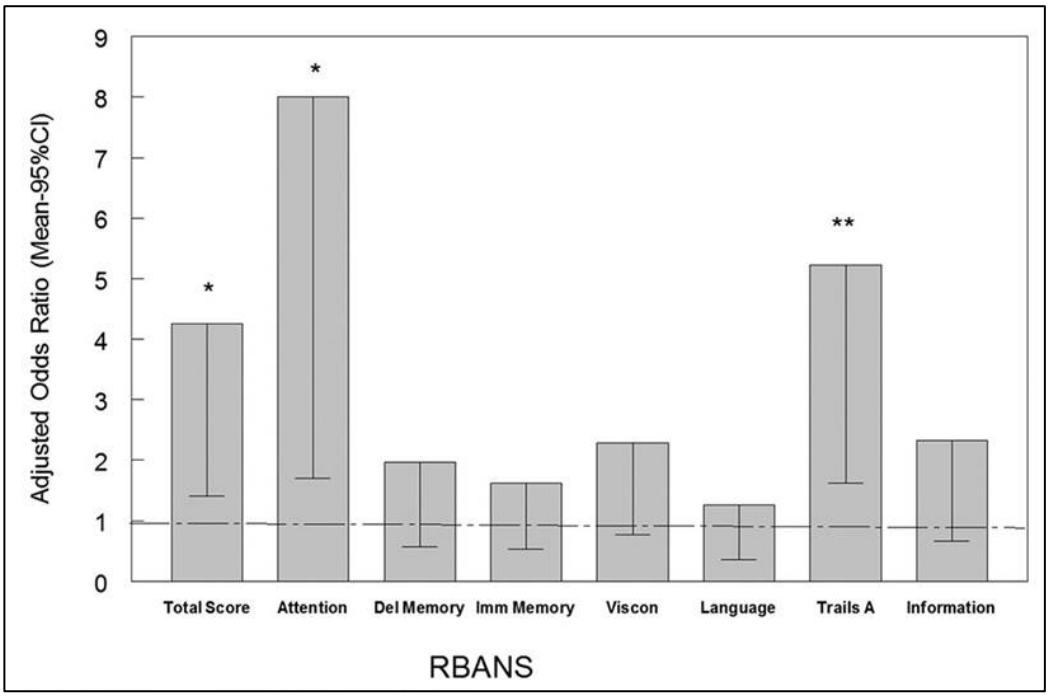


**Figure 6**: Odds of detecting *Chlorovirus* ATCV-1 in the pharynx by percentile of score on cognitive testing in humans without a psychiatric disorder. Bars represent the mean and 95% confidence interval odds of detecting ATCV-1 DNA in the oropharynx in individuals with the indicated test. The odds ratios are adjusted for the demographic variables of age, sex, race, maternal education, educational status, and place of birth in the United States. Trails A and Information are separate tests and not part of the RBANS. $^{**}p < 0.005$, $^{*}p < 0.01$, adjusted for the same covariates. (Figure reproduced from *Yolken* et al. (2014).

(*Wang* et al., 2016). While it has many biological effects its specific mode of action in schizophrenia is not known with certainty. Previous studies have shown that the administration of valproate alters the microbiome in animal models perhaps related to its homology with fatty acids (*de Theije* et al., 2014). Our finding suggests that valproate may also alter the phage composition of the virome, probably due to alterations in the levels of the bacterial hosts. The possibility that valproate and perhaps other medications employed in schizophrenia exert at least some of their effect through alterations in the microbiome should be the topic of future studies. Further analyses of the role of the immune response to phage in terms of cognitive functioning in individuals with schizophrenia are on-going.

We also measured exposure to φadh by immunoassays that measured the immune response to viral proteins in 620 blood samples obtained from 323 individuals with schizophrenia not of recent onset. We found widespread serological evidence of exposure to φadh as revealed by the prevalence of antibody. Differences in IgG class antibodies did not correlate with available clinical or demographic data. However, the level of IgA class antibodies expressed as normalized scores, were associated with lower levels of cognitive functioning in individuals with schizophrenia as measured by the

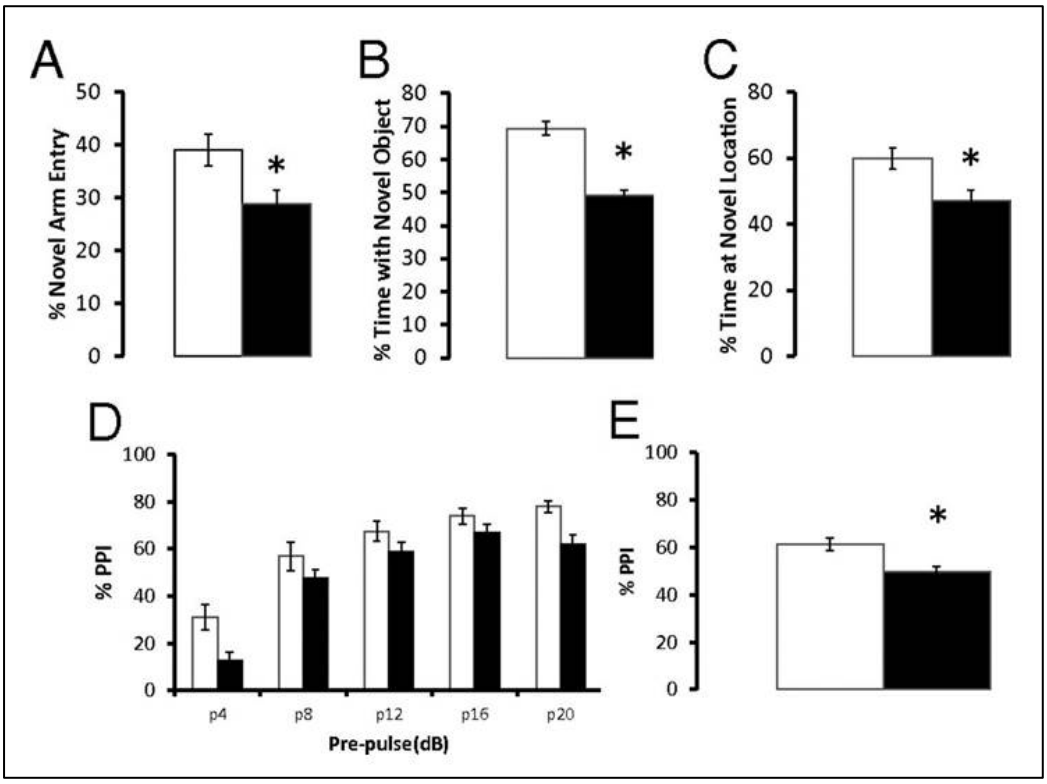


**Figure 7**: Behavioural effects of oral ATCV-1 exposure. Mice were orally infected with the alga *Chlorella heliozoae* alone (open bars) or with ATCV-1–infected *C. heliozoae* (solid bars) as described in the text. (A) Spatial recognition memory; the y-axis displays the percentage of the previously blocked (i.e., novel) arm entries; *p = 0.015 measured by one-way ANOVA. (B) Novel object recognition; the y axis depicts the percentage of time spent exploring the novel object; *p < 0.001 measured by one-way ANOVA. (C) Place recognition memory recognition; the y axis depicts the percentage of time spent exploring the new location of the familiar object; *p < 0.008 measured by one-way ANOVA. (D) Impaired PPI; mice were exposed to presentation of pulse alone (120 dB) and prepulse–pulse combinations across different prepulse intensities; for example, p4 indicates pairing of the prepulse (4 dB above the background noise of 70 dB) with the pulse alone (120 dB) (see the text for more details); the y axis displays the percentage of PPI. (E) Impaired average PPI; the y axis displays the percentage of PPI; *p < 0.015 measured by post hoc test. (Figure reproduced from *Yolken* et al. (2014).

RBANS total score (regression coefficient -1.21, 95% confidence interval -2.35, -0.06, p = 0.039) and the RBANS visuospatial/constructional score (regression coefficient -1.253208, 95% confidence interval -2.29, -.21, p=0.019) both coefficients adjusted for age, gender, race, level of maternal education and multiple samples per individual. These findings suggest that there is widespread exposure to φadh at mucosal surfaces where IgA antibodies can be generated and that this exposure may be associated with lower levels of cognitive functioning in some individuals with schizophrenia. Analysis of larger sample sizes will be required to explore the relationship between antibodies to phage proteins and cognitive functioning in other populations.

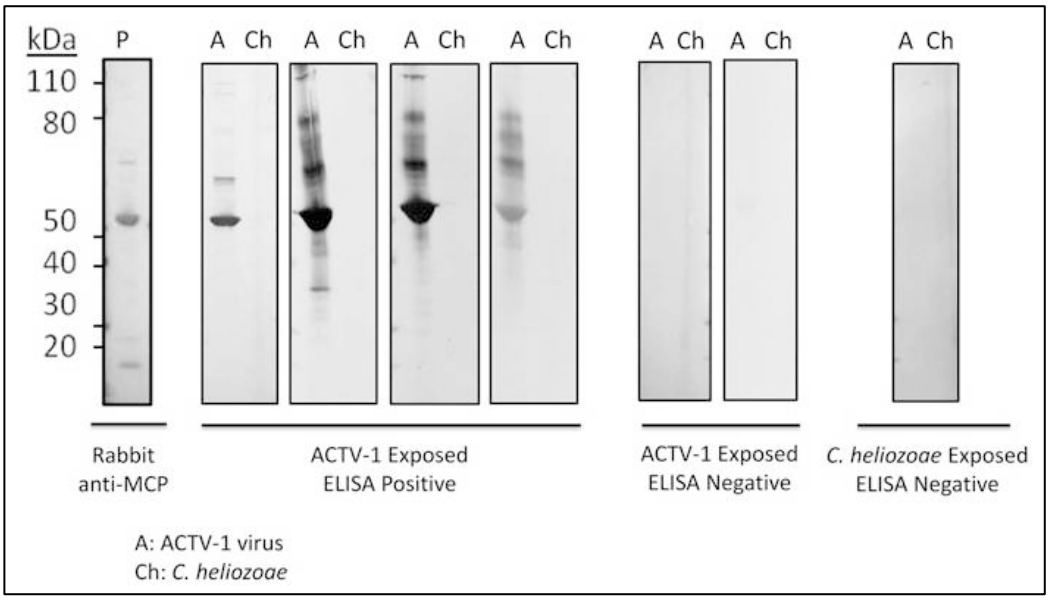


**Figure 8**: Western Blot assays performed with antigens derived from purified ATCV-1 (A) and *C. heliozoae* (Ch). The first lane (labelled P) is from rabbit antibody prepared against the major capsid protein (A430L) of chlorovirus PBCV-1 as a reference. ATCV-1 exposed ELISA positive, reactivity of sera from mice exposed to ATCV-1 and reactive to ATCV-1 antigens by ELISA. ATCV-1 exposed ELISA negative, reactivity of sera from mice exposed to ATCV-1 and not reactive to ATCV-1 antigens by ELISA. *C. heliozoae* exposed ELISA negative, reactivity of serum from a mouse exposed to *C. heliozoae* in the absence of ATCV-1. All mice with this exposure were nonreactive by ELISA. (Figure reproduced from *Yolken* et al. (2014).

