This is a previous version of the article published in Nature Reviews Microbiology. 2020, 18: 705-716. https://doi.org/10.1038/s41579-020-00444-0

Single-virus genomics and beyond

Joaquín Martínez Martínez[1], Francisco Martinez-Hernandez[2] and Manuel Martinez-Garcia[2†]

1. Bigelow Laboratory for Ocean Sciences, , East Boothbay, Maine, USA.

2. Department of Physiology, Genetics, and Microbiology, University of Alicante, , Alicante, Spain.

[†]email: m.martinez@ua.es

Abstract

Viruses are extremely diverse and modulate important biological and ecological processes globally. However, much of viral diversity remains uncultured and yet to be discovered. Several powerful culture-independent tools, in particular metagenomics, have substantially advanced virus discovery. Among those tools is single-virus genomics, which yields sequenced reference genomes from individual sorted virus particles without the need for cultivation. This new method complements virus culturing and metagenomic approaches and its advantages include, for example, targeted investigation of specific virus groups and investigation of genomic microdiversity within viral populations. In this Review, we provide a brief history of single-virus genomics, outline how this emergent method has facilitated advances in virus ecology, and discuss its current limitations and future potential. Finally, we address how this method may synergistically intersect with other single-virus and single-cell approaches.

Introduction

The majority of viruses are assumed to infect microorganisms, which are the most abundant cellular hosts on Earth, both free-living and those associate with animals and plants[1–4]. Viruses are ubiquitous and likely the most numerous and genetically diverse biological entities in nature[5]. Commonly, the investigation of viruses focuses on their role in causing disease in humans and commercially important plants and animals, although the impact of viruses in nature goes beyond disease and mortality alone. Viruses modulate microbial mortality and community dynamics, horizontal gene transfer, metabolic reprogramming in infected hosts, and biogeochemical cycles in the environment[6,7].

For many years, cultivation of virus-host pairs was the only approach available to provide insights into viral characteristics and the gold standard in taxonomy. Cultivation facilitates producing clonal virus particles in relatively large quantities, which is typically a requirement for detailed ultrastructural and molecular analysis. Additionally, culturing enables experimental investigation of viral-host dynamics, mechanisms of infection, and responses of the host under controlled conditions in the laboratory; however, these conditions often do not represent those present in nature[8–13]. Furthermore, cultured virus isolates only represent a small fraction of the total virus diversity [14–18]. The mismatch between viral diversity in the environment and in the number of laboratory isolates is also evident from the number of sequenced viral genomes. As of February 2020, the DOE-Joint Genome Institute IMG/VR database[19] contained 8,392 genomes of isolated viruses, three times less than the

number of the uncultured `high-quality draft viral genomes´ (n=25,659) and almost 100-fold less than the number of the viral genomic fragments (n=735,112) predicted from assembled shotgun metagenomes. Expanding the cultivable viral diversity is limited by the inability to grow the majority of ecologically relevant hosts or to generate host-derived cell lines[14–18]. Furthermore, there is an issue of scale. Despite some controversy regarding the extent of global viral diversity[3,20–24], even considering the most conservative estimates for dsDNA viruses with ~4 million different viral proteins[22], it is unimaginable that all viruses will ever be cultured.

Metagenomics is a powerful culture-independent approach to discover novel, uncultured microbial and viral genetic diversity[4,17,18,25–31]. The use of metagenomics provided a first glimpse at the level of diversity within uncultured marine viral communities and made it evident that much of that diversity had yet to be characterized[32]. More recent large-scale metagenomics studies continue to shed light on global genetic viral diversity. For example, the *Tara* expedition enabled assembly of genome fragments and complete or almost complete genomes from 15,222 marine viral populations [25]. Another study of ~3,000 metagenomes sampled in diverse environments around the globe, including human microenvironments, yielded 125,000 partial DNA viral genomes[4]. Metagenomics is also expanding our understanding of viral genetic diversity and function in terrestrial soils. The analysis of 668 global terrestrial metagenomes revealed the distribution of over 24,000 viral sequences and identified a suite of cosmopolitan virally encoded auxiliary metabolic genes that are potentially involved in the metabolism of organic carbon in soil. Furthermore, metagenomics has provided insights of the virus communities in extreme environments (reviewed in Ref [33]). A recent study led by researchers at the DOE Joint Genome Institute reconstructed 2,074 genomes of eukaryotic nucleocytoplasmic large DNA viruses (NCLDVs) from global metagenome data[30]. Although NCLDV genome sequences can be found within metagenomes that target cellular microbes[26,34], they typically represent a low-abundance but diverse fraction of the community, which complicates assembling large contigs and capturing most of their genetic diversity [26]. This new NCLDV dataset obtained by flow cytometry sorting and sequencing of multiple sets of 100 sorted viruses combined with metagenomics revealed novel phylogenetic and functional diversity and connected viral lineages to potential eukaryotic hosts, highlighting the ecological relevance of NCLDVs across environments. Metagenomics has been proposed as a framework to develop a sequence-based virus taxonomy to complement taxonomy based on virus isolates alone[35]. There is agreement that much of that diversity remains to be discovered[4,22,23], especially at the genus and species

level[36]. However, current metagenomics and bioinformatics methods typically miss important viral populations mainly owing to biases and limitations in sampling or assembly steps [36–39].

Viral genetic information can also be obtained by single-cell genomics (SCG) from uncultivated single sorted, infected cells (further discussed in detail below)[40–46]. Most commonly, individual cells are sorted by flow cytometry, chemically lysed and their genomes amplified (single-amplified genomes, SAGs) by multiple displacement amplification prior to sequencing[47–50]. Although sorting is not targeted for infected cells, infected cells often are present at high prevalence, leading to meaningful virus detection[40–46]. Single-virus genomics (SVG) is a new approach that is complementary to metagenomics and SCG. The basic SVG workflow (Fig. 1) comprises sample collection (typically less than 1-5 ml of sample containing 10^4-10^6 virions/ml is enough, although samples with lower particle concentrations take longer to sort); fluorescent staining of viral DNA (for example, SYBR dyes); separation or sorting of individual viral particles (for example, by flow cytometry); lysis of virion capsids (for example, through a combination of temperature and chemical shock); whole-genome amplification (for example, multiple displacement amplification); and sequencing and data analysis. For more details on SVG methods and protocols refer to[51]. SVG can uncover uncultured, abundant and cosmopolitan viral populations that encompass microdiversity overlooked by metagenomics[36]. Missing predominant viruses limits our understanding of the diversity and interactions of viral and microbial communities and thus the building of accurate ecological models. In this Review, we discuss the road leading from SCG to SVG and how SCG helps disentangle genome diversity. We also discuss current limitations of SVG and its potential, in particular, complementing other tools for interrogating individual cells and virus particles.

The road to single virus genomics

With every new discovery in virology it becomes clearer that there is little room for making generalizations about the ecological roles of viruses and that our knowledge of the biology and ecology of viral populations is limited. Almost every virus-host system has its own mode of interaction, and outcomes of infection. Whereas the genomes of some viruses have as little as two genes, others have over one thousand genes including diverse auxiliary metabolic genes [18,25,52,53], which modulate the metabolism of host cells during infection, so that the virus can replicate more efficiently. Diverse viral capsid morphologies and sizes and genome types (single stranded (ss) and double-stranded (ds) DNA or RNA) hinder the development of a single method for studying all viruses in a community. In addition, there is

no universal genetic marker for taxonomic identification and molecular quantification of viruses[54], complicating virus investigation.

SVG requires very low sample volumes and less than 1 ml volume often is enough[51], whereas metagenomics typically requires large volumes (liters). Sample volume is not a limitation for SVG, what matters is viral concentration. The minimum detectable viral concentration is an important factor for fluorescence-activated virus sorting. The minimum sample volume that can be analysed and sorted depends on the technical specifications of the flow cytometry sorter, and can be as low as ≈200 µl without introducing air bubbles in the fluidic lane of the instrument. Two previous studies analysed low-concentration, cryopreserved samples (~10^4 virions/ml) collected at 4,000 m depth at the Atlantic and Pacific Oceans[36,55], sorting one or up to 10,000 particles per well and amplifying and sequencing the whole genome. Samples with even lower viral concentrations could also be studied by SVG, yet it may be necessary to increase the sample volume and/or the sorting time to obtain a meaningful number of individual particles for analysis.

Standard procedures for most viral metagenomic (hereinafter referred to as `viromics´) studies include filtration through filters with 0.2 – 0.45µm pore size to exclude cellular organisms[17,25,32,39,56–59] and to enrich for viral particles. Often, filtration and purification do not successfully remove small microorganisms and all cellular nucleic acids[60]. As a result, a large fraction of publicly available viromes are not efficiently enriched in viral sequences[61]. Furthermore, filtration cannot discriminate cellular gene transfer agents or membrane vesicles from real virus particles.

Bioinformatic discovery and identification of viral genome fragments are complex (**Box 1**) and there is no common standard procedure widely accepted by the community. Moreover, bioinformatic tools often can only partially detect and remove cellular contaminants[60,62–64]. Although it is now well-established that many viruses carry and express metabolic genes that are homologs of host genes, such as auxiliary metabolic genes[18,25,52,65,66], assessing the metabolic potential of environmental viral communities through viromics can be hindered by cellular contamination in viromes, which might confound the assignment of predicted genes as viral[60]. Consequently, the best way to validate the viral origin of sequences is by their unequivocal placement in assembled virus genome fragments (contigs) or, ideally, in complete virus genomes. However, viromics data are often fragmented and yield consensus sequences hampering the assembly of complete discrete genomes and the study of viral microdiversity in uncultured assemblages. Uncovering microdiversity is important for

fully understanding the structure of viral populations and specific host-virus interactions, which often involve strain-specific lineages. Although ultra-deep sequencing and improvements in analysis pipelines, including in quality control and trimming of sequence reads, assembler algorithms, binning, and manual curation, improve recovery of viral genomes from metagenomes[31,67,68], genetic heterogeneity and co-occurrence of viral strains and variants of hypervariable islands are difficult to resolve bioinformatically[36–38,69]. Often, a large fraction of reads (up to 80%) remains unassembled in environmental viromes that are obtained by short-read sequencing[25,36]. Recent advances in long-read sequencing technologies[70] , which yield large assembly-free genomes, are promising. Hybrid assembly of long Nanopore and short Illumina reads can minimize the high error rate normally associated with long-read sequencing. This approach enabled recovery more complete viral genomes and capturing longer genomic island than typically achieved with short-read sequencing[71,72]. Additionally, metagenomics based on large-insert cloning of genomic DNA from virus-enriched samples into fosmids is another effective strategy for obtaining genetic information from uncultured viruses, although the insert size is limited by the size of the cloning vector system, preventing sequencing viruses with genomes larger than 50 Kb[73–76].

Although metagenomics is extremely powerful, additional culture-independent approaches, such as SCG, can be complementary. In 2007, SCG emerged as a cutting-edge technique to provide genomic information from individually sorted uncultured archaeal, bacterial and eukaryotic cells[42,49,77–83]. Mining of SAGs from SCG is a helpful cultivation-independent approach to discover new bacteria and archaea[84–89]. Following the success and existing methods for SCG, SVG has now been added to the toolkit available to study viruses. SVG circumvents several of the limitations of culturing or viromics discussed above[36,51,88,90–92]. However, SVG has its own biases and technical challenges. For example, the detection of viruses with very small capsid size and/or ssDNA and RNA genomes is virtually impossible with current flow cytometers, which are mostly designed for targeting cells instead of nanoparticles[93]. The low fluorescence derived from the staining of ssDNA and RNA viral genomes with commercially available dyes and their low side and forward scatter signals are below detection limit or overlap with background signal and electronic noise[94,95]. The development of flow cytometry instruments[94,95] that can detect very small, low-fluorescence particles would help pushing the boundaries of SVG to capture hidden viral diversity. Other strategies for virus detection and sorting based on microfluidic nano-devices and lab-on a chip with optics integration[96] are becoming very attractive alternatives to flow cytometry. Despite the well-established protocols for detecting and targeting dsDNA viruses using

fluorescent nucleic acid-binding stains[36,90,97,98], to the best of our knowledge, current commercial dyes with high affinity for ssDNA and RNA do not fully discriminate against dsDNA, further complicating the distinction of ssDNA and RNA containing particles. Furthermore, although Phi29 DNA polymerase and variations of it can amplify DNA, we are not aware of a commercially available enzyme with the required sensitivity for whole genome amplification (WGA) of a single copy of a RNA virus genome. Also, no manufactured enzyme and molecular reagents are contaminant-free[99], which is crucial when working at the level of single-copy cell or virus genomes. Other sources of contamination are from the sorting instrument or environmental DNA co-sorted within single-virion containing droplets.

It is imperative that the same strict practices for decontaminating and preventing DNA contamination used for SCG are implemented throughout the SVG pipeline[36,50,51,100]. An additional recommended practice for SVG is to sequence several WGA reaction control blanks without any sorted particles. During bioinformatics analyses, these blanks function as negative controls for subtracting potential contaminant reads from sequence libraries. Further advances in WGA chemistry are improving the outcomes of SCG and SVG. For example, WGA-X uses a thermostable mutant of the phi29 polymerase[101], which improves the speed, genome recovery and size of new assemblies, even for GC-rich genomes of bacteria, archaea, protists and virions[88], compared to the regular phi29 polymerase[102].

In addition to technological and chemical advances and strict decontamination procedures and quality controls[36,50,51,100], we argue that the full potential of SVG for viral discovery may be achieved through broad use of this tool and combined efforts by the wider microbiology and virology community. Large-scale SVG studies, analogous to previous global metagenomics studies, in which hundreds or thousands of single viral particles are sequenced from samples collected at high spatial and temporal resolution from diverse terrestrial, aquatic and animal environments, could redefine the role of this tool for understanding viral and microbial ecology. A limitation to this ambitious proposal may be the cost and limited access to SVG technology for many researchers. However, several affordable commercial kits for WGA that also work for single dsDNA viral genomes are now available. Additionally, commercial services can provide relatively affordable, high-yield sequencing. A flow cytometry sorter or compatible microfluidic devices that can sort individual virions are the most expensive requirement for SVG. Based on our experience, a dedicated instrument for SCG or SVG is not needed as long as there is a thorough instrument-cleaning step between samples. If research groups do not have access to their own sorter, we

suggest outsourcing virus sorting to well-established fee-for-service flow cytometry facilities, which exist in many universities, are fully equipped and have experienced technical personnel. So far, Influx™ and FACSAria™ sorters (BD Biosciences) have proven suitable for sorting giant viruses[90,103] and bacteriophages[36,88,92] for SVG. Other flow cytometry sorters may also have the sensitivity to detect and sort viral particles. However, we have not tested these sorters ourselves and there are no published studies so far. We encourage testing[36,92] additional sorters by fluorescently staining a control viral sample, sorting several virus particles onto a glass microscope slide and confirming the presence of single virus particles by confocal or super-resolution fluorescence microscopy.

Bioinformatic analysis (sequencing trimming, genome assembly and annotation) is crucial for a successful SVG workflow. Genome assembly from viral sequence data is one of the most important steps. Although, different algorithms and genome assemblers exist, SPAdes[104], using the `single-cell´ option for dealing with uneven genome coverage introduced during WGA step, probably shows the best performance. With the currently technologies, obtaining complete genomes from whole-genome amplified material is unrealistic, as demonstrated in SCG, with only one reported example of a complete genome from a single-cell[105]. In our experience and looking at available data, similar issues apply to SVG. Based on the recently proposed criteria on the minimum information about an uncultivated virus genome (MIUViG)[106] (REF), we conclude that genomes assembled from single-virus sequences are not finished genomes, rather `genome fragments´ (recovery of <90% of the expected genome length) or `high-quality draft genomes´ (≥90% of the expected genome sequence). In metagenomics, nearly 95% of recovered viral contigs (average contig length of 16.9 kb)[19] are classified as `genome fragments (<90% of the expected genome length).

Insights from single-virus genomics

Although only few SVG studies have been published so far, they provide important insights into viral diversity and ecology. We hope the SVG case studies presented below entice other researches to use SVG approaches in their work.

Sequencing of the genome from a single virus particle was first reported in 2011 by a team at the J. Craig Venter Institute[92]. The researchers sorted virions of lambda and T4 phages of *Escherichia coli* by flow cytometry[107,108] and they used multiple displacement amplification to produce enough DNA for sequencing[92]. Despite the limitations of the study,

that is, use of a simple mixture of two phage isolates and a low-throughput, it demonstrated the feasibility of this approach to study uncultured viruses.

A few years later, a similar fluorescence-activated virus sorting and WGA strategy was used to study virus particles in a one-milliliter surface seawater sample collected off the Patagonian Shelf [103]. However, this study did not investigate single-virus genomes but a pool of ≈5,000 virions. Flow cytometry resolved three distinct virus clusters and they were sorted into three separate tubes prior to genome amplification. The study recovered sequences of uncultured giant viruses, validating the effectiveness of this approach to target specific viral groups such as eukaryotic NCLDVs. Ecologically important NCLDV particles[109] are commonly removed during the filtering step of conventional viromics owing to their big genomes and capsids. Interestingly, another targeted flow-cytometric bulk sorting study discovered 16 novel soil NCLDVs, including novel lineages and the largest currently known viral genome in the Mimiviridae with 2.4 Mb[26].

In 2017, SVG was successfully applied in a high-throughput manner to marine environmental samples[36,90]. The study sorted 2,234 single virus particles from seawater samples collected from the Atlantic Ocean (4,000 m depth) and the Mediterranean Sea (surface and deep chlorophyll maximum) using fluorescence-activated virus sorting and performed WGA. Subsequently, 44 of these viral single-amplified genomes (vSAGs) were randomly selected for Illumina sequencing. None of the 44 vSAGs matched known virus isolates or metagenomics databases. Indeed, they represented 36 novel viral species and 7 new genera or families, which were cosmopolitan and abundant in the ocean[36]. These findings suggest that SVG likely recovered dsDNA viral populations that dominate the oceans. Remarkably, the study also showed that the newly discovered virus vSAG 37-F6 probably is one of the most abundant and cosmopolitan marine viral species and that it is present both as free virus and associated with host cells, as indicated by its high abundance in marine cellular metagenomes. Furthermore, 37-F6 virus-like species were also detected in the deep ocean. A capsid protein of 37-F6 was the most abundant viral protein in marine ecosystems, as determined by viral proteomics[110]. A recent metatranscriptomic study reported high transcription levels of 37-F6 viral genes in surface microbial communities[111]. Remarkably, metagenomic data from recruitment patterns and virome simulation data showed that the 37-F6 viral population was highly microdiverse and that this microdiversity had hindered metagenomic assembly, which likely explained why 37-F6 had not been identified before. A recent single-cell genomic study demonstrated that several members of 37-F6 viral population were present in uncultured sorted single cells belonging to the

candidate genus Pelagibacter from different oceans[45]. Fine ecogenomic analyses indicated that virus 37-F6 was a distinct, unrelated genetic population to previously described pelagiphage isolates[45,112]. Now, we know from SCGs that 37-F6 likely is lytic and infects Pelagibacter, which is one of the most abundant bacteria on Earth[45] and thus, this virus is expected to have a major impact on carbon cycling. Recent estimates from digital PCR suggest that 10-400 cells per ml are infected at any given time[113] with a potential carbon release from 124 fg to 4.9 pg per ml (assuming total carbon content in oceanic bacterial cells of 12.4 fg per cell [114]. As this virus is ubiquitous in all *Tara* samples[36], it is reasonable to think that an enormous amount of carbon might enter the viral shunt because of this virus.

Intriguingly, these discoveries would have been possible without looking at the most elemental component of viral communities, the single viral particle through the lens of SVG. Furthermore, considering that only 44 viruses, a tiny fraction of total sorted viruses, were sequenced, leading to the discovery of abundant viral species, future work should address whether other major abundant dsDNA viral populations, in addition to 37-F6, have been overlooked by viromics so far. Large-scale sequencing of sorted vSAGs from different oceanic regions from pole to pole, complemented with metagenomic long-read sequencing, might help to close this gap**.**

Another marine study[90] used SVG to target and sequence 12 individual giant ocean viruses directly recovered from a coastal seawater sample collected at Boothbay Harbor, ME, USA. A wide viral diversity was identified, including a member of the *Iridoviridae*, several members of the *Mimiviridae* and a taxonomically novel (unresolved) giant virus . The discovery of a putative viral metacaspase gene in the genome of one of these giant viruses suggested a mechanism by which the virus could influence host metabolism to promote viral infection and led to the demonstration that viral metacaspases are widespread in the ocean. The study also found a putative mimivirus with both a reverse transcriptase and a transposase, suggesting a novel mechanism of latent propagation[90]. The experiments for this study took place in 2011, and it was, to the best of our knowledge, the first study that sorted viruses (with large genomes) from an environmental sample. A caveat of this study, in contrast to current sequencing platforms, is that it primarily relied on 454 sequencing using Titanium chemistry, which was a common method at the time but suffers from non-random error distribution (mainly homopolymer errors) leading to poor genome assemblies and annotations, as reflected by the overall highly fragmented genomes (16-1,051 contigs, with a mean contig length of ~1,000 bp). Despite this limitation the findings support the use of SVG for deepening the genomic understanding of specific virus groups such as NCLDVs.

Demonstrating the progress in SVG methods, two recent SCG papers reported the retrieval of more complete and less fragmented genomes of giant marine viruses[115,116] from single sorted cells. One of the studies discovered the first viruses of choanoflagellates (ChoanoViruses) and revealed genomes enriched in enzymes that modify organic compounds, for example, for degrading chitin, and the presence of a viral rhodopsin photosystem with distinct evolutionary history from those that capture sunlight in cellular organisms[116]. As the giant viruses where found in individually-sorted cells, it is possible that they represented active infections with multiple copies of the same viral clone in the infected cells, which would have facilitated better genome coverage and assembly than in the study from 2011[90], which had single copy genomes as the starting material were single copy genomes.

Although so far SVG has been mainly used to study marine environmental microbial ecology, SVG has huge potential for application in plant, animal and human virology. In a recent pilot SVG study in humans[91], saliva samples from three volunteers were analyzed using SVG combined with viromics. Results showed a high proportion of uncharacterized viruses in the oral cavity. A total of 12 vSAGs were recovered, and one of them, vSAG 92-C13, was a putative *Streptococcus* spp. virus and one of the most abundant viruses in the oral virome.

[H1] Beyond single virus genomics

[H2] Non-genomic approaches to study viral particles

Interrogating single viral particles by SVG is only one of several techniques **(Box 2)** that provide biologically meaningful information about single or quasi-individual viral particles. We envision that continued refinement and standardization of sample processing and data analysis in SVG together with large-scale, high-throughput SVG studies will lead to transformative discoveries in virology; especially, when combined with other culture-independent and non-genomic approaches. For example, combining biorthogonal non-canonical amino acid tagging (BONCAT) [117–119] and SVG could potentially identify active viruses that were recently produced and released. Proteomics, high-resolution imaging, mass spectrometry, and Raman spectroscopy are other techniques that can provide a wealth of information on the viral architecture, morphology, chemical composition and structure. Viruses and many other large biomolecular complexes are in a mass range and size that are challenging to measure with conventional methods. Recent technological advances in mass spectrometry have, for example, enabled the first measurements of the molecular mass of individual viral particles of *Escherichia coli* bacteriophage T5 (ref.[120]). Furthermore, Raman

nanospectroscopy[121] can provide chemical and structural information at the single-viral particle level, which has been used to study different viruses, including human and plant viruses[122]. Altogether, these techniques that focus on the non-genetic molecular components of virions complement SVG approaches and open a new exciting era to explore (**Box 2**).

[H2] Single virus technology to study cellular vesicles

Optimized single virus technologies may also be adapted for the investigation of cellular vesicles. Accumulating empirical evidence indicates that vesicles have important roles in communication between cells within and across microorganisms and multicellular organisms[123–127]. For example, viruses take advantage of communication through extracellular vesicles between cells of the ecologically important, bloom-forming microalgae *Emiliania huxleyi* to promote infection[123]. Vesicles may package nucleic acids and other biomolecules from the host cell[128]. However, it is unknown whether the pool of vesicles in a natural sample such as seawater contains specific genes or other biomolecules that the producing microorganisms have packaged as a response to specific environmental queues, quorum sensing or intracellular factors. The study of vesicles has similar technical limitations as SVG . Furthermore, vesicles and viruses cannot be discriminated easily. Indeed, both types of particles overlap in size (ranging from a few dozen nanometers to over one micrometer) and they have diverse contents, for example, ss or ds DNA or RNA. Viruses can be further distinguished by the presence or absence of a lipid envelope. SVG methods could be used for the investigation of vesicles in aquatic, terrestrial and animal environments. Specifically, fluorescence-activated virus sorting methods[36,88,90,103] can be adapted to detect and sort vesicles, for example, by using different or additional fluorescence dyes . Lipophilic fluorescent dyes (such as FM4-64) could stain the membrane of vesicles and they could be combined with dyes targeting DNA (such as SYBR Gold), protein or cell metabolites in different vesicle types.

Pairing viruses with their hosts

More than sixty years ago Jacob and Wollman wrote that “viruses may exist in three states: the extracellular infectious state, the vegetative state of autonomous replication and finally the proviral state”[129]. Nevertheless, viruses are often seen as extracellular virions comprised by genetic material enclosed in a capsid protein. As formidably exposed by Patrick Forterre, individual “viral particles reveal their viral nature only if they encounter a host”[130]. In other words, a virus without its host is likely `convicted´ to irreversible decay. A virus-centric concept of infected cells (`virocells´) has been proposed, which refers to infected cells whose function is to produce virions[131]. Similarly, a recent theoretical cell-centric framework

proposed to quantify the fitness of viruses in relation to the proliferation of viral genomes inside cells instead of enumerating free viral particles outside cells[132]. In summary, the study of infected cells is key for understanding virus-host dynamics, and dissociating viruses (either at the community, population or individual level) from their host cells only provides a partial view. `Virocells´ can be studied at the level of individual cells, as exemplified by a study[42] that paired uncultivated viruses with their hosts. The authors sequenced three individual protist cells from seawater, which belonged to the clade of Picobiliphytes, and sequences from one of these cells were dominated by reads assigned to a widespread single-stranded DNA virus that contains a putative replication-associated protein. In 2012, a broader screening of different uncultured marine protists showed prey preferences and symbiotic interactions between the protists and different bacteria and archaea[133]. Recently, SCG of 65 individual marine protist cells from 11 essentially uncultured stramenopiles lineages that are widely distributed around the oceans identified 64 non-redundant viral contigs, only seven of which could be assigned to known viral families, including an endogenous *Mavirus* virophage [43]. Another SCG study[44] revealed two novel circular DNA viruses in the genome of eukaryotic SAGs from the *Micromonas* and *Ostreococcus* genera.

There is a wealth of information on viruses and their microbial hosts from different environments, ranging from extreme ecosystems[134] to the human microbiome[135]. For example, in 2014 a study identified 69 viral genome fragments representing five new genera of dsDNA and ssDNA phages in 127 SAGs of the uncultivated bacteria belonging to the SUP05 clade from the oxygen minimum zone at the Saanich Inlet in western Canada[41]. One year later, another study identified 20 novel phages (18 draft genomes and 2 complete genomes) analyzing SAGs of phylogenetically diverse marine bacteria and archaea from several geographical locations[40]. Notably, this study discovered the first examples of viruses infecting *Thaumarchaeota*, *Marinimicrobia*, *Verrucomicrobia* and *Gammaproteobacteria* clusters SAR86 and SAR92. Although it is widely assumed that one phage infects one host cell, we have recently learnt that a 'ménage à trois', in which more than one virus infects (nearly) every cell, might be more frequent than previously thought in nature[136]. Recently, SCG combined with metagenomics have been applied to unveil host-virus interactions in hot spring biofilms[137]. Around 26% of analyzed single cells contained a viral contig and data suggested that most of these viruses had a predominantly lysogenic lifestyle with limited diffusion of viruses between the different layers in the biofilm.

Other more targeted approaches, such as PhageFISH or viral tagging have proven useful to link phages to their hosts in natural microbial communities[135,138,139]. Viral tagging

identifies host-virus pairs by adding environmental virions stained with a generic nucleic acid-binding fluorochrome to cultured[138] or uncultured[135] host cells. 'Tagged' cells and the 'tagging' viruses can then be identified by fluorescence-activated cell sorting, whole-genome amplification and sequencing. With this method, a study[135] identified 363 unique phage-host pairs in the human gut demonstrating that most phages only bind to one bacterial species limiting the risk of horizontal gene transference between species, for example, of antibiotic resistance genes. In addition, a high level of cross-reactivity between phages and bacteria from different subjects was observed despite inter-individual variability in the human microbiome. PhageFISH[139] or VirusFISH uses fluorescence in situ hybridization with specific probes targeting intracellular and extracellular viral DNA to monitor infection dynamics and viral production. Although host cells can be identified using rRNA probes, prior knowledge of virus genetic information is required to design target-specific DNA probes.

Studying virus-hosts pairs goes beyond DNA; for example, single-cell RNA sequencing has reveals valuable biological insights into the interactions of single cells infected with one virus in different human infections, such as dengue, herpes simplex or influenza[140–142]. Single-cell RNA-seq of herpex simplex virus 1-infected cells has shed some light into early stages of infection and connected metabolic activation of viral pathways with antiviral programs[142].

In this way we can identify how a virus modifies the transcriptome of the host cells, and how the host responds to the infection at single-cell level. Finally, microfluidic and on-chip investigation strategies are also promising[143], and one of the best examples is the characterization of molecular inhibitors of human enteroviruses after screening thousands of individual virus-cell interactions[144]. Studying free single viral particles is very informative, however, looking at individual infected cells with different techniques provides complementary insights into the biology of viruses that are difficult to obtain when viruses are `decontextualized´ from their hosts.

Conclusions and outlook

Currently, most studies of uncultured viral communities rely heavily on *de novo* assembly of sequencing reads to recover diversity and functional information from metagenomic data. However, metagenomic assembly is challenging for virome data and can result in fragmented assemblies and poor recovery of viral community members (see for example the last comprehensive report[69]). Consequently, in most viromics applications and environments, a

large fraction of reads remain unassembled for different reasons[36–38], some discussed above such as microdiversity, although some of the limiting factors are unclear as yet. SVG, although relatively new and with only a handful of studies, has demonstrated the power of this approach to complement the toolkit in viral ecology and to simplify the complexity of viral diversity. Development and democratization of this technique largely depend on technological advances and automatization, such as nanofluidics lab-on-a chip and/or friendly flow nanoparticle platforms, which will happen sooner rather than later in the era of nanorobotics and nanotechnology, pushing the technique to unforeseen limits. New techniques, such as SVG, frequently open new research avenues beyond the one for which they were initially developed. Looking back at 2002, when multiple-displacement amplification was initially used for amplifying extracted human DNA[47,145], no one anticipated then that a few years later, this technique would be key to launch the SCG field, which has provided so many valuable insights not only in the microbial world[49,79,82,83,146–148] but also in neurobiology[149,150], stem cell differentiation, pathogenesis[151] and cancer[152,153].

SVG has the potential to change our understanding of viral genetic diversity and viral infection in the fields of environmental, agricultural and medical virology. Specifically, high-throughput SVG could speed up discovery by potentially delivering close to complete genomes of uncultured viruses; improving the coverage of viruses underrepresented in other datasets, such as giant viruses and microdiverse viruses systematically removed during standard virome preparation or not captured in metagenomic assembly; enabling the investigation of microdiversity in viral species and populations, including pathogens, without the need for previous genetic information and molecular markers; facilitating unequivocal or quasi-unequivocal identification of the viral origin of sequencing reads through selective sorting of virus particles removing gene transfer agents, membrane vesicles and other cellular components . The combination of next-generation short-read and novel long-read sequencing technologies[71] for producing dsDNA viromes and vSAGs from clinical or environmental samples will generate a wealth of data, which will enable addressing questions about viral diversity (and microdiversity), evolution, adaptation, and ecology in a more effective and comprehensive manner than previously possible.

References

1. Koonin, E. V. The wonder world of microbial viruses. in *Expert Review of Anti-Infective Therapy* **8**, 1097–1099 (2010).
2. Yong, E. *I contain multitudes : the microbes within us and a grander view of life*. (Ecco, an imprint of HarperCollinsPublishers, 2016).
3. Breitbart, M. & Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends in microbiology* **13**, 278–284 (2005).
4. **Paez-Espino, D. *et al.* Uncovering Earth's virome. *Nature* 536, 425–430 (2016). Massive metagenomic study on global viral diversity, distribution and host specificty of viruses. A total of 125,000 partial DNA viral genomes were discovered**
5. Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nature Reviews Microbiology* **3**, 504–510 (2005).
6. Suttle, C. A. Marine viruses-major players in the global ecosystem. *Nature reviews. Microbiology* **5**, 801–812 (2007). **A fundamental *"must read"* review on the role of viruses in marine ecosystems**
7. Abedon, S. T. *Bacteriophage ecology: population growth, evolution, and impact of bacterial viruses*. (Cambridge University Press, 2008).
8. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature* **424**, 1047–1051 (2003).
9. Sullivan, M. B. *et al.* Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental Microbiology* **12**, 3035–3056 (2010).
10. Kauffman, K. M. *et al.* A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature* **554**, 118–122 (2018).
11. Atanasova, N. S., Roine, E., Oren, A., Bamford, D. H. & Oksanen, H. M. Global network of specific virus-host interactions in hypersaline environments. *Environmental Microbiology* **14**, 426–440 (2012).
12. Marston, M. F. *et al.* Rapid diversification of coevolving marine Synechococcus and a virus. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 4544–4549 (2012).
13. Enav, H., Kirzner, S., Lindell, D., Mandel-Gutfreund, Y. & Béjà, O. Adaptation to sub-optimal hosts is a driver of viral diversification in the ocean. *Nature Communications* **9**, 1–11 (2018).
14. **Rappé, M. S. & Giovannoni, S. J. The Uncultured Microbial Majority. *Annual***

***Review of Microbiology* 57, 369–394 (2003).**

A comprehensive review addressing a fundamental question in Microbial Ecology on the difficulty for culturing most of microbes in the laboratory and how this bias and impact microbial discovery

15. Pedrós-Alió, C. The Rare Bacterial Biosphere. *Annual Review of Marine Science* **4**, 449–466 (2012).

16. Brum, J. R., Schenck, R. O. & Sullivan, M. B. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *The ISME Journal* **7**, 1738–1751 (2013).

17. Brum, J. R. *et al.* Patterns and ecological drivers of ocean viral communities. *Science* 348, (2015).

A pioneer, comprehensive metagenomic study on global marine viral diversity from hundred of samples collected during *Tara expedition.*

18. Trubl, G. *et al.* Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. *mSystems* **3**, (2018).

19. Paez-Espino, D. *et al.* IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Research* 47, D678–D686 (2019).

Description of the most comprehensive genome database of uncultured viruses recovered by metagenomics from different ecosystems including our human body with >700,000 viral genome fragments.

20. Carlson, C. J., Zipfel, C. M., Garnier, R. & Bansal, S. Global estimates of mammalian viral diversity accounting for host sharing. *Nature Ecology and Evolution* **3**, 1070–1075 (2019).

21. Carroll, D. *et al.* The Global Virome Project. *Science* **359**, 872–874 (2018).

22. Cesar Ignacio-Espinoza, J., Solonenko, S. A. & Sullivan, M. B. The global virome: Not as big as we thought? *Current Opinion in Virology* 3, 566–571 (2013).

Authors address a hot topic in viral ecology about how big is the viral diversity in nature and estimate the total number of different viral proteins; a proxy for quantifying number of different existing viruses.

23. Rohwer, F. Global phage diversity. *Cell* **113**, 141 (2003).

24. Suttle, C. A. Environmental microbiology: Viral diversity on the global stage. *Nature Microbiology* **1**, 1–2 (2016).

25. Roux, S. *et al.* Ecogenomics and potential biogeochemical impacts of globally

abundant ocean viruses. *Nature* **537**, 689–693 (2016).

26. Schulz, F. *et al.* Hidden diversity of soil giant viruses. *Nature communications* 9, 1–9 (2018).

Discovery of several relevant giant viruses -one of those with 2.4 Mb genome size- by employing metagenomics and a similar methodology used in single-virus genomics, but in this case targeting multiple sets of 100 viruses, instead of single-viral particles.

27. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).

28. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications* **7**, 1–11 (2016).

29. Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* **2**, 1533–1542 (2017).

30. Schulz, F. *et al.* Giant virus diversity and host interactions through global metagenomics. *Nature* **578**, 432–436 (2020).

31. Al-Shayeb, B. *et al.* Clades of huge phages from across Earth's ecosystems. *Nature* **578**, 425–431 (2020).

32. Breitbart, M. *et al.* Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences* **99**, 14250–14255 (2002).

33. Dávila-Ramos, S. *et al.* A review on viral metagenomics in extreme environments. *Frontiers in Microbiology* **10**, 2403 (2019).

34. Chatterjee, A., Sicheritz-Pontén, T., Yadav, R. & Kondabagil, K. Genomic and metagenomic signatures of giant viruses are ubiquitous in water samples from sewage, inland lake, waste water treatment plant, and municipal water supply in Mumbai, India. *Scientific Reports* **9**, 1–9 (2019).

35. Simmonds, P. *et al.* Consensus statement: Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology* **15 (3)**, 161–168 (2017).

36. Martinez-Hernandez, F. *et al.* Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nature Communications* 8, 1–13 (2017).

A pioneer reference high-troughput single virus genomic study that unveiled extremely abundant and ubiqutous uncultured marine viruses overlooked for years by current state-of the art, standard metagenomic based studies.

37. Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A. & Sullivan, M. B. Benchmarking

viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5, e3817 (2017).

This in silico study performs a through bioinformatic comparison of different tools used commonly in viral metagenomics and aims to provide some useful recommendations and standards for the scientific community.

38. Aguirre de Cárcer, D., Angly, F. E. & Alcamí, A. Evaluation of viral genome assembly and diversity estimation in deep metagenomes. *BMC Genomics* **15**, 1–12 (2014).

39. López-Pérez, M., Haro-Moreno, J. M., Gonzalez-Serrano, R., Parras-Moltó, M. & Rodriguez-Valera, F. Genome diversity of marine phages recovered from Mediterranean metagenomes: Size matters. *PLoS Genetics* **13**, e1007018 (2017).

40. Labonté, J. M. *et al.* Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *The ISME journal* 9, 2386–2399 (2015).

The screening of sequencing data from hundred of single cells obtained from seawater unveiled virus-host interactions in different ecologically important prokaryote groups

41. Roux, S. *et al.* Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife* **2014**, e03125 (2014).

42. Yoon, H. S. *et al.* Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 332, 714–717 (2011).

First report of single-cell genomics in uncultivated widespread microbial eukaryotes showing complex viral interactions and metabolic insights into phycobiliphyte groups

43. Castillo, Y. M. *et al.* Assessing the viral content of uncultured picoeukaryotes in the global-ocean by single cell genomics. *Molecular Ecology* **28**, 4272–4289 (2019).

44. Benites, L. F. *et al.* Single cell ecogenomics reveals mating types of individual cells and ssDNA viral infections in the smallest photosynthetic eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20190089 (2019).

45. Martinez-Hernandez, F. *et al.* Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean. *The ISME Journal* **13**, 232–236 (2019).

46. Brussaard, C. P. D., Noordeloos, A. A. M., Sandaa, R. A., Heldal, M. & Bratbak, G. Discovery of a dsRNA virus infecting the marine photosynthetic protist Micromonas pusilla. *Virology* **319**, 280–291 (2004).

47. Dean, F. B. *et al.* Comprehensive human genome amplification using multiple

displacement amplification. *Proceedings of the National Academy of Sciences* **99**, 5261–5266 (2002).

48. Raghunathan, A. *et al.* Genomic DNA Amplification from a Single Bacterium. *Applied and Environmental Microbiology* **71**, 3342–3347 (2005).

49. Stepanauskas, R. & Sieracki, M. E. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proceedings of the National Academy of Sciences* **104**, 9052–9057 (2007).

50. Rinke, C. *et al.* Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nature protocols* **9**, 1038–1048 (2014).

51. Martinez-Garcia, M., Martinez-Hernandez, F. & Martínez Martínez, J. Single-Virus Genomics: Studying Uncultured Viruses, One at a Time. in *Reference Module in Life Sciences* (Elsevier, 2020). doi:10.1016/b978-0-12-809633-8.21497-0

Authors provide methodological details and protocols for implementing single-virus genomics to complement other existing methdologies in viral ecology

52. Lindell, D. *et al.* Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proceedings of the National Academy of Sciences* **101**, 11013–11018 (2004).

53. Breitbart, M., Thompson, L., Suttle, C. & Sullivan, M. Exploring the Vast Diversity of Marine Viruses. *Oceanography* **20**, 135–139 (2007).

54. Brum, J. R. & Sullivan, M. B. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology* 13, 147–159 (2015).

A recommended review for readers that would like to introduce in recent technological advances in marine virology

55. De Corte, D. *et al.* Viral communities in the global deep ocean conveyor belt assessed by targeted viromics. *Frontiers in Microbiology* **10**, 1801 (2019).

56. Aylward, F. O. *et al.* Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 11446–11451 (2017).

57. Luo, E., Aylward, F. O., Mende, D. R. & Delong, E. F. Bacteriophage distributions and temporal variability in the ocean's interior. *mBio* **8**, (2017).

58. Angly, F. E. *et al.* The marine viromes of four oceanic regions. *PLoS Biology* **4**, 2121–2131 (2006).

59. Coutinho, F. H., Rosselli, R. & Rodríguez-Valera, F. Trends of Microdiversity Reveal Depth-Dependent Evolutionary Strategies of Viruses in the Mediterranean. *mSystems*

4, 1–17 (2019).

60. Roux, S., Krupovic, M., Debroas, D., Forterre, P. & Enault, F. Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biology* **3**, 130160 (2013).

61. Zolfo, M. *et al.* Detecting contamination in viromes using ViromeQC. *Nature Biotechnology* **37**, 1408–1412 (2019).

62. Amgarten, D., Braga, L. P. P., da Silva, A. M. & Setubal, J. C. MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. *Frontiers in Genetics* **9**, 304 (2018).

63. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).

64. Ponsero, A. J. & Hurwitz, B. L. The Promises and Pitfalls of Machine Learning for Detecting Viruses in Aquatic Metagenomes. *Frontiers in Microbiology* **10**, 806 (2019).

65. Crummett, L. T., Puxty, R. J., Weihe, C., Marston, M. F. & Martiny, J. B. H. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **499**, 219–229 (2016).

66. Pagarete, A., Allen, M. J., Wilson, W. H., Kimmance, S. A. & de Vargas, C. Host-virus shift of the sphingolipid pathway along an *Emiliania huxleyi* bloom: survival of the fattest. *Environmental Microbiology* **11**, 2840–2848 (2009).

67. Gregory, A. C. *et al.* Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177**, 1109–1123 (2019).

68. Kavagutti, V. S., Andrei, A. Ş., Mehrshad, M., Salcher, M. M. & Ghai, R. Phage-centric ecological interactions in aquatic ecosystems revealed through ultra-deep metagenomics. *Microbiome* **7**, 1–15 (2019).

69. Sutton, T. D. S., Clooney, A. G., Ryan, F. J., Ross, R. P. & Hill, C. Choice of assembly software has a critical impact on virome characterisation. *Microbiome* **7**, 12 (2019).

70. Madoui, M.-A. *et al.* Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics* **16**, 327 (2015).

71. Warwick-Dugdale, J. *et al.* Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* 7, e6800 (2019).

This pioneer study succesfully combines long- and short-read sequnecing data to improve viral metagenomic assemblies and shows the potential of Nanopore

sequencing data to advance viral discovery.

72. Beaulaurier, J. *et al.* Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Research* **30**, 437–446 (2020).

73. Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E. & Ghai, R. Expanding the marine virosphere using metagenomics. *PLoS genetics* **9**, e1003987 (2013).

74. Garcia-Heredia, I. *et al.* Reconstructing Viral Genomes from the Environment Using Fosmid Clones: The Case of Haloviruses. *PLoS ONE* **7**, e33802 (2012).

75. Chow, C. E. T., Winget, D. M., White, R. A., Hallam, S. J. & Suttle, C. A. Combining genomic sequencing methods to explore viral diversity and reveal potential virus-host interactions. *Frontiers in Microbiology* **6**, 265 (2015).

76. Mizuno, C. M., Ghai, R., Saghaï, A., López-García, P. & Rodriguez-Valera, F. Genomes of abundant and widespread viruses from the deep ocean. *mBio* **7**, e00805-16 (2016).

77. Martinez-Garcia, M. *et al.* High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *The ISME Journal* **6**, 113–123 (2012).

78. Stepanauskas, R. Single cell genomics: an individual look at microbes. *Current Opinion in Microbiology* **15**, 613–620 (2012).

79. Sieracki, M. E. *et al.* Single cell genomics yields a wide diversity of small planktonic protists across major ocean ecosystems. *Scientific Reports* **9**, 1–11 (2019).

80. Lasken, R. S. Genomic sequencing of uncultured microorganisms from single cells. *Nature Reviews Microbiology* **10**, 631–640 (2012).

81. López-Escardó, D. *et al.* Evaluation of single-cell genomics to address evolutionary questions using three SAGs of the choanoflagellate Monosiga brevicollis. *Scientific Reports* **7**, 1–14 (2017).

82. Mangot, J. F. *et al.* Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. *Scientific Reports* **7**, 1–12 (2017).

83. Seeleuthner, Y. *et al.* Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nature Communications* **9**, 1–10 (2018).

84. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437 (2013).

This is an excellent example of the power of single-cell technologies to provide biological

insights into uncultured microbes.

85. Swan, B. K. *et al.* Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proceedings of the National Academy of Sciences* **110**, 11463–11468 (2013).

86. Garcia, S. L. *et al.* Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *The ISME Journal* **7**, 137–147 (2013).

87. Martinez-Garcia, M. *et al.* Capturing Single Cell Genomes of Active Polysaccharide Degraders: An Unexpected Contribution of Verrucomicrobia. *PLoS ONE* **7**, e35314 (2012).

88. Stepanauskas, R. *et al.* Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nature Communications* 8, 1–10 (2017).

Lead authors in single-cell field managed to implement these technologies to flow cytometry sort uncultured single viruses and amplified their genomes with a new variant of an efficient phi29 enzyme, commonly used in single-cell and -virus genomics. This study represents another single-virus genomic example targeting uncultured viruses

89. Ghylin, T. W. *et al.* Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. *ISME Journal* **8**, 2503–2516 (2014).

90. Wilson, W. H. *et al.* Genomic exploration of individual giant ocean viruses. *The ISME Journal* 11, 1736–1745 (2017).

A reference single-virus genomic study targeting for the first time uncultured giant viruses in nature that commonly are ignored with standard metagenomic techniques

91. de la Cruz Peña, M. *et al.* Deciphering the Human Virome with Single-Virus Genomics and Metagenomics. *Viruses* **10**, 113 (2018).

First study on single-virus genomics applied to the Human Virome. Authors implement this novel technology combined with metagenomics in salivary human samples and discovered important, abundant phages.

92. Allen, L. Z. *et al.* Single Virus Genomics: A New Tool for Virus Discovery. *PLoS ONE* 6, e17722 (2011).

First report showing the feasibility of single-virus genomics as a new tool for viral

discovery. Authors sucessfully employed this technology to sequence several single sorted viral particles of viral isolates T4 and lambda of *Escherichia coli*

93. Holmfeldt, K., Odić, D., Sullivan, M. B., Middelboe, M. & Riemann, L. Cultivated single-stranded DNA phages that infect marine bacteroidetes prove difficult to detect with DNA-binding stains. *Applied and Environmental Microbiology* **78**, 892–894 (2012).

94. Pospichalova, V. *et al.* Simplified protocol for flow cytometry analysis of fluorescently labeled exosomes and microvesicles using dedicated flow cytometer. *Journal of Extracellular Vesicles* **4**, 1–15 (2015).

95. Giesecke, C. *et al.* Determination of background, signal-to-noise, and dynamic range of a flow cytometer: A novel practical method for instrument characterization and standardization. *Cytometry Part A* **91**, 1104–1114 (2017).

96. Schmidt, H. & Hawkins, A. R. Single-virus analysis through chip-based optical detection. *Bioanalysis* **8**, 867–870 (2016).

97. Brussaard, C. P. D., Payet, J. P., Winter, C. & Weinbauer, M. G. Quantification of aquatic viruses by flow cytometry. *Manual of aquatic viral ecology* **11**, 102–109 (2010).

98. Mojica, K. D. A. & Brussaard, C. P. D. Factors affecting virus dynamics and microbial host-virus interactions in marine environments. *FEMS Microbiology Ecology* **89**, 495–515 (2014).

99. Blainey, P. C. & Quake, S. R. Digital MDA for enumeration of total nucleic acid contamination. *Nucleic Acids Research* **39**, e19 (2011).

100. Woyke, T. *et al.* Decontamination of MDA Reagents for Single Cell Whole Genome Amplification. *PLoS ONE* **6**, e26161 (2011).

101. Povilaitis, T., Alzbutas, G., Sukackaite, R., Siurkus, J. & Skirgaila, R. In Vitro Evolution of phi29 DNA Polymerase Using Isothermal Compartmentalized Self Replication Technique. *Protein engineering, design & selection : PEDS* **29**, 617–628 (2016).

102. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics* 17, 175–188 (2016).

One of the most complete technical and scientific reviews on single-cell genomic technologies of unicellular and multicellular organisms that discuss how this technologies has enabled new biology discovery in multiple fields from microbiology to cancer or inmunology.

103. Martínez Martínez, J., Swan, B. K. & Wilson, W. H. Marine viruses, a genetic reservoir revealed by targeted viromics. *The ISME journal* 8, 1079–1088 (2014).

This study employs similar technologies used in single-virus genomics to discover giant viruses and other relevant uncultured viruses from a sorted pool of marine uncultured viruses

104. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology* **19**, 455–77 (2012).

105. Woyke, T. *et al.* One Bacterial Cell, One Complete Genome. *PLoS ONE* **5**, e10314 (2010).

106. Roux, S. *et al.* Minimum information about an uncultivated virus genome (MIUVIG). *Nature Biotechnology* **37**, 29–37 (2019).

107. Hercher, M., Mueller, W. & Shapiro, H. M. Detection and discrimination of individual viruses by flow cytometry. *Journal of Histochemistry and Cytochemistry* **27**, 350–352 (1979).

108. Lippé, R. Flow Virometry: a Powerful Tool To Functionally Characterize Viruses. *Journal of Virology* **92**, (2017).

109. Koonin, E. V. & Yutin, N. Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. in *Advances in Virus Research* **103**, 167–202 (Academic Press Inc., 2019).

110. Brum, J. R. *et al.* Illuminating structural proteins in viral 'dark matter' with metaproteomics. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2436–2441 (2016).

111. Alonso-Sáez, L., Morán, X. A. G. & Clokie, M. R. Low activity of lytic pelagiphages in coastal marine waters. *The ISME Journal* **12**, 2100–2102 (2018).

112. Zhao, Y. *et al.* Abundant SAR11 viruses in the ocean. *Nature* **494**, 357–360 (2013).

113. McMullen, A., Martinez-Hernandez, F. & Martinez-Garcia, M. Absolute quantification of infecting viral particles by chip-based digital polymerase chain reaction. *Environmental Microbiology Reports* **11**, 855–860 (2019).

114. Fukuda, R., Ogawa, H., Nagata, T. & Koike, I. Direct determination of carbon and nitrogen contents of natural bacterial assemblages in marine environments. *Applied and Environmental Microbiology* **64**, 3352–3358 (1998).

115. Needham, D. M. *et al.* Targeted metagenomic recovery of four divergent viruses reveals shared and distinctive characteristics of giant viruses of marine eukaryotes.

Philosophical Transactions of the Royal Society B: Biological Sciences **374**, 20190086 (2019).

116. Needham, D. M. *et al.* A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 20574–20583 (2019).

117. Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9482–9487 (2006).

118. Hatzenpichler, R. *et al.* In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. *Environmental Microbiology* **16**, 2568–2590 (2014).

119. Pasulka, A. L. *et al.* Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. *Environmental Microbiology* **20**, 671–692 (2018).

120. Dominguez-Medina, S. *et al.* Neutral mass spectrometry of virus capsids above 100 megadaltons with nanomechanical resonators. *Science* **362**, 918–922 (2018).

121. Hermelink, A. *et al.* Towards a correlative approach for characterising single virus particles by transmission electron microscopy and nanoscale Raman spectroscopy. *Analyst* **142**, 1342–1349 (2017).

122. Ruokola, P. *et al.* Raman Spectroscopic Signatures of Echovirus 1 Uncoating. *Journal of Virology* **88**, 8504–8513 (2014).

123. Schatz, D. *et al.* Communication via extracellular vesicles enhances viral infection of a cosmopolitan alga. *Nature Microbiology* **2**, 1485–1492 (2017).

124. Berleman, J. & Auer, M. The role of bacterial outer membrane vesicles for intra- and interspecies delivery. *Environmental Microbiology* **15**, 347–354 (2013).

125. Van Niel, G., D'Angelo, G. & Raposo, G. Shedding light on the cell biology of extracellular vesicles. *Nature Reviews Molecular Cell Biology* **19**, 213–228 (2018).

126. Machtinger, R., Laurent, L. C. & Baccarelli, A. A. Extracellular vesicles: Roles in gamete maturation, fertilization and embryo implantation. *Human Reproduction Update* **22**, 182–193 (2016).

127. Biller, S. J. *et al.* Membrane vesicles in sea water: Heterogeneous DNA content and implications for viral abundance estimates. *ISME Journal* **11**, 394–404 (2017).

128. Kulp, A. & Kuehn, M. J. Biological Functions and Biogenesis of Secreted Bacterial

Outer Membrane Vesicles. *Annual Review of Microbiology* **64**, 163–184 (2010).
129. Jacob, F. & Wollman, E. L. Viruses and genes. *Scientific American* **204**, 93–107 (1961).
130. Forterre, P. The virocell concept and environmental microbiology. *ISME Journal* **7**, 233–236 (2013).
131. Forterre, P. Manipulation of cellular syntheses and the nature of viruses: The virocell concept. in *Comptes Rendus Chimie* **14**, 392–399 (Elsevier Masson SAS, 2011).
132. Weitz, J. S., Li, G., Gulbudak, H., Cortez, M. H. & Whitaker, R. J. Viral invasion fitness across a continuum from lysis to latency. *Virus Evolution* **5**, vez006 (2019).
133. Martinez-Garcia, M. *et al.* Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. *ISME Journal* **6**, 703–707 (2012).
134. Martínez-García, M., Santos, F., Moreno-Paz, M., Parro, V. & Antón, J. Unveiling viral–host interactions within the 'microbial dark matter'. *Nature Communications* **5**, 1–8 (2014).
135. Džunková, M. *et al.* Defining the human gut host–phage network through single-cell viral tagging. *Nature Microbiology* 4, 2192–2203 (2019).
Probably one of the most comprehensive single-cell genomic studies within the context of human gut microbiome that unveiled a total of 363 unique *host–phage* pairings, expanding the known *host–phage network* of the *gut* microbiome136. Munson-Mcgee, J. H. *et al.* A virus or more in (nearly) every cell: Ubiquitous networks of virus-host interactions in extreme environments. *ISME Journal* **12**, 1706–1714 (2018).
137. Jarett, J. K. *et al.* Insights into the dynamics between viruses and their hosts in a hot spring microbial mat. *The ISME Journal* 1–15 (2020). doi:10.1038/s41396-020-0705-4
138. Deng, L. *et al.* Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. *Nature* **513**, 242–245 (2014).
139. Allers, E. *et al.* Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environmental Microbiology* **15**, 2306–2318 (2013).
140. Zanini, F. *et al.* Virus-inclusive single-cell RNA sequencing reveals the molecular signature of progression to severe dengue. *Proceedings of the National Academy of Sciences* **115**, E12363–E12369 (2018).
141. Steuerman, Y. *et al.* Dissection of Influenza Infection In Vivo by Single-Cell RNA Sequencing. *Cell Systems* **6**, 679–691 (2018).
142. Wyler, E. *et al.* Single-cell RNA-sequencing of herpes simplex virus 1-infected cells

connects NRF2 activation to an antiviral program. *Nature Communications* **10**, 1–14 (2019).

143. Guo, Q., Duffy, S. P., Matthews, K., Islamzada, E. & Ma, H. Deformability based Cell Sorting using Microfluidic Ratchets Enabling Phenotypic Separation of Leukocytes Directly from Whole Blood. *Scientific Reports* **7**, 1–11 (2017).
144. Liu, W. *et al.* More than efficacy revealed by single-cell analysis of antiviral therapeutics. *Science Advances* **5**, eaax4761 (2019).
145. Lasken, R. S. Single-cell genomic sequencing using Multiple Displacement Amplification. *Current Opinion in Microbiology* **10**, 510–516 (2007).
146. Marcy, Y. *et al.* Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences* **104**, 11889–11894 (2007).
147. Swan, B. K. *et al.* Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* **333**, 1296–1300 (2011).
148. Ahrendt, S. R. *et al.* Leveraging single-cell genomics to expand the fungal tree of life. *Nature Microbiology* **3**, 1417–1428 (2018).
149. McConnell, M. J. *et al.* Mosaic copy number variation in human neurons. *Science* **342**, 632–637 (2013).
150. Poulin, J. F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J. M. & Awatramani, R. Disentangling neural cell diversity using single-cell transcriptomics. *Nature Neuroscience* **19**, 1131–1141 (2016).
151. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
152. Sandberg, R. Entering the era of single-cell transcriptomics in biology and medicine. *Nature Methods* **11**, 22–24 (2014).
153. Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).
154. Lindell, D. *et al.* Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**, 83–86 (2007).
155. Roux, S., Tournayre, J., Mahul, A., Debroas, D. & Enault, F. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC bioinformatics* **15**, 1–12 (2014).
156. Watson, M., Schnettler, E. & Kohl, A. viRome: an R package for the visualization and analysis of viral small RNA sequence datasets. *Bioinformatics (Oxford, England)* **29**,

1902–1903 (2013).
157. Jurtz, V. I., Villarroel, J., Lund, O., Voldby Larsen, M. & Nielsen, M. MetaPhinder—Identifying Bacteriophage Sequences in Metagenomic Data Sets. *PLOS ONE* **11**, e0163111 (2016).
158. Zheng, T. *et al.* Mining, analyzing, and integrating viral signals from metagenomic data. *Microbiome* **7**, 1–15 (2019).
159. Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A. & Sun, F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* **5**, 69 (2017).
160. Fang, Z. *et al.* PPR-Meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. *GigaScience* **8**, giz066 (2019).
161. Tampuu, A., Bzhalava, Z., Dillner, J. & Vicente, R. ViraMiner: Deep learning on raw DNA sequences for identifying viral genomes in human samples. *PLOS ONE* **14**, e0222271 (2019).
162. Bin Jang, H. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology* **37**, 632–639 (2019).
163. Schleyer, G. *et al.* In plaque-mass spectrometry imaging of a bloom-forming alga during viral infection reveals a metabolic shift towards odd-chain fatty acid lipids. *Nature Microbiology* **4**, 527–538 (2019).
164. Van Etten, J. L., Burbank, D. E., Kuczmarski, D. & Meints, R. H. Virus infection of culturable Chlorella-like algae and development of a plaque assay. *Science* **219**, 994–996 (1983).
165. Maxwell, K. L. & Frappier, L. Viral Proteomics. *Microbiology and Molecular Biology Reviews* **71**, 398–411 (2007).
166. Lum, K. K. & Cristea, I. M. Proteomic approaches to uncovering virus-host protein interactions during the progression of viral infection. *Expert Review of Proteomics* **13**, 325–340 (2016).
167. Cheng, W. & Schimert, K. A method for tethering single viral particles for virus-cell interaction studies with optical tweezers. in *Optical Trapping and Optical Micromanipulation XV* 107233B (SPIE-Intl Soc Optical Eng, 2018). doi:10.1117/12.2500050
168. Ekeberg, T. *et al.* Three-dimensional reconstruction of the giant mimivirus particle with an X-ray free-electron laser. *Physical Review Letters* **114**, 098102 (2015).
169. Cheng, Y. Single-particle Cryo-EM at crystallographic resolution. *Cell* **161**, 450–457

(2015).
170. Lyumkis, D. Challenges and opportunities in cryo-EM single-particle analysis. *Journal of Biological Chemistry* **294**, 5181–5197 (2019).
171. Subramaniam, S., Bartesaghi, A., Liu, J., Bennett, A. E. & Sougrat, R. Electron tomography of viruses. *Current Opinion in Structural Biology* **17**, 596–602 (2007).
172. Gamage, S. *et al.* Probing structural changes in single enveloped virus particles using nano-infrared spectroscopic imaging. *PLoS ONE* **13**, e0199112 (2018).
173. Martínez Martínez, J., Schroeder, D. C., Larsen, A., Bratbak, G. & Wilson, W. H. Molecular dynamics of Emiliania huxleyi and cooccurring viruses during two separate mesocosm studies. *Applied and Environmental Microbiology* **73**, 554–562 (2007).
174. Martínez Martínez, J. *et al.* New lipid envelope-containing dsDNA virus isolates infecting Micromonas pusilla reveal a separate phylogenetic group. *Aquatic Microbial Ecology* **74**, 17–28 (2015).

Acknowledgements

This work has been supported by Gordon and Betty Moore Foundation (grant 5334), the National Science Foundation (NSF-OPP 1644155, NSF-OCE 1933289), Spanish Ministry of Economy and Competitiveness (refs CGL2013-40564-R, RTI2018-094248-B-I00 and SAF2013-49267-EXP) and Generalitat Valenciana (ref. ACOM/2015/133 and ACIF/2015/332).

Author contributions

All authors researched data for the article, contributed to the discussion of the content, and reviewed and edited the manuscript before submission. M. M.-G. and J.M.M wrote the article.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews XXX thanks [Referee#1 name], [Referee#2 name] and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Figure 1. Methods to study viruses. Culturing has been the standard and most commonly used technique for many years to study viruses and their hosts (top)[112,154]. Although undoubtedly this method is very powerful, it restricted to the few virus-host pairs that can be cultured and maintained in the laboratory. Viromics (middle)[25,32,58], which usually requires high volumes of sample, sequences nucleic acids obtained from bulk environmental samples (enriched for viruses or cells) to study uncultured microbial and viral genetic diversity. Metagenomic assembly of viruses often is challenging[36–38]. Single-virus genomics and single-cell genomics sequence the genome of individual viruses[36,88,90,103] or viruses infecting individual cells[40,45], respectively. As sequenced genