Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Virology

Metagenomics and future perspectives in virus discovery

John L Mokili[1], Forest Rohwer[1,2] and Bas E Dutilh[1,3]

Monitoring the emergence and re-emergence of viral diseases with the goal of containing the spread of viral agents requires both adequate preparedness and quick response. Identifying the causative agent of a new epidemic is one of the most important steps for effective response to disease outbreaks. Traditionally, virus discovery required propagation of the virus in cell culture, a proven technique responsible for the identification of the vast majority of viruses known to date. However, many viruses cannot be easily propagated in cell culture, thus limiting our knowledge of viruses. Viral metagenomic analyses of environmental samples suggest that the field of virology has explored less than 1% of the extant viral diversity. In the last decade, the culture-independent and sequence-independent metagenomic approach has permitted the discovery of many viruses in a wide range of samples. Phylogenetically, some of these viruses are distantly related to previously discovered viruses. In addition, 60–99% of the sequences generated in different viral metagenomic studies are not homologous to known viruses. In this review, we discuss the advances in the area of viral metagenomics during the last decade and their relevance to virus discovery, clinical microbiology and public health. We discuss the potential of metagenomics for characterization of the normal viral population in a healthy community and identification of viruses that could pose a threat to humans through zoonosis. In addition, we propose a new model of the Koch's postulates named the 'Metagenomic Koch's Postulates'. Unlike the original Koch's postulates and the Molecular Koch's postulates as formulated by Falkow, the metagenomic Koch's postulates focus on the identification of metagenomic traits in disease cases. The metagenomic traits that can be traced after healthy individuals have been exposed to the source of the suspected pathogen.

Addresses
[1] Department of Biology, San Diego State University, San Diego, CA 92182, USA
[2] Center for Microbial Sciences, San Diego State University, San Diego, CA 92182, USA
[3] Centre for Molecular and Biomolecular Informatics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands

Corresponding author: Mokili, John L (jmokili@gmail.com)

Current Opinion in Virology 2012, **2**:63–77

This review comes from a themed issue on
Environmental virology
Edited by Christiane Wobus and Helen Nguyen

Available online 20th January 2012

1879-6257/$ – see front matter

DOI 10.1016/j.coviro.2011.12.004

Introduction

Direct-count epifluorescence and transmission electron microscopy have shown that viruses are highly abundant in most environments. Bergh *et al.* demonstrated that 1 l of seawater can contain as many as 10^{10} virus-like particles (VLPs) [1], approximately 10 times more than the number of prokaryotes. Terrestrial environments often have 10^9 VLPs per gram. By extrapolation from the estimated number of prokaryotes in different environments [2], viruses are the most abundant entities in the biosphere totaling an estimated number of 1.2×10^{30}, 2.6×10^{30}, 3.5×10^{31}, and $0.25–2.5 \times 10^{31}$ in the open ocean, in soil and in oceanic and terrestrial subsurfaces, respectively.

In the human holobiont, the 10^{13} human cells are outnumbered 10-fold by bacteria and 100-fold by viruses. Viral acquisition starts early in life *in utero* or perinatally during the first few weeks after birth as demonstrated by studies of the gut viral communities in infants. While no VLPs could be detected in the earliest infant stool samples, there were $\sim 10^8$ virus particles per gram wet weight of feces by the end of the first week [2]. The majority of these VLPs appear to be bacteriophages, the bacteria-infecting viruses [2–4].

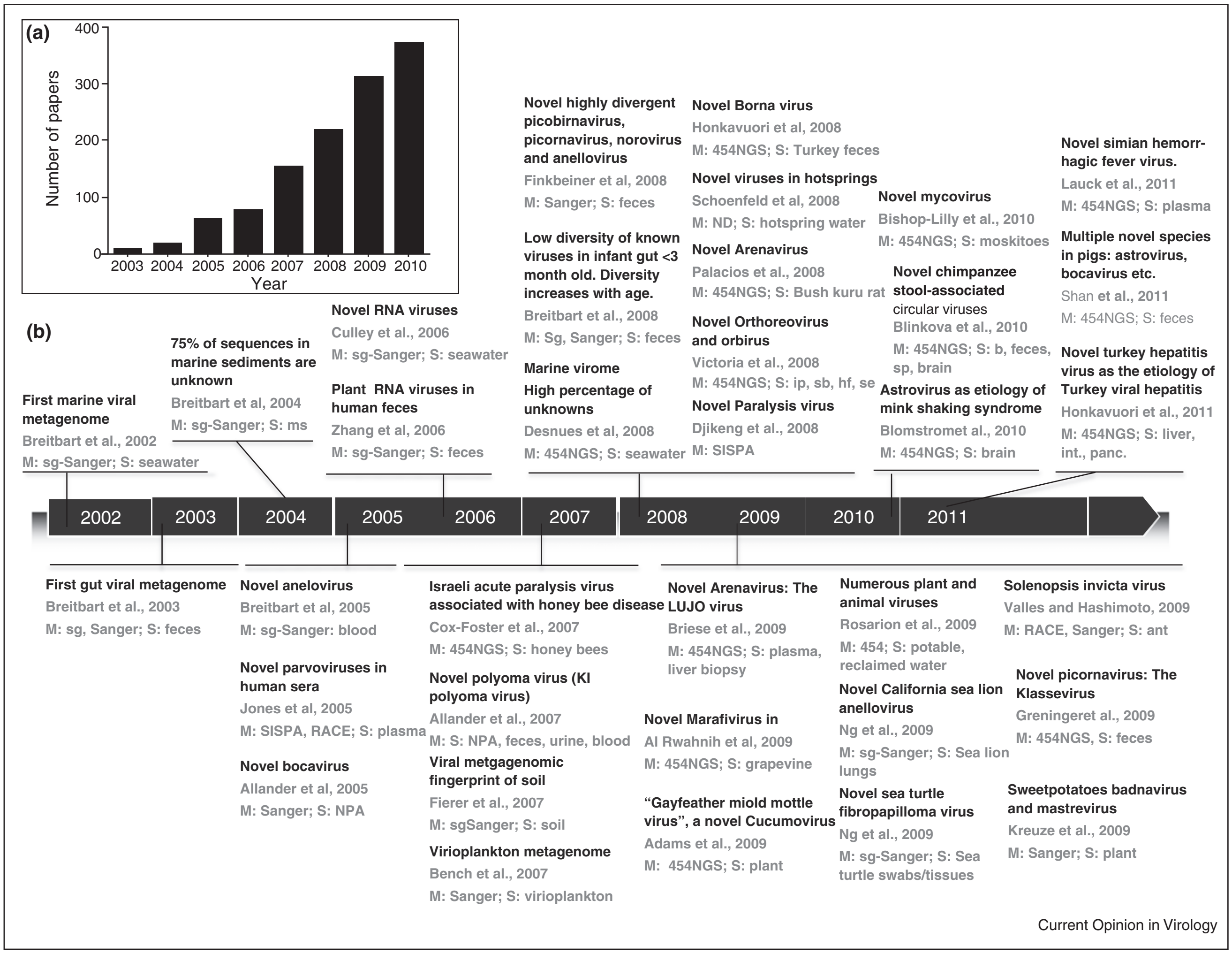
Culture techniques have been the gold standard for the detection of viruses for over a century. Despite the knowledge gained using the cultivation of viruses in cell culture, the consensus is that we have barely begun to chart the viral world, which is the 'dark matter' of the biological universe and a rich source of future discoveries [3]. Since the vast majority of viruses are not easily cultivatable, exploration of this dark matter requires culture-independent methods with larger detection coverage than culture.

While the sequencing of the 16S fragment of the small subunit of the ribosomal RNA (rRNA) gene has a proven track record for the detection of known and novel cellular organisms [4–10], this technique is not applicable to viruses because they lack the gene. Indeed, viruses do not share any common gene that could similarly qualify as a unified phylogenetic marker [11].

Metagenomics is an alternative culture-independent and sequence-independent approach that does not rely on the presence of any particular gene in all the subject entities. This approach was originally developed as a tool for 'functional and sequence-based analysis of collective microbial genomes contained in environmental samples' [12,13]. Early metagenomic studies analyzing the genetic content of environmental samples yielded the identification of metabolic traits, the characterization of organisms and the discovery of new antibiotics and enzymes [12–16]. Metagenomic studies now encompass a wide scope of research fields including marine environmental research, plant and agricultural biotechnology, human genetics and diagnostics of human diseases. Accordingly, the number of metagenomics papers in peer-reviewed journals has increased greatly since 2002 (Figure 1a). The scope of applications for metagenomics will likely widen from environmental microbiome studies to routine clinical diagnostics for palliative care of patients, public health, industry and beyond.

The first application of metagenomics to the field of virology was in the analysis of the viral communities sampled at two near-shore marine locations in San Diego [17••]. Since then, it has been used to survey viruses in numerous environments including freshwater, marine sediment, soil and the human gut. Figure 1b shows an overview of diverse areas where the metagenomic approach has been applied for virus discovery since 2002. The success of these studies relied upon the advances observed in the past decade in the area of sequencing technology and in bioinformatics. Although the fundamental concept of metagenomics has not changed, several technical advances have proven valuable for the discovery of previously unidentified, uncultured viruses. While metagenomics originally depended upon cloning for the analysis of double-stranded DNA genomes [17••,18,19,20•], high-throughput sequencing technologies can now be applied to all types of genomes, including single-stranded DNA and RNA [21].

Figure 1



Current Opinion in Virology

Overview of viral metagenomic studies between 2002 and 2011. **(a)** The number of published papers on metagenomics (2003–2010), as determined by Pubmed searches using the keywords ‘metagenome OR metagenomics’. **(b)** Timeline of important landmarks and virus discoveries achieved with the metagenomic approach between 2002 and 2011. The following studies were used to generate the figure: [17••,18,20•,62,66,70,71,74,84–86,88,104–123,148,161,162]. M: Main characterization method used: 454NGS, 454 high-throughput sequencing using GS FLX or GS titanium platform; sg-Sanger, shotgun library with Sanger sequencing method. S: sample, Symbols used for sample type: ip, insect pool; sb, skunk brain; int: intestine; panc: pancreas; hf: human feces; se, sewer effluent; ms: marine sediment; nasopharyngeal aspirates (NPA).

Historically, diseases caused by viruses have been known before the discovery of their causative agents. The acquired immunodeficiency syndrome (AIDS), poliomyelitis, cervical cancers, and Burkitt's lymphoma were identified before their causative agents. Whereas poliomyelitis was documented in ancient Egyptian literature as early as approximately 3700 BC [22], poliomyelitis virus was not discovered by Landsteiner and Popper until 1909 [23]. Descriptions of clinical conditions likely to have smallpox have been found in ancient literature from Egypt (1100–1580 BC), China (1122 BC) and India (1500 BC)—long before both Jenner's discovery of smallpox vaccination and the later isolation of variola virus [24–26].

The future perspectives in virology appear that, the metagenomic approach will generate a plethora of genetic information from unknown and potentially infectious agents, some of which could be associated with human diseases. The discovery of viruses will start to precede the characterization of the diseases they cause, well before the pathogenicity of these agents is defined.

At this turning point in history, important questions need to be answered. For example, how far has this new viral metagenomics discipline evolved in its first decade? What has been learned so far that can be applied to viral discovery and the forecasting of future viral outbreaks? In this article, we review virus discovery techniques with a focus on metagenomic approaches that employ high-throughput sequencing technologies to characterize novel viruses.

Traditional techniques for virus discovery

Before the advent of molecular methods, many techniques including filtration, tissue culture, electron microscopy (EM), serology and vaccination have been used for the detection of viruses. In 1892, Ivanovski demonstrated the presence of infectious agents, coined 'virus' by Beijerinck in 1898, in filtrate of infected leaves passed through a Chamberland filter. This marks the discovery of the tobacco mosaic virus [27] and the birth of a new era in virology. Until then, the field of virology was not clearly defined. The instrumentation, from the discovery of tissue culture to modern molecular biology methods, has shaped the field and helped to discover many viruses. Since the invention of the technique of tissue culture in 1907 and the propagation of poliovirus in animal cells in 1909, cultivation of viruses has remained the gold standard for virus discovery for over a century [28–30]. Despite the achievements made by the culture technique, several limitations have hindered the discovery and detection of viruses in routine laboratory settings. Virus propagation requires the development of controlled conditions that mimic the natural ecosystem shared between viruses and their hosts [31•].

The invention of the electron microscope in 1933 provided the first visual proof of a virus. However, this technique is relatively expensive, tedious and lacks both sensitivity and specificity. Alternatively, serology can provide a hint of the acquisition of novel viruses — as was the case for hepatitis C virus [32,33] — before the viral agents have been cultured or viewed by electron microscopy. The immune sera method has shown little value for virus discovery. The inoculation method, however, not only helped to identify novel viruses, but also was used as an immunization method to confer cross-protection against closely related viruses. Indeed, the cowpox-based inoculation developed by Jenner in 1796 was the first effective vaccine against an infectious disease. Nearly two centuries later, this strategy was used to eradicate smallpox. However, it is unlikely that Jenner's method would pass the scrutiny of modern ethical review boards for vaccine or virus discovery [34].

Molecular methods for virus discovery

The trends in clinical virology practices show gradual substitution of the traditional virus discovery methods with novel molecular biology technology. Nevertheless, traditional and the newer molecular biology techniques to isolate, identify, and characterize viruses play complementary roles in the viral discovery effort. For a comprehensive list and detailed description of molecular methods used for virus discovery, readers are referred to reviews by Delwart [31•] and Tang [35•]. Here, we focus on the viruses discovered using these methods and their future applications in clinical microbiology and public health settings.

Two types of molecular methods have been used for the virus discovery effort: sequence-dependent and sequence-independent methods.

Sequence-dependent methods, including PCR using consensus primers and hybridization methods such as microarrays, require the knowledge of the nucleic acid for the detection of novel viruses. Indeed, consensus sequences of previously known viruses have been used to identify novel viruses including highly divergent clades of human immunodeficiency virus [36], simian retroviruses [37–40], and hepatitis E virus [41]. However, PCR using consensus primers based on previously characterized viruses have little or no value in detecting completely novel viruses. The microarray techniques were first introduced in 1995 to monitor the expression of multiple genes simultaneously [42]. For virus discovery, microarrays can be prepared with probes that hybridize known viral sequences and potentially novel viruses with sufficient sequence similarity. The method has been applied to detect a wide range of known viruses as well as novel highly divergent viral taxa [43]. Microarray screening has led to the identification and characterization of a novel gammaretrovirus, xenotropic murine leukemia virus-related virus (XMRV), in prostate tumors [43,44]. Subsequent studies did not confirm these initial findings

[45,46], which points to potential limitations of the method. Another example of a well-known virus discovered with microarrays is SARS-CoV, a highly divergent coronavirus discovered amid a worldwide outbreak of the severe acute respiratory syndrome (SARS) in 2003 [43]. Reproducibility of results between microarray tests is frequently poor [47].

Unlike PCR and microarrays, the sequence-independent viral metagenomic approaches do not rely on prior knowledge of viruses in the samples. The suppression subtractive hybridization (SSH) and representational difference analysis (RDA) are examples of sequence-independent virus discovery methods. SSH was used first to study gene expression [48] and was later applied to investigate the etiology of diseases of unknown origin [49]. By hybridizing DNA obtained from patients and control subjects, nucleic acid from an unknown pathogen(s) can be detected [49–51]. Use of RDA led to the discovery of human herpes simplex virus type 8 (HHV8) [52], Torque Teno virus (TTV) [53], GBV-A, GBV-B viruses [54] and a novel highly divergent murine norovirus [55]. This method lacks sufficient sensitivity to detect viruses when the viral burden is low or when the DNA sequence of the suspected etiological agent is not clearly distinguishable from the control sample [31].

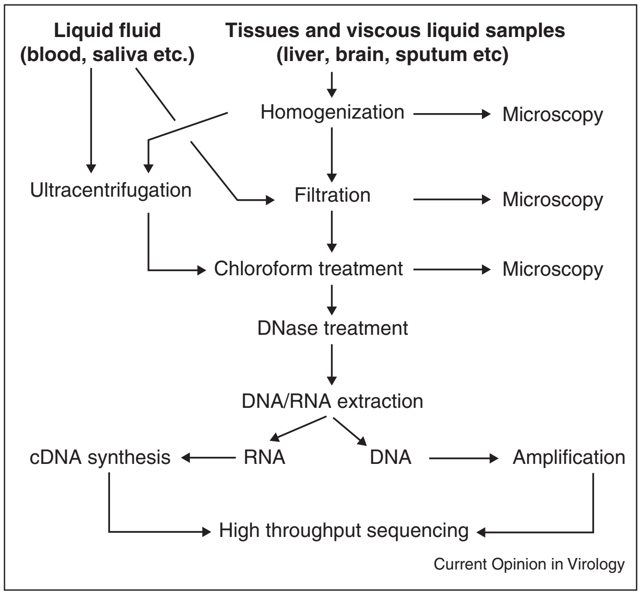
Sequence-independent single-primer amplification (SISPA) circumvents the viral load limitation of SSH. Although there are several variations to the original protocol published by Reyes *et al.* [56], the main strategy of SISPA is to exploit the sensitivity and the specificity of PCR amplification using primers that bind oligonucleotide fragments ligated to any putative viral DNA materials in the sample. SISPA has been modified to allow the detection of both DNA and RNA viruses after the removal of genomic and contaminating nucleic acids [57]. The SISPA method was used successfully for the discovery of Hepatitis E virus [58,59], Norwalk virus [60], Human astrovirus [61,62], and Parvoviruses 2 and 3 [63]. Another sequence-independent technique, the viral metagenomics (described in detail below), provides superior capability to detect known and unknown viruses than the traditional and molecular sequence-dependent and sequence-independent methods.

Viral metagenomics

Compared to virus discovery approaches outlined above, viral metagenomics is less biased. Potentially, any viruses in the samples, culturable or unculturable, known or novel can be readily detected with the viral metagenomic approach.

Viral metagenomic methods have evolved significantly since they were first developed. In early studies [17••,18,19,20•], preliminary sample preparation involved shearing of DNA and cloning. These steps were required in order to obtain sufficient DNA given the low amount of viral DNA in environmental samples (~10 μg/100 l of sea water). Because viral DNA often contains modified nucleotides and because some viral genes (e.g. holins and lysozymes) are toxic to cells, the DNA was randomly sheared to produce small fragments before cloning [17••,18,19,20•]. The process of sample preparation has since been streamlined and the sequencing speed increased with the advent of high-throughput sequencing technologies. The replacement of cloning with high-throughput methods has revolutionized metagenomics.

Figure 2



Flow chart for the generation of a viral metagenome using high-throughput sequencing.

There are several high-throughput sequencing platforms commercially available that vary by the sequencing principle, the sequencing speed, the cost and read length. An overview of a typical viral metagenomic protocol that can be used in a virus discovery study is provided in Figure 2. Essentially, a metagenomic analysis involves three main steps: (1) sample preparation, (2) high-throughput sequencing and (3) bioinformatic analysis. Below we provide an outline of each of these steps. More detailed descriptions have been previously published [64••].

Sample preparation. Theoretically, any type of sample can be analyzed using the metagenomic approach, including seawater [65], blood [66], horse feces [67], stool [20•,68–71], marine sediments [18], coral tissues [72,73], and hot springs [74]. Because viral genomes are relatively short, bacterial or eukaryotic nucleic acids can severely interfere with the isolation and detection of viral DNA or RNA that typically represents only a small fraction. Thus, removal of non-viral nucleic acid is necessary [64••,75]. Homogenization, filtration and ultracentrifugation are often necessary to concentrate the viral particles present in the sample (Figure 2). To ensure that viruses are not lost during the virus preparation, epifluorescence microscopy with SYBR-gold staining is used on aliquots of samples obtained after the homogenization, filtration, and chloroform treatments to monitor the presence of VLPs [64••].

Chloroform treatment followed by DNase digestion is used to remove contaminating DNA. The chloroform disrupts mitochondrial, bacterial and eukaryotic membranes, thereby exposing non-viral DNA to the subsequent nuclease treatment [76,77]. Unfortunately, chloroform treatment may also cause enveloped viruses to lose their protective lipid membrane, thereby rendering their DNA subject to DNase digestion [66]. Moreover, DNase treatment does not always completely eliminate non-viral DNA in the sample [63,64••]. After extraction, DNA may need to be amplified with random primers [78,79]. The Whole Transcriptome Amplification (WTA) kit can be used for the synthesis of cDNA from viral RNA [80].

Single virus genomics (SVG) was introduced by Allen and collaborators to selectively isolate viruses before sequencing [81]. SVG uses flow cytometry to sort viruses based on a method originally described by Brussard *et al.* [82]. Following the sorting, DNA of different sizes is immobilized in agarose gel, and then amplified using the multiple displacement amplification (MDA) method. The SVG approach can also be applied to RNA viruses provided a reverse-transcription step is inserted between the flow cytometry and MDA.

High-throughput sequencing. Early metagenomic applications involved the generation of shotgun libraries and direct sequencing of the total DNA content using the Sanger enzymatic dideoxy-sequencing method. This approach permitted the discovery of novel phages in marine environments [61,66]. The Sanger technique had been the standard method for sequencing since it was first described in 1977 [83]. Development of the 'next-generation' sequencing platforms offered the combined advantages of speed, automation and high-throughput, thereby increased sequencing capabilities by a factor of 100 to a million relative to the Sanger technology.

The Illumina/Solexa and Roche 454 next-generation sequencing platforms have been used most often in virus discovery (Figure 1). The Illumina/Solexa method is based on sequencing-by-synthesis chemistry using fragments of the sample DNA ligated to oligonucleotide adapters. The adapters on a solid support act as primers for DNA polymerase to incorporate reversible terminator nucleotides, each labeled with a different fluorescent dye. A typical sequencing run can generate up to 18 gigabases of data with an average read length of 75–100 nucleotides [21]. The Sweetpotato badnavirus and the Sweetpotato mastrevirus are examples of viruses discovered using the Illumina/Solexa sequencing platform [84].

The 454 FLX titanium pyrosequencer commercialized by Roche has been the most used for the discovery and characterization of novel viruses (http://www.454.com/publications-and-resources/publications.asp?postback=true). This platform was used for the identification of an uncharacterized mycovirus [85], Solenopsis invicta virus 3 [86], Merino Walk virus and a new arenavirus [87,88], among others (Figure 1b). For sequencing, DNA is fragmented and ligated to biotinylated specific linkers. The complex DNA/linkers fragment is attached to streptavidin-coated beads that anchor the DNA inside a droplet of water and PCR reagents in oil emulsion. Each fragment is first amplified to produce the template for sequencing reaction. Sequencing is carried out by annealing primers to the linker portion of the template complex, followed by the incorporation of nucleotides by DNA polymerase, which facilitates the extension of the complementary DNA. The pyrophosphate released by this process is measurable by the production of light [89,90]. The Roche 454 system measures the pyrophosphate released as the result of nucleotide incorporation during DNA synthesis mediated by DNA polymerase. The amount of light released is proportional to the intensity of the light signal captured by a charge-coupled device (CCD) camera, which then converts light signals into digital data [91,92]. A typical optimum run using a 454 pyrosequencer yields about one million reads with an average length of 350–450 nucleotides, totaling about 0.4 gigabases.

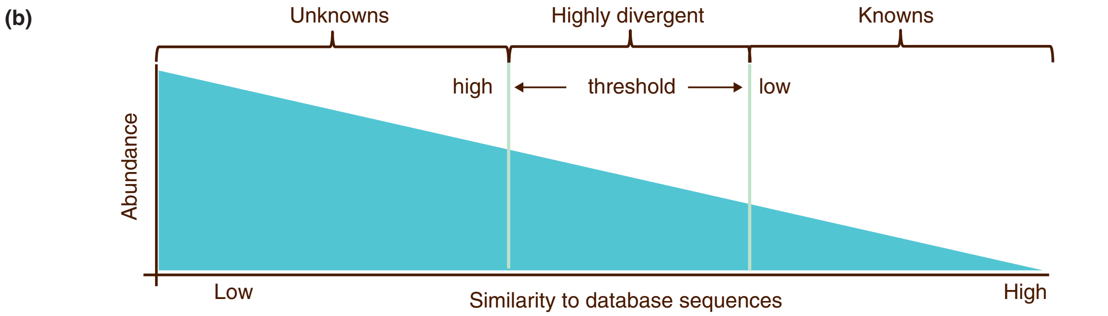
Bioinformatic analyses. The analysis of the copious data generated by high-throughput sequencing is the most challenging aspect of metagenomics. An inherent difficulty in assigning taxonomic designations to viral sequences is that there is no universally homologous nucleic acid component present in all viruses that can be used to build phylogenetic trees — a factor that also fuels the debate over whether or not viruses belong in the tree of life [11,93–96]. In most metagenomic studies, sequences generated by high-throughput sequencing are queried by homology search tools to previously documented sequences stored either in a local database or in public databases such as the Genbank. Unfortunately, homology searches against known sequences in Genbank cannot characterize unknown viruses (Figure 3).

The analysis of metagenomic libraries requires fast computation and the right algorithms to characterize sequences as belonging to putative viruses. To ensure that bioinformatic analyses are performed only on high quality data, the reads are typically processed through a software pipeline to remove any background sequences including host and bacterial DNA that had not been

Figure 3

(a)

Reference	Sample type	Method	Cut off	Unknown
Breitbart et al., 2002	Seawater	TBLASTX	E value =0.001	65%
Breitbart et al., 2003	Human feces	TBLASTX	E value = 0.001	59%
Breitbart et al., 2004	Marine sediment	TBLASTX	E value = 0.001	75%
Zhang et al., 2006	Human feces	TBLASTX	E value = 0.001	91%
Benchet al., 2007	Virioplankton	TBLASTX	E value = 0.00	61%
Breitbart et al., 2008	Infant feces	TBLASTX	E value = 0.001	66%
Rosarion et al, 2009	Potable water	TBLASTX	E value = 0.001	44%
Rosarion et al, 2009	Reclaimed water (DNA)	TBLASTX	E value = 0.001	70%
Rosarion et al, 2009	Reclaimed water (RNA)	TBLASTX	E value = 0.001	57%
Desnues et al., 2008	Stromatolite	TBLASTX	E value = 0.01	99%



Current Opinion in Virology

The unknowns: sequences with no detectable homologs in Genbank. **(a)** Proportion of the unknowns reported in viral metagenomic studies of diverse environments. **(b)** Diagram illustrating the abundance of unknown and known sequences in the environment. The distinction between known and unknown depends on the thresholds used.

removed by the filtration, chloroform, and DNase I treatments [97–99]. The resulting sequence reads are assembled with strict parameters to generate contigs, each made of sequences derived from the same organism quasi-species. Using a stringent assembly parameter is critical to avoid sequence chimerization. The contigs sequences are then compared to the Genbank non-redundant nucleotide database using BLAST [100] or USEARCH [101]. Note that using a database containing only viral sequences will not be able to identify bacterial, archaeal or eukaryotic sequences and lead to an overestimation of the fraction of unknowns (see below).

With the increasing number of data generated from different studies, there is a need for a cross-metagenome meta-analysis [102,103]. This is particularly important because of the diversity of different viral metagenomic protocols and the lack of standard algorithm for downstream data analysis. The following items should be included in any report on viral metagenomic studies: firstly, the sequencing platform and its version number; secondly, raw sequence data accession numbers in a public database; thirdly, details about the bioinformatic analysis, including the homology search tool and the database being used to assign the taxonomy, and their versions; fourthly, a list of known and previously unknown viruses found, clearly showing if the 'novel' viruses are new strains of a previously described species or completely different viruses; and fifthly, causality evidence if any.

The challenge with the unknown sequences

The most intriguing aspect of viral metagenomics is the fact that a large number — usually the majority — of sequences has no significant similarity to anything known. In this review, we refer to these sequences as the 'unknown' (Figure 3a). A typical human or environmental viral metagenome can contain between 60% and 99% unknown sequences (Figure 3) [17••,18,20•,62,66,70,71,74,84–86,88,104–123]. Factors contributing to this variation include the sample type, the length of the sequence reads, the homology search method (BLASTn, tBLASTx, etc.), the similarity threshold (*E*-value cutoff), the database and version of that database used for the homology searches (Figure 3b). Depending on how they are viewed, the unknowns can represent either a formidable challenge or a treasure trove for virus discovery. Although researchers often tend to consider the unknowns as 'junk,' these sequences could be a valuable blueprint for the discovery of novel viruses [112,124,125]. Thus far, there is a lack of suitable bioinformatic methods to characterize the unknown sequences.

A tentative solution is to compare the sequences between samples in order to at least gain some insight about the viral entities that are shared between them. A program such as PHACCS (PHAge Communities from Contig Spectra) can be used to assess the biodiversity of uncultured viral communities by mathematically modeling the community structure using the contig spectrum of metagenome assemblies [126]. This method was extended to assess cross-assemblies of reads from different samples [65], providing a homology-independent tool for the comparison of metagenomes with a high proportion of unknown sequences. Although PHACCS may provide a glimpse of the composition and difference between metagenomes, it has limited value for the characterization of novel viruses. Two tools can be used to predict whether unknown sequences are from bacteriophages undergoing lytic and lysogenic lifestyles. One such tool described by Deschavanne *et al.* [127] compares the genome signatures of query sequences against those of their host genome in order to identify host–phage relationship and information about the phage lifestyle. The second method, PHACTS, depends on residual homology between the putative unknown sequence and sets of randomly selected viral proteins from known viruses (K McNair *et al.*, *PHACTS: a computational approach to classifying the lifestyle of phages*, unpublished data). Alternatively, viruses may be classified by basic sequence properties. For instance, the circularity of the contig, its oligonucleotide profile [128], and the open reading frame (ORF) structure (S Akhter *et al.*, *PhiSpy: A novel algorithm for finding prophages in bacterial genomes that combines similarity-based and composition-based strategies*, under review) may all provide clues whether the unknown sequence could be from a potential novel virus. These properties can be combined into a prediction network used to classify viruses into lifestyle groups or taxonomic clades.

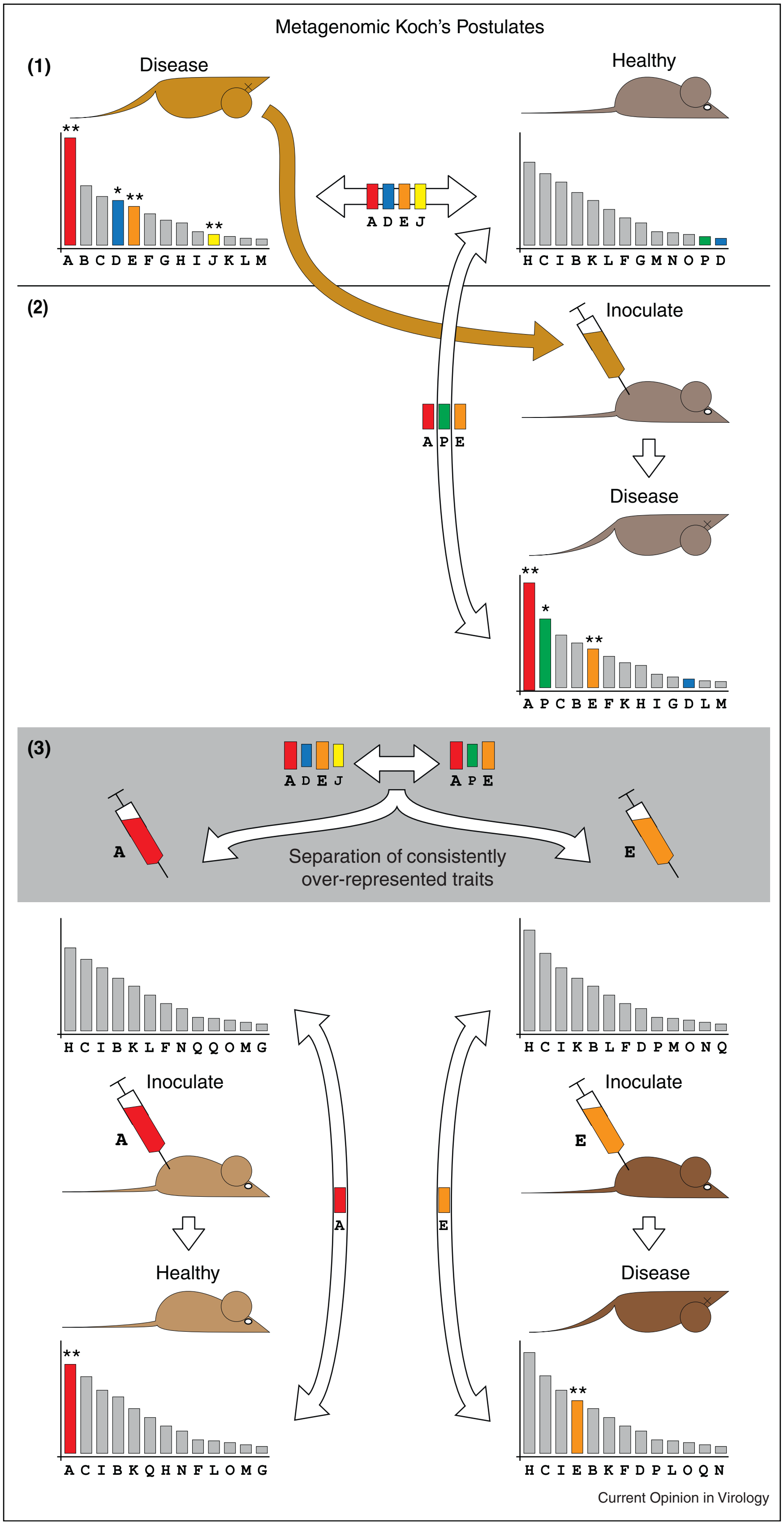
Although newly discovered viruses are often labeled 'novel,' the question remains whether these sequences represent truly novel viruses or ancient viruses that simply have never been observed before. The age of a sequence has traditionally been determined by multiple alignments of query sequences with their homologs and by calculating the divergence times from a common ancestral node on a phylogenetic tree. Dates can be estimated using either a molecular clock [129] or by assigning a calibration date to a specific node in the tree based on fossil or other evidence [130–132]. For viral metagenomic sequences, however, building a phylogenetic tree is itself problematic because often the sequenced reads may represent non-overlapping subregions of an unknown viral genome. Moreover, there is no fossil data available to calibrate the age of nodes in the tree. A promising approach might be to estimate divergence times from assembled viral contigs. *De novo* assembly allows non-overlapping regions to be combined into a single consensus sequence. For a given molecular clock, SNP analysis of the contributing reads could provide an estimation of how long ago the sequenced reads diverged. Such estimates may be critical when addressing the question of the origin of a newly identified infectious agent.

Koch's postulates, metagenomics and viral etiology of diseases

Until recently, virus discoveries were made in the context of disease etiology. Thus, virus discovery studies were biased mainly because of the use of convenient samples available from patients. Because of the difficulties involved, the investment of efforts and resources required to isolate viruses often could not be justified outside the disease context. It is likely that the context of the diseases has also led to the misconception that all viruses are pathogenic. This dogma was challenged by the discovery of viruses such as *Torque Teno virus* (TTV) and hepatitis G virus (GBV-C), originally associated with post-transfusion hepatitis [53,133–135], and then were subsequently shown be classical examples of viral commensals [136,137]. The widely accepted notion that viruses act as obligatory pathogens is beginning to give way to the concept that viruses can be part of the normal flora of the human body. Considering their high abundance in the gastrointestinal tract, on skin and even in blood and lungs [138] it is unlikely that viruses could only be pathogenic without any benefits for their hosts. The abundance of viruses, particularly phages, in the lung — an environment previously thought to be sterile — may reflect their beneficial role in keeping bacterial populations in check [138]. The pathogenicity of the GBV-C has shifted to a more radical designation as a 'good' virus in cases of co-infection with HIV. Indeed, GBV-C has been associated with a more favorable prognosis for patients with HIV infection by slowing the progression to AIDS [139,140]. Similarly, dengue virus, a known pathogen, has been shown to limit HIV-1 replication and to reduce the viral load [141]. These examples need to be taken into account when metagenomic approach is applied to virus discovery. The characterization of a novel virus can be easily achieved *in silico* with limited bioinformatics tools but the determination of causation may not always be trivial.

The causality is not always conclusive even when the suspect virus is found in the scene of the crime. In other words, finding a virus in a sample from a patient with an illness of unknown etiology and even demonstrating the association does not always prove causation. For this reason, strict guidelines proposed by Robert Koch and later modified by Rivers [142] have been used to assign

Figure 4


Metagenomic Koch's Postulates
(1)
Disease
Healthy
**
*
**
**
A B C D E F G H I J K L M
A D E J
H C I B K L F G M N O P D
(2)
Inoculate
A P E
Disease
**
*
**
A P C B E F K H I G D L M
(3)
A D E J
A P E
A
E
Separation of consistently
over-represented traits
H C I B K L F N Q Q O M G
H C I K B L F D P M O N Q
Inoculate
A
Inoculate
E
A
E
Healthy
Disease
**
**
A C I B K Q H N F L O M G
H C I E B K F D P L O Q N

Current Opinion in Virology

causality to infectious agents. One of Koch's postulates requires that the candidate etiological agent be isolated from a diseased organism and grown in pure culture. However, many viruses cannot be propagated by culture techniques [143].

New molecular biology techniques have been used for virus discovery bypassing the prerequisite of the Koch's postulates. For instance, the Merkel cell polyomavirus (MCV) was identified as the causative agent of Merkel's cell carcinoma without satisfying all of the requisites of Koch's postulates [144]. Similarly, the sea turtle tornovirus 1 was associated with fibropapillomatosis using a culture-independent metagenomic approach [118].

The methodological shift, from culture to metagenomics, will likely create a paradigm shift in the demonstration of disease causation. In many instances Koch's postulates will no longer be satisfied if culture techniques are used to prove causality. Falkow [145••] proposed the modified Koch's postulates which uses molecular methods to monitor the role played by genes in distinct bacterial virulence. To satisfy the revised molecular Koch's postulates, a strong association must be established between the phenotype or property under investigation and the pathogenic members of a genus or pathogenic strains of a species. The gene of interest should be found in all pathogenic members of the genus or species but be absent in nonpathogenic strains. At best, the nonpathogenic strains could carry the gene with critical mutations that could render the strain non-virulent. However, new molecular methods do not always distinctively characterize virulence genes and make a clear association with a disease of unknown etiology. This could be because genes can be expressed at different time-points during infection. Genes can be turned on and off and may require intrinsic factors in order to trigger the disease process.

Alternatively, we propose the metagenomic Koch's postulates, which focus on the identification of metagenomic traits in disease subjects. The metagenomic traits are molecular markers such as sequence reads, assembled contigs, genes or full-genomes that can uniquely distinguish diseased metagenomes from those obtained from matched healthy control subjects (Figure 4). The metagenomic traits found in diseased patients can be monitored in healthy individuals exposed to the suspected infectious agent. Although this novel approach requires separation or isolation of remaining co-occurring disease candidates (Figure 4.3), it does not necessarily require the isolation of the pathogen in tissue culture or pure culture media unlike the original Koch's postulates. Therefore, the genetic make-up of the agent responsible for a disease can provide early clues before its isolation by tissue culture.

The modified metagenomic Koch's postulates proposed in this paper require that: Firstly, the diseased metagenome be significantly different from the metagenome constructed with the same sample type obtained from a healthy matched control subject. The suspected metagenomic traits must be present and more abundant in the diseased subject compared to matched control (Figure 4.1). Secondly, inoculating a healthy individual with a sample from a diseased subject must result in disease state (Figure 4.2). Differential metagenomic traits in step (1) recovered in the newly induced diseased subject may be the biomarker of the candidate etiological agent; and finally, selective inoculation of samples from the disease subject (in step 2) must induce disease in another healthy control subject if the metagenomic contains the trait associated with the etiological agent of the disease, or phenotype under investigation (Figure 4.3). Assuming that the metagenomic trait 'E' (Figure 4.3) is a contig sequence from a previously unknown and unculturable virus, its early identification using the metagenomic approach could spearhead the effort to generate diagnostic assays such as ELISA and PCR, well before the isolation and the characterization of the viruses by culture techniques.

Fulfilling this metagenomic model of the Koch's postulates is possible when one or multiple viral agents are involved in disease causation. With the original Koch's postulates or the modified molecular Koch's postulates, it is difficult enough to prove causality with one suspected agent using the culturing prerequisite. The complexity is even greater when multiple viruses are involved in the causation of a disease.

A similar approach, the siRNA-ome used previously by Kreuze *et al.* [84] led to the detection of etiological viruses causing diseases in plants despite the low copy number of the suspected traits [84]. The modified metagenomic Koch's postulates could also be tested in human diseases such as the murine mink cell leukemia caused by a C-type retrovirus, named the mink cell focus-inducing virus (MCFIV) [146]. MCFIV requires the cooperative interaction with other viruses to increase its propensity to cause leukemia [146]. The Burkitt's lymphoma caused by others Epstein-Barr virus (EBV) in regions holoendemic

(Figure 4 Legend) Metagenomic Koch's postulates. Comparison between a diseased and healthy control animal shows a significant difference between the metagenomic libraries (depicted by the histograms of relative abundance reads). In order to fulfill the metagenomic Koch's postulates: (1) The metagenomic traits in diseased subject must be significantly different from healthy subject. For example traits A, D, E and J found in the disease animal that are not present in the healthy control; (2) Inoculation of samples from the disease animal into the healthy control must lead to the induction of the disease state. Comparison of the metagenomes before and after inoculation should suggest the acquisition or increase of new metagenomic traits (A, E and P). New traits can be purified by methods such as serial dilution or time-point sampling of specimens from a disease animal. (3) Inoculation of the suspected purified traits into a healthy animal will induce disease if the traits form the etiology of the disease.

for *Plasmodium falciparum*, the etiology of malaria [147]. Metagenomics could become the future method of choice enabling the simultaneous analysis of multiple agents in a sample and assessment of the association and disease causality without the limitations imposed by culture techniques [138,148,149].

Future application of metagenomics to public health

The etiology of many diseases remains unknown. These ailments are collectively defined as diseases of unknown etiology when all conventional testing laboratory techniques are unsuccessful. Yet, the diseases with unknown origin have high rates of morbidity and mortality. For example, as many as 40% of cases of the infantile diarrhea, which alone claims ~1.8 million fatalities annually, have no known specific causative agent [112]. Infantile diarrhea, the pyrexia of unknown origin, influenza-like illnesses, chronic fatigue syndrome, Alzheimer's disease, various forms of tumors such as diffuse large B-cell lymphoma and many other diseases of unknown origin can benefit directly from the metagenomic technology.

The success of metagenomics in identifying novel viruses in a wide variety of samples opens doors to new application areas particularly in public health and the prevention of infectious diseases. Although the metagenomic technology is not yet part of the routine diagnostics, results from clinical virology research provides valuable proof of concepts for a new era in clinical virology practices. For example, Finkbeiner *et al.* analyzed samples from 12 children using metagenomics and identified a large number of known eukaryotic viruses as well as sequences from putatively novel viruses [112]. Another study identified a corona-like virus, the Human Cosavirus E1 (HcoSV-E1), in a child with acute diarrhea [150]. These initial studies identified promising viral candidates to establish the etiology in these cases of diarrhea. The 2009 pandemic of influenza A (2009 H1N1) provided proof of concept in that metagenomics was effective to rapidly characterize the full genome of the flu virus [151]. Using the metagenomic approach, Palacios *et al.* discovered an arenavirus in samples which had tested negative by culture, PCR, serology and a microarray assay using oligonucleotide probes from a wide range of infectious agents [87], suggesting a potential causative agent for unexplained cases of post-transplantation death. In another study, Towner *et al.* described a new Ebola virus responsible for an outbreak of a hemorrhagic fever in the District of Bindibugyo, Uganda [152]. Rapid identification of these agents would provide the blueprint for the development of therapeutic regimen or preventive vaccine.

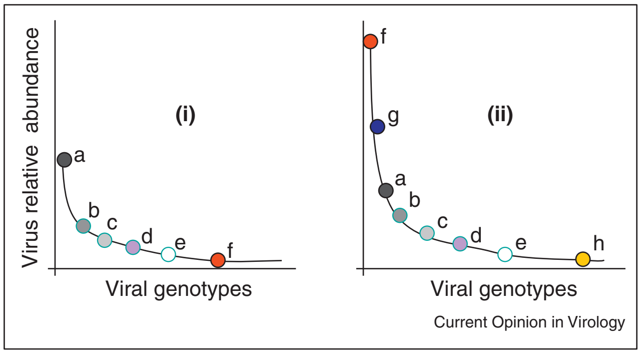
Prevention is better than cure. Potentially, a single or multiple jump of an animal virus to humans can have serious consequences. One way to prevent infectious diseases is through vaccine development. But the development of a vaccine takes time and demands a huge amount of resources. Preventing the introduction of an unknown virus to human populations is rather a far-reaching goal unless the methods of virus identification and characterization are put in place. A simple and practical strategy would be to assess the danger posed by viruses that thrive in animals and could cross to human through zoonosis.

Zoonosis is a source of up to 75% of emerging infectious diseases in humans [153]. As such, cross-species transfer from animals to humans has serious repercussions not only in public health but also in the socio-economical and political stability [68,154–158]. The detection and characterization of novel viruses are of paramount importance in the forecasting of future outbreaks of viral diseases in humans. Surveying natural reservoirs for potential zoonotic infection [69] and human populations such as bush meat hunters who are exposed to animals could help prevent major outbreaks before the wide spread of viruses to human population. Data obtained in early identification of viruses are valuable for forecasting new emerging and re-emerging viral epidemics.

The experience gained from studying marine environments and hostile mine environments can be applied in public health programs that seek to determine the normal viral population and monitor changes in different geographical settings. We have termed such an approach as Public Health Viral Metagenomics Surveillance (PHVMS). Viral metagenomics surveillance is defined as the survey of the functional and taxonomic signatures representing the viruses normally circulating within that population in the absence of noticeable epidemics. In the event of a zoonotic outbreak, these functional and taxonomic signatures of the virome will likely show detectable shifts. Figure 5 shows a hypothetical rank abundance curve for six viruses (a–f). The introduction of a highly pathogenic species (g) can be expected to result in a disruption of the normal virome, including the appearance of opportunistic viral infections (h).

Using PHACCS analysis [126], several parameters can be compared between the normal and disturbed viromes including the total number of viral species (richness) and their relative abundance (evenness). Another approach would be to determine the normal virome, a background viral metagenome to refer to in case of an outbreak. Lessons learned from studies of bacterial microbial metagenomes suggest that different environments often have different microbial signatures [159], including the functional metabolic information, the nucleotide usage, proportion of different species. Disrupting key metabolic processes of an environment can lead to disruption of the balance in that ecosystem. Similarly, the viromes in different human populations in different locations may display functional profiles characteristic of their respective environment, lifestyle and viruses

Figure 5



Monitoring of emerging infectious diseases using a metagenomic approach. A hypothetical example of the potential use of the Public Health Viral Metagenomics Surveillance (PHVMS) approach for virus discovery based on comparison of viromes sampled before (I) and during (II) an epidemic. Depicted here are the rank abundance curves for viral species (a–h), where g represents a newly introduced, highly pathogenic species and h a less virulent virus.

circulating in each region. The magnitude of disturbance of the virome profile will depend on the fitness and virulence of the newly introduced pathogens and the immune fitness of the host. The viral communities in two different metagenomes can be compared using XIPE [160]. This statistical approach was developed for comparing metagenomic sequences derived from samples collected from the Sargasso Sea and from acid mine drainage and was able to accurately predict the physiology, metabolic potential and ecology of each ecosystem [160].

Conclusion

During the last decade, we have witnessed the emergence of metagenomics as a powerful novel tool with endless areas of applications in virology. Epidemiological data suggest that novel viruses are likely to be introduced into the human population through zoonosis [153,158]. Also, the danger of intentional introduction of viruses through bioterrorism cannot be ignored. Viral metagenomics is a powerful, fast and sensitive technique available for identifying viruses including those that cannot be detected by conventional culture and sequence-dependent methods.

Acknowledgements

We are grateful to Merry Youle for helpful suggestions and editing of the manuscript. JM was supported by a grant from the UCSD Center for AIDS Research (NIAID 5 P30 AI36214) and Moores UCSD Cancer Center (NCI 5P30 CA23100). BED was supported by the Dutch Science Foundation (NWO) Veni grant (016.111.075).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Bergh O, Borsheim KY, Bratbak G, Heldal M: **High abundance of viruses found in aquatic environments**. *Nature* 1989, **340**:467-468.

2. Whitman WB, Coleman DC, Wiebe WJ: **Prokaryotes: the unseen majority**. *Proc Natl Acad Sci U S A* 1998, **95**:6578-6583.

3. Rohwer F, Youle M: **Consider something viral in your research**. *Nat Rev Microbiol* 2011, **9**:308-309.

4. Schmidt TM, DeLong EF, Pace NR: **Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing**. *J Bacteriol* 1991, **173**:4371-4378.

5. Shah N, Tang H, Doak TG, Ye Y: **Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics**. *Pac Symp Biocomput* 2011:165-176.