## A CHLOROVIRUS ASSOCIATED WITH COGNITIVE BEHAVIOUR?

The above studies indicate that a virus not generally considered important to human health is associated with increased risk of a psychiatric disorder and a decreased level of cognitive functioning. Metagenomic sequencing performed on throat swab samples obtained from 92 individuals without a psychiatric diagnosis revealed other viruses, which are generally not considered to infect humans or animals. Of particular interest in this regard are a group of viruses called Chloroviruses, the natural hosts of which are eukaryotic green algae (*Van Etten* and *Dunigan*, 2012). One of these chloroviruses, called Acanthocystis turfacea chlorella virus 1 (ATCV-1) was associated with lower scores on some cognitive tests in individuals without a psychiatric diagnosis. The most affected tests were ones that measured visual processing and visual motor speed (Figure 6). In order to verify an association between ATCV-1 and lower cognitive functioning we exposed mice to ATCV-1 by gavage and noted alterations in several cognitive domains, including ones involving recognition memory and sensory gating (Figure 7) (*Yolken* et al., 2014). In a subsequent experiment intracranial inoculation of ATCV-1 into mice resulted in impaired memory as well as altered levels of immune markers (*Petro* et al., 2016). Additional studies indicated that ATCV-1 achieves partial replication in mouse macrophages resulting in the expression of

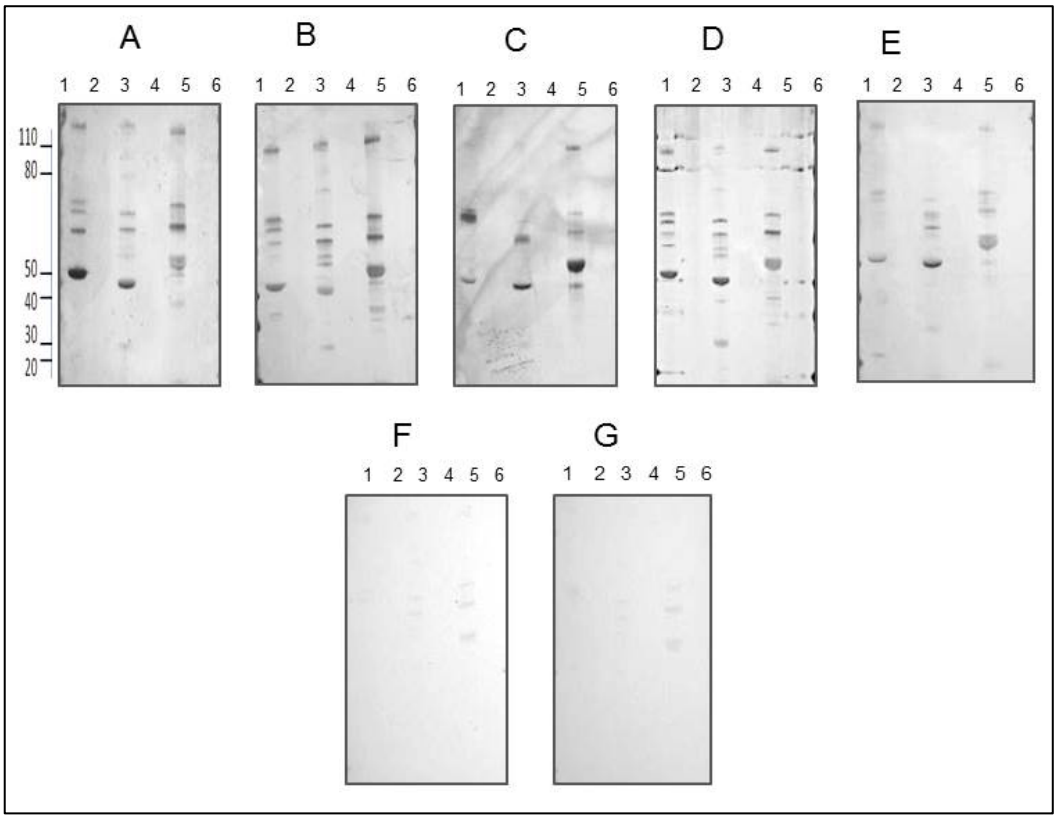


**Figure 9**: Western Blot reactivity of human sera to proteins derived from ATCV-1 as well as 2 additional chloroviruses, Paramecium bursaria chlorella virus 1 (PBCV-1) and Paramecium bursaria chlorella virus CVM-1 (CVM-1) as well as their host algae.
Antigens contained in the lanes are as follows: 1. ATCV-1 chlorella virus; 2. *C. heliozoae* (SAG-3) host for ATCV-1 3; CVM-1 chlorella virus; 4. *Micractinium conductrix* Pbi host for CVM-1; 5. PBCV-1 chlorella virus; 6. *Chlorella variabilis* NC64A host for PBCV-1.
The samples tested were: A: 46 year-old female with Bipolar Disorder born in Maryland. B: 20 year old male with recent onset psychosis born in Maryland. C: 31-year-old male with recent onset psychosis born in New York. D: 40-year-old female with schizophrenia born in New York. E: 26-year-old woman with mania born in Maryland. F: 29-year-old female with schizophrenia born in Pennsylvania. Panel G depicts the reactivity to labelled anti-human IgG in the absence of added serum.

immune mediators (*Petro* et al., 2015). These studies suggest that ATCV-1 may be exerting its effect on cognition through immune activation, a process that has previously been found to link other infectious agents with altered cognitive functioning in humans (*Rempel* et al., 2013).

We have further investigated the immune response to ATCV-1 in humans. Initial studies were performed using Western Blotting. Antibodies to multiple proteins were detected in mice exposed to ATCV-1 (Figure 8). Furthermore, antibodies to several ATCV-1 proteins and antigenically related chloroviruses were detected in individuals with a psychiatric disorder (Figure 9) as well as in individuals without a psychiatric disorder. Mice and human sera also reacted with two recombinant, hypothetical ATCV-1 proteins, Z227L

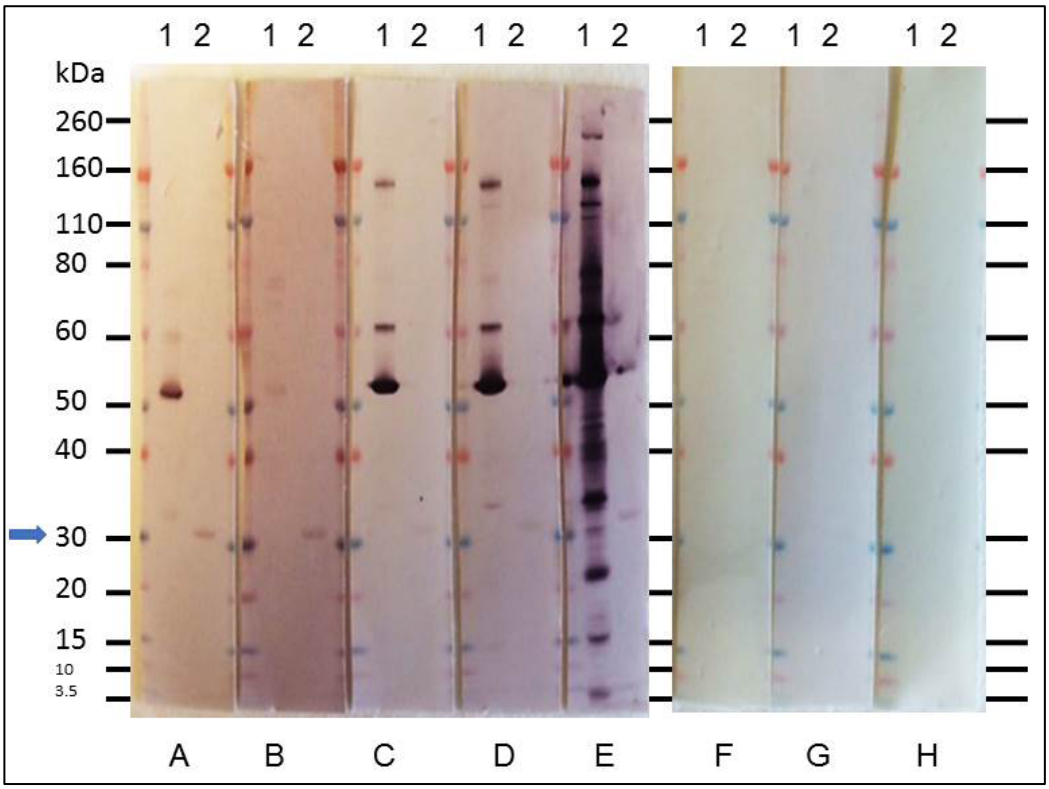


**Figure 10**: Reactivity of human and mouse sera to proteins derived from ATCV-1 virion- (lanes marked "1") and recombinant protein ATCV-1_Z227L (lanes marked "2"). The expected position of the cloned recombinant protein ATCV1_Z227L is shown by the arrow.
The samples tested were: A. 58 year old female without a psychiatric disorder born in Maryland; B. 49 year old female with schizophrenia born in Maryland; C. Mouse exposed to ATCV-1 through the gastrointestinal tract; D. Mouse exposed to ATCV-1 through the gastrointestinal tract. E. Rabbit immunized with ATCV-1. F-H. Samples without added serum processed with secondary antibody directed at human, mouse and rabbit immunoglobulins, respectively.

(GenBank: ABT16361.1) and Z223L (NCBI Reference Sequence: YP_001426704.1). These antigenically related proteins were originally selected because they have no homologs in GenBank and they are not part of the virus proteome. Therefore, the host immune system would presumably only be exposed to the proteins during viral replication (*Van Etten* and *Dunigan*, 2012) We found that human and mouse sera with antibodies to other ATCV-1 proteins react to hypothetical ATCV-1 protein Z227L cloned into a baculovirus vector (Figure 10). We also discovered that some human sera recognize epitopes in these proteins as measured by reactivity to overlapping synthetic peptides derived from hypothetical ATCV-1 protein Z223L (Figure 11). These findings indicate that for some humans there is a clear interaction between the systemic immune system and proteins antigenically related to ATCV-1 encoded proteins. Additional studies using Western Blot assays and synthetic peptide techniques are on-going.

We have performed additional studies in mice in an attempt to identify the pathological consequences of ATCV-1 exposure. We found that mice infected with ATCV-1 administered by the

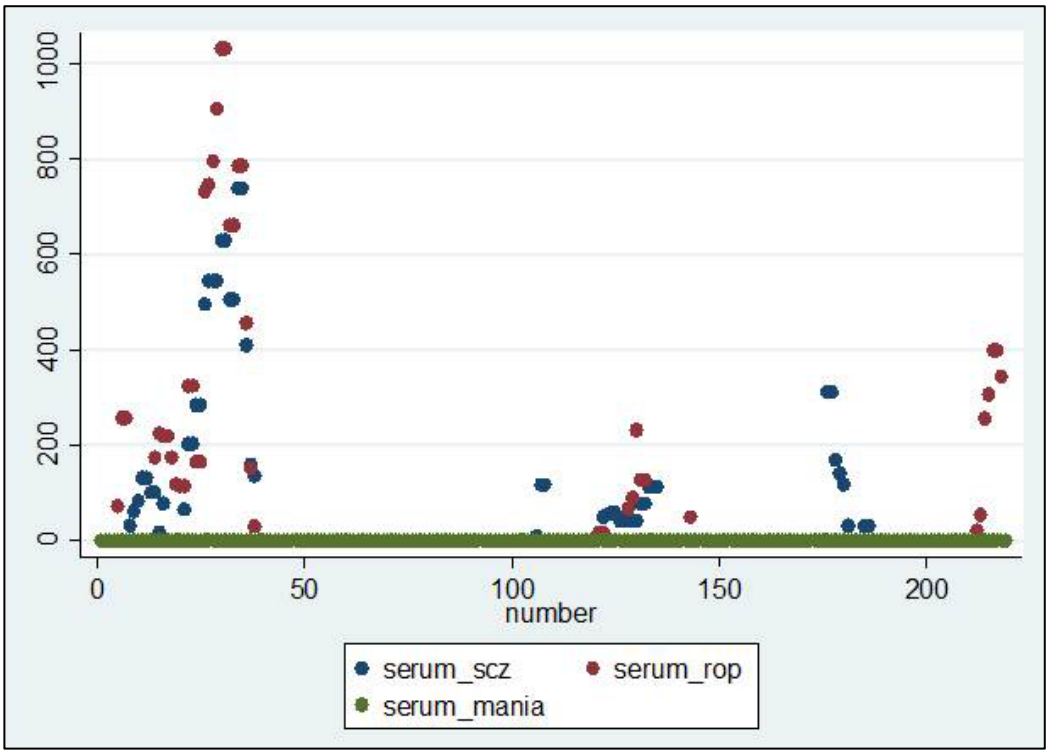


**Figure 11**: Reactivity of 3 human sera to overlapping peptides spanning the entire length of hypothetical protein ATCV-1_Z223L. The numbers on the x-axis indicate the peptide number based on the target sequence (NCBI Reference Sequence: YP_001426704). The numbers on the y-axis indicate the relative activity in arbitrary units. Each dot represents the reactivity to a single peptide starting at the indicated position. "Serum_scz" indicates sample from an individual with established schizophrenia, "serum_rop" indicates a sample from an individual with recent onset psychosis and "serum_mania" indicates a sample from an individual with mania.

intraperitoneal route are associated with periportal inflammation in the liver (Figure 12). We also identified ATCV-1 antigens in the livers of mice infected by this route, most likely in cells with phagocytic activity (Figure 13). The testing of samples from humans with liver diseases is on-going.

