Chapter 4
The Human Virome

Matthew Haynes and Forest Rohwer

Background

Until fairly recently, it has been customary, in the absence of clinically significant infection, to view the human organism as an isolated entity. In fact, the healthy human body always contains a large number of foreign cells and viruses (Virgin et al., 2009; Dethlefsen et al., 2007; Relman, 2002). There are more viral particles in the human body than microbial cells, which are ten times more numerous than eukaryotic (human) cells. Similarly, only about 1.5% of the human genome encodes recognizable "human proteins," whereas approximately 45% our genome is retro-transposons, DNA transposons, and viral sequences. Most of the human-associated microbes and viruses, often found on "external" surfaces lining the lumens of organs such as the gut and oral/nasal cavities, participate in complex commensal or mutualistic relationships with their human host (Dethlefsen et al., 2007; Relman, 2002). Therefore, it is not advantageous to attempt to eradicate every virus and microbial cell from the body in response to infection. A new medical paradigm is emerging: an illness may be defined by a disruption of the normal "healthy" microbiome and/or virome, and that restoration of this state, not elimination of all nonhuman organisms, should be the goal of medical treatment (Harrison, 2007). Current interest in the human microbiome reflects the increasing acceptance of the view that the microbiota per se should not be seen merely as invasive disease vectors but are in fact an intrinsic part of the human supra-organism (Dethlefsen et al., 2007).

The classical method of viral isolation is by culturing. Koch's postulates (Rivers, 1937) dictate the conditions under which a virus cultured in vitro should be regarded as the cause of an infectious disease; human viruses are usually cultured only in this context. In addition, culturing will be successful only for the small fraction of viruses for which appropriate culture conditions can be determined. To break from the limited view that all viruses are intrinsically harmful requires new methodologies that enable us to characterize entire uncultured viral communities. A

M. Haynes (✉)
Department of Biology, San Diego State University, 5500 Campanile Dr., San Diego, CA 92182, USA
e-mail: mhaynes@projects.sdsu.edu

culture-independent metagenomics approach to viral community analysis will yield a broader view of the human virome, just as metagenomic sequencing has revealed a wider range of bacteria in the human microbiome than culture-based methods (Harris et al., 2007; Rogers et al., 2004).

Metagenomics

A viral metagenome or virome is the total genetic (DNA and RNA) sequence derived from a viral community. Mathematically, the structure of a community may be represented by a graph whose functional form (lognormal, power function, etc.) reflects the relative abundance distribution of its members. The *evenness* of the distribution (fractional contribution of each genotype), along with the *richness* (the total number of genotypes), are often combined to denote the *diversity* of a community, as in the Shannon–Wiener index (H'),

$$H' = - \sum_{i=1}^{S} r_i \ln r_i \tag{4.1}$$

where S is the sample richness, and r_i is the relative abundance of genotype i. Viral communities tend to be unevenly distributed, with a small number of species or genotypes dominating in abundance (Fig. 4.1).
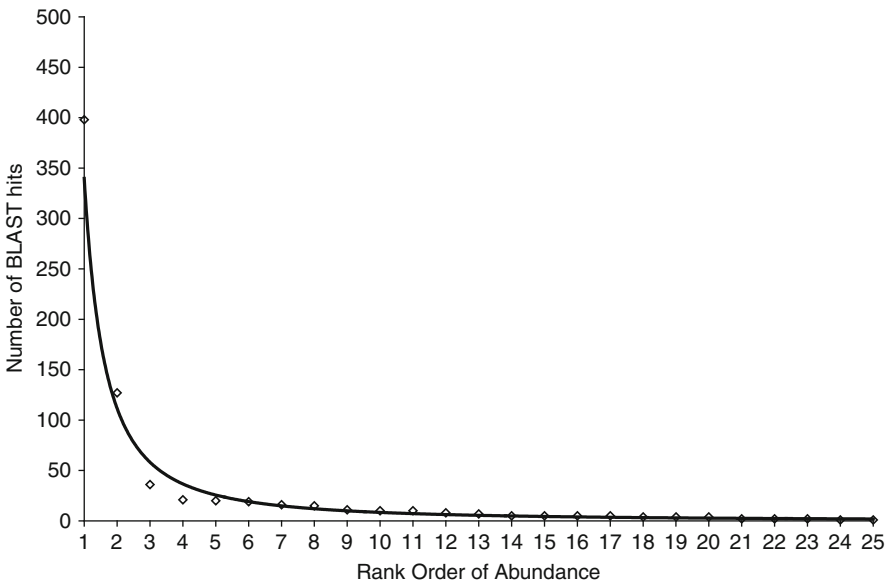