6. Gross EL, Leys EJ, Gasparovich SR, Firestone ND, Schwartzbaum JA, Janies DA, Asnani K, Griffen AL: **Bacterial 16S sequence analysis of severe caries in young permanent teeth**. *J Clin Microbiol* 2010, **48**:4121-4128.

7. Tringe SG, Hugenholtz P: **A renaissance for the pioneering 16S rRNA gene**. *Curr Opin Microbiol* 2008, **11**:442-446.

8. Manichanh C, Chapple CE, Frangeul L, Gloux K, Guigo R, Dore J: **A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library**. *Nucleic Acids Res* 2008, **36**:5180-5188.

9. Streit WR, Schmitz RA: **Metagenomics — the key to the uncultured microbes**. *Curr Opin Microbiol* 2004, **7**:492-498.

10. Liles MR, Manske BF, Bintrim SB, Handelsman J, Goodman RM: **A census of rRNA genes and linked genomic sequences within a soil metagenomic library**. *Appl Environ Microbiol* 2003, **69**:2684-2691.

11. Rohwer F, Edwards R: **The Phage Proteomic Tree: a genome-based taxonomy for phage**. *J Bacteriol* 2002, **184**:4529-4535.

12. Riesenfeld CS, Schloss PD, Handelsman J: **Metagenomics: genomic analysis of microbial communities**. *Annu Rev Genet* 2004, **38**:525-552.

13. Schloss PD, Handelsman J: **Biotechnological prospects from metagenomics**. *Curr Opin Biotechnol* 2003, **14**:303-310.

14. Krause DO, Denman SE, Mackie RI, Morrison M, Rae AL, Attwood GT, McSweeney CS: **Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics**. *FEMS Microbiol Rev* 2003, **27**:663-693.

15. Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C *et al.*: **Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms**. *Appl Environ Microbiol* 2000, **66**:2541-2547.

16. Brady SF, Clardy J: **Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water**. *J Nat Prod* 2004, **67**:1283-1286.

17. •• Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F: **Genomic analysis of uncultured marine viral communities**. *Proc Natl Acad Sci U S A* 2002, **99**:14250-14255.

First proof-of-concept paper applying the metagenomic on marine viruses. This is regarded as the first viral metagenomic study that described the high abundance and high diversity of marine viruses and documentation of a high number of sequences with no homologs in existing databases including Genbank.

18. Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P, Rohwer F: **Diversity and population structure of a near-shore marine-sediment viral community**. *Proc Biol Sci* 2004, **271**:565-574.

19. Breitbart M, Wegley L, Leeds S, Schoenfeld T, Rohwer F: **Phage community dynamics in hot springs**. *Appl Environ Microbiol* 2004, **70**:1633-1640.

20. • Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, Rohwer F: **Metagenomic analyses of an uncultured viral community from human feces**. *J Bacteriol* 2003, **185**:6220-6223.

First application of the metagenomic approach to study viral communities in human feces, spearheading the virus discovery effort on clinical samples. Most sequences were unrelated to previously documented taxa in Genbank.

21. Ansorge WJ: **Next-generation DNA sequencing techniques**. *N Biotechnol* 2009, **25**:195-203.

22. Paul JR: **A history of poliomyelitis**. *Yale Studies in the History of Science and Medicine*. Yale University Press; 1971.

23. Kristensson K: **The discovery of the poliovirus**. *Brain Res Bull* 1999, **50**:461.

24. Dixon CW: *Smallpox*. London: Churchill; 1962.

25. Fenner F, Henderson DA, Arita I, Jezek Z, Ladnyi ID: In *Smallpox and Its Eradication.* Edited by Fenner F. WHO; 1988.

26. Hopkins DR: *The Greatest Killer — Smallpox in History*. Chicago University of Chicago Press; 1983.

27. Lecoq H: **Discovery of the first virus, the tobacco mosaic virus: 1892 or 1898?** *C R Acad Sci III* 2001, **324**:929-933.

28. Hamza IA, Jurzik L, Uberla K, Wilhelm M: **Methods to detect infectious human enteric viruses in environmental water samples**. *Int J Hyg Environ Health* 2011, **214**:424-436.

29. Leland DS, Ginocchio CC: **Role of cell culture for virus detection in the age of technology**. *Clin Microbiol Rev* 2007, **20**:49-78.

30. Spector SA, Dankner WM: **Rapid viral diagnostic techniques**. *Adv Pediatr Infect Dis* 1986, **1**:37-59.

31. Delwart EL: **Viral metagenomics**. *Rev Med Virol* 2007,
• **17**:115-131.
A very comprehensive description of metagenomic methods and important benchmarks achieved in the virus discovery effort.

32. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M: **Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome**. *Science* 1989, **244**:359-362.

33. Kuo G, Choo QL, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, Miyamura T, Dienstag JL, Alter MJ, Stevens CE *et al.*: **An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis**. *Science* 1989, **244**:362-364.

34. Davies H: **Ethical reflections on Edward Jenner's experimental treatment**. *J Med Ethics* 2007, **33**:174-176.

35. Tang P, Chiu C: **Metagenomics for the discovery of novel**
• **human viruses**. *Future Microbiol* 2010, **5**:177-189.
A comprehensive review describing metagenomic methods and important benchmarks achieved.

36. Mokili JL, Rogers M, Carr JK, Simmonds P, Bopopi JM, Foley BT, Korber BT, Birx DL, McCutchan FE: **Identification of a novel clade of human immunodeficiency virus type 1 in Democratic Republic of Congo**. *AIDS Res Hum Retroviruses* 2002, **18**:817-823.

37. Takemura T, Ekwalanga M, Bikandou B, Ido E, Yamaguchi-Kabata Y, Ohkura S, Harada H, Takehisa J, Ichimura H, Parra HJ *et al.*: **A novel simian immunodeficiency virus from black mangabey (*Lophocebus aterrimus*) in the Democratic Republic of Congo**. *J Gen Virol* 2005, **86**:1967-1971.

38. Barlow KL, Ajao AO, Clewley JP: **Characterization of a novel simian immunodeficiency virus (SIVmonNG1) genome sequence from a mona monkey (*Cercopithecus mona*)**. *J Virol* 2003, **77**:6879-6888.

39. Osterhaus AD, Pedersen N, van Amerongen G, Frankenhuis MT, Marthas M, Reay E, Rose TM, Pamungkas J, Bosch ML: **Isolation and partial characterization of a lentivirus from talapoin monkeys (*Myopithecus talapoin*)**. *Virology* 1999, **260**:116-124.

40. Clewley JP, Lewis JC, Brown DW, Gadsby EL: **A novel simian immunodeficiency virus (SIVdrl) pol sequence from the drill monkey, *Mandrillus leucophaeus***. *J Virol* 1998, **72**:10305-10309.

41. Reyes GR, Purdy MA, Kim JP, Luk KC, Young LM, Fry KE, Bradley DW: **Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis**. *Science* 1990, **247**:1335-1339.

42. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray**. *Science* 1995, **270**:467-470.

43. Wang D, Urisman A, Liu YT, Springer M, Ksiazek TG, Erdman DD, Mardis ER, Hickenbotham M, Magrini V, Eldred J *et al.*: **Viral discovery and sequence recovery using DNA microarrays**. *PLoS Biol* 2003, **1**:E2.

44. Urisman A, Molinaro RJ, Fischer N, Plummer SJ, Casey G, Klein EA, Malathi K, Magi-Galluzzi C, Tubbs RR, Ganem D *et al.*: **Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant**. *PLoS Pathog* 2006, **2**:e25.

45. Farley SJ: **Prostate cancer: XMRV — contaminant, not cause?** *Nat Rev Urol* 2011, **8**:409.

46. Switzer WM, Jia H, Zheng H, Tang S, Heneine W: **No association of xenotropic murine leukemia virus-related viruses with prostate cancer**. *PLoS ONE* 2011, **6**:e19065.

47. Draghici S, Khatri P, Eklund AC, Szallasi Z: **Reliability and reproducibility issues in DNA microarray measurements**. *Trends Genet* 2006, **22**:101-109.

48. Morales P, Thurston CF: **Efficient isolation of genes differentially expressed on cellulose by suppression subtractive hybridization in *Agaricus bisporus***. *Mycol Res* 2003, **107**:401-407.

49. Ambrose HE, Clewley JP: **Virus discovery by sequence-independent genome amplification**. *Rev Med Virol* 2006, **16**:365-383.

50. Dios S, Poisa-Beiro L, Figueras A, Novoa B: **Suppression subtraction hybridization (SSH) and macroarray techniques reveal differential gene expression profiles in brain of sea bream infected with nodavirus**. *Mol Immunol* 2007, **44**:2195-2204.

51. Diatchenko L, Lukyanov S, Lau YF, Siebert PD: **Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes**. *Methods Enzymol* 1999, **303**:349-380.

52. Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS: **Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma**. *Science* 1994, **266**:1865-1869.

53. Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M: **A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology**. *Biochem Biophys Res Commun* 1997, **241**:92-97.

54. Simons JN, Pilot-Matias TJ, Leary TP, Dawson GJ, Desai SM, Schlauder GG, Muerhoff AS, Erker JC, Buijk SL, Chalmers ML *et al.*: **Identification of two flavivirus-like genomes in the GB hepatitis agent**. *Proc Natl Acad Sci U S A* 1995, **92**:3401-3405.

55. Karst SM, Wobus CE, Lay M, Davidson J, Virgin HWt: **STAT1-dependent innate immunity to a Norwalk-like virus**. *Science* 2003, **299**:1575-1578.

56. Reyes GR, Kim JP: **Sequence-independent, single-primer amplification (SISPA) of complex DNA populations**. *Mol Cell Probes* 1991, **5**:473-481.

57. Bexfield N, Kellam P: **Metagenomics and the molecular identification of novel viruses**. *Vet J* 2010, **190**:191-198.

58. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI: **Viruses in the faecal microbiota of monozygotic twins and their mothers**. *Nature* 2010, **466**:334-338.

59. Reyes GR, Yarbough PO, Tam AW, Purdy MA, Huang CC, Kim JS, Bradley DW, Fry KE: **Hepatitis E virus (HEV): the novel agent responsible for enterically transmitted non-A, non-B hepatitis**. *Gastroenterol Jpn* 1991, **26(Suppl 3)**:142-147.

60. Matsui SM, Kim JP, Greenberg HB, Su W, Sun Q, Johnson PC, DuPont HL, Oshiro LS, Reyes GR: **The isolation and characterization of a Norwalk virus-specific cDNA**. *J Clin Invest* 1991, **87**:1456-1461.

61. Finkbeiner SR, Li Y, Ruone S, Conrardy C, Gregoricus N, Toney D, Virgin HW, Anderson LJ, Vinje J, Wang D *et al.*: **Identification of a novel astrovirus (astrovirus VA1) associated with an outbreak of acute gastroenteritis**. *J Virol* 2009, **83**:10836-10839.

62. Blomstrom AL, Widen F, Hammer AS, Belak S, Berg M: **Detection of a novel astrovirus in brain tissue of mink suffering from shaking mink syndrome by use of viral metagenomics**. *J Clin Microbiol* 2010, **48**:4392-4396.

63. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J: **A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species**. *Proc Natl Acad Sci U S A* 2001, **98**:11609-11614.

64. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F:
•• **Laboratory procedures to generate viral metagenomes**. *Nat Protoc* 2009, **4**:470-483.
An excellent compilation of standard operating procedures to perform metagenomic analysis on different types of samples.

65. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H *et al.*: **The marine viromes of four oceanic regions**. *PLoS Biol* 2006, **4**:e368.

66. Breitbart M, Rohwer F: **Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing**. *Biotechniques* 2005, **39**:729-736.

67. Cann AJ, Fandrich SE, Heaphy S: **Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes**. *Virus Genes* 2005, **30**:151-156.

68. Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, Shaukat S, Masroor MA, Wilson ML, Ndjango JB, Peeters M *et al.*: **Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces**. *J Virol* 2009, **84**:1674-1682.

69. Li L, Victoria JG, Wang C, Jones M, Fellers GM, Kunz TH, Delwart E: **Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses**. *J Virol* 2010, **84**:6955-6965.

70. Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, Mahaffy JM, Mueller J, Nulton J, Rayhawk S *et al.*: **Viral diversity and dynamics in an infant gut**. *Res Microbiol* 2008, **159**:367-373.

71. Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SW, Hibberd ML, Liu ET, Rohwer F, Ruan Y: **RNA viral community in human feces: prevalence of plant pathogenic viruses**. *PLoS Biol* 2006, **4**:e3.

72. Marhaver KL, Edwards RA, Rohwer F: **Viral communities associated with healthy and bleaching corals**. *Environ Microbiol* 2008, **10**:2277-2286.

73. Vega Thurber R, Willner-Hall D, Rodriguez-Mueller B, Desnues C, Edwards RA, Angly F, Dinsdale E, Kelly L, Rohwer F: **Metagenomic analysis of stressed coral holobionts**. *Environ Microbiol* 2009, **11**:2148-2163.

74. Schoenfeld T, Patterson M, Richardson PM, Wommack KE, Young M, Mead D: **Assembly of viral metagenomes from yellowstone hot springs**. *Appl Environ Microbiol* 2008, **74**:4164-4174.

75. Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC Jr, Rohwer F: **Using pyrosequencing to shed light on deep mine microbial ecology**. *BMC Genomics* 2006, **7**:57.

76. Willner D, Furlan M, Schmieder R, Grasis JA, Pride DT, Relman DA, Angly FE, McDole T, Mariella RP Jr, Rohwer F *et al.*: **Microbes and health sackler colloquium: metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity**. *Proc Natl Acad Sci U S A* 2010.

77. Lee S, Hallam SJ: **Extraction of high molecular weight genomic DNA from soils and sediments**. *J Vis Exp* 2009:1569.

78. Dean FB, Nelson JR, Giesler TL, Lasken RS: **Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification**. *Genome Res* 2001, **11**:1095-1099.

79. Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, Leamon JH: **Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing**. *BMC Genomics* 2006, **7**:216.

80. Tomlins SA, Mehra R, Rhodes DR, Shah RB, Rubin MA, Bruening E, Makarov V, Chinnaiyan AM: **Whole transcriptome amplification for gene expression profiling and development of molecular archives**. *Neoplasia* 2006, **8**:153-162.

81. Allen LZ, Ishoey T, Novotny MA, McLean JS, Lasken RS, Williamson SJ: **Single virus genomics: a new tool for virus discovery**. *PLoS ONE* 2011, **6**:e17722.

82. Brussaard CP, Marie D, Bratbak G: **Flow cytometric detection of viruses**. *J Virol Methods* 2000, **85**:175-182.

83. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors**. *Proc Natl Acad Sci U S A* 1977, **74**:5463-5467.

84. Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I, 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* 2009, **388**:1-7.

85. Bishop-Lilly KA, Turell MJ, Willner KM, Butani A, Nolan NM, Lentz SM, Akmal A, Mateczun A, Brahmbhatt TN, Sozhamannan S *et al.*: **Arbovirus detection in insect vectors by rapid, high-throughput pyrosequencing**. *PLoS Negl Trop Dis* 2010, **4**:e878.

86. Valles SM, Hashimoto Y: **Isolation and characterization of *Solenopsis invicta* virus 3, a new positive-strand RNA virus infecting the red imported fire ant, *Solenopsis invicta***. *Virology* 2009, **388**:354-361.

87. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan PL, Hui J, Marshall J *et al.*: **A new arenavirus in a cluster of fatal transplant-associated diseases**. *N Engl J Med* 2008, **358**:991-998.

88. Palacios G, Savji N, Hui J, Travassos da Rosa A, Popov V, Briese T, Tesh R, Lipkin WI: **Genomic and phylogenetic characterization of Merino Walk virus, a novel arenavirus isolated in South Africa**. *J Gen Virol* 2008, **91**:1315-1324.

89. Meyer M, Stenzel U, Hofreiter M: **Parallel tagged sequencing on the 454 platform**. *Nat Protoc* 2008, **3**:267-278.

90. Meyer M, Stenzel U, Myles S, Prufer K, Hofreiter M: **Targeted high-throughput sequencing of tagged nucleic acid samples**. *Nucleic Acids Res* 2007, **35**:e97.

91. Nyren P: **The history of pyrosequencing**. *Methods Mol Biol* 2007, **373**:1-14.

92. Hyman ED: **A new method of sequencing DNA**. *Anal Biochem* 1988, **174**:423-436.

93. Brussow H: **The not so universal tree of life or the place of viruses in the living world**. *Philos Trans R Soc Lond B Biol Sci* 2009, **364**:2263-2274.

94. Hegde NR, Maddur MS, Kaveri SV, Bayry J: **Reasons to include viruses in the tree of life**. *Nat Rev Microbiol* 2009, **7**:615 author reply 615.

95. Ludmir EB, Enquist LW: **Viral genomes are part of the phylogenetic tree of life**. *Nat Rev Microbiol* 2009, **7**:615 author reply 615.

96. Raoult D: **There is no such thing as a tree of life (and of course viruses are out!)**. *Nat Rev Microbiol* 2009, **7**:615 author reply 615.

97. Schmieder R, Edwards R: **Quality control and preprocessing of metagenomic datasets**. *Bioinformatics* 2011, **27**:863-864.

98. Schmieder R, Edwards R: **Fast identification and removal of sequence contamination from genomic and metagenomic datasets**. *PLoS ONE* 2011, **6**:e17288.

99. Schmieder R, Lim YW, Rohwer F, Edwards R: **TagCleaner: identification and removal of tag sequences from genomic and metagenomic datasets**. *BMC Bioinformatics* 2010, **11**:341.

100. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: **GenBank**. *Nucleic Acids Res* 2011, **38**:D46-D51.

101. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**:403-410.

102. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV *et al.*: **The minimum information about a genome sequence (MIGS) specification**. *Nat Biotechnol* 2008, **26**:541-547.

103. Raes J, Foerstner KU, Bork P: **Get the most out of your metagenome: computational analysis of environmental sequence data**. *Curr Opin Microbiol* 2007, **10**:490-498.

104. Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B: **Cloning of a human parvovirus by molecular screening of respiratory tract samples**. *Proc Natl Acad Sci U S A* 2005, **102**:12891-12896.

105. Culley AI, Lang AS, Suttle CA: **Metagenomic analysis of coastal RNA virus communities**. *Science* 2006, **312**:1795-1798.

106. Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, Persson MA, Dalianis T, Ramqvist T, Andersson B: **Identification of a third human polyomavirus**. *J Virol* 2007, **81**:4130-4136.

107. Bench SR, Hanson TE, Williamson KE, Ghosh D, Radosovich M, Wang K, Wommack KE: **Metagenomic characterization of Chesapeake Bay virioplankton**. *Appl Environ Microbiol* 2007, **73**:7629-7641.

108. Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan PL, Briese T, Hornig M, Geiser DM *et al.*: **A metagenomic survey of microbes in honey bee colony collapse disorder**. *Science* 2007, **318**:283-287.

109. Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S *et al.*: **Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil**. *Appl Environ Microbiol* 2007, **73**:7059-7066.

110. Desnues C, Rodriguez-Brito B, Rayhawk S, Kelley S, Tran T, Haynes M, Liu H, Furlan M, Wegley L, Chau B *et al.*: **Biodiversity and biogeography of phages in modern stromatolites and thrombolites**. *Nature* 2008, **452**:340-343.

111. Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N, Afonso C, Zhang X, Anderson NG, Ghedin E *et al.*: **Viral genome sequencing by random priming methods**. *BMC Genomics* 2008, **9**:5.

112. Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D: **Metagenomic analysis of human diarrhea: viral detection and discovery**. *PLoS Pathog* 2008, **4**:e1000011.

113. Honkavuori KS, Shivaprasad HL, Williams BL, Quan PL, Hornig M, Street C, Palacios G, Hutchison SK, Franca M, Egholm M *et al.*: **Novel borna virus in psittacine birds with proventricular dilatation disease**. *Emerg Infect Dis* 2008, **14**:1883-1886.

114. Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL: **Rapid identification of known and new RNA viruses from animal tissues**. *PLoS Pathog* 2008, **4**:e1000163.

115. Adams IP, Glover RH, Monger WA, Mumford R, Jackeviciene E, Navalinskiene M, Samuitiene M, Boonham N: **Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology**. *Mol Plant Pathol* 2009, **10**:537-545.

116. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, Khristova ML, Weyer J, Swanepoel R, Egholm M *et al.*: **Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa**. *PLoS Pathog* 2009, **5**:e1000455.

117. Greninger AL, Runckel C, Chiu CY, Haggerty T, Parsonnet J, Ganem D, DeRisi JL: **The complete genome of klassevirus — a novel picornavirus in pediatric stool**. *Virol J* 2009, **6**:82.

118. Ng TF, Manire C, Borrowman K, Langer T, Ehrhart L, Breitbart M: **Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics**. *J Virol* 2009, **83**:2500-2509.

119. Ng TF, Suedmeyer WK, Wheeler E, Gulland F, Breitbart M: **Novel anellovirus discovered from a mortality event of captive California sea lions**. *J Gen Virol* 2009, **90**:1256-1261.

120. Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M: **Metagenomic analysis of viruses in reclaimed water**. *Environ Microbiol* 2009, **11**:2806-2820.

121. Blinkova O, Victoria J, Li Y, Keele BF, Sanz C, Ndjango JB, Peeters M, Travis D, Lonsdorf EV, Wilson ML *et al.*: **Novel circular DNA viruses in stool samples of wild-living chimpanzees**. *J Gen Virol* 2010, **91**:74-86.

122. Honkavuori KS, Shivaprasad HL, Briese T, Street C, Hirschberg DL, Hutchison SK, Lipkin WI: **Novel picornavirus in Turkey poults with hepatitis, California, USA**. *Emerg Infect Dis* 2011, **17**:480-487.

123. Shan T, Li L, Simmonds P, Wang C, Moeser A, Delwart E: **The fecal virome of pigs on a high-density farm**. *J Virol* 2011, **85**:11697-11708.

124. Gomez-Alvarez V, Teal TK, Schmidt TM: **Systematic artifacts in metagenomes from complex microbial communities**. *ISME J* 2009, **3**:1314-1317.

125. Wooley JC, Ye Y: **Metagenomics: facts and artifacts, and computational challenges**. *J Comput Sci Technol* 2009, **25**:71-81.

126. Angly F, Rodriguez-Brito B, Bangor D, McNairnie P, Breitbart M, Salamon P, Felts B, Nulton J, Mahaffy J, Rohwer F: **PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information**. *BMC Bioinformatics* 2005, **6**:41.

127. Deschavanne P, DuBow MS, Regeard C: **The use of genomic signature distance between bacteriophages and their hosts displays evolutionary relationships and phage growth cycle determination**. *Virol J* 2010, **7**:163.

128. Willner D, Thurber RV, Rohwer F: **Metagenomic signatures of 86 microbial and viral metagenomes**. *Environ Microbiol* 2009, **11**:1752-1766.

129. Babkin IV, Babkina IN: **Molecular dating in the evolution of vertebrate poxviruses**. *Intervirology* 2011, **54**:253-260.

130. Gilbert C, Feschotte C: **Genomic fossils calibrate the long-term evolution of hepadnaviruses**. *PLoS Biol* 2010:8.

131. Heinemann J, Maaty WS, Gauss GH, Akkaladevi N, Brumfield SK, Rayaprolu V, Young MJ, Lawrence CM, Bothner B: **Fossil record of an archaeal HK97-like provirus**. *Virology* 2011, **417**:362-368.

132. Sanderson MJ: **r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock**. *Bioinformatics* 2003, **19**:301-302.

133. Simmonds P, Davidson F, Lycett C, Prescott LE, MacDonald DM, Ellender J, Yap PL, Ludlam CA, Haydon GH, Gillon J *et al.*: **Detection of a novel DNA virus (TTV) in blood donors and blood products**. *Lancet* 1998, **352**:191-195.

134. Tanaka M, Nishiguchi S, Tanaka T, Enomoto M, Fukuda K, Takeda T, Nakajima S, Shiomi S, Kuroki T, Monna T *et al.*: **Prevalence of GBV-C and hepatitis G virus variants in patients with fulminant hepatic failure in Japan**. *J Hepatol* 1997, **27**:966-972.

135. Wang JT, Tsai FC, Lee CZ, Chen PJ, Sheu JC, Wang TH, Chen DS: **A prospective study of transfusion-transmitted GB virus C infection: similar frequency but different clinical presentation compared with hepatitis C virus**. *Blood* 1996, **88**:1881-1886.

136. Bernardin F, Operskalski E, Busch M, Delwart E: **Transfusion transmission of highly prevalent commensal human viruses**. *Transfusion* 2010, **50**:2474-2483.

137. Tang S, Lai KN: **Chronic viral hepatitis in hemodialysis patients**. *Hemodial Int* 2005, **9**:169-179.

138. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, Tammadoni S, Nosrat B, Conrad D, Rohwer F: **Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals**. *PLoS ONE* 2009, **4**:e7370.

139. George SL: **Persistent GB virus C infection is associated with decreased HIV-1 disease progression in the Amsterdam Cohort Study**. *J Infect Dis* 2005, **191**:2156-2157 author reply 2158–2160.

140. Lefrere JJ, Ferec C, Roudot-Thoraval F, Loiseau P, Cantaloube JF, Biagini P, Mariotti M, LeGac G, Mercier B: **GBV-C/hepatitis G virus (HGV) RNA load in immunodeficient individuals and in immunocompetent individuals**. *J Med Virol* 1999, **59**:32-37.

141. Watt G, Kantipong P, Jongsakul K: **Decrease in human immunodeficiency virus type 1 load during acute dengue fever**. *Clin Infect Dis* 2003, **36**:1067-1069.

142. Rivers TM: **Viruses and Koch's postulates**. *J Bacteriol* 1937, **33**:1-12.

143. Fredericks DN, Relman DA: **Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates**. *Clin Microbiol Rev* 1996, **9**:18-33.

144. Feng H, Shuda M, Chang Y, Moore PS: **Clonal integration of a polyomavirus in human Merkel cell carcinoma**. *Science* 2008, **319**:1096-1100.

••145 Falkow S: **Molecular Koch's postulates applied to microbial pathogenicity**. *Rev Infect Dis* 1988, **10(Suppl 2)**:S274-S276.
Significant paradigm change and a challenge to the existing Koch's postulates and the proposal to use molecular methods to assign etiology to infectious agents.

146. Wensel DL, Li W, Cunningham JM: **A virus-virus interaction circumvents the virus receptor requirement for infection by pathogenic retroviruses**. *J Virol* 2003, **77**:3460-3469.

147. Olweny CL: *Etiology of Endemic Burkitt's Lymphoma*. IARC Sci Publ; 1984:. 647–653.

148. Al Rwahnih M, Daubert S, Golino D, Rowhani A: **Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus**. *Virology* 2009, **387**:395-401.

149. Metzger MJ, Holguin CJ, Mendoza R, Miller AD: **The prostate cancer-associated human retrovirus XMRV lacks direct transforming activity but can induce low rates of transformation in cultured cells**. *J Virol* 2010, **84**:1874-1880.

150. Holtz LR, Finkbeiner SR, Kirkwood CD, Wang D: **Identification of a novel picornavirus related to cosaviruses in a child with acute diarrhea**. *Virol J* 2008, **5**:159.

151. Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, Kim E, Pillai DR, Guyard C, Mazzulli T *et al.*: **A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America**. *PLoS ONE* 2010, **5**:e13381.

152. Towner JS, Sealy TK, Khristova ML, Albarino CG, Conlan S, Reeder SA, Quan PL, Lipkin WI, Downing R, Tappero JW *et al.*: **Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda**. *PLoS Pathog* 2008, **4**:e1000212.

153. Taylor LH, Latham SM, Woolhouse ME: **Risk factors for human disease emergence**. *Philos Trans R Soc Lond B Biol Sci* 2001, **356**:983-989.

154. Nayar A: **Emerging disease: looking for trouble**. *Nature* 2009, **462**:717-719.

155. Wolfe ND, Daszak P, Kilpatrick AM, Burke DS: **Bushmeat hunting, deforestation, and prediction of zoonoses emergence**. *Emerg Infect Dis* 2005, **11**:1822-1827.

156. Wolfe ND, Heneine W, Carr JK, Garcia AD, Shanmugam V, Tamoufe U, Torimiro JN, Prosser AT, Lebreton M, Mpoudi-Ngole E *et al.*: **Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters**. *Proc Natl Acad Sci U S A* 2005, **102**:7994-7999.

157. Wolfe ND, Switzer WM, Carr JK, Bhullar VB, Shanmugam V, Tamoufe U, Prosser AT, Torimiro JN, Wright A, Mpoudi-Ngole E *et al.*: **Naturally acquired simian retrovirus infections in central African hunters**. *Lancet* 2004, **363**:932-937.

158. Reperant LA: **Applying the theory of island biogeography to emerging pathogens: toward predicting the sources of future emerging zoonotic and vector-borne diseases**. *Vector Borne Zoonotic Dis* 2010, **10**:105-110.

159. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L *et al.*: **Functional metagenomic profiling of nine biomes**. *Nature* 2008, **452**:629-632.

160. Rodriguez-Brito B, Rohwer F, Edwards RA: **An application of statistics to comparative metagenomics**. *BMC Bioinformatics* 2006, **7**:162.

161. Jones MS, Kapoor A, Lukashov VV, Simmonds P, Hecht F, Delwart E: **New DNA viruses identified in patients with acute viral infection syndrome**. *J Virol* 2005, **79**:8230-8236.

162. Lauck M, Hyeroba D, Tumukunde A, Weny G, Lank SM, Chapman CA, O'Connor DH, Friedrich TC, Goldberg TL: **Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan red colobus monkey discovered using direct pyrosequencing**. *PLoS ONE* 2011, **6**:e19056.