We have also examined other body sites for the presence of DNA sequences related to ATCV-1. We did not find any sequences in faecal samples obtained from a small cohort of individuals with a psychiatric disease or controls (*Schwarz* et al., 2017). We also examined plasma samples for the presence of ATCV-1 DNA using a previously described quantitative PCR method (*Yolken* et al., 2014). Testing of more than 200 samples from individuals with psychiatric disorders and controls resulted in the identification of one sample containing detectable ATCV-1 DNA. This sample was obtained from a 51 year old man from Maryland with schizophrenia and lung cancer. Metagenomic sequencing of the sample identified 304 sequence reads, which had high identity to ATCV-1 (Figure 14). A skin swab sample taken from this individual at the same time did not have any detectable DNA sequences homologous to ATCV-1 indicating that the detection in the plasma

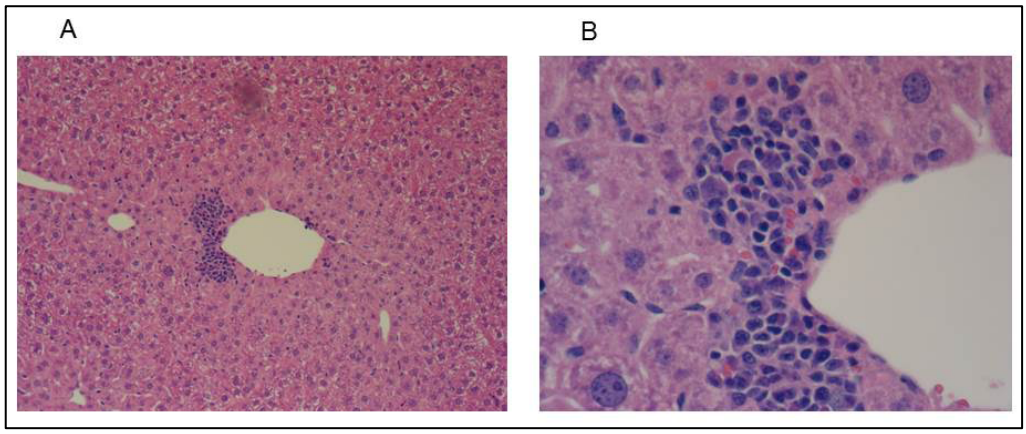


**Figure 12**: Haematoxylin and eosin stained sections of liver from mice exposed to ATCV-1 by the intraperitoneal injection route displaying cellular infiltration around the portal tract. The magnifications are 10x (A) and 40x (B). Significant infiltration adjacent to the portal tract was present on approximately 50% of hepatic portal veins in all ATCV-1 exposed animals and was not observed at all in ATCV-1 naive animals.

was not simply due to skin contamination. These findings indicate that ATCV-1 may in some cases result in systemic infection.

## CONCLUSIONS

Our studies indicate that exposure to infectious agents contribute to the pathogenesis of human psychiatric disorders. Similarly, exposure to microbial agents is also associated with lower levels of cognitive functioning in individuals with psychiatric disorders as well as individuals without a psychiatric disorder. Some of the microbial agents are well-recognized human pathogens. However, other viral agents identified by metagenomic sequencing and associated with these states are not normally considered to be "human" viruses or even animal viruses because their known primary hosts are either bacteria or algae. The surprising finding that "non-human" viruses can produce an immune response in humans and alter behaviour in animal models suggests that even though the viruses are considered to be "non-human" they need to be considered as factors affecting human health. The mechanism(s) by which these "non-human" viruses affect humans is not known with certainty but they are likely to exert their affect through changes in the immune system or alterations in the microbiome. The exact mechanisms by which these "non-human" viruses can affect human health remain an exciting area for future research.

## ACKNOWLEDGEMENTS

The work was partially supported by funding from the Stanley Medical Research Institute and National Science Foundation-Experimental Program to Stimulate Competitive Research Grant EPS-1004094 (JVE)

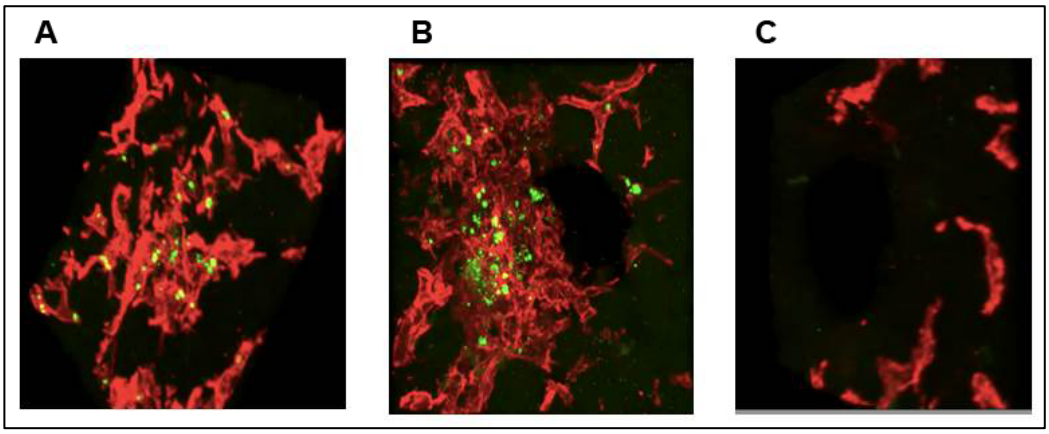


**Figure 13**: Reactivity of fluorescent-labelled rabbit antibody to ATCV-1 (green) and rat antibody to mouse IBA-1 (red). (A) IP injected ATCV-1 exposed mouse, section of mid liver. (B) ATCV-1 exposed mouse, section of the peri-portal region of the liver. (C) Peri-portal region of an unexposed mouse. Magnifications are at 100x. Three-dimensional analysis confirms the presence of foci of ATCV-1 antigen clustered inside IBA-1 positive cells, indicating that the viral antigen is found intracellularly almost exclusively in phagocytic cells. These cells may be infiltrative phagocytic cells, or consistent with their location within the liver tissue, are likely to include hepatic resident macrophages.

## LITERATURE

Aiello, A.E., Haan, M.N., Pierce, C.M., Simanek, A.M., and Liang, J.: Persistent infection, inflammation, and functional impairment in older Latinos. J. Gerontol. A Biol. Sci. Med. Sci. 63, 610-618 (2008).

Barnes, L.L., Capuano, A.W., Aiello, A.E., Turner, A.D., Yolken, R.H., Torrey, E.F., and Bennett, D.A.: Cytomegalovirus infection and risk of Alzheimer disease in older black and white individuals. J. Infect. Dis. 211, 230-237 (2014).

Becker, Y.: HSV-1 brain infection by the olfactory nerve route and virus latency and reactivation may cause learning and behavioral deficiencies and violence in children and adults: a point of view. Virus Genes. 10, 217-226 (1995).

Beers, D.R., Henkel, J.S., Kesner, R.P., and Stroop, W.G.: Spatial recognition memory deficits without notable CNS pathology in rats following herpes simplex encephalitis. J. Neurol. Sci. 131, 119-127 (1995).

Bergen, S.E., O'Dushlaine, C.T., Lee, P.H., Fanous, A.H., Ruderfer, D.M., Ripke, S., International Schizophrenia Consortium, Swedish Schizophrenia Consortium, Sullivan, P.F., Smoller, J.W., Purcell, S.M., and Corvin, A.: Genetic modifiers and subtypes in schizophrenia: investigations of age at onset, severity, sex and family history. Schizophr. Res. 154, 48-53 (2014).

Buchanan, R.W., Davis, M., Goff, D., Green, M.F., Keefe, R.S., Leon, A.C., Nuechterlein, K.H., Laughren, T., Levin, R., Stover, E., Fenton, W., and Marder, S.R.: A summary of the FDA-NIMH-MATRICS workshop on clinical trial design for neurocognitive drugs for schizophrenia. Schizophr. Bull. 31, 5-19 (2005).

Chentoufi, A.A., Kritzer, E., Yu, D.M., Nesburn, A.B., and Benmohamed, L.: Towards a rational design of an asymptomatic clinical herpes vaccine: the old, the new, and the unknown. Clin. Dev. Immunol. 2012, 187585 (2012).

D'Aiuto, L., Di Maio, R., Heath, B., Raimondi, G., Milosevic, J., Watson, A.M., Bamne, M., Parks, W.T., Yang, L., Lin, B., Miki, T., Mich-Basso, J.D., Arav-Boger, R., Sibille, E., Sabunciyan, S., Yolken, R., and

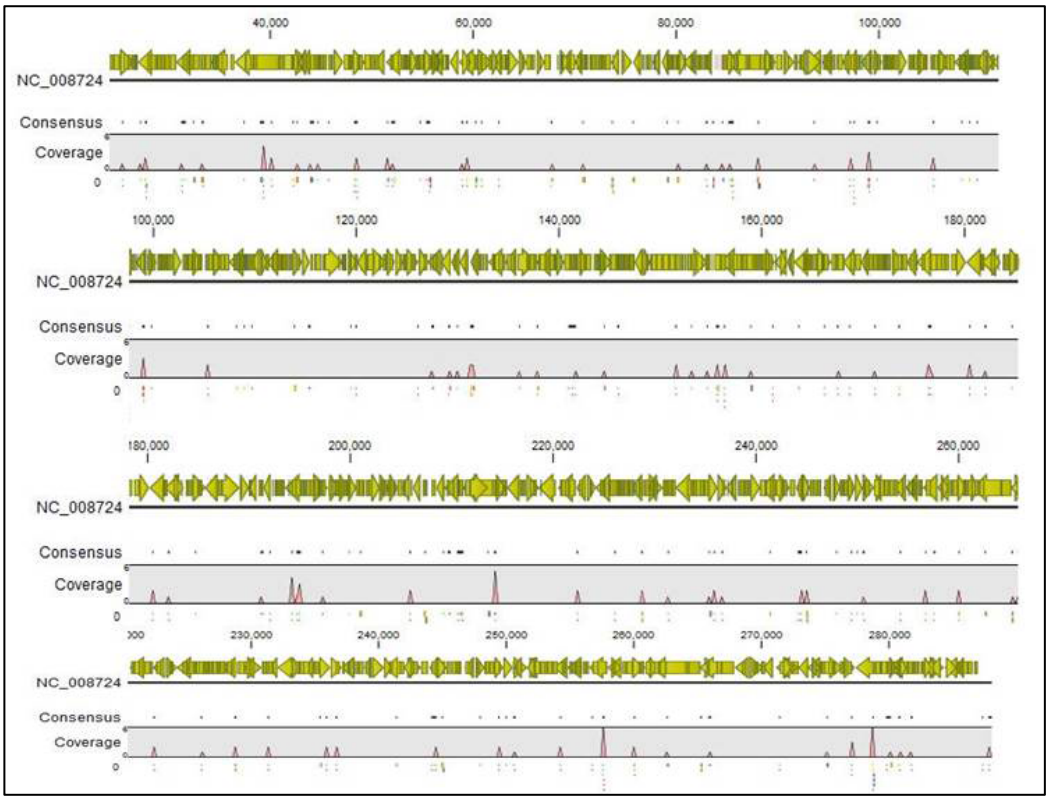


**Figure 14**: Sequence matches to ATCV-1 found in the blood of a 51 year old man with schizophrenia and lung cancer. The methods used were similar to those employed for ATCV-1 sequences derived from throat swab samples as described in *Yolken* et al. (2014).

Nimgaonkar, V.: Human induced pluripotent stem cell-derived models to investigate human cytomegalovirus infection in neural cells. PLoS One 7, e49700 (2012).