Fig. 4.1 Example of a human viral community rank-abundance curve using sequences from a human oropharyngeal metagenome. Here the relative frequencies of BLAST n hits to a viral sequence database follow a relationship that can be approximated by a power-law equation of the type $y = a\, x^{-b}$ (Willner et al., 2010)

Metagenomics has been greatly facilitated by recent advances in sequencing technology. Pyrosequencing (Roche/454 Life Sciences), as well as other technologies (e.g., Solexa, SOLiD), enable routine DNA sequencing on the scale of 10^8 bp. All of these new high-throughput methods replace traditional cloning in bacteria with mechanical separation of DNA molecules by some means (e.g., emPCR with DNA immobilized on beads for 454). This requires the creation of a minimally biased DNA library that better reflects the viral community in the sample. At present, the original DNA sample must often be amplified before sequencing, increasing the opportunity for artificial over- or underrepresentation of particular sequences. Despite this limitation, these methods appear to avoid most of the problems associated with conventional cloning, which is subject to strong sequence bias against some "unclonable" sequences. This phenomenon appears to be particularly pronounced in attempts to clone viral sequences. Microarrays, as well, remain semiquantitative detection methods because it is impossible to simultaneously optimize the hybridization of thousands of individual sequences (Table 4.1).

Table 4.1 Examples of viruses detected in human samples by metagenomic methods

Method	Sample source	Examples of viruses found
DNA recovery from microarrays	Nasopharynx (Ksiazek et al., 2003)	SARS Coronavirus
	Nasopharynx (Kistler et al., 2007)	Coronavirus, Rhinovirus
Random RT-PCR	Plasma (Jones et al., 2005)	PARV4, SAV-1, SAV-2
	Nasopharynx (Allander et al., 2005)	HBoV
	Stool (Victoria et al., 2009)	Cosavirus HCoSV
	Nasopharynx (Nakamura et al., 2009)	Influenza A, polyomavirus
	Stool (Zhang et al., 2005)	PMMV
Metagenomic shotgun libraries	Stool (adult) (Breitbart et al., 2003)	phages
	Blood (Breitbart and Rohwer, 2005)	TTV, HHV3, SEN virus, phages
	Stool (infant) (Breitbart et al., 2008)	phages
	Respiratory tract (Willner et al., 2009a)	HHV-1, HHV-2, phages
	Oropharynx (Willner et al., 2010)	Epstein–Barr Virus
	Stool (Reyes et al., 2009)	Phages

For further details See "Viral Metagenomics Methods"

Approximate Number and Distribution of Viruses in the Human Body

How Many Viruses Are There in a Human?

We can approach this question from two directions: estimation of the number of phages expected based on the size of the human microbiome and the typical viral (phage):host ratio, or by direct counts of viruses in samples from healthy

individuals. The human body is composed of about 10^{13} cells (Savage, 1977). There are about 10 times this number of microbial cells associated with the healthy human body (Savage, 1977). The observed ratio of 7–10 viral-like particles per microbial cell in environmental (Rohwer, 2003) and human samples (Furlan, 2009) means that we could expect to find about 10^{15} phages in the body. It is possible to compare this prediction with results from recent studies. The data in Table 4.2 are from direct counts of viruses using epifluorescence microscopy. These data indicate the presence of approximately 3×10^{12} viruses in the body.

Table 4.2 The diversity and estimated viral load of nonviremic humans

Organ system	VLPs per ml fluid	VLPs per cm^2	Total VLPs	Number of genotypes
Oral/nasal/pharynx	2.1×10^8		1.2×10^{10}	250
Tooth associated	8.5×10^7		3.0×10^9	?
Lower respiratory tract	2.8×10^8		1.4×10^{11}	250
Distal gut	4.8×10^8		3.0×10^{12}	1,000
Blood	1.0×10^5		1.0×10^8	20
Skin		1.2×10^6	2.1×10^{10}	?
Vagina	6.6×10^7		1.2×10^9	?
Urinary tract	3.7×10^6		1.9×10^9	?
Body total			3.0×10^{12}	1,500

Data sources: Willner et al. (2009a, 2010), Reyes et al. (2009), Furlan (2009), Furlan and Hanson (2009, personal communication)

Abundance of Viruses at Specific Body Sites

Wherever microbes (bacteria and archaea) are present, their viruses will be found. Thus in the human body, the regions of high microbial levels, in particular the gut, also have the highest abundance of viruses. Other organ systems with mucus membranes, such as the nasal and oral cavities and vagina, harbor a smaller but significant viral community.

What Types of Viruses Inhabit the Human Host?

Compared with environmental viral communities, the diversity of the human virome is low. We estimate that there are 1,500 viral genotypes in a typical healthy, human virome. By contrast, 1 kg of marine sediment will contain at least ten thousand, and perhaps a million, viral genotypes. The human-associated viruses are unevenly distributed, with the bulk of the virome composed of a handful of dominant species (Table 4.3). In the limited data available to date, it appears that a disease state is correlated with an increase in the diversity of the virome (Willner et al., 2009a).

Table 4.3 Comparison of diversity indices of typical environmental and human metagenomes. Values calculated using PHACCS (Angly et al., 2005) utilize all data, not merely identified sequences

Metagenome	Shannon index (nats)	Richness	Evenness	% most abundant genotype
Human oropharynx	5.0	248	0.92	6.5
Human lower resp. tract	4.8	243	0.92	6.2
Environmental (Angly et al., 2005, 2006)	6.0–10.8	>3000	0.85–1.00	2.3–13

Most of the viruses are phages. There are also certain eukaryotic viruses, such as herpesviruses, anelloviruses, and papillomaviruses, that are ubiquitous in the human virome and tend to cause few problems considering their abundance (Virgin et al., 2009). See also Fig. 4.3.

Phage Community

Commensal microbes are ubiquitous in the healthy human body (Dethlefsen et al., 2007; Wilson, 2005), occupying niches on skin (Grice et al., 2008, 2009), distal gut (Gill et al., 2006; Turnbaugh et al., 2009), vagina (Hyman et al., 2005). As a result, viruses that infect microbes (phages) are numerous (Letarov and Kulikov, 2009) and have been found in the gut (Reyes et al., 2009), nasopharynx (Allander et al., 2005), oropharynx (Willner et al., 2010), oral cavity (Hitch et al., 2004), blood (Breitbart and Rohwer, 2005), and lung secretions (Willner et al., 2009a).

Phages comprise by far the majority of the human virome (Willner et al., 2009a, 2010) and can be expected to exert an influence on the human microbial community (Gill et al., 2006; Hendrix, 2005) that parallels the interactions observed in a variety of environmental samples (Letarov and Kulikov, 2009; Weinbauer, 2006; Rodriguez-Mueller et al., 2010; Breitbart et al., 2005). By killing specific host organisms, phages regulate the absolute and relative abundance of microbial species (Breitbart et al., 2005). Genetic variation in the hosts is therefore favored as a means of escaping phage predation (Kunin et al., 2008). In addition, phages are major vehicles of DNA transfer to and from host cells (*horizontal gene transfer*) through both *lytic* and *lysogenic* pathways (Little, 2005), potentially conferring new phenotypes that can increase the pathogenicity (Breitbart et al., 2005) or the fitness (Sharon et al., 2009; Wagner and Waldor, 2002) of the host. Analysis of the phage metagenome can thus provide information not only about potential host taxonomy, but also reveal potential metabolic pathways available to the microbial community (Willner et al., 2009a; Sharon et al., 2009). Box 4.1 shows the "core" phage metagenome found in the human lower respiratory tract: 19 phage types that were all present in five normal control subjects and five cystic fibrosis patients (Willner et al., 2009a) (Fig. 4.2).

Box 4.1 Core phage community of the human lower respiratory tract. Phage types in the order of abundance that were present in all 10 samples regardless of disease state (tBLASTx hits with E < 10^{-5}) (Willner et al., 2009a)

Aeromonas hydrophila phi Aeh1
Aeromonas phi 31
Bacillus cereus phi phBC6A51
Bacillus subtilis phi 105
Bacillus subtilis phi SPBc2
Brucella melitensis 16 M phi Bruc 1 prophage
Escherichia coli phi CP073-4 prophage
Escherichia coli phi CP4-6 prophage
Escherichia coli phi QIN prophage
Escherichia coli phi Sp18 prophage
Haemophilus influenzae phi HP1
Lactobacillus plantarum phi LP65
Mycobacterium phi Bxz1
Mycobacterium phi CJW1
Pseudomonas phi KZ
Shigella flexneri phi Flex4 prophage
Staphylococcus phi Twort
Vibrio parahaemolyticus phi KVP40
Xylella fastidiosa phi Xpd5 prophage

a   **b**
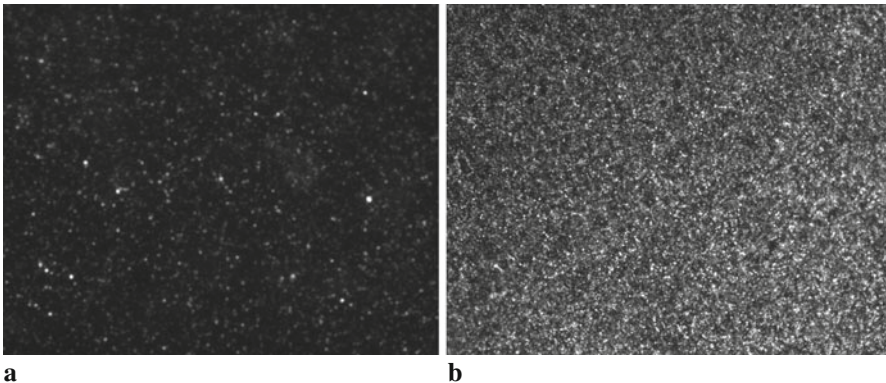
Fig. 4.2 Virus-like particles (VLPs) from asymptomatic individuals. (**a**) respiratory tract; (**b**) gut. Viruses were purified and concentrated by CsCl density gradient centrifugation as described in Breitbart et al. (2003). The VLPs were visualized by capturing on a 0.02-μm Anodisc filter, SYBR Gold staining, and viewing by epifluorescence microscopy

Eukaryotic Viruses

Viruses capable of infecting the human host ("eukaryotic viruses"), while obviously present in diseased individuals, can also be found in healthy subjects (Virgin et al., 2009; Willner et al., 2009a, 2010). In asymptomatic subjects, the abundance of these viruses is far lower than that of phages in the healthy human body (Willner et al., 2009a, 2010). Depending on the area of the body under examination, the presence of eukaryotic viruses will be due to either transient environmental exposure of accessible regions (e.g., the lungs) or chronic infections that do not give rise to recognizable clinical symptoms. The lack of symptoms might reflect a low-level viral infection that is successfully suppressed by the immune system at an early stage, or perhaps a commensal virus that causes no apparent harm (Virgin et al., 2009; Stapleton et al., 2004; Okamoto, 2009; Antonsson et al., 2000). An example of the latter is Torque Teno Virus (TTV), which was originally thought to be associated with a form of hepatitis, but now seems likely to be a ubiquitous but benign commensal virus (Okamoto, 2009). Instances of true viral–human mutualism in this context are not yet well understood, but it has been suggested that co-infection with GB Virus Type C (originally termed Hepatitis G virus) reduces mortality in HIV-infected individuals (Stapleton et al., 2004). Box 4.2 shows the "core" eukaryotic viral metagenome found in the human lower respiratory tract: 20 viruses that were all present in five normal control subjects and five cystic fibrosis patients (Willner et al., 2009a).

Box 4.2 Core eukaryotic viral community of the human respiratory tract. Viruses that were present in all 10 samples regardless of disease state (tBLASTx hits with E < 10^{-5}) (Willner et al., 2009a)

Acanthamoeba polyphaga mimivirus
Aedes taeniophyncus iridescent virus
Amsacta moorei entomopoxvirus "L"
Bovine Adenovirus A
Bovine Adenovirus 5
Cercopithicine herpesvirus 1
Cercopithicine herpesvirus 16
Cercopithicine herpesvirus 2
Cercopithicine herpesvirus 9
Chlorella virus ATCV-1
Chlorella virus FR483
Ectocarpus siliculosis virus 1
Frog virus 3
Human herpesvirus 1
Human herpesvirus 2

Melanoplus sanguinipes entomopoxvirus
Paramecium bursaria Chlorella virus AR158
Suid herpesvirus 1
Taterapox virus
Trichoplusia ni ascovirus

Residence Time and Pathogenicity

We can characterize viruses by their persistence (residence time in the body) and the degree of mutualism they exhibit (Fig. 4.3). The viruses that comprise the core human virome are relatively persistent (never cleared from the body). This distinguished them from pathogenic viruses causing acute and short-lived infections. There are, however, a number of pathogenic viruses such as herpesviruses that may persist in the body in an intracellular form, only to cause sporadic shedding of viral particles. Still other viruses are transient but common members of the human virome. Plant viruses such as PMMV are taken in with food and pass directly through the digestive tract (Zhang et al., 2005).

Fig. 4.3 Residence time vs. symbiotic modality of selected viruses. *EBV* epstein–barr virus, *HIV* human immunodeficiency virus, *HSV* herpes simplex virus, *TTV* Torque Teno virus, *PMMV* pepper mild mottle virus
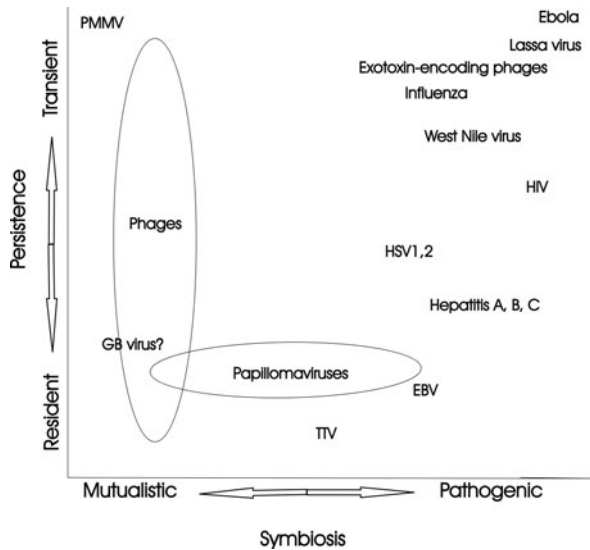
Viral Metagenomics Methods

Investigation of the human virome has recently been accelerated by technological and methodological developments. The methods fall into three categories: viral nucleic acid isolation, DNA sequencing, and data analysis. For a review of methods in viral metagenomics see Delwart (2007).

Recovery from Microarrays

The SARS coronavirus was discovered by hybridizing nucleic acids to an array (Virochip) that contained sequences representing all fully sequenced viruses, physically removing the annealed DNA from the array, and PCR amplifying this DNA using primers complementary to linkers that had been added (Kistler et al., 2007; Wang et al., 2002; Chiu et al., 2008). The prime example of this approach is the cloning and sequencing of the SARS coronavirus (Ksiazek et al., 2003). Limitations of the method are that it will only succeed with viruses that share significant homology with previously known viruses and that simultaneous optimization of multiple hybridizations on an array may be impossible.

Random RT-PCR

There are several variations of randomly primed reverse-transcription PCR (RT-PCR) for amplification of RNA viral sequences. Viral RNA is converted to cDNA using primers containing random octamers for both first- and second-strand synthesis, followed by PCR amplification. These methods have been successful in identifying many RNA viruses from human samples. Examples can be found in Victoria et al. (2009), Nakamura et al. (2009), and Jones et al. (2005). The method may be limited by PCR amplification bias, but it is highly sensitive.

Virus Purification and Phi29 Amplification

DNA viral metagenomes, including many phages, have been sequenced by purification of viral particles by CsCl density gradient centrifugation, DNase treatment, DNA isolation, and random amplification with Phi 29 DNA polymerase. Examples are respiratory tract metagenomes (mostly phages) from CF and non-CF subjects (Willner et al., 2009a) and an oropharyngeal metagenome from pooled samples from 19 healthy individuals (Willner et al., 2010). Limitations are potential amplification bias (Phi29 polymerase favors small circular and large linear genomes). This method has proved more successful for DNA than for RNA viruses.

Sequencing Methods

Due to the "untargeted" nature of metagenomics, and the often unavoidable contamination of viral nucleic acids with large amounts of human DNA, high-throughput sequencing has been essential. To date, the Roche/454 Life Sciences GS-FLX platform has been at the forefront of this technology, particularly because long sequence reads are necessary for shotgun sequencing. Sequencing technology is currently experiencing an unprecedented expansion, however, and it would not be surprising to see a series of further significant changes in sequencing methodology in the near future.

Bioinformatics

Data analysis is often the most challenging aspect of metagenomics research because the results are not pre-filtered by culturing or another selection process. The desired information must be extracted from a very large data set. Bioinformatics methods can be divided overall into two categories: similarity-based and similarity-independent approaches.

Similarity-Dependent Analyses

The original and more conventional means of sequence data analysis is to find segments of similarity to known sequences by searching databases. The most common tools are the various versions of BLAST (McGinnis and Madden, 2004), which will find local similarities based on the nucleic acid sequence or the deduced amino acid sequence. Microarray hybridization patterns have also been used to characterize novel viral nucleic acids (Urisman et al., 2005). These approaches are limited when the sample contains novel viruses that share little similarity with known viruses. Viruses in particular are subject to great variations in sequence composition. A large percentage of the sequences in a typical viral metagenome will not resemble any known sequences with any significance.

Metabolic Pathways

A metagenome can be characterized not only by taxonomy, but also by the cumulative metabolic potential encoded by the metagenome (Meyer et al., 2008). In the case of viral sequence data derived from lung sputum from CF patients and healthy subjects, the disease state of individuals correlated more strongly with the metabolic potential of viral metagenomes than with the taxonomic analysis (Willner et al., 2009a). In many cases the phage community appears to carry genes that complement the functions of the microbial community. In particular, phages often seem to use genes for proteins that will increase the short-term energy output of the host cells, either to increase viability (lysogeny) or to boost the production of viral particles (lytic). Some bacteria, such as cholera, are dependent on phage infection to achieve their virulence.

### *Similarity-Independent Analyses*

More recently, similarity-independent methods have been developed that do not require database searches. For example, PHACCS (Angly et al., 2005) uses contig spectra derived from the sequence data to infer the diversity of genotypes present in the original sample. Other methods enable the comparison of one metagenome to another on the basis of relative abundance of shared sequences. These methods will not identify the unknown viruses, but they can help to characterize the sample by defining the overall complexity of the community. Other methods involve analysis based on the percent G/C content of genomes or the relative frequency of various

dinucleotide combinations (Karlin et al., 1997; Burge et al., 1992; Karlin, 1998; Willner et al., 2009b), which in some cases is diagnostic of particular taxa.

Uncharacterized Viral Diversity

When viruses are purified from any human or environmental sample, the extracted DNA inevitably yields a large number of sequences (usually 70–99%) that show no significant similarity to any known sequences (Fig. 4.4) (Willner et al., 2009a, 2010; Jones et al., 2005).

Fig. 4.4 Unidentifiable sequences dominate the typical human viral metagenome. A similar phenomenon is observed in viral metagenomes from environmental samples

unknown
90.0%

phages
8.0%

euk virus
2.0%

■ unknown ■ phages ■ euk virus
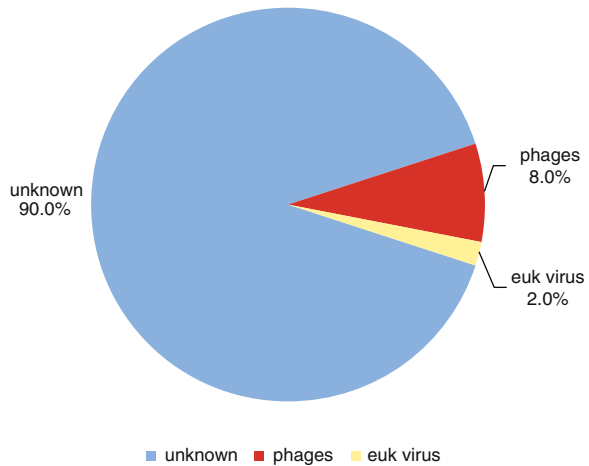
Provided that adequate precautions have been taken to avoid contamination with nonviral nucleic acids, this suggests that a very large fraction of the existing viral diversity remains uncharacterized. One of the strengths of the "untargeted" approach to viral metagenomics is that these sequences are obtained, but understanding the origin and significance of the "unknown" viral sequences is a substantial bioinformatic challenge that has yet to be solved. If a sequence has no similarity to the DNA of known organisms as defined by BLAST (McGinnis and Madden, 2004) or similar search algorithms, other methods must be developed for this purpose. For example, genome organization patterns such as large-scale arrangements of open reading frames or regulatory elements (promoters, enhancers, and origins of replication) may be signatures that would identify sequences as being of viral origin. This approach would likely require long sequences or even complete genomes to be successful.

Implications for Medical Care

An accurate assessment of the normal human virome provides a reference point from which to detect any novel viruses. This will serve as a background against

which an emerging pathogen or bioterrorism agent would appear in the human population through suitable screening programs. The health of the human subject should be judged by variation from the true "community" that it is, not by the assumption that no nonhuman entities should be present. This is analogous to restoration of a disturbed ecosystem. Knowledge of the normal viral community and assessment of any perturbations found in patients may enable physicians to diagnose disturbances of the microbiome.

Glossary

amplification bias   Inaccurate representation of the true relative abundances of genotypes in a DNA sample has been subjected to nonspecific amplification methods such as **MDA** or PCR.

BLAST   (Basic Local Alignment Search Tool) An algorithm used to search nucleic acid and protein databases for sequences similar to a query sequence (McGinnis and Madden, 2004).

commensalism   A form of symbiosis that benefits one partner while providing no apparent benefit to the other.

community   A set of interacting populations in an ecosystem.

diversity   A measure of the range of variation in a community, frequently represented as a combination of **richness** (number of variants) and **evenness** (skewness of the distribution).

emPCR   PCR performed in a water-in-oil emulsion, so that each micelle functions as a microreactor containing a single amplicon.

evenness   An index of the skewness of variation: an evenness value close to 0 implies that a community is dominated by one or very few members; a value of 1 implies equal abundance of every member.

genome   The nucleic acid (DNA or RNA) that constitutes genetic information from a single organism.

genotype   A genetic subtype that can be distinguished in a sample. In practical terms, two sequences will often be considered to legitimately represent the same genotype if they overlap at least 35 base pairs with 98% identity.

hybridization   (molecular biology) The annealing of complementary single-stranded DNA or RNA.

MDA   (multiple displacement amplification) DNA amplification using random primers in an isothermal reaction with a polymerase with helicase activity (Phi29 DNA polymerase), capable of nonspecific replication of double-stranded DNA.

metagenome   The total genomic nucleic acid (DNA and/or RNA) derived from a community.

mutualism   A form of symbiosis that benefits both partners.

population   The total set of members of a genetically distinguishable species or genotypes in a defined biome.

sequence read   A term frequently used to describe a sequence obtained by high-throughput methods

richness   The total number of distinct species or genotypes that can be distinguished in a community.

Shannon–Wiener index   One of the several measures of community diversity. A high value is associated with high richness and evenness values.

species   A genomic subtype that constitutes a genetic lineage or **population** that exists in a sample or biome. Due to the genomic plasticity of viruses and microbes it can be challenging to define a species, hence the use of the term **genotype** in a DNA sample when species definition or identification is problematic.

symbiosis   Any association between two organisms.

viremia   The presence of viruses in the blood.

virome   The cumulative viral community in an ecosystem.

References

Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B (2005) Cloning of a human parvovirus by molecular screening of respiratory tract samples. PNAS 102(36):12891–12896

Angly F, Rodriguez-Brito B, Bangor D, McNairnie P, Salamon P, Felts B, Nulton J, Mahaffy J, Rohwer F (2005) PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. BMC Bioinformatics 2(6(1)):41

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F (2006) The marine viromes of four oceanic regions. PLoS Biol 4(11):2121–2131

Antonsson A, Forslund O, Ekberg H, Sterner G, Hansson BG (2000) The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensalic nature of these viruses. J Virol 74(24):11636–11641

Breitbart M, Rohwer F (2005) Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. BioTechniques 39:729–736

Breitbart M, Rohwer F, Abedon ST (2005) Phage ecology and bacterial pathogenesis. In: Waldor MK, Friedman DI, Adhya SL (eds) Phages: their role in bacterial pathogenesis and biotechnolgy. ASM Press, Washington, DC, pp 66–92

Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, Mahaffy JM, Mueller J, Nultonc J, Rayhawk S, Rodriguez-Brito B, Salamon P, Rohwer F (2008) Viral diversity and dynamics in an infant gut. Res Microbiol 159(5):367–373

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

Burge C, Campbell AM, Karlin S (1992) Over- and under-representation of short oligonucleotides in DNA sequences. PNAS 89(4):1358–1362

Chiu CY, Greninger AL, Kanada K, Kwok T, Fischer KF, Runckel C, Louie JK, Glaser CA, Yagi S, Schnurr DP, Haggerty TD, Parsonnet J, Ganem D, DeRisi JL (2008) Identification of

cardioviruses related to Theiler's murine encephalomyelitis virus in human infections. PNAS 105(37):14124–14129

Delwart EL (2007) Viral metagenomics. Rev Med Virol 17(2):115–131

Dethlefsen L, McFall-Ngai M, Relman DA (2007) An ecological and evolutionary perspective on human–microbe mutualism and disease. Nature 449:811–818

Furlan M (2009) Viral and microbial dynamics in the human respiratory tract. Biology. San Diego State University, San Diego, CA

Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE (2006) Metagenomic analysis of the human distal gut microbiome. Science 312(5778):1355–1359

Grice E, Kong H, Renaud G, Young A, Bouffard G, Blakesley R, Wolfsberg T, Turner M, Segre J (2008) A diversity profile of the human skin microbiota. Genome Res 18(7):1043–1050

Grice E, Kong H, Conlan S, Deming C, Davis J, Young A, Bouffard G, Blakesley R, Murray P, Green E, Turner M, Segre J (2009) Topographical and temporal diversity of the human skin microbiome. Science 324(5931):1190–1192

Harris JK, Groote MAD, Sagel SD, Zemanick ET, Kapsner R, Penvari C, Kaess H, Deterding RR, Accurso FJ, Pace NR (2007) Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. PNAS 104(51):20529–20533

Harrison F (2007) Microbial ecology of the cystic fibrosis lung. Microbiology 153(Part 4):917–923

Hendrix RW (2005) Bacteriophage evolution and the role of phages in host evolution. In: Waldor MK, Friedman DI, Adhya SL (eds) Phages: their role in bacterial pathogenesis and biotechnology. ASM Press, Washington, DC, pp 55–65

Hitch G, Pratten J, Taylor PW (2004) Isolation of bacteriophages from the oral cavity. Lett Appl Microbiol 39:215–219

Hyman RW, Fukushima M, Diamond L, Kumm J, Giudice LC, Davis RW (2005) Microbes on the human vaginal epithelium. PNAS 102(22):7952–7957

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

Karlin S (1998) Global dinucleotide signatures and analysis of genomic heterogeneity. Curr Opin Microbiol 1(5):598–610

Karlin S, Mrazek J, Campbell A (1997) Compositional biases of bacterial genomes and evolutionary implications. J Bacteriol 179(12):3899–3913

Kistler A, Avila PC, Rouskin S, Wang D, Ward T, Yagi S, Schnurr D, Ganem D, DeRisi JL, Boushey HA (2007) Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. J Infect Dis 196:817–825

Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T (2003) A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 348:1953–1966

Kunin V, He S, Warnecke F, Peterson SB, Martin HG, Haynes M, Ivanova N, Blackall LL, Breitbart M, Rohwer F, McMahon KD, Hugenholtz P (2008) A bacterial metapopulation adapts locally to phage predation despite global dispersal. Genome Res 18:293–297

Letarov A, Kulikov E (2009) The bacteriophages in human- and animal body-associated microbial communities. J Appl Microbiol 107(1):1–13

Little JW (2005) Lysogeny, prophage induction, and lysogenic conversion. In: Waldor MK, Friedman DI, Adhya SL (eds) Phages: their role in bacterial pathogenesis and biotechnology. ASM Press, Washington, DC, pp 37–54

McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20–W25

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386

Nakamura S, Yang C-S, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J,

Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T (2009) Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One 4(1):e4219 [online only]

Okamoto H (2009) History of discoveries and pathogenicity of TT viruses. Curr Top Microbiol Immunol 331:1–201–220

Relman DA (2002) The human body as microbial observatory. Nat Genet 30:131–133

Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon J (2009) Phages in the distal human gut. Nature 466:334–338

Rivers TM (1937) Viruses and Koch's postulates. J Bacteriol 33(1):1–12

Rodriguez-Mueller B, Li LL, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues C, Dinsdale E, Edwards R, Felts B, Haynes M, Liu H, Lipson D, Mahaffy J, Martin-Cuadrado AB, Mira A, Nulton J, Pasic L, Rayhawk S, Rodriguez-Mueller J, Rodriguez-Valera F, Salamon S, Thingstad TF, Tran T, Willner D, Youle M, Rohwer F (2010) Viral and microbial community dynamics in four aquatic environments. ISME J 4(6):739–751

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Bruce KD (2004) Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. J Clin Microbiol 42(11): 5176–5183

Rohwer F (2003) Global phage diversity. Cell 113(2):141

Savage DC (1977) Microbial ecology of the gastrointestinal tract. Ann Rev Microbiol 31:107–133

Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, Atamna-Ismaeel N, Pinter RY, Partensky F, Koonin EV, Wolf YI, Nelson N, Béjà O (2009) Photosystem I gene cassettes are present in marine virus genomes. Nature 461:258–262

Stapleton JT, Williams CF, Xiang J (2004) GB virus type C: a beneficial infection? J Clin Microbiol 42(9):3915–3919

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI (2009) A core gut microbiome in obese and lean twins. Nature 457:480–484

Urisman A, Fischer KF, Chiu CY, Kistler AL, Beck S, Wang D, DeRisi JL (2005) E-Predict: a computational strategy for species identification based on observed DNA microarray hybridization patterns. Genome Biol 6: R78 [online only]

Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, Naeem A, Zaidi S, Delwart E (2009) Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J Virol 83:4642–4651

Virgin HW, Wherry EJ, Ahmed R (2009) Redefining chronic viral infection. Cell 138:30–50

Wagner PL, Waldor MK (2002) Bacteriophage control of bacterial virulence. Infect Immun 70(8):3985–3993

Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL (2002) Microarray-based detection and genotyping of viral pathogens. Proc Natl Acad Sci USA 99:15687–15692

Weinbauer MG (2006) Ecology of prokaryotic viruses. FEMS Microbiol Rev 28(2):127–181

Willner D, Furlan M, Haynes M, Schmieder R, Angly F, Silva J, Tammadoni S, Nosrat B, Conrad D, Rohwer F (2009a) Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. PLoS One 4(10):e7370

Willner D, Furlan M, Schmieder R, Grasis J, Pride D, Relman D, Angly FE, McDole T, Mariella R, Rohwer F, Haynes M (2010) Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. PNAS Early Edition.

Willner D, Thurber RV, Rohwer F (2009c) Metagenomic signatures of 86 microbial and viral metagenomes. Environ Microbiol 11(7):1752–1766

Wilson M (2005) Microbial inhabitants of humans: their ecology and role in health and disease. Cambridge University Press, New York, NY

Zhang T, Breitbart M, Lee WH, Run J-Q, Wei CL, Soh SWL, Hibberd ML, Liu ET, Rohwer F, Ruan Y (2005) RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol 4(1):e3[online only]