

Günther Witzany *Editor*

# Viruses: Essential Agents of Life

 Springer

# Viruses: Essential Agents of Life



Günther Witzany

Editor

# Viruses: Essential Agents of Life

 Springer

*Editor*  
Günther Witzany  
Telos – Philosophische Praxis  
Bürmoos, Austria

ISBN 978-94-007-4898-9                      ISBN 978-94-007-4899-6 (eBook)  
DOI 10.1007/978-94-007-4899-6  
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012951571

© Springer Science+Business Media Dordrecht 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

## Preface

When I finished my studies in philosophy of science in the early 1980s, focusing on the outcomes of the scientific discourse on the philosophy of language and communication between 1920 and 1980, some editions of *Scientific American* happened to attract my attention. In a variety of articles on cellular and genetic processes by different authors, some of them Nobel Laureates, I found an astonishing vocabulary: *genetic code, code without commas, misreading of the genetic code, coding, copying, open reading frame, genetic storage medium DNA, genetic information, genetic alphabet, genetic expression, messenger RNA, cell-to-cell communication, immune response, transcription, translation, nucleic acid language, amino acid language, recognition sequences, recognition sites, protein coding sequences, repeat sequences, signal transduction, signalling pathways*. All these terms combine a linguistic and communication theoretical vocabulary with one that is biological.

In parallel, I read a book by Manfred Eigen and Ruthild Winkler about the molecular syntax of the genetic code in which they proposed that the genetic code should be taken seriously as a real language, and not as a metaphor (“At any rate one can say that the prerequisite for both great evolutionary processes of nature – the origin of all forms of life and the evolution of the mind – was the existence of a language.”). This interested me, because they described in detail all the typical features of languages/codes that are present in the language of nucleic acid. The only deficit to the philosophy of science discourse about how to define a real language was that the authors did not understand, or ignored, the concept that any real language must encompass a third level of rules: not solely syntax and semantics, but also pragmatics. If one level of rules is missing, one cannot take a language to be a real language. Syntax defines how to combine the variety of signs (alphabet) into larger content units such as words and sentences. Semantics defines how these signs can designate real objects. Pragmatics defines how actual living agents use these signs to coordinate and organize their lives in real-life situations, i.e. the context within which the signs are used.

If someone is familiar with the process of the philosophy of science discussions in the twentieth century, it is interesting how logical empirism (neo-positivism) and later on critical rationalism tried to install the model of an exact scientific language,

i.e. the mathematical theory of language, from 1920 to 1960, consistently with the thoughts and theories of Wittgenstein (*Tractatus Logico Philosophicus*), Carnap, Neurath, Tarski, Gödel, Russel (*Principia Mathematica*) and Popper.

Manfred Eigen followed this school of thought in the early 1970s: although at this time it had been proved that the fundamentals of this theoretical framework were false, they agreed that the syntax of a real scientific language, i.e. the combinatorial rules, represents the material reality of physics and chemistry. Therefore the meaning (semantics) of the signs of a language, and their combination in sequences such as sentences, is the result of the sign order (“The relative arrangement of the individual genes, the gene map, as well as the syntax and semantics of the molecular language are (...) largely known today”). The information processing occurs like this: a sender processes signals coherent to the material reality of the brain, i.e. the signals represent the neuronal combination logic of the brain organ. He sends the signals through an information channel to a receiver. The receiver senses these signals and his brain organ decodes the signalling sequence and extracts meaning using his inherent value programme, which is coherent with the neuronal molecular structure of the brain organ (“...a universal regularity evidently originating in the organization of the human brain”). Because mathematics depicts material reality in the natural laws of physics and chemistry, the exact scientific language must be formalizable. Exact science has to describe investigated objects with formalizable procedures such as algorithms.

For several reasons this model of language was falsified (in line with Wittgenstein’s “Philosophical investigations”, Austin’s “How to do things with words”, Searle’s “Speech Acts”, Apel’s “transcendental pragmatics” and Habermas’ “universal pragmatics”).

Certainly the most important was that the former proponents overlooked the third level of rules inherent in every natural language, i.e. the relation of signs to the real-life sign user. Pragmatics is the term used to designate this level, defining that the context in which a sign-using agent is involved determines the meaning of the sign sequence, and not the syntax. This makes sense in animated nature, which has regard to energy costs, because real sign users need only a limited number of signs (alphabet) and a limited number of rules (syntax, semantics, pragmatics) to generate an unlimited number of correct sign sequences. In contrast to artificial languages that may only follow formalizable procedures, natural languages have properties that are not formalizable in principle.

The second consequence was that the sender-receiver narrative no longer met reality. Natural languages do not emerge as system properties, but as a social phenomenon. This means that, wherever consortia of related living agents are present, there exists population-based signalling to coordinate interaction and reproduction. Language use occurs if individuals-in-population share signals and rules to coordinate between themselves. Language is a social property, not a *solus ipse* principle. Natural languages are not a 1:1 depiction of a universal grammar that is inherent in our neuronal brain order, but serve to coordinate behaviour in order to be able to adapt appropriately to changing situational circumstances. According to Gödel’s “incompleteness theorem”, language is therefore not a closed system, but is

indefinitely open: competent natural language users are able to generate *de novo* sequences that do not derive from previous sequences but are completely new.

Based on this knowledge, between 1987 and 1990 I developed a theory of communicative nature. Living organisms communicate to coordinate and organize behaviour and to reproduce, and the genetic code is a real language according to syntactic, pragmatic and semantic rules.

The remaining unknown factor was the agents that use the genetic code, i.e. those that: (i) generate nucleic acids into sequences with content; and (ii) combine these nucleic acids correctly (according to Chargaff's rules), integrating them into pre-existing nucleic acid sequences without destroying the previous content that codes for proteins. This means that such agents must be competent to identify the semantics of such sequences and identify appropriate integration sites.

I tried to identify these agents between 1990 and 2005, but I must confess I did not find them. I argued that somehow there must be an innovation code, or evolution code, that functions in evolutionary relevant situations such as environmental changes or stress situations, and which starts by changing the order of genetic content in populations. But natural codes do not code themselves, just as no natural language speaks itself. In any case, empirical data indicate that there must be consortia of living agents that are competent to generate and use natural languages/codes according to syntactic, semantic and pragmatic rules. So, what and where are these agents of the genetic code?

In 2005 I read the book *Viruses and the Evolution of Life*. It described how viruses colonize every cell of an organism in a persistent, non-lytic way. In most cases, they are not widely functional ("defectives") and serve as species-specific (and most often tissue-specific) co-opted adaptations, i.e. regulatory elements that are part of an integrated network of gene regulation. Within this virus-first perspective, viruses are the most abundant genetic sequences on this planet, and cellular genomes, their natural habitat, are a limited resource for this abundance.

From this point, what are the essential agents of life became immediately clear to me, i.e. living agents that are competent to edit the genetic code in a manner coherent with the rules of molecular syntax (Chargaff's rules), pragmatics (context) and semantics (content). But the question arose, why have these agents been ignored or underestimated for so long?

Sixty years ago viruses took the centre stage of biological research, when phages were detected and viruses were first used as transporters and tools in industrial realms to recombine genetic sequences for generating vaccines. Viruses have been viewed as the simplest components of life and genetics. Experiments with viruses led to a fundamental understanding of the molecular mechanisms of living organisms and the foundation of molecular biology. What followed was the success story of molecular biology, which investigated genetic properties in the light of physical and chemical laws. The physicist Max Delbrück defined genetic variations as statistical molecular random changes (mutations). Evolution was therefore the accumulation of statistical errors, or damage and its selection. After this, viruses became simple chemicals. In 1943, Luria and Delbrück performed their famous experiment to prove the fact of random mutations. Although their experiment did provide evidence



of this, it did not exclude the possibility of non-random variability, i.e. natural genome editing. However, only the assumption of random mutations took centre stage in theories of molecular evolution. Natural genome editing by competent agents was not an object of investigation. This led to the exclusive scientific value of mechanistic explanations for the origin of genetic variation and represents a hallmark of mechanistic molecular biology.

From this point viruses became seen as dangerous disease-causing parasites which had escaped from cellular life. As molecular genetic parasites, they have not been considered to have any relevance in evolutionary or developmental processes. In evolutionary biology, they take part as footnotes, at best.

For the next 60 years, genetic structures were investigated by statistical methods and quantitative analyses; these were consistent with the theoretical approach of the Turing and von Neumann cybernetic systems and its mechanical explanation for the origin of information, and its inherent mathematical theory of language as a quantifiable set of signs. Genetic “code” was seen as helpful linguistic metaphor but, in light of the empiristic logic of science, it was also a structure for mathematical exact computation. Its syntactic structure determines the meaning of the stored information that leads to the coded proteins.

With Barbara McClintock’s proof of mobile genetic elements, molecular biology and its central dogma of “DNA-RNA-Protein-anything else” became more dynamic. It became increasingly clear that cellular DNA is not a fixed structure, but is dynamically constituted. In parallel, it also became increasingly clear that there are many regulatory elements, vital for expression patterns and silencing of genes. The discovery of epigenetic marking opened the perspective of the whole genome being marked for transcription and translation, and that these markings can change according to changing environmental conditions or stress-related experiences.

Today, we are at the edge of a main turning point in understanding biological processes. The prevailing central dogma of molecular biology of the last 50 years is no more than a subordinate clause, relevant only to a small fraction of reality. The main role of DNA was relativized through the detection of the early RNA world and its abundance of RNA agents and ribozymes that cooperated and competed in consort. Today, we can consider the increasing knowledge of the important roles played by RNA-agents in all regulatory processes of translation, transcription, recombination, epigenetics and repair, as well as its regulation of all the developmental processes of cellular life. The more complex the living organisms, the more abundant are the involved non-coding RNAs. In some organisms, such as humans, non-coding RNAs represents the most abundant part of the genome.

Now, the new renaissance of viruses is taking centre stage. Research data from the last decade indicate the important roles of viruses, both in the evolution of all life and as symbionts or co-evolutionary partners of host organisms. There is increasing evidence that all cellular life is colonized by exogenous and/or endogenous viruses in a non-lytic but persistent lifestyle. Viruses and viral parts form the most numerous genetic matter on this planet.

A persistent lifestyle in cellular life-forms most often seems to derive from an equilibrium status reached by at least two competing genetic settlers and the immune

function of the host that keeps them in balance. Persistent settlement of host genomes means that, if we postulate agent-driven genetic text editing, we then have to look at their *in vivo* life strategies to understand their habits and the situational contexts that determine the arrangements of their content. On this basis we can reconstruct nucleic acid sequences that function as a code, not as a statistically random mixture of nucleotides, but as informational content in a syntactic order that is coherent with the whole sequence space generated by agents that are linguistically competent in nucleic acid language, that is, the genetic code. As in every language, each character, word, and sentence, together with starts, stops, commas, and spaces in-between, has content and a text-formatting function and is generated by competent agents.

If we imagine that humans and one of the simplest animals, *Caenorhabditis elegans*, share a nearly equal number of genes (ca. 20,000) it becomes obvious that the elements creating the enormous diversity are not the protein coding genes but their higher order regulatory network processed by the mobile genetic elements, such as transposons, retrotransposons and the non-coding RNAs, such as microRNAs, piwiRNAs, tRNAs and rRNAs. If we consider the important role of the highly structured and ordered regulatory network of non-coding RNAs as not being randomly derived, one of the most favourable models with explanatory power is the virus-first hypothesis. This supposes that the evolution of the non-coding RNA world in cellular genomes is the result of persistent viral life strategies. The whole range of mobile genetic agents that are competent to edit the genetic code/nucleic acid language not only edit, but also regulate the key cellular processes of replication, as well as transcription, translation, recombination, repair, and even inventions via a wide variety of small RNA molecules. In this respect, DNA is not only an information-storing archive but a habitat for linguistically-competent RNA agents, most of them seemingly of viral or subviral descent.

To understand their competence in natural genome editing, we have to look not only at their linguistic competence in editing and regulating correct nucleotide sequences, but also at their communicative competence, that is, how they interact with each other, how they compete within host organisms, how they symbiotically interact with host organisms to ward off competing parasites, how they generate *de novo* sequences and what life strategies they share. Exactly these features are presented in this volume. Persistent infection lifestyles that do not harm hosts, and symbiotic, cooperating viral swarms, may be more successful in evolutionary terms for integrating advantageous phenotypes into host organisms than are “selfish” agents.

Increasing empirical data about the abundance of viruses and virus-derived parts in the ecosphere of this planet, and their roles in the evolution and developmental processes of cellular life forms at the level of the microscopic processes of replication, transcription, translation, alternative splicing, RNA-editing, epigenetics and repair, raise a fundamental question concerning a crucial decision about how to define and explain life, as follows.

Those that want to continue a reductionist view of life will rely on the mathematical order of the universe, as determined by the fundamental mechanics of thermodynamics and the resulting mechanisms based on the key elements of this universe and its everlasting unchangeable natural laws. In this respect, evolutionary

statistical mechanisms will remain a driving method in measurements, experimental set-ups and the assembly of quantitative data. Other methods of mathematical language theory will investigate features and processes of nucleotide sequence assemblies, such as bioinformatics, which can evaluate sequence similarities to help in the detection of sequence-determined functions of genes and genomes.

Those who want to leave reductionism and its mechanisms need not move backwards to vitalism or creationism. There is a third way that better fits the available empirical data and that spans agent-based competencies to natural genome editing.

The agent-based perspective is evident in the observation that every coordination process between cells, tissues, organs and organisms depends on signs that function as signals between signalling agents. Signalling and communication does not occur by signals alone, but by living agents that are competent to use signs. In all cases, the participating agents share a competence to generate signs, to receive appropriate messages and to interpret their content. In contrast to former opinions of information theory, the sequence order, that is the syntax of the message, does not determine the meaning of the signals, which is rather determined by the context, the situational set up, or the *in vivo* situation. In this respect, one identical sequence order (syntactic structure) can transport different and, in extreme cases, even contradictory messages. If we look at a single recent example, the use of Auxin in plant communication, we can identify six different purposes of messages that can be transported. This depends on the varying contexts in which this signal molecule is generated, transported, received and interpreted and—of most importance—can trigger varying behavioural responses. I should note that for artificial machines, constructed by humans, this is impossible in principle. Context-dependent interpretation is not possible for algorithm-based programmes that determine machine functions. “The shooting of the hunters”, which every language-competent child can play with various meanings, is not unequivocal and cannot transport contradictory meanings for a computer.

The competent genetic editing, the natural genetic engineering perspective (or natural genome editing), additionally integrates all the currently available knowledge on how genetic sequence orders have evolved, changed, varied (as being crucial for evolutionary variation) and changed dynamically in all adaptational purposes: for example, in the organisation of adaptive immunity. The crucial difference from the reductionist and mechanistic perspective of the last century is that random mutation (copying error or damage), when considered as the most prominent reason for genetic variation, cannot incorporate all the available empirical data. This is the data on viral integration into host genomes (e.g. phages, plasmids and DNA viruses in prokaryotes, retroviruses in mammals, RNA viruses in plants, etc.) that remain either as fully functional viruses, or as defectives that act in an exapted function, such as non-coding RNAs for gene regulation and all the currently known “mobile genetic elements”. It is an empirical fact that random mutations occur, but their role for evolutionary novelty has been overestimated for more than half a century because of the predominance of mechanistic molecular biology.

This book could help to decide which of the two alternatives are chosen. It is important to note that the agent-based perspective does not contradict physical laws, because all the agent-based competencies are consistent with physical laws. In contrast

to the reductionist approach, the agent-based approach can integrate newly available data on signalling, cell–cell communication and natural genome editing that occurs non-mechanically but communicatively. Cell–cell communication and agent-based natural genome editing are both absent in inanimate nature. There is no syntax, semantics and pragmatics when water freezes to form ice. Viruses play a vital role in all cellular and genetic functions, and we can therefore define viruses as essential agents of life.

Buermoos, Austria

Günther Witzany



# Contents

<b>Revolutionary Struggle for Existence: Introduction to Four Intriguing Puzzles in Virus Research .....</b>	<b>1</b>
Matti Jalasvuori	
<b>Quasispecies Dynamics of RNA Viruses .....</b>	<b>21</b>
Miguel Angel Martínez, Gloria Martrus, Elena Capel, Mariona Parera, Sandra Franco, and Maria Nevot	
<b>The Origin of Virions and Virocells: The Escape Hypothesis Revisited .....</b>	<b>43</b>
Patrick Forterre and Mart Krupovic	
<b>Scratching the Surface of Biology’s Dark Matter .....</b>	<b>61</b>
Merry Youle, Matthew Haynes, and Forest Rohwer	
<b>Virus Universe: Can It Be Constructed from a Limited Number of Viral Architectures .....</b>	<b>83</b>
Hanna M. Oksanen, Maija K. Pietilä, Ana Sencilo, Nina S. Atanasova, Elina Roine, and Dennis H. Bamford	
<b>The Addiction Module as a Social Force.....</b>	<b>107</b>
Luis P. Villarreal	
<b>Viral Integration and Consequences on Host Gene Expression .....</b>	<b>147</b>
Sébastien Desfarges and Angela Ciuffi	
<b>Persistent Plant Viruses: Molecular Hitchhikers or Epigenetic Elements? .....</b>	<b>177</b>
Marilyn J. Roossinck	
<b>The Concept of Virus in the Post-Megavirus Era .....</b>	<b>187</b>
Jean-Michel Claverie and Chantal Abergel	
<b>Unpacking the Baggage: Origin and Evolution of Giant Viruses.....</b>	<b>203</b>
Jonathan Filée and Michael Chandler	

<b><i>Megavirales</i> Composing a Fourth Domain of Life: <i>Mimiviridae</i> and <i>Marseilleviridae</i> .....</b>	217
Philippe Colson and Didier Raoult	
<b>On Viruses, Bats and Men: A Natural History of Food-Borne Viral Infections .....</b>	245
Harald Brüssow	
<b>LTR Retroelement-Derived Protein-Coding Genes and Vertebrate Evolution .....</b>	269
Domitille Chalopin, Marta Tomaszekiewicz, Delphine Galiana, and Jean-Nicolas Volf	
<b>Koala Retrovirus Endogenisation in Action .....</b>	283
Rachael E. Tarlinton	
<b>The Evolutionary Interplay Between Exogenous and Endogenous Sheep Betaretroviruses.....</b>	293
Alessia Armezzani, Lita Murphy, Thomas E. Spencer, Massimo Palmarini, and Frédérick Arnaud	
<b>Endogenous Retroviruses and the Epigenome .....</b>	309
Andrew B. Conley and I. King Jordan	
<b>From Viruses to Genes: Syncytins .....</b>	325
Philippe Pérot, Pierre-Adrien Bolze, and François Mallet	
<b>Hepatitis G Virus or GBV-C: A Natural Anti-HIV Interfering Virus .....</b>	363
Omar Bagasra, Muhammad Sheraz, and Donald Gene Pace	
<b>Salutary Contributions of Viruses to Medicine and Public Health .....</b>	389
Stephen T. Abedon	
<b>From Molecular Entities to Competent Agents: Viral Infection-Derived Consortia Act as Natural Genetic Engineers .....</b>	407
Günther Witzany	
<b>Index.....</b>	421

# Corresponding Authors

**Stephen T. Abedon** Department of Microbiology, The Ohio State University, Mansfield, OH, USA

**Frédéric Arnaud** UMR754 UCBL INRA ENVL, EPHE, École Nationale Vétérinaire de Lyon, Lyon, France

**Omar Bagasra** Department of Biology, South Carolina Center for Biotechnology, Claflin University, Orangeburg, SC, USA

**Dennis H. Bamford** Institute of Biotechnology and Department of Biosciences, Biocenter 2, University of Helsinki, Helsinki, Finland

**Harald Brüssow** BioAnalytical Science Department, Food and Health Microbiology, Nestlé Research Centre, Nestec Ltd, Lausanne 2, Switzerland

**Jean-Michel Claverie** Structural & Genomic Information Laboratory (UMR7256), Mediterranean Institute of Microbiology, Centre National de la Recherche Scientifique, Aix-Marseille University, Marseille Cedex 09, France

CNRS – UPR2589, Institut de Microbiologie de la Méditerranée (IMM, IFR-88), Parc Scientifique de Luminy, Marseille Cedex 09, France

**Sébastien Desfarges** Institute of Microbiology (IMUL), University Hospital Center and University of Lausanne, Lausanne, Switzerland

**Jonathan Filée** Laboratoire Evolution, Génomes et Spéciation, CNRS UPR 9034, Gif sur Yvette Cedex, France

**Patrick Forterre** Institut Pasteur, Paris, France

**Matti Jalasvuori** Division of Ecology, Evolution and Genetics, Research School of Biology, Centre of Excellence in Biological Interactions, Australian National University, Canberra, Australia

Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland



**I. King Jordan** Georgia Institute of Technology, School of Biology, Atlanta, GA, USA

PanAmerican Bioinformatics Institute, Santa Marta, Colombia

**François Mallet** Laboratoire Commun de Recherche Hospices Civils de Lyon-bioMérieux, Cancer Biomarkers Research Group, Pierre Bénite Cedex, France

**Miguel Angel Martínez** Fundació irsiCaixa, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Spain

Laboratori de Retrovirologia, Fundació irsiCaixa, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Spain

**Didier Raoult** Unité de Recherche sur les Maladies Infectieuses et Tropicales Émergentes (URMITE), Centre National de la Recherche Scientifique (CNRS), Unité Mixte de Recherche (UMR) 6,236, Institut de Recherche pour le Développement (IRD) 3R198INSERM U1095, Méditerranée Infection, Facultés de Médecine et de Pharmacie, Aix-Marseille Université, Marseille, France

Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Centre Hospitalo-Universitaire Timone, Assistance publique des hôpitaux de Marseille, Marseille, France

**Forest Rohwer** Department of Biology, San Diego State University, San Diego, CA, USA

**Marilyn J. Roossinck, Ph.D.** Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, OK, USA

Plant Pathology and Biology, Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA

**Rachael E. Tarlinton** School of Veterinary Medicine and Science, University of Nottingham, Loughborough, UK

**Luis P. Villarreal** Center for Virus Research, Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, USA

**Jean-Nicolas Volff** Equipe Génomique Evolutive des Poissons, Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, CNRS, Université de Lyon, Lyon Cedex 07, France

**Günther Witzany** Telos – Philosophische Praxis, Buermoos, Austria

# Revolutionary Struggle for Existence: Introduction to Four Intriguing Puzzles in Virus Research

Matti Jalasvuori

**Abstract** Cellular life is immersed into an ocean of viruses. Virosphere forms the shadow of this cell-based tree of life: completely dependent on the tree for existence, yet, the tree is equally unable to escape its ever evolving companion. How important role has the shadow played in the evolution of life? Is it a mere ethereal partner or a constitutive factor? In this chapter four puzzles in virus research are taken under the scope in order to probe some of the intriguing ways by which viruses can help us understand life on Earth. These puzzles consider the origin of genetic information in viruses, viruses as symbiotic partners, the structural diversity of viruses and the role of viruses in the origin of cellular life. More than providing answers, this introduction exemplifies how viruses can be approached from various angles and how each of the angles can open up new ways to appreciate their potential contributions to life.

## 1 Introduction

Life on Earth is composed of multitude of cellular organisms, some of them being as tiny as bacteria, others as complex as humans. Yet, this cellular way of living is overwhelmed in both number and genetic diversity by non-cellular entities, each of which is capable of enforcing cellular organisms to fulfill their selfish needs. A word *virus*, a Latin term for poison, commonly refers to this strategy for survival.

---

M. Jalasvuori (✉)

Division of Ecology, Evolution and Genetics, Research School of Biology,  
Centre of Excellence in Biological Interactions,  
Australian National University, Canberra, Australia

Department of Biological and Environmental Science, University of Jyväskylä,  
P.O. Box 35,40014, Jyväskylä, Finland  
e-mail: matti.jalasvuori@jyu.fi

And for a poison they are often treated. This is of no surprise, given that the apparent simplicity and inanimate nature of deadly viruses (van Regenmortel 2000; Moreira and López-García 2009) may lead us to intuitively neglect or completely ignore them in our approaches to understand the evolutionary spectacle that living things have to offer. Yet, while being relatively simple in comparison to cells, there is much that we do not know about viruses or their roles in evolutionary processes. Viruses have been here for a long time (Forterre and Prangishvili 2009a), and studies suggest that viruses appear to have played a part in events such as the origin of cellular life (Koonin et al. 2006) and the evolution of mammals (Gifford 2012). But what has their role been exactly? When does the inclusion of viruses into the frame of analysis lead to evolutionary insights? Or even breakthroughs?

Unfortunately in many instances we are still after on a mere hunch. For this reason, instead of providing you with a set of scientifically chewed and grounded answers, I introduce you to a four selected puzzles in virus research in an attempt to scope where the limits of some of our contemporary knowledge lies. The presented questions revolve around themes such as the origin of new genetic information, the origin of new types of symbiotic relationships, and even the origin of life as we know it. Naturally profound puzzles as these are horribly difficult ones to address in a complete and comprehensive manner. Yet, in the spirit of this book, these puzzles can help determining whether viruses could be considered truly as essential agents of life.

### *1.1 Viruses and Virions: What Is the Difference?*

First, however, a relatively commonly adopted misconception on what a biological virus actually is must be resolved because it has been behind many of the misunderstandings on viruses. The heart of the issue lies in the notion that a virus often refers only to the protein-formed protective capsid, which encloses viral genomic information in the extracellular environment (see discussion in Jacob and Wollman 1961; Forterre and Prangishvili 2009b; Villarreal and Witzany 2010; Moreira and López-García 2009; Jalasvuori 2012). This infectious particle is known as a virion and they are generally regarded to be dead (in many depressingly unfruitful discussions). Virions are entities that intrude and assume the control of cellular organisms in order to produce more virions. But should this dead virion actually be considered equal to a virus? And what then would a virus be, if not a virion? The seemingly trivial difference between a virus and a virion needs to be tackled as it allows us to appreciate viruses as evolutionary players, or even as living organisms (Forterre and Prangishvili 2009b; Villarreal and Witzany 2010; Forterre 2011; Jalasvuori 2012). In any case and regardless of our opinions on their living status, viruses are part of the evolving biosphere and therefore a relevant factor in various evolutionary processes.

Virion is the extracellular step in the life cycle of a virus. Virion is the traditional picture that every book offers for depicting a virus. Virion is the transient stage by which the viral genetic information gets from one host organism to another. This virion, however, lacks the *life* of the virus since it is only the dormant and inactive form of viral genetic information (Brüssow 2009). For this reason viruses might appear as toxic substances that have the capability to occasionally cause the demise of cellular organisms but that are essentially just another environmental factor of only minor interest from evolutionary point of view.

However, arguably, the actual virus is more than its dead shell in the environment. Virus is part of a living organism when it is inside a host cell. And the phenotype of this organism is partly expressed by the virus (Forterre and Prangishvili 2009a; Forterre 2010; Jälasvuori 2012). Many viruses maintain the potential for producing inanimate virions during their endure within the cellular organism, but virus itself should be considered to be its full reproductive cycle including both external and internal parts (Villarreal and Witzany 2010; Jälasvuori 2012). Yet, strictly speaking, only the within-cell reproductive cycle is required for the survival of the viral genetic information (Krupovic and Bamford 2010; Jälasvuori 2012). And this requirement lets us approach viruses as a genuine form of life that can exploit foreign cell-vehicles for preserving and propagating their genetic information (Forterre 2010, 2011).

In other words, virus should not be mistaken only for their non-essential extracellular form given that viruses are equally dependent on cells with all other genetic replicators – being those chromosomes, plasmids or anything else. Virus just is not dependent on any particular cell due to their capability transfer themselves from one cell to another via virions. And due to this extracellular form of existence, viruses are not terminated even if their replication causes the demise of the current host organism. However, jumping from this notion to the conclusion that viruses are dead and thus irrelevant partners of evolutionary processes is unwarranted. Naturally, our definitions of viruses include the infectious extracellular part, but for thorough understanding of viral life it must be noted that any such definitions are in the end artificial. Virus is one of the ways by which genetic information have adapted to survive in this biosphere. From the viewpoint of cellular organisms, this way of struggle for existence is much more complex than the presence of chemical substances in the environment would be. Viruses, unlike poisons, are capable of evolving genetically and going extinct. Sometimes they can also form more or less permanent mutually benefiting relationships with their hosts.

Now this perhaps more allowing perspective to viral life sets a more appropriate stage to consider any virus related puzzles. Each of the presented questions approach viruses from different angles and hopefully provide an intriguing introduction to the diversity of ways by which viruses may help us understand the evolution of our biosphere. However, I wish to note that I consciously retained from drowning the reader in supporting evidence in order to keep the text fast pacing and relatively easy to digest.

## 2 Can Genes Emerge in Viruses?

Novel sequencing and sampling techniques have made it possible to determine the overall genetic information in any particular sample. Moreover, sequences of complete organisms have revealed the true genetic diversity of living entities. These studies have led to the revelation that many organisms harbor a variety of genes that are unknown to science (Mocali and Benedetti 2010). In other words, our biosphere is abundant with genetic information for which we cannot assign a role, function or evolutionary origin (Cortez et al. 2009). Interestingly, a fair portion of these novel genes are found from viral genomes (Yin and Fischer 2008; Prangishvili et al. 2006) or belong to genome integrating genetic elements (Cortez et al. 2009). How did these genes end up in viruses?

### 2.1 *Are Viruses Only Hitchhiking on Genetic Information?*

Viruses are completely dependent on cellular resources for reproduction. Viruses use cellular amino acids to make viral proteins and some acquire lipids from cellular membranes to assemble functional virions. All viruses embrace cellular nucleotides to produce copies of viral genetic information. Given the profoundly parasitic nature of viruses, it seems reasonable to assume that viruses are also completely dependent on cellular genes for evolution. Indeed, many viral genes appear to have been acquired from their hosts and thus viruses could be considered as genetic burglars, hitchhikers on the highway of genetic information. Viruses are something that themselves are not evolving but which are evolved by cells (Moreira and López-García 2009). The actual *de novo* origin of genetic information would happen within stable cellular beings such as bacteria.

However, many viral genes appear to have no cellular counterparts (Yin and Fischer 2008; Forterre and Prangishvili 2009b). Why is this? Do we need to sequence more bacterial genomes in order to find the common ancestor form a cellular chromosome? Yet, as the number of sequenced bacterial chromosomes has increased, the number of unknown genes in viruses has remained unchanged (Forterre and Prangishvili 2009b). Sometimes when some rare types of virus genes are finally discovered from host chromosomes, it turns out that the genes in the chromosomes actually belong to genome integrated viruses (Jalasvuori et al. 2009, 2010). Therefore the sequencing of bacterial chromosomes does not seem to provide an easy way out of the puzzle. Perhaps the genetic novelty of viruses is of genuine nature and there are no cellular homologies to be found. Or could it just be that the rapid evolutionary rates of genes in viruses is simply making the homology with cellular genes untraceable?

In principle, it is possible that majority of genes evolve in such a fast pace in viruses that the sequence can no longer be recognized to be of cellular origin (Forterre and Prangishvili 2009b). Indeed, general analyses of the divergences of amino acid sequences propose that even the most conserved proteins in our biosphere have not

discovered all potential ways to encode their function (Povolotskaya and Kondrashov 2010). Therefore there appears to be room in the sequence space into which the host-derived genes can evolve to in viral genomes.

However, comparison of nucleotide or amino acid sequences is not the only mean by which gene divergences can be studied. While the sequence on DNA or amino acid level may evolve rapidly, the three dimensional structure of the gene product, usually a protein, can remain relatively unchanged. Indeed, generally there is no selection to preserve any certain amino acid sequence but only the (whatever) function that is associated with the three dimensional conformation of the protein. Save for amino acids mediating chemical reactions, the same structural conformation can be acquired with a variety of different sequences.

Viruses seem to have genes that produce structurally and functionally conserved proteins, which have no apparent cellular ancestors (Bamford et al. 2005; Koonin et al. 2006; Keller et al. 2009). These genes have been within (relatively) independently evolving viral genomes perhaps for as long as billions of years and they can still be shown to share a common ancestry. Did these genes emerge in virus genomes in the first place? It seems possible, given that many of these conserved “hallmark” virus genes (Koonin et al. 2006) encode for viruses specific tasks such as capsid proteins or packaging enzymes that facilitate the transfer of viral genome into the capsid.

## ***2.2 If Gene Emerges Within a Cell But Survives in Viral Genome, Is It a Viral Gene?***

Naturally, the emergence of a gene in a virus does not indicate that the gene popped into existence within the protective capsid in an extracellular environment (Forterre and Prangishvili 2009b; Forterre 2010; Jalasvuori 2012). Rather, it would mean that a virus, while replicating in a cell, ended up having an altered genetic sequence. This altered sequence opened the road for the emergence and evolution of a new gene. In practice the gene would form through point mutations and other genetic changes similarly with any other emerging genes (Forterre and Prangishvili 2009b).

But if the new gene would emerge within a cell, is it not rather a cellular gene than a viral one (Moreira and López-García 2009)? Doesn't this indeed only enforce the view of cellular origin of viral genetic information? No, it does not, if we allow ourselves to consider viruses to be more than just their encapsulated extracellular forms (Forterre 2010). If the gene formed through mutations in a viral genome and the new gene was able to survive due to its benefits to the virus and not to the host, then it would seem only reasonable to consider the gene to be of viral origin (Jalasvuori 2012). Therefore, even if a cell serves the function of a vessel for the development of a new gene, the gene would remain in the global gene pool because of viruses. Eventually, when metagenomic studies, for example, are performed, these novel genes could be discovered from capsid enclosed genomes of viruses with no apparent counterparts in any cellular organisms.

Even if the *de novo* origin of genes actually occurred in viruses, it would be only a starting point from which to approach other interesting questions. What do these novel genes do? There are countless unique genes in viruses, but are they also encoding countless unique functions. Or is it possible that they only have unique sequences while affecting very similar cellular processes? And what would that indicate?

Viruses of bacteria, also known as bacteriophages, can have genes for very different types of functions. Some phages encode transfer RNAs and other essential cellular functions (Miller et al. 2003). Others can carry genetic information for mediating photosynthesis (Mann et al. 2003) or producing lethal toxins (O'Brien et al. 1984). Much of the phage genes, however, affect genetic regulation, virion assembly and host-virus interactions. Yet, other viruses (like Mimivirus) have genes that were earlier considered to be only part of cellular chromosomes and thus blurred the line between what viruses can and what they can not do (Raoult et al. 2004).

Nevertheless, in principle, it seems possible that the product of a viral gene can influence any thinkable biological process. Some truly novel genetically encoded functions allowing, for example, exploitation of completely new types of resources or inhabit previously uninhabitable environments, may come into existence in the genome of a virus. Perhaps viral innovations can open new niches for cellular organisms to occupy: many of the novel genes in bacteria are taxonomically restricted and ecologically important (Wilson et al. 2005).

### 3 Can Viruses Become Symbionts?

Viruses are generally seen as parasites of cellular organisms. Viruses enter the host cell, utilize cellular resources for creating new viruses and then sacrifice (or damage) their temporary slaves in order to escape the scene of crime. How could this violent strategy ever turn into a mutually benefiting symbiosis?

In a mutualistic relationship the fitness of the two entities together is (often) higher than the fitness of either of the components alone. In other words, both of the symbionts would suffer from abandoning its partner. Therefore, if a virus was ever to be appreciated as a mutually benefiting partner, it should be counterproductive for the host cell to get rid of a virus that has integrated into genome of the host. This seems to be a problematic approach, given that the avoidance of parasites is considered to be one of the key drivers of evolution and responsible (at least partly) for the maintenance of such fundamental traits as sexual reproduction (Hamilton et al. 1990; King et al. 2011).

#### 3.1 Endogenous Viruses: Fossils or Something More?

Nevertheless, viral genetic information is often found to be incorporated to cellular genomes (Holmes 2011). For example, human chromosomes contain more viral DNA than actual human genes. In fact, remnants of viruses are abundant in genomes

of many different organisms, ranging from animals to bacteria (Casjens 2003; Katzourakis and Gifford 2010; Jalasvuori et al. 2010). How did these viral elements get into all these organisms? What types of evolutionary processes may be responsible for these genomic fusions, and could they be of evolutionary importance?

Are the existing viral remnants in genomes mere evolutionarily insignificant left-overs of previous virus infections (Jern and Coffin 2008)? Were they so insignificant to the fitness of the hosting cell that there simply was no selection to get rid of the element? Many of the endogenous viruses are relatively conserved and have persisted over evolutionary times in various species, such as humans and our primate cousins, suggesting that the relatively error-free host polymerases that are used to replicate the endogenous viruses are able to preserve these sequences as viral fossils over evolutionary times (Duffy et al. 2008). However, many of the virus elements have also shown to accumulate inactivating mutations and thus they are evolving only as non-encoding pseudogenes in animal genomes (Katzourakis and Gifford 2010). Yet, other virus genes have remained functional, suggesting that there has been a purifying selection to maintain the correct sequence.

### ***3.2 What Benefits Can Viral Elements Provide to the Host?***

Could it be possible that some of these viral elements in cellular chromosomes resulted essentially from mutually benefiting although aggressive genetic fusions (Ryan 2009)? Can the symbioses of viruses with cells be evolutionarily favorable steps, not mere coincidences?

In order to be more precise, the question is not whether genetic fusions of the genomes of viruses and cells can improve the reproductive rate of cells *per se*. There are clear examples for this to be true. As a tragic example several viruses are known to cause the uncontrolled multiplication of human cells, which results in the formation of tumors. These virus-containing cells out-reproduce other human cells and thus they end up having much more descendants than the virus-free cells. Within this limited framework the virus-cell symbiotic can have the highest fitness. But by extending our perspective we notice that this short-term benefit rapidly backfires due to the demise of the hosting animal. The selfish behavior of some cells leads to a tragedy of commons, where the gain of few is decreasing the fitness of both host and the virus. Therefore, the real question is whether viruses and their hosts may form symbiotic relationship that can increase the fitness of the whole organism within a large-enough evolutionary frame. In other words, we can ask, for example, if the virus-host symbiont could invade a population of virus-free hosts because of the advantages that the virus provides to its hosts.

Some viruses that infect bacteria are known to form temporary mutually benefiting symbiotic relationships with bacterial cells (Roossinck 2011). These viruses enter the host cell and, instead of producing vast number of virions and destroying the cell, they take up residence within the host. During this latent infection temperate viruses replicate their genomes along with the cell but deter from making



virions. Only in the distress of their hosts they ignite the production of virions and they do it in order to escape the potentially doomed bacterium.

These temperate bacterial viruses may carry genes (e.g. for producing toxins) that can significantly improve the performance and thus the reproduction of their host bacteria. The combination of the bacterial virus and the bacterium can end up being the evolutionary winner in a competition against bacteria that did not have the latent viral infection. Therefore, among bacterial organisms such straightforward mutualistic relationships may emerge on regular basis (Roossinck 2011). Moreover, the short-term benefit provided by the phage does not backfire in the same sense as the spreading tumors do within animal hosts. But then, bacteria and humans are quite different in multiple respects. Are these symbioses limited only to single-celled beings or can such relationships emerge among more complex organisms that reproduce via specific germ cells? Indeed, despite of the all the movies, we do not know of any viruses that carry bacteriophage-like toxin genes, which would grant us some sort of superpowers. Therefore this bacterial approach may simply be ill-suited to understand symbiotic relationships in animals.

However, there is another way by which temperate viruses of bacteria boost the survival of their hosts. Whenever a bacterial virus resides within a bacterium, it renders the cell immune to infections by similar viruses. And this quality of viruses, the incapability of a single virus type to multiply infect an already-infected cell (i.e. the resistance of superinfection), appears to be very common among all viruses and therefore also applicable to other organisms (Berngruber et al. 2010). Prevention of superinfection allows viruses to establish latent infections that are especially important under conditions where chances for horizontal transfer of the virus are limited.

Among bacterial populations that are subjected to temperate viruses, the most rapid mean by which resistant host cells emerge are due to the latent infections by temperate viruses themselves. The presence of the virus therefore selects the bacterial population to become prevalent with integrated viruses. When there are both susceptible hosts and infective virions in the same environment, the resistant hosts have an apparent advantage (Roossinck 2011). Moreover, the genome integrated viruses sometimes produce virions and thus maintain the selection for the presence of the latent virus. The fact that viruses themselves contain genetic means to make host cells immune to the virus may prove to be the evolutionary superpower that can facilitate the formation of a symbiotic relationship also between a virus and its animal host.

However, even if viral infections can make the host animal resistant to further infections by similar types of viruses, it is not a heritable symbiosis. We are immune to chickenpox after an infection, but our children still need to get infected themselves in order to become resistant (or, alternatively, be vaccinated against the virus). Is it possible that the resistance would become inheritable so that the progeny of an infected individual would not need to face the severe effects of an infection?

Complex multi-cellular animals develop from a fertilized cell. This single cell divides and the divided cells specialize to different functions eventually producing a complete organism. The genetic information in all animal cells remains essentially

the same throughout the life of the organism even if the phenotypes of cells can vary tremendously. Therefore, if the virus was integrated already in the original germ cell, it would become inherited to every cell of the multi-cellular organism, including those that eventually become the germ cells of the next generation. In such a case the virus could both protect the organism from the external versions of the virus and be transmitted vertically to the next generation.

### 3.3 *Taming the Enemy into an Ally*

During a roaming virus epidemic, this integration of a virus to germ line cells could provide an advantage to an individual (Jern and Coffin 2008). Indeed, in many cases endogenous viruses appear to protect their hosts against exogenous viruses (Maori et al. 2007; Katzourakis and Gifford 2010). However, such endogenous viruses themselves seem to be able to reinfect the germ line cells (Belshaw et al. 2004). Nevertheless, the endogenous virus may be able to make the host organism to be able to ignore the ill-effects that the epidemic causes to other individuals. Naturally inheritable resistance against chickenpox is not a significant advantage but resistance against a more severe virus could be.

So, in principle and under certain conditions, germ line infection could prove to be a favorable trait within a population (Maori et al. 2007). The new *virus alleles* may even be able to invade the whole population, if the maintenance of the virus remains to improve the fitness of the virus-containing individuals over their virus-free counterparts (Katzourakis and Gifford 2010). Indeed, as with bacteriophages, endogenous viruses of animals can remain partly active even after endogenization (Coffin et al. 1997; Tarlinton et al. 2006) and thus the virus itself can maintain the pressure to retain the virus allele within the population.

In such a case, is it possible to consider that the virus has established a mutually benefiting relationship with its animal host. Maybe, given that it would be disadvantageous for the organism to get rid of the virus since it would make the organism susceptible to infections. Of course, this symbiotic partnership would exist mainly on the level of genetic information (Ryan 2009), but it would still emerge through a fusion of two distinct genetically reproducing entities. In the end, very little is still known about the endogenization process. Even if viruses could be considered to form symbiotic relationships via whatever mechanisms, several interesting questions remain. How does this new integrated virus affect the subsequent evolution of their hosts? Endogenous virus changes the genetic composition of the chromosomes and can, for example, regulate the expression of host genes (Jern and Coffin 2008). Some of the viruses are active elements and cannot be dismissed as irrelevant components of organisms. Indeed, some virus derived genes in mammals and other animals appear to have remained active for over tens of millions of years (Katzourakis et al. 2005; Katzourakis and Gifford 2010). But even then, it is difficult to say for certain how significant role did these viruses play in the evolution of their hosts. However, we are free to do little speculation.

Endogenous viruses can integrate repeatedly into various places within and among host chromosomes (Katzourakis et al. 2007). The number of elements and the site of integration can have significant effects on the phenotype of the host cell. The establishment of the viral genome into the host chromosome appears to be followed by in-genome evolution (Tarlinton et al. 2006; Katzourakis et al. 2007). Does this evolution select for the viruses to be integrated in positions where they induce the lowest possible cost on the host or, perhaps, even induce changes that increase the host fitness?

Sexual reproduction effectively filters genetic information to produce beneficial combinations. Could sexually reproducing individuals become favored over asexually reproducing phenotypes as the sexual recombination of genetic material allows the integrated virus to more rapidly settle within fixed beneficial locations in chromosomes? Or perhaps allow the hosts to tame the uncontrollably proliferating endogenous viruses (Katzourakis et al. 2005)? Could the subsequent evolution after virus endogenization induce notable changes in the phenotype of the organism as the genome stabilizes to cope with the presence of the new element?

Some or even most of the endogenous viruses may be just insignificant remnants of previous infections and as such they would not much affect the evolution of their host species. But other symbiotic viruses probably made a real difference. As an example of such, a virus derived gene, labeled as syncytin, appears to be crucially important for the morphogenesis of placenta (Mi et al. 2000). Did pregnancy as humans and other placental mammals experience it emerge as a result of viral endogenization?

## 4 Why Are There Only Few Types of Bacteriophages?

Viruses are known to evolve rapidly and viral genomes often contain unique genes for which no homologues can be determined. But are virions, the extracellular forms of viruses, composed of similarly diverse structures? Is there a novel structural design waiting whenever we pick up any of the  $10^{31}$  or so virions (Suttle 2007) from the environment?

The proteins on the virion dictate whether or not viruses are able to attach to a suitable host cell and therefore there should be constant selection driving the evolution of these proteins (as well as their host counterparts) Weitz et al. 2005. This is indeed what has been observed: the genes responsible for encoding virion proteins that mediate host-cell attachment are the ones that evolve most rapidly (Saren et al. 2005; Paterson et al. 2010). Even closely related viruses may have completely different genes for producing the host-recognizing spikes on the virion (Jaakkola et al. 2012).

But virion is more than a mean to mediate host recognition. The capsid serves as the protective shell for genetic information in the extracellular environment and therefore viruses must also encode proteins (or other means) to produce this shell. Are the genes and the architectural principles for forming capsids equally diverse with host recognition genes?

While virions are extremely abundant and the genetic information they enclose can be very diverse, the capsids of a significant portion of virions in this biosphere may be arranged into just few conserved and homologous lineages (Krupovic and Bamford 2011). Given the astronomical number of virions on earth, this appears to be worth of a closer look.

#### ***4.1 Astronomical Number of Bacteriophages in a Handful of Lineages***

Bacteria are the most abundant type of a cellular organism on earth and their viruses are equally common. Bacteriophages almost exclusively form virions with a spherical head on which a tail is attached to. The head beholds the genetic information of the virus whereas the tail serves as a tool for attaching onto new host cells and, sometimes, as an injection needle during the infection process. This homologous group of viruses is known as *Caudovirales* (Ackermann 1998). Other types of bacterial viruses also exist, but they are not many (Ackermann 2001): there are icosahedral viruses with inner – and outer membranes, amorphous viruses and helical viruses (Oksanen et al. 2010). Altogether, we have discovered only less than ten truly different types of virion-architectures from all currently known bacteriophages.

What is this architectural conservation trying to tell us? Why are there not a 100 different types of bacterial viruses, or 100 billion types? Even if there were 100 billion unique types of viruses, each of them would still have over billion billion virions. And such a large number of individuals could indeed retain a stable population over evolutionary times. This, however, is not the case. You can calculate the virion architectures of bacteriophages with your fingers. Viruses are generally considered to be of polyphyletic origin, indicating that there are multiple viral ancestor and not a single common one. Still, the apparently limited number of architectural types suggests that new virus types are not emerging on regular basis, since, if they were, we would be likely to find new viruses all the time. This leads to a question: when did these existing structural types emerge and why did they cease emerging?

We know that mankind may be facing a completely new and highly lethal epidemic any given day. HIV, SARS, Ebola and other doomsday candidates emerged out of the blue just to bring destruction to the world. Is it only bacterial viruses that are no longer emerging whereas higher organisms, like humans, can still have completely novel viruses? But are human viruses actually unique?

#### ***4.2 Deep Evolutionary Connections Between Viruses***

In 1999 when the major structural proteins of bacterial virus PRD1 and human Adenovirus were compared on structural level, it was noticed, surprisingly, that they were highly similar (Benson et al. 1999). Despite of the sequence dissimilarity, both

viruses used a unique but respectively common type of interlinked protein-barrels (so-called double beta-barrels) for composing their protective capsids. The obvious question emerged: are these two viruses that infect very distantly related hosts (bacteria and humans) actually related to each other? Or is this just another case of convergent evolution where two entities independently evolved towards the same direction (Moreira and López-García 2009)?

Closer analysis of both of these viruses and their other relatives revealed more things in common (Krupovic and Bamford 2008). Vast majority of them had an inner lipid membrane beneath the protein capsid, a generally rare trait among viruses. Moreover, these viruses encode related ATPases (with certain specific motifs) which have been shown to facilitate the transfer of the viral genome into empty capsids. Later on similar viruses were found to infect thermophilic crenarchaea (Khayat et al. 2005) and reside in the genomes of thermophilic euryarchaea (Krupovic and Bamford 2008). In terms of genetic exchange, the Archaeal phylum of Crenarchaeota consists of deep-branching organisms that appear to have been evolving relatively isolated from all other life forms since the emergence of cellular life (Gribaldo and Brochier-Armanet 2006). Together these characteristics suggested that convergence appears to be an improbable cause to explain all the common features and thus it is reasonable to assume the existence of a common ancestor in some distant past. But this leads us to the same question as before: how distant are we actually talking about? 100 million years? A billion? Four billion?

Several analyses suggest that Bacteria and Eukaryote (a domain that includes us humans along with baking yeast) had their last common ancestor about four billion years ago. The same branching time applies to the divergence of Bacteria from Archaea. In other words, these double beta-barrel viruses infected all the domains of life and many deep branches within those domains. But are these viral lineages as old as their cellular hosts? Or is it possible that these viruses emerged later on just to spread to infect all domains of life? We know that viruses are very host specific and usually the viral tree of life corresponds quite well with the evolutionary tree of their hosts (McGeoch et al. 2005). However, there are exceptions and therefore this line of reasoning does not provide a way out of the problem.

Interestingly, several other domain-spanning lineages have been discovered. Herpes viruses have the same peculiar way to produce their capsids as do the extremely abundant tailed viruses that infect bacteria and archaea. Certain RNA-viruses such as bacterial cystoviruses and eukaryal reoviruses appear to be of common origin due to unique genome and capsid organization. There are also other lineages.

It seems that many viruses can have representatives infecting all basic cell types, but these representatives themselves have no recent common ancestors. Moreover, viruses appear to harbor genes that does seem to have been derived from none of the three domains of cellular life but which are very conserved and prevalent among viruses (Koonin et al. 2006). One possible way to explain all these features is to assume that the ancestor of these viruses may have emerged already before the separation of Bacteria, Archaea and Eukaryote into their independent domains.

Recently it was discovered that the double beta-barrel viruses appear to have evolved from a novel viral lineage, so-called single beta-barrel viruses, which

themselves form an independent domain spanning lineage (Krupovic and Bamford 2008; Jalasvuori et al. 2009; Ilona Rissanen personal communication). It is possible that these two viral lineages diverged already before the emergence of contemporary cellular domains. This on the other hand means that by studying viral lineages it might be possible to reach back to some past evolutionary events that occurred before the last universal common ancestor of cells. That period in the evolution of life is generally shrouded in unknown, given that the last common ancestor of cells have been considered as the ultimate boundary beyond which we cannot go by comparing differences between existing living organisms. But if we are not solely dependent on cells in our analyses, then this boundary may be breachable. Study of viral lineages and their origins can give us unique clues about the very first steps of life on Earth.

### 4.3 *Structural Diversity of Hot Archaeal Viruses*

Interestingly, while bacteriophages are either head-tail viruses or one of the few other types, the virions infecting hyperthermophilic crenarchaeal hosts are structurally very diverse (Prangishvili and Garrett 2004; Pina et al. 2011). There are lemon-shaped viruses, tulip-shaped viruses, bottle-shaped viruses, there are sticks with hooks and pleomorphic-viruses along with all sorts of globular, icosahedral and filamentous morphologies. Why is there such a variation especially among archeal viruses? Bacteria and archaea are so similar to each other that it was only recently that we were even able to distinguish them from one another.

Hyperthermophilic crenarchaea are very deeply branching organisms in the tree of life and their viruses are equally unique (Ortmann et al. 2006). They also inhabit extremely hot environments. Are these clues relevant for understanding the diversity of viral phenotypes? Indeed, when the viruses of less thermophilic archaeal organisms have been studied, they were found to be less diverse morphologically. Could it be possible that there was wider diversity of viral phenotypes during the early steps of the evolution of life? And has this diversity been somehow better prevailing among hyperthermophilic crenarchaeal organisms whereas it was lost among other prokaryotes (Jalasvuori and Bamford 2009)? The viruses of most deep-branching hyperthermophile bacterial families (like *Thermotoga* or *Aquifex*) have not been studied. It would be interesting to see if their viruses resemble only the usual head-tail viruses or whether they are more like the ones infecting crenarchaea – or something totally different.

It is likely that all contemporary life forms on earth have evolved from thermophilic ancestors (Di Giulio 2003). There are at least two potential explanations for this, both of which can be correct. First, life may have emerged within a hot habitat such as hydrothermal vents on the ocean floor. Second, life may have faced multiple near-extinction level catastrophes in which all the surviving organisms were thermophiles. Indeed, earth is known to have been under heavy bombardment of massive comets and asteroids during the Hadean period (ending about 3.8 billion years ago).