D'Aiuto, L., Prasad, K.M., Upton, C.H., Viggiano, L., Milosevic, J., Raimondi, G., McClain, L., Chowdari, K., Tischfield, J., Sheldon, M., Moore, J.C., Yolken, R.H., Kinchington, P.R., and Nimgaonkar, V.L.: Persistent infection by HSV-1 is associated with changes in functional architecture of iPSC-derived neurons and brain activation patterns underlying working memory performance. Schizophr. Bull. 41, 123-132 (2014).

D'Aiuto, L., Williamson, K., Dimitrion, P., McNulty, J., Brown, C.E., Dokuburra, C.B., Nielsen, A.J., Lin, W.J., Piazza, P., Schurdak, M.E., Wood, J., Yolken, R.H., Kinchington, P.R., Bloom, D.C., and Nimgaonkar, V.L.: Comparison of three cell-based drug screening platforms for HSV-1 infection. Antiviral Res. 142, 136-140 (2017).

Dantzer, R., O'Connor, J.C., Freund, G.G., Johnson, R.W., and Kelley, K.W.: From inflammation to sickness and depression: when the immune system subjugates the brain. Nat. Rev. Neurosci. 9, 46-56 (2008).

de Theije, C.G., Wopereis, H., Ramadan, M., van Eijndthoven, T., Lambert, J., Knol, J., Garssen, J., Kraneveld, A.D., and Oozeer, R.: Altered gut microbiota and activity in a murine model of autism spectrum disorders. Brain Behav. Immun. 37, 197-206 (2014).

Dickerson, F.B., Boronow, J.J., Stallings, C., Origoni, A.E., Ruslanova, I., and Yolken, R.H.: Association of serum antibodies to

herpes simplex virus 1 with cognitive deficits in individuals with schizophrenia. Arch. Gen. Psychiatry 60, 466-472 (2003).

Dickerson, F.B., Boronow, J.J., Stallings, C., Origoni, A.E., Cole, S., Krivogorsky, B., and Yolken, R.H.: Infection with herpes simplex virus type 1 is associated with cognitive deficits in bipolar disorder. Biol. Psychiatry 55, 588-93 (2004).

Dickerson, F.B., Boronow, J.J., Stallings, C., Origoni, A.E., Cole, S., Leister, F., Krivogorsky, B., and Yolken, R.H.: The catechol O-methyltransferase Val158Met polymorphism and herpes simplex virus type 1 infection are risk factors for cognitive impairment in bipolar disorder: additive gene-environmental effects in a complex human psychiatric disorder. Bipolar Disord. 8, 124-132 (2006).

Dickerson, F., Stallings, C., Sullens, A., Origoni, A., Leister, F., Krivogorsky, B., and Yolken, R.: Association between cognitive functioning, exposure to Herpes Simplex Virus type 1, and the COMT Val158Met genetic polymorphism in adults without a psychiatric disorder. Brain Behav. Immun. 22, 1103-1107 (2008).

Dickerson, F.B., Stallings, C.R., Boronow, J.J., Origoni, A.E., Sullens, A., and Yolken, R.H.: Double blind trial of adjunctive valacyclovir in individuals with schizophrenia who are seropositive for cytomegalovirus. Schizophr. Res. 107, 147-149 (2009).

Dickerson, F., Stallings, C., Origoni, A., Vaughan, C., Khushalani, S., and Yolken, R.: Additive effects of elevated C-reactive protein and exposure to Herpes Simplex Virus type 1 on cognitive impairment in individuals with schizophrenia. Schizophr. Res. 134, 83-88 (2012).

Dickerson, F., Adamos, M.B., Katsafanas, E., Khushalani, S., Origoni, A., Savage, C.L., Schroeder, J., Schweinfurth, L.A., Stallings, C., Sweeney, K., and Yolken, R.: The association among smoking, HSV-1 exposure, and cognitive functioning in schizophrenia, bipolar disorder, and non-psychiatric controls. Schizophr. Res. 176, 566-571 (2016).

Fruchter, E., Goldberg, S., Fenchel, D., Grotto, I., Ginat, K., and Weiser, M.: The impact of Herpes simplex virus type 1 on cognitive impairments in young, healthy individuals - A historical prospective study. Schizophr. Res. 168, 292-296 (2015).

Gale, S.D., Erickson, L.D., Berrett, A., Brown, B.L., and Hedges, D.W.: Infectious disease burden and cognitive function in young to middle-aged adults. Brain Behav. Immun. 52, 161-168 (2016).

Gerber, S.I., Krienke, U.J., Biedermann, N.C., Grunze, H., Yolken, R.H., Dittmann, S., and Langosch, J.M.: Impaired functioning in euthymic patients with bipolar disorder - HSV-1 as a predictor. Prog. Neuropsychopharmacol. Biol. Psychiatry 36, 110-116 (2012).

Hahn, K., Schildmann, E.K., Baumeister, C., Seggern, I., and Schielke, E.: Cognitive impairment after acute encephalitis: an ERP study. Int. J. Neurosci. 122, 630-636 (2012).

Hamdani, N., Daban-Huard, C., Godin, O., Laouamri, H., Jamain, S., Attiba, D., Delavest, M., Lepine, J.P., Le Corvoisier, P., Houenou, J., Richard, J.R., Yolken, R.H., Krishnamoorthy, R., Tamouza, R., Leboyer, M., and Dickerson, F.B.: Effects of Cumulative Herpesviridae and Toxoplasma gondii Infections on Cognitive Function in Healthy, Bipolar, and Schizophrenia Subjects. J. Clin. Psychiatry 78, e18-e27 (2017).

Hill, A.: The environment and disease: association or causation? Proc. R. Soc. Med. 58, 295-300 (1965).

Itzhaki, R.F.: Herpes simplex virus type 1 and Alzheimer's disease: increasing evidence for a major role of the virus. Front. Aging Neurosci. 6, 202 (2014).

Khandaker, G.M., Cousins, L., Deakin, J., Lennox, B.R., Yolken, R., and Jones, P.B.: Inflammation and immunity in schizophrenia: implications for pathophysiology and treatment. Lancet Psychiatry 2, 258-270 (2015).

Li, H., Zhang, J., Kumar, A., Zheng, M., Atherton, S.S., and Yu, F.S.: Herpes simplex virus 1 infection induces the expression of proinflammatory cytokines, interferons and

TLR7 in human corneal epithelial cells. Immunology 117, 167-176 (2006).

Meyer, U., Feldon, J., and Fatemi, S.H.: In-vivo rodent models for the experimental investigation of prenatal immune activation effects in neurodevelopmental brain disorders. Neurosci. Biobehav. Rev. 33, 1061-1079 (2009).

Millier, A., Schmidt, U., Angermeyer, M.C., Chauhan, D., Murthy, V., Toumi, M., and Cadi-Soussi, N.: Humanistic burden in schizophrenia: A literature review. J. Psychiatr. Res. 54, 85-93 (2014).

Nimgaonkar, V.L., Yolken, R.H., Wang, T., Chang, C.C., McClain, L., McDade, E., Snitz, B.E., and Ganguli, M.: Temporal Cognitive Decline Associated With Exposure to Infectious Agents in a Population-based, Aging Cohort. Alzheimer Dis. Assoc. Disord. 30, 216-222 (2016).

Patterson, S.L.: Immune dysregulation and cognitive vulnerability in the aging brain: Interactions of microglia, IL-1β, BDNF and synaptic plasticity. Neuropharmacology 96, 11-18 (2015).

Petro, T.M., Agarkova, I.V., Zhou, Y., Yolken, R.H., Van Etten, J.L., and Dunigan, D.D.: Response of Mammalian Macrophages to Challenge with the Chlorovirus Acanthocystis turfacea Chlorella Virus 1. J. Virol. 89, 12096-12107 (2015).

Petro, M.S., Agarkova, I.V., and Petro, T.M.: Effect of Chlorovirus ATCV-1 infection on behavior of C57Bl/6 mice. J. Neuroimmunol. 297, 46-55 (2016).

Prasad, K.M., Shirts, B.H., Yolken, R.H., Keshavan, M.S., and Nimgaonkar, V.L.: Brain morphological changes associated with exposure to HSV1 in first-episode schizophrenia. Mol. Psychiatry 12, 105-113 (2007).

Prasad, K.M., Watson, A.M., Dickerson, F.B., Yolken, R.H., and Nimgaonkar, V.L.: Exposure to herpes simplex virus type 1 and cognitive impairments in individuals with schizophrenia. Schizophr. Bull. 38, 1137-1148 (2012).

Prasad, K.M., Eack, S.M., Keshavan, M.S., Yolken, R.H., Iyengar, S., and Nimgaonkar, V.L.: Antiherpes virus-specific treatment and cognition in schizophrenia: A test-of-concept randomized double-blind placebo-controlled trial. Schizophr. Bull. 39, 857-866 (2013).

Raya, R.R., Kleeman, E.G., Luchansky, J.B., and Klaenhammer, T.R.: Characterization of the temperate bacteriophage phi-adh and plasmid transduction in Lactobacillus acidophilus ADH. Appl. Environ. Microb. 55, 2206-2213 (1989).

Rempel, H., Sun, B., Calosing, C., Abadjian, L., Monto, A., and Pulliam, L.: Monocyte activation in HIV/HCV coinfection correlates with cognitive impairment. PLoS One 8, e55776 (2013).

Schretlen, D.J., Vannorsdall, T.D., Winicki, J.M., Mushtaq, Y., Hikida, T., Sawa, A., Yolken, R.H., Dickerson, F.B., and Cascella, N.G.: Neuroanatomic and cognitive abnormalities related to herpes simplex virus type 1 in schizophrenia. Schizophr. Res. 118, 224-231 (2010).

Schwarz, E., Maukonen, J., Hyytiainen, T., Kieseppa, T., Oresic, M., Sabunciyan, S., Mantere, O., Saarela, M., Yolken, R., and Suvisaari, J.: Analysis of microbiota in first episode psychosis identifies preliminary associations with symptom severity and treatment response. Schizophr. Res. in press (2017).

Severance, E.G., Yolken, R.H., and Eaton, W.W.: Autoimmune diseases, gastrointestinal disorders and the microbiome in schizophrenia: more than a gut feeling. Schizophr. Res. 176, 23-35 (2016).

Shirts, B.H., Prasad, K.M., Pogue-Geile, M.F., Dickerson, F., Yolken, R.H., and Nimgaonkar, V.L.: Antibodies to cytomegalovirus and Herpes Simplex Virus 1 associated with cognitive function in schizophrenia. Schizophr. Res. 106, 268-274 (2008).

Steiner, I., Kennedy, P.G., and Pachner, A.R.: The neurotropic herpes viruses: herpes simplex and varicella-zoster. Lancet Neurol. 6, 1015-1028 (2007).

Strandberg, T.E., Pitkala, K.H., Linnavuori, K.H., and Tilvis, R.S.: Impact of viral and bacterial burden on cognitive impairment in

elderly persons with cardiovascular diseases. Stroke 34, 2126-2131 (2003).

Tarter, K.D., Simanek, A.M., Dowd, J.B., and Aiello, A.E.: Persistent viral pathogens and cognitive impairment across the life course in the third national health and nutrition examination survey. J. Infect. Dis. 209, 837-844 (2014).

Tas, C., Brown, E., Cubukcuoglu, Z., Aydemir, O., Danaci, A.E., and Brune, M.: Towards an integrative approach to understanding quality of life in schizophrenia: the role of neurocognition, social cognition, and psychopathology. Compr. Psychiatry 54, 262-268 (2013).

Thomas, P., Bhatia, T., Gauba, D., Wood, J., Long, C., Prasad, K., Dickerson, F. B., Gur, R.E., Gur, R.C., Yolken, R.H., Nimgaonkar, V.L., and Deshpande, S.N.: Exposure to herpes simplex virus, type 1 and reduced cognitive function. J. Psychiatr. Res. 47, 1680-1685 (2013).

Van Etten, J.L. and Dunigan, D.D.: Chloroviruses: not your everyday plant virus. Trends Plant Sci. 17, 1-8 (2012).

Vanyukov, M.M., Nimgaonkar, V.L., Kirisci, L., Kirillova, G.P., Reynolds, M.D., Prasad, K., Tarter, R.E., and Yolken, R.H.: Association of cognitive function and liability to addiction with childhood herpesvirus infections: A prospective cohort study. Dev. Psychopathol. Apr 19, 1-10 (2017a).

Vanyukov, M.M., Nimgaonkar, V.L., Kirisci, L., Kirillova, G.P., Reynolds, M.D., Prasad, K., Tarter, R.E., and Yolken, R.H.: Association of cognitive function and liability to addiction with childhood herpesvirus infections: A prospective cohort study-ERRATUM. Dev. Psychopathol. May 24, 1-2 (2017b).

Wang, Y., Xia, J., Helfer, B., Li, C., and Leucht, S.: Valproate for schizophrenia. Cochrane Database Syst. Rev. 11, CD004028 (2016).

Watson, A.M., Prasad, K.M., Klei, L., Wood, J.A., Yolken, R.H., Gur, R.C., Bradford, L.D., Calkins, M.E., Richard, J., Edwards, N., Savage, R.M., Allen, T.B., Kwentus, J., McEvoy, J.P., Santos, A.B., Wiener, H.W., Go, R.C., Perry, R.T., Nasrallah, H.A., Gur, R.E., Devlin, B., and Nimgaonkar, V.L.: Persistent infection with neurotropic herpes viruses and cognitive impairment. Psychol. Med. 43, 1023-1031 (2013).

Westman, G., Studahl, M., Ahlm, C., Eriksson, B.M., Persson, B., Ronnelid, J., Schliamser, S., and Aurelius, E.: N-methyl-d-aspartate receptor autoimmunity affects cognitive performance in herpes simplex encephalitis. Clin. Microbiol. Infect. 22, 934-940 (2016).

Williamson, L.L., Sholar, P.W., Mistry, R.S., Smith, S.H., and Bilbo, S.D.: Microglia and memory: modulation by early-life infection. J. Neurosci. 31, 15511-15521 (2011).

Yolken, R.H., Torrey, E.F., Lieberman, J.A., Yang, S., and Dickerson, F.B.: Serological evidence of exposure to Herpes Simplex Virus type 1 is associated with cognitive deficits in the CATIE schizophrenia sample. Schizophr. Res. 128, 61-65 (2011).

Yolken, R.H., Jones-Brando, L., Dunigan, D.D., Kannan, G., Dickerson, F., Severance, E., Sabunciyan, S., Talbot, C.C., Jr., Prandovszky, E., Gurnon, J.R., Agarkova, I.V., Leister, F., Gressitt, K.L., Chen, O., Deuber, B., Ma, F., Pletnikov, M.V., and Van Etten, J.L.: Chlorovirus ATCV-1 is part of the human oropharyngeal virome and is associated with changes in cognitive functions in humans and mice. Proc. Natl. Acad. Sci. USA 111, 16106-16111 (2014).

Yolken, R.H., Severance, E.G., Sabunciyan, S., Gressitt, K.L., Chen, O., Stallings, C., Origoni, A., Katsafanas, E., Schweinfurth, L.A., Savage, C.L., Banis, M., Khushalani, S., and Dickerson, F.B.: Metagenomic Sequencing Indicates That the Oropharyngeal Phageome of Individuals With Schizophrenia Differs From That of Controls. Schizophr. Bull. 41, 1153-1161 (2015).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 89-96 (2017).

# READING THE ANTIBODY REPERTOIRE AGAINST VIRUSES, MICROBES, AND THE SELF

URI LASERSON

Department of Genetics and Genomic Sciences and Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

## SUMMARY

Accessing the information in the antibody repertoire promises to provide an unprecedented window into the exposure history and disease susceptibility of an individual. However, the non-heritability and massive diversity of the receptors are major hurdles to interpreting the repertoire. We have developed the phage immunoprecipitation sequencing assay (PhIP-seq) that inexpensively generates high-dimensional functional profiles of a complex mixture of antibodies of unknown specificity. The assay leverages the recent advances in next-generation DNA sequencing along with synthesis of large pools of DNA oligonucleotides. We have successfully applied PhIP-seq to characterize auto-antigen profiles in patients with autoimmune diseases along with a global survey of antiviral immune responses. We are further developing the assay to apply to host-microbiome interactions, among other application areas.

## INTRODUCTION

The repertoire of antibodies present in a single drop of blood should reveal an enormous amount of information about an individual, such as their history of infection, travel, vaccines, allergies, and even their susceptibility to future diseases (*Lerner* et al., 1991). However, multiple hurdles have obstructed our ability to glean such insights from the repertoire: proteins are more difficult to work with than nucleic acids, the non-germline nature of the repertoire complicates population studies (since most antibodies are rare antibodies), and the enormous sequence diversity has historically exceeded the capabilities of any measurement technology. The introduction of next-generation DNA sequencing technology (NGS) a decade ago (*Shendure* et al., 2017) has reinvigorated excitement in studying antibody and T cell receptor (TCR) repertoires (*Georgiou* et al., 2014). The use of NGS to study the immune repertoire has resulted in many insights about the core function of immunity and its role in virtually every disease, including infections (*Parameswaran* et al., 2013), cancer (*Li* et al., 2016; *Liu* and *Mardis*, 2017), and autoimmunity (*Dekosky* et al., 2015).

A holy grail in the field of adaptive immunity is the creation of a mapping between antibody/TCR sequences and their cognate epitopes, such that an antigen/epitope can be determined purely from inspecting the antibody sequence and vice versa. But despite the massive increase in antibody sequence data (*Breden* et al., 2017), most studies typically report either aggregate statistics (e.g., repertoire diversity met-

rics, germline gene usage) or compare them across experimental conditions to discover biomarkers (*Yaari* and *Kleinstein*, 2015). Alternatively, some studies have characterized antibody repertoires in the context of a single antigen and require low-throughput sorting of antigen-specific cells. The ability to predict immune receptor function from its sequence has only been demonstrated in very narrow situations and primarily for TCRs (*Emerson* et al., 2017; *Glanville* et al., 2017).

Here we describe phage immunoprecipitation sequencing (PhIP-seq) (*Larman* et al., 2011), an NGS-based assay that can determine the specificity of a repertoire of antibodies against hundreds of thousands of antigens simultaneously. The assay makes use of phage display libraries containing a set of antigens, each of which will be measured for binding to some antibody of interest. Because the universe of possible antigens grows exponentially with peptide length, such peptide libraries have typically been constructed as random oligomers with lengths <20 amino acids. However, libraries of random peptides "waste" clones on sequences that are not observed in nature. Furthermore, the use of short peptides precludes their ability to exhibit natural structure. To circumvent these limitations, we have leveraged improvements in the ability to synthesize large, complex pools of DNA oligonucleotides (*Kosuri* and *Church*, 2014). Similarly to pooled-screen experiments, the ability to synthesize large pools of DNA oligos to-order allows us to "synthesize" many experiments and execute them in parallel (*Gasperini* et al., 2016).

As described below, PhIP-seq has been used successfully to generate high-dimensional profiles of antibody immunity in the context of autoimmunity and viral infection. In the following sections, we review some of the results of these efforts, along with preliminary methods for expanding the scope of PhIP-seq to bacterial antigens.

## PHAGE IMMUNOPRECIPITATION SEQUENCING

Given a monoclonal antibody or a complex mixture of antibodies (e.g., serum) of unknown specificity, the PhIP-seq assay can determine the specificity of the antibodies from a large set of candidate epitopes that have been selected in advance. Specifically, the "query antibody" is used to immunoprecipitate a phage display library expressing many antigens of interest. The phage clones that are pulled down by the antibody are quantified using next-generation DNA sequencing. Therefore, in a single ~$30 assay, small amounts of antibody (nanogram to microgram quantities) can be assayed against hundreds of thousands of candidate epitopes in parallel. In principle, the diversity of the epitope library is only limited by what can be cloned into the phage display library.

For example, the set of human protein sequences can be tiled with a set of ~400,000 unique peptide 36-mers. The DNA oligos that code for the 36-mers can be synthesized by a commercial DNA vendor (e.g., Twist Bioscience or Agilent Technologies) and cloned into a commercially available bacteriophage T7 protein display system (e.g., Novagen T7Select) to yield a phage library containing every possible human protein 36-mer. The phage library can then be used to identify unknown auto-antigens using serum samples from patients with autoimmune diseases.

It is estimated that a majority of

antibody epitopes are "conformational" rather than linear (*Sela* et al., 1967; *Barlow* et al., 1986). A clear limitation of this method is that the epitopes are limited to the length of DNA oligonucleotides that can be synthesized en masse. As of this writing, synthesis of large oligo pools is generally limited to 200-300 bp sequences, in which some of the sequence must be used for common adaptor sequence for cloning (allowing ~56 aa peptides). Even if we could synthesize or assemble significantly larger pieces of DNA, we would eventually reach the capacity of the phage display system to display larger proteins without interfering with the assembly of infective phage particles. Furthermore, because the peptides are processed in *E. coli*, we cannot include any mammalian post-translational modifications when applicable. These limitations ultimately lead to a loss of sensitivity and make PhIP-seq somewhat unsuitable for diagnostic applications.

It is our experience that robust immune responses generally include at least some responses to linear epitopes. Therefore, PhIP-seq is highly suitable for hypothesis generation experiments, similarly to genome wide association studies (GWAS). Instead of testing SNPs for association with a phenotype, we can use PhIP-seq to test whether particular antibody specificities are associated with a phenotype. Despite its limitations, PhIP-seq provides an inexpensive means to generate a very high-dimensional characterization of the antibody specificity repertoire.

## DISCOVERY OF AUTO-ANTIGENS

PhIP-seq was first demonstrated by building a library of ~400,000 36-mer peptides tiling >24,000 human open reading frames with 7 aa overlaps (*Larman* et al., 2011). Our initial test searched for auto-antigens in several cancer patients with paraneoplastic neurological disorder (PND), a cancer-induced autoimmune syndrome. After successfully detecting known and novel auto-antigens in these patients, we undertook a larger survey of ~300 subjects with multiple sclerosis (MS), type 1 diabetes (T1D), or rheumatoid arthritis (RA), along with healthy controls (*Larman* et al., 2013A). Analysis of the repertoire of autoantibodies in healthy controls revealed that most individuals exhibit antibodies against a variety of self-epitopes. However, the vast majority of these self-specificities were "private" in the sense that they were observed in only a single individual, likely a result of a cross-reactive antibody that is not causing any pathology. Interestingly, some auto-antigen specificities were observed in >20 healthy control subjects, such at MAGEE1, ACVR2B, and TTN. But given the healthy status of these individuals, the most likely explanation is that they represent common cross-reactivities. Using the samples from T1D patients, we analysed the sensitivity of the PhIP-seq assay by comparing peptide enrichments to RIA assays for known autoantibodies against insulin, ZnT8A, GAD2, and PTPRN. PhIP-seq with the 36-mer library was only able to recover reactivities against GAD2 and PTPRN in some patients. Analysis of RA patients did not find any new auto-antigens; however, we were able to cluster a subset of patients into one of two specificity profiles, which was not correlated with their seropositivity status. Finally, by performing a motif analysis on the positive peptide enrichments of

the MS cohort, we were able to recapitulate a previously discovered reactivity to the BRRF2 protein. In a separate PhIP-seq study, the auto-antigen for inclusion body myositis was discovered to be cytosolic 5'-nucleotidase 1A (*Larman* et al., 2013B).

## CHARACTERIZING GLOBAL VIRAL IMMUNITY

Following the demonstration and use of the human PhIP-seq library, the "VirScan" PhIP-seq library was constructed containing 93,904 56-mer peptides (28 aa overlap) derived from 206 viral species that infect humans (e.g., HIV, HCV, EBV, influenza) (*Xu* et al., 2015). Sera from 569 individuals of various ages primarily from the United States but also including sera from Thailand, South Africa, and Peru were characterized with VirScan to generate a global survey of immunological memory/repertoire against viral infection. A subset of samples were tested with ELISA against HIV1, HCV, HSV1, and HSV2, and compared to the VirScan results. When aggregating VirScan epitopes to the virus level, VirScan exhibited >90% sensitivity for all viruses with 100% specificity for HIV1, HSV1, and HSV2. The VirScan-computed fraction of individuals with immunity to CMV (48.5%) and EBV (87.1%) recapitulated known prevalences of the diseases. In contrast, the computed prevalences of Influenza A (53.4%), Poliovirus (33.7%), and VZV (24.3%) were all lower than expected, possibly reflecting a narrowing of the immune response due to absence of the antigens, or the low specificity of PhIP-seq due to the limitation to linear epitopes. Most interestingly, the population analysis of VirScan provided a dramatic display of immunodominance across a large number of viral proteins. When individuals displayed antibodies against a 56-mer derived from a viral protein, they generally all showed specificity for the same 56-mer tile. However, some cases showed that the immunodominant epitopes varied with geolocale. Because the epitope mapping resolution of VirScan is only 28 amino acids, several 56-mer tiles were chosen for alanine scanning mutagenesis in order to pinpoint the location of the epitope. In most cases, individuals specific for a particular 56-mer tile also showed that the sera was specific for the same ~10-mer epitope. Overall, VirScan has been shown to provide an unprecedentedly high-dimensional view of global immunity against the human virome.

## BUILDING LIBRARIES FOR THE MICROBIOME

The influence of the microbiome on human health and disease has been receiving increasing attention (*Cho* and *Blaser*, 2012). This interaction is largely mediated through the mucosal immune system (*McCoy* et al., 2017). Indeed, the majority of antibody synthesized by the human body is IgA that is secreted into the gut lumen to interact with the commensal flora and protect against pathogens (*Fagarassan* and *Honjo*, 2003). However, the specific mechanisms and antigens that drive this interaction are currently unknown (*Kubinak* and *Round*, 2016). Significant progress has been made with the development of the IgA-seq assay, in which gut microbes are sorted into

IgA-bound and IgA-free fractions, followed by 16S rRNA sequencing to determine which taxa are targeted by secretory IgA (sIgA) (*Palm* et al., 2014; *Kau* et al., 2015). While this technique has generated useful hypotheses in the pathogenesis of gut inflammation and other disorders, the specific antigens involved cannot be determined from the relatively low-resolution 16S data.

We have been developing library construction methods for bacterial antigens in order to characterize the sIgA repertoire at antigen-resolution. While we have designed synthetic libraries for particular bacterial species (*Staphylococcus aureus* and *Streptococcus pyogenes*), the microbiome as a whole is too genetically diverse to practically synthesize as DNA oligos using available technology. As an alternative, we have developed a protocol for cloning shotgun metagenomic libraries into the T7 phage display system. The two main advantages of this approach are that we are not limited by current knowledge of the genetic contents of the microbiome and we can also clone larger fragments than can be synthesized as oligonucleotides. However, because the genomic DNA of the microorganisms is randomly sheared, we expect only one sixth of the clones to be in the correct reading frame. While out-of-frame clones may show elevated levels of background noise, analysis of the sequences should distinguish which clones represent true ORF enrichments. Alternatively, a recent method has been developed based on padlock probes in order to capture large DNA fragments in the correct reading frame (*Tosi* et al., 2017). We hope that these libraries will have many applications, such as understanding the role of the mucosal immune system in conferring protection against infection, or understanding the host immune response in inflammatory bowel disease.

## ALTERNATIVE TECHNOLOGIES

A number of assays similar to PhIP-seq have been developed that have different trade-offs. For ease of library construction and non-biasing of sequences, some have used random peptides in display systems (*Pantazes* et al., 2016). However, these libraries tend to "waste" many clones on epitopes that are not observed in nature. Others have used alternative protein display technologies to phage display (*Boder* and *Wittrup*, 1997; *Spatola* et al., 2013; *Zhu* et al., 2013). Most intriguingly, a system for characterizing TCR specificities has been developed by expressing large libraries of peptide-MHC complexes using yeast display (*Birnbaum* et al., 2014). However, the system is limited to MHC-II, and each MHC allele must be individually optimized and engineered to fold correctly. To overcome the size limitation of phage display, the Elledge group has also developed PLATO, a system analogous to PhIP-seq that uses ribosome display to express full ORFs (*Zhu* et al., 2013; *Larman* et al., 2014). In contrast to direct measurement of specificities, the recently published GLIPH technique can perform semi-supervised clustering of TCRs in order to determine their specificities (*Glanville* et al., 2017). Finally, protein microarrays are a well-characterized alternative to protein display-based methods (*Templin* et al., 2002). While they can sometimes accommodate full ORFs, they suffer some solid-state kinetics, generally require larger amounts of input material, and are relatively expensive and lower-throughput than NGS-based assays.

## DISCUSSION

The simultaneous improvement of next-generation DNA sequencing technology and large-scale DNA synthesis technology has enabled the development of highly multiplexed assays at low costs. PhIP-seq provides the ability to assay antibody specificities with unprecedented breadth and will provide new insights into the landscape of host immune responses to their environments. One of the primary bottlenecks is the availability of libraries containing antigens of interest. While we described the use of libraries containing auto-antigens and viral antigens, we are actively developing methods for cloning bacterial libraries. Furthermore, we have designed libraries for allergens and toxins, among other classes of antigens of interest.

By generating a large number of PhIP-seq profiles, our data sets could be used to train B cell epitope predictors (e.g., BepiPred; *Larsen* et al., 2006). Most intriguingly, coupling the PhIP-seq assay to single-cell RNA-seq methods (*Stubbington* et al., 2017) opens the possibility of generating training data sets that could allow us to achieve the holy grail of predicting antibody specificity from primary sequence.

## ACKNOWLEDGEMENTS

I would like to thank Ben Larman, Meimei Shan, and Tim O'Donnell for helpful discussions.

## LITERATURE

Barlow, D.J., Edwards, M.S., and Thornton, J.M.: Continuous and discontinuous protein antigenic determinants. Nature 322, 747-748 (1986).

Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J., Ozkan, E., Davis, M.M., Wucherpfennig, K.W., Garcia, K.C.: Deconstructing the peptide-MHC specificity of T cell recognition. Cell 157, 1073-1087 (2014).

Boder, E.T. and Wittrup, K.D.: Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. 15, 553-557 (1997).

Breden, F., Luning Prak, E.T., Peters, B., Rubelt, F., Schramm, C.A., Busse, C.E., Vander Heiden, J.A., Christley, S., Bukhari, S.A.C., Thorogood, A., Matsen, F.A. IV, Wine, Y., Laserson, U., Klatzmann, D., Douek, D.C., Lefranc, M.P., Collins, A.M., Bubela, T., Kleinstein, S.H., Watson, C.T., Cowell, L.G., Scott, J.K., and Thomas B. Kepler, T.B.: Reproducibility and Reuse of Adaptive Immune Receptor Repertoire Data. Front. Immunol. 8, 1418 (2017).

Cho, I. and Blaser, M.J.: The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13, 260-270 (2012).

Dekosky, B.J., Kojima, T., Rodin, A., Charab, W., Ippolito, G.C., Ellington, A.D., and Georgiou, G.: In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. Nat. Med. 21, 86-91 (2015).

Emerson, R.O., DeWitt, W.S., Vignali, M., Gravley, J., Hu, J.K., Osborne, E.J., Desmarais, C., Klinger, M., Carlson, C.S., Hansen, J.A., Rieder, M., and Robins, H.S.: Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell

repertoire. Nat. Genet. 49, 659-665 (2017).
Fagarasan, S.S. and Honjo, T.T.: Intestinal IgA synthesis: regulation of front-line body defences. Nat. Rev. Immunol. 3, 63-72 (2003).
Gasperini, M., Starita, L., and Shendure, J.: The power of multiplexed functional analysis of genetic variants. Nat. Protoc. 11, 1782-1787 (2016).
Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., and Quake, S.R.: The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat. Biotechnol. 32, 158-168 (2014).
Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C., Haas, N., Arlehamn, C.S.L., Sette, A., Boyd, S.D., Scriba, T.J., Martinez, O.M., and Davis, M.M.: Identifying specificity groups in the T cell receptor repertoire. Nature 547, 94-98 (2017).
Kau, A.L., Planer, J.D., Liu, J., Rao, S., Yatsunenko, T., Trehan, I., Manary, M.J., Liu, T.C., Stappenbeck, T.S., Maleta, K.M., Ashorn, P., Dewey, K.G., Houpt, E.R., Hsieh, C.S., and Gordon, J.I.: Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. Sci. Transl. Med. 7, 276ra24 (2015).
Kosuri, S. and Church, G.M.: Large-scale de novo DNA synthesis: technologies and applications. Nat. Methods 11, 499-507 (2014).
Kubinak, J.L. and Round, J.L.: Do antibodies select a healthy microbiota? Nat. Rev. Immunol. 16, 767-774 (2016).
Larman, H.B., Zhao, Z., Laserson, U., Li, M.Z., Ciccia, A., Gakidis, M.A.M., Church, G.M., Kesari, S., Leproust, E.M., Solimini, N.L., and Elledge, S.J.: Autoantigen discovery with a synthetic human peptidome. Nat. Biotechnol. 29, 535-541 (2011).
Larman, H.B., Laserson, U., Querol, L., Verhaeghen, K., Solimini, N.L., Xu, G.J., Klarenbeek, PL., Church, G.M., Hafler, D.A., Plenge, R.M., Nigrovic, P.A., De Jager, P.L., Weets, I., Martens, G.A., O'Connor, K.C., and Elledge, S.J.: PhIP-Seq characterization of autoantibodies from patients with multiple sclerosis, type 1 diabetes and rheumatoid arthritis. J. Autoimmun. 43, 1-9 (2013A).
Larman, H.B., Salajegheh, M., Nazareno, R., Lam, T., Sauld, J., Steen, H., Kong, S.W., Pinkus, J.L., Amato, A.A., Elledge, S.J., and Greenberg, S.A.: Cytosolic 5′-nucleotidase 1A autoimmunity in sporadic inclusion body myositis. Ann. Neurol. 73, 408-418 (2013B).
Larman, H.B., Liang, A.C., Elledge, S.J., and Zhu, J.: Discovery of protein interactions using parallel analysis of translated ORFs (PLATO). Nat Protoc. 9, 90-103 (2014).
Larsen, J.E.P., Lund, O., and Nielsen, M.: Improved method for predicting linear B-cell epitopes. Immunome Res. 2, 2 (2006).
Lee, J.Y., Artiles, K.L., Zompi, S., Vargas, M.J., Simen, B.B., Hanczaruk, B., McGowan, K.R., Tariq, M.A., Pourmand, N., Koller, D., Balmaseda, A., Boyd, S.D., Harris, E., and Fire, A.Z.: Convergent antibody signatures in human dengue. Cell Host Microbe 13, 691-700 (2013).
Lerner, R.A., Barbas, C.F. 3rd., Kang, A.S., and Burton, D.R.: On the use of combinatorial antibody libraries to clone the “fossil record” of an individual’s immune response. Proc. Natl. Acad. Sci. USA 88, 9705-9706 (1991).
Li, B., Li, T., Pignon, J.C., Wang, B., Wang, J., Shukla, S.A., Dou, R., Chen, Q., Hodi, F.S., Choueiri, T.K., Wu, C., Hacohen, N., Signoretti, S., Liu, J.S., and Liu, X.S.: Landscape of tumor-infiltrating T cell repertoire of human cancers. Nat. Genet. 48, 725-732 (2016).
Liu, X.S. and Mardis, E.R.: Applications of Immunogenomics to Cancer. Cell 168, 600-612 (2017).
McCoy, K.D., Ronchi, F., and Geuking, M.B.: Host-microbiota interactions and adaptive immunity. Immunol. Rev. 279, 63-69 (2017).
Palm, N.W., De Zoete, M.R., Cullen, T.W.,

Barry, N.A., Stefanowski, J., Hao, L., Degnan, P.H., Hu, J., Peter, I., Zhang, W., Ruggiero, E., Cho, J.H., Goodman, A.L., and Flavell, R.A.: Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell 158, 1000-1010 (2014).

Pantazes, R.J., Reifert, J., Bozekowski, J., Ibsen, K.N., Murray, J.A., and Daugherty, P.S.: Identification of disease-specific motifs in the antibody specificity repertoire via next-generation sequencing. Sci. Rep. 6, 30312 (2016).

Sela, M., Schechter, B., Schechter, I., and Borek, F.: Antibodies to Sequential and Conformational Determinants. Cold Spring Harbor Symp. Quant. Biol. 32, 537-545 (1967).

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., and Waterston, R.H.: DNA sequencing at 40: past, present and future. Nature 550, 345-353 (2017).

Spatola, B.N., Murray, .JA., Kagnoff, M., Kaukinen, K., and Daugherty, P.S.: Antibody repertoire profiling using bacterial display identifies reactivity signatures of celiac disease. Anal. Chem. 85, 1215-1222 (2013).

Stubbington, M.J.T., Rozenblatt-Rosen, O., Regev, A., and Teichmann, S.A.: Single-cell transcriptomics to explore the immune system in health and disease. Science 358, 58-63 (2017).

Templin, M.F., Stoll, D., Schrenk, M., Traub, P.C., Vöhringer, C.F., and Joos, T.O.: Protein microarray technology. Drug Discov. Today 7, 815-822 (2002).

Tosi, L., Sridhara, V., Yang, Y., Guan, D., Shpilker, P., Segata, N., Larman, H.B., and Parekkadan, B.: Long-adapter single-strand oligonucleotide probes for the massively multiplexed cloning of kilobase genome regions. Nat. Biomed. Eng. 1, 0092 (2017).

Xu, G.J., Kula, T., Xu, Q., Li, M.Z., Vernon, S.D., Ndung'u, T., Ruxrungtham, K., Sanchez, J., Brander, C., Chung, R.T., O'Connor, K.C., Walker, B., Larman, H.B., and Elledge, S.J.: Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. Science 348, aaa0698 (2015).

Yaari, G. and Kleinstein, S.H.: Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Med. 7, 121 (2015).

Zhu, J., Larman, H.B., Gao, G., Somwar, R., Zhang, Z., Laserson, U., Ciccia, A., Pavlova, N., Church, G., Zhang, W., Kesari, S., and Elledge, S.J.: Protein interaction discovery using parallel analysis of translated ORFs (PLATO). Nat. Biotechnol. 31, 331-334 (2013).

Old Herborn University Seminar Monograph 31:
Evolutionary biology of the virome, and impacts in human health and disease.
Editors: Peter J. Heidt, Pearay L. Ogra, Mark S. Riddle and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 97-104 (2017).

# EVOLUTIONARY BIOLOGY OF THE VIROME AND IMPACTS IN HUMAN HEALTH AND DISEASE: SUMMARY OF THE 31ST OLD HERBORN UNIVERSITY SEMINAR AND THE STRUCTURED DISCUSSION

PEARAY L. OGRA[1] and MARK S. RIDDLE[2]

[1]Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, NY, USA;
[2]Uniformed Services University of Health Sciences, Bethesda, MD, USA

In view of the relatively limited knowledge available about the functional characteristics of the human virome, the 31st Old Herborn University Seminar (OHUS) was planned to acquire a more comprehensive, broad based and interdisciplinary current understanding of the environmental virome. It was also hoped to explore its role in the human homeostatic processes and their biologic functional development during normal physiologic as well as pathologic disease states.

This seminar, based on 10 formal presentations followed by a comprehensive discussion, addressed several important aspects of environmental virome including, the virome of unicellular and multicellular organisms, and the nature of human virome and its impact on human health and disease.

Dr. Mark Riddle from the Uniformed Services University of Health Sciences, Bethesda, Maryland provided a brief overview of the meeting plans relative to the specific areas identified above. He discussed briefly the possible role played by environmental virome in the evolution of life, and in the synthetic aspects of biology, emerging infectious diseases, and the development of some of the relevant tools of modern biotechnology to examine these issues in further detail. He also discussed the framework of the virome proposed recently as an instrument to acquire additional knowledge about human viruses (*Parker*, 2016). This model suggests that virome functions in commensal, parasitic and or in mutualistic interactions with the specific cellular hosts, which may result in, no benefits, deleterious effects, or mutually beneficial effects to the virome as well as the host.

Professor Patrick Forterre from the Institute Pasteur in Paris, France provided a provocative and elegant introduction to the controversies surrounding the definition of the virus, and the development of the environmental virome in unicellular organisms. He started with the complex question of "what is life" and how recent technological advances have challenged the definition of viruses and of life itself. Prof. Forterre reviewed the existing data on viruses in different cellular domains; including archaea, bacteria and eukaryotic life forms, and the discovery of giant viruses whose genomes are bigger than many bacteria and archaea. Based on this information, Prof. Forterre proposed that viruses be defined as cellular organisms producing virions. Virions are products that involve transformation between different transient states: the virus, the virocell, the integrated genome, and others. The concept of virocell is designed to focus on the cellular steps of the viral replication cycle that involves the transformation of all or part of the infected host into a "living" viral organism. He also

provided evidence to suggest that most viral genes originate de novo in the viral genome. Finally he raised the possibility that the interactions between viruses and the host cells are the major engines of biologic evolution.

Dr. Kimberly Seed of the University of California at Berkeley provided a comprehensive discussion of the cellular aspects of the infection of *Vibrio cholerae* with several bacteriophages. Both biotic and abiotic factors critically influence the growth and abundance of *V. cholerae* in many parts of the world. Specific bacteriophages appear to be responsible for the growth patterns and the pathogenesis of this bacterial organism. Based on the studies carried out by Dr. Seed and her colleagues with *V. cholerae* specific phages, ICP1, ICP2 and ICP3, it appears that the bacterium can evolve specific mechanisms of resistance to phage infection. Interestingly enough, the cholera phages can counter-adapt to overcome the defence barriers of the bacterial host organism. She has shown that virtually all *V. cholerae* coexisting with ICP2 phage are resistant to other phage infections. The mechanisms underlying such resistance include induction of phage-inducible chromosomal island like elements (PLE). The induction of PLE appears to be highly specific in its activity to abort phage infections in the bacterial organism. Although, the studies of Dr. Seed are limited to the bacteriophage of *V. cholerae* alone, their implications are clearly applicable to other phage-bacterium interactions**.**

Dr. Joao Margues from the University Federal de Minas Gerais in Brazil provided an overview of the virome of the insect world, and their impact on human health and disease. Insects are one of the most successful life forms, representing over 70% of the animal species and over 50% of all life forms on this planet. Insects can adapt rapidly and are highly successful at colonizing virtually every niche of the earth. Insect virome is as diverse as the vast kingdom of the insects themselves.

Viruses associated with plant insects (phytophagous vectors) include topovirus, cytohabdovirus, nucleohabdovirus, emarovirus, and tenuivirus. Blood-feeding insects of the animal world act as vectors, and carry many viruses. These include, nairovirus, phlebovirus, orthobunyavirus, ephimerovirus and vesiculovirus. Other insect viruses associated with vertebrates include filovirus, alphavirus, bornavirus, flavivirus, hantavirus, influenza virus, paramyxovirus, lyssavirus, and possibly other viruses. Dr. Margues provided an extensive review of the mosquito-borne arbovirus infections, especially in the flavivirus and alphavirus families, which are of major public health importance for humans. At least 100 human arboviruses have been identified to date. These include Eastern, Western and Venezuelan equine encephalitis, chikungunya virus, dengue, Japanese, St. Louis encephalitis virus, West Nile virus, and yellow fever virus. Dr. Margues subsequently provided extensive characterization of the virome of the mosquitoes in Caratinga area in Brazil, as a model for exploring the spectrum of insect virome in parts of the world which are endemic for mosquito and other insect-borne disease.

Following the presentations on bacteriophage, and insect viruses, the virome of corals and the development of immunity in other metazoans were discussed in some detail by Dr. Steven Quistad of the San Diego State University in California. The Cnidarian phyla which includes sea anemones, jellyfish, Hydra, and corals is considered to be

phylogenetically basal and similar to Bilateria (files, worms and humans). Due to their relative position within the tree of life, Cnidarians provide insight into the genetic toolkit of the last common metazoan ancestor. Their simple body plan consists of only two epithelial cell layers connected by a jelly-like mesoglea. Despite their morphological simplicity, the Cnidarian immune system is surprisingly complex, with multiple immune components conserved from the earliest Cnidarian life forms to the modern humans. For example, the tumor necrosis factor receptor superfamily (TNFRSF) is a central mediator of apoptosis in humans. Interestingly, corals have been found to possess the most complex TNFR repertoire ever reported. In addition to TNFRSF, the Acropora genome has been shown to possess all the central components of the canonical apoptotic cascades. These include 3 TRAF (TNF receptor associated factor) proteins, 30 caspases, and 8 members of the Bcl-2 family. Remarkably, exposure of coral cells to human TNF-α was found to cause apoptosis-induced formation of blebs, caspase activation, cell death, and the resultant coral bleaching. Subsequently, exposure of human T-cell lymphocytes to a member of the coral TNFSF was found to cause apoptosis in human cells, suggesting a conservation of TNF-induced apoptotic mechanisms across the 550 million year span of evolution.

In the second portion of his talk, Dr. Quistad outlined a novel approach to identify predicted immune proteins in Cnidarians that are missed by using traditional protein annotation methods alone. Briefly, this approach is based on the identification of similarities between viral gene products and host proteins and can be applied to virtually any system. Viruses are extracted from the non-model organism of interest using a virus-like particle extraction protocol of choice and nucleic acid is sequenced. If a gene model is not available, total RNA is also extracted from host tissue and an assembled transcriptome is prepared. First, viral gene segments are compared to the translated transcriptome through tBLASTn to identify matches to host proteins. These proteins are further analysed through comparison to a well-characterized immune system using BLASTp (e.g., human or mouse) and domain prediction database (e.g., Conserved Domain CDD, Pfam). Proteins that lack hits to either database represent predicted immune genes with unknown function and are selected for further biochemical investigations using *in vitro* and *in vivo* experimentation.

Dr. Tom Lachnit from the Christian-Albrecht's-University at Kiel, Germany presented an elegant discussion of bacteriophage in the fresh water polyp Hydra and their role in the homeostasis of holobiont. A holobiont is a host organism (animal or plant) which is in constant contact with all its associated microorganisms as an entity for selection of evolution. Hydra and possibly humans are model organisms as holobiont. A series of very imaginative studies carried out by Dr. Lachnit, Dr. Thomas Bosch and their colleagues have shown that different species of Hydra are colonized by different sets of bacteria in a highly (species)-specific manner. The bacteria in Hydra exhibit unique bacterial-bacterial interaction between different species via glycocalyx with the associated microbes. This interaction appears to be critically mediated by the presence of bacteriophage in the microbes. The microbial organisms possess specific viral communities. These are predominantly from nine families of eukaryotic viruses, and to a smaller extent from the four families of prokaryotic viruses.

Also, there appears to be on-going interactions between the prokaryotic and eukaryotic associated bacteria and the host. Studies carried out with prophage of Curvibacter species in this model system have suggested that environmental stress may potentiate induction of bacteriophage, and Curvibacter phages can also influence bacterial colonization *in vivo*. Genome sequencing of 13 Hydra species- associated bacteria have revealed the presence of phages as, intact prophage, partial prophage or no prophage in, 54%, 23% and 27% of the bacterial species respectively. These observations suggest that phages are an important part of the metaorganism. The host appears to influence bacterial-bacterial interaction via the induction of prophages. Finally, it was suggested that mechanisms of host immune defence can specifically regulate the activation of prophages in this model holobiont species.

The final component of this seminar focused on the interaction of the virome in human health and disease. Professor Frederick Bushman from University of Pennsylvania at Philadelphia discussed via a previously recorded video, several important aspects of human virome. These included composition and targeted hyper-variation of the virome and the mechanisms underlying the individual variations in harbouring distinct and different viral communities by different subjects. Professor Bushman provided a very comprehensive database to suggest that human virome is composed of three sets of distinct viral communities:

1) **Persistent and latent human viruses**.

These include, among others, Epstein Barr virus (EBV), Varicella-Zoster virus (VZV), Herpesviruses including Herpes simplex virus, HSV1, papillomaviruses, Cytomegalovirus (CMV), and possibly other viruses. Up to 60-100% of humans are seropositive for these viral agents during their lifetime. Other human viruses such as HSV2, Human Immune deficiency virus (HIV), Hepatitis C virus (HCV) are less common and exhibit lifetime infection rates ranging from up to 22% for HSV and to about 1% for HIV respectively.

**2) Endogenous retroviruses**.

These agents constitute up to 8% of human DNA.

**3) Bacteriophages**.

Predators of bacteria and archaea associated with human skin and mucosal surfaces (especially in gut). It is estimated that over $10^{10\text{-}11}$ viral particles are present in each gram of human faeces. Based on Solexa/Illumina HiSeq data analysis of viral sequences, it appears that as many as 500-1000 virus species may be present in each individual. Most viral sequences identified to date probably represent new viruses, with little or no resemblance to isolates from other subjects.

The composition of faecal virome in healthy subjects is mostly derived from the bacteriophage. These resident phages have profound influence on the functional attributes of the bacterial host. These include development of toxins (*Shigella, V. cholerae*), virulence determinants, metabolic capacity and development of antibiotic resistance. Furthermore, bacteriophages function as parasites and the host bacteria devote much of their coding capacity toward the assembly of the phage. During diseased or immunologically compromised states, the virome appears to be quite different. In one such situation, metagenomics sequencing have identified a virus, a new bocavirus in the parvovirus family, from several patients with severe combined immunodeficiency disorder.

Targeted hypervariation appears to be another distinct feature of human gut bacteriophage. This phenomenon is mediated by an abundant class of reverse transcriptase enzyme. Another bacteriophage driven process described recently is the CRISPER (Clusters of Regularly Interspaced Short Palindromic Repeats) system, which targets viruses for the generation of escape mutants. Other genomic alterations result in rapid evolution of ssDNA phages in the gut referred to as microviridae. These biologic phenomena in the gut environment appear to be responsible for generation and harbouring of newly evolved viral communities in the human virome.

Dr. Uri Laserson from the Mount Sinai Medical Center, New York provided a comprehensive report on an important recently developed immunological tool, the reading of antibody repertoire to environmental and human microbiome and auto-immune (self) antigens. This approach is based on Phage immunoprecipitation sequencing (PhIP-seq). It can evaluate binding to several hundred epitopes of given antibodies with unknown specificity. The PhIP-seq system requires antigen library cloned into phage, and the antibodies to be profiled. The system has been employed to, identify short synthetic peptides, scan for mutagenesis, and more recently to identify known variants of viruses and, personalize malignancy induced neopeptides.

The final presentation considered the role of viral infections on the longitudinal cognitive functioning in man. Professor Robert Yolken from the Johns Hopkins School of Medicine, Baltimore, Maryland reviewed psychiatric disorders such as schizophrenia, bipolar disorder, and major depression, as causes of mortality and serious morbidity worldwide, particularly in younger individuals. These disorders are associated with both alterations in mood and cognitive impairment. Several genetic, epidemiological and pathophysiological studies indicate a possible role for viruses and other microbial agents in the pathogenesis of some of these disorders. However no specific causative agents have been identified to date. He and his colleagues have employed a number of techniques in an attempt to identify any possible viral agents associated with such psychiatric disorders and cognitive impairment. Using standard enzyme immunoassays they have found that increased levels of antibodies to the neurotropic herpesvirus, Herpes simplex Virus Type 1 (HSV-1) are associated with decreased levels of cognitive functioning in individuals with schizophrenia or bipolar disorder, as well as in individuals without any defined psychiatric disorder. This dysfunction was seen largely in the domains assessing memory and was not seen in association with antibodies to other herpesviruses.

In a series of recent studies, Professor Yolken and his colleagues have explored the role of viruses in the aetiology of these conditions by the use of metagenomics sequencing of throat swab samples. These studies also did not find any recognized human virus to be associated with cognitive impairment or a psychiatric diagnosis. However, a number of phage and other non-human viruses were differentially present in many samples. In particular, the Lactobacillus phage Phiadh was significantly increased in samples obtained from individuals with schizophrenia and was associated with an increased rate of autoimmune disorders and a differential response to medications. It was also found that an algae virus, Acanthocystis turfacea chlorella

virus 1 (ATCV-1), was associated with significantly lower performance in cognitive tests of motor ability in individuals without any defined psychiatric disorder. Feeding this virus to mice resulted in altered cognitive performance and changes in the expression of relevant transcripts within the brain. These observations indicate that both human and non-human viruses may affect human behaviour and cognition. A greater understanding of the role of these viruses and other infectious agents may lead to new methods for the prevention and treatment of psychiatric disorders and cognitive dysfunction.

The formal session of scientific presentations concluded with a presentation by Professors Peter Heidt and Volker Rusch. This special presentation was a tribute to Professor Dirk van der Waaij who passed away in 2016 after an illustrious academic career.

Professor van der Waaij was the Professor and Chairman of the Department of Bacteriology and Serology at the University of Groningen from 1975 until his retirement in 1992. However, he remained a true scholar even after his retirement. During his lifetime, Professor van der Waaij made seminal contributions to the understanding of human microbiome in human health and disease (*van Bekkum* et al., 1974; *van der Waaij*, 1977; *Heidt* et al., 1983; *Vossen* et al., 2014). These include, the introduction of concepts of complete gastrointestinal decontamination, reverse isolation, use of human donor flora for reconstitution after complete decontamination (Julia flora), microbial intervention in immunodeficiency states, characterization of microflora by morphometry and many other innovative contributions which are still in use in current patient management.

This tribute to Professor van der Waaij during the 31st Seminar is especially personal to the OHUS organization. He was one of the founders of the OHUS seminar series. He was instrumental as a major planner and as co-editor of the first twenty-one volumes of Seminar publications. For his scientific contribution and for his personal commitment, we are profoundly grateful to our late friend.

**Concluding Remarks**

Based on the information summarized during this seminar, it is clear that our understanding of the environmental virome and its impact on human health and disease is still in its infancy, but continues to evolve rather rapidly. A modest consensus is emerging about the definition of the virus as an intracellular organism that involves transformation between different transient states, the virion, the virocell, the integrated genome and possibly other cellular events. Most viral genes seem to originate *de novo* in the viral genomes. The origin of viruses and their interaction with the host cells may represent the single most important trigger for the selection in the on-going evolutionary biology.

Virtually all cellular life forms are colonized or parasitized by a specific viral entity. Yet, relatively little is known about the impact of the virome on the development and cellular homeostasis of the host. The interaction between the human host and the virome environment must include the relative impact of viruses on the particular strategies of the life forms which have adapted to their environments and reproductive success. For example, the impact of a particular virus on a multicellular regeneration-capable sponge is quite different than the impact on a human with a neuron that, once damaged, may be catastrophic to that organism. As such, one might expect that the host response repertoires and pathways to

deal with viruses may be similarly divergent in their tolerance to such external threats. Thus, to understand the different adaptive responses to viruses, host interactions from the context of evolutionary principles should be considered. For example, one must consider the question of how bacterial and phage interactions provide any fitness benefits to the host.

Little information is currently available about the possible use of any intervention measures directed at different components of human virome including bacteriophage to prevent or treat human disease. During the open discussion session period of the seminar, Prof. Yolken introduced some exciting additional studies from his laboratory on the role of microbiome in psychiatric disorders relative to gut-brain-immune interaction and opportunities for prevention and treatment. Prof. Yolken suggested, based on large nationwide survey, that infections and repeated use of anti-infective agents exhibit a strong correlation with the risk of severe mental disorders especially affective disorders and possibly schizophrenia (*Kohler*, et al, 2017). A significant association was observed between the use of antibiotics and bipolar mania in hospitalized patient settings. These observations suggest in patients hospitalized with mania a possible role of altered microbiome, associated with mucosal inflammation of the gut secondary to the use of antibiotics. Based on these observations, a longitudinal study employing orally administrated Lactobacillus GG and *Bifidobacterium lactis* ($>10^8$ CFU) for a period of 6 months was undertaken in a large number of subjects. Preliminary analysis of the results suggests that such treatment improved the clinical course of patients with mania, but did not significantly alter the psychotic symptoms in patients with schizophrenia.

Other recognized concepts of importance, but also uncertainty, relate to understanding the effect of predatory phage in the aquatic environment. Few preliminary studies have begun to examine this relationship in the human host. However, large gaps remain in our understanding about the nature of specific interactions in the marine estuarial environments. Could such events also shape factors that are involved in human disease? Furthermore, very little is known about virome-virome interactions, as well as about specific virome dynamics in insects which are vectors in a host of human diseases. The adaptive immune system and impacts on host-virome interactions is an additional area of future research needs.

During the past two decades the incidence of microbial antibiotic resistance and emergence of new infectious disease states has increased at an alarming rate. Significant effort is currently underway to utilize bacteriophage components of human virome as a potential tool to preserve a healthy human microbiome. Although the evidence of possible phage-induced antibacterial activity in bacteria-free filtered water of river Ganges in India was observed by Hankin as early as 1896, the concepts of phage therapy are now being re-examined in order to develop alternate approaches to the use of antibiotics. Currently, several potential bacteriophage-based therapeutic products are being explored for their antibacterial activity and for possible use in humans. These include phage lysine for *Staphylococcus aureus* bacteraemia, and natural phage cocktails for *E. coli* and Shigella enteric infections, chronic ulcers, skin infections and prosthetic infections (*Madhusoodan*, 2016).

The human virome is relatively specific in each individual and significant hypervariation is associated with the constant generation of novel viral parti-

cles in the human gut. However, it remains to be determined if the introduction of foreign phages as probiotics, or via faecal transplants will be safer and more physiologic than the use of antibiotics.

Finally since their discovery, viruses have been viewed negatively, mostly as nasty disease-producing organisms. Unfortunately this perception is not always based on facts. Of the trillions of viruses and viral particles which form the resident environmental virome, only a few hundred produce a symptomatic infection and or a fatal clinical disease. Viruses have been recovered from all mucosal surfaces and the skin, and are acquired in a very distinct pattern before and immediately after birth. The role of viral infections in the anatomical development of mammalian mucosal surfaces (such as intestinal villi), regulation of mucosal inflammation and maturation and functional development of immune system are well known. However, it is also well recognized that a small minority of viruses are pathogenic and the immune system is designed to promote tolerance towards the vast majority of non-threatening antigens that it sees on a daily basis. Thus, the precise roles of the virome in the functional development of the diverse homeostatic mechanisms of the human host are critical to understand, and still remain to be clearly defined.

## LITERATURE


Heidt, P.J., van der Waaij, D., Vossen, J.M., and Hendriks, W.D.H.: The use of a human donor flora for recontamination following antibiotic decontamination. Prog. Food Nutr. Sci. 7, 55-63 (1983).

Kohler, O., Petersen, L., Mors, O., Mortensen, P.B., Yolken, R.H., Gasse, C., and Benros, M.E.: Infections and exposure to anti-infective agents and the risk of severe mental disorders: a nationwide study. Acta Psychiatr. Scand. 135, 97-105 (2017).

Madhusoodan, J.: Viral soldiers. Scientist 30, 27-33 (2016).

Parker M.: An ecological framework of the human virome provides classification of current knowledge and identifies areas of forthcoming discovery. Yale. J. Biol. Med. 89, 339-351 (2016).

van Bekkum, D.W., Roodenburg, J., Heidt, P.J., and van der Waaij, D.: Mitigation of secondary disease of allogeneic mouse radiation chimeras by modification of the intestinal microflora. J. Natl. Cancer Inst. 52, 401-404 (1974).

van der Waaij, D., Vossen, J.M., Korthals Altes, C., and Hartgrink, C.: Reconventionalization following antibiotic decontamination in man and animals. Am. J. Clin. Nutr. 30, 1887-1895 (1977).

Vossen, J.M., Guiot, H.F.L., Lankester, A.C., Vossen, A.C.T.M., Bredius, R.G.M., Wolterbeek, R., Bakker, H.D.J., and Heidt, P.J.: Complete suppression of the gut microbiome prevents acute graft-versus-host disease following allogeneic bone marrow transplantation. PloS One 9, e105706 (2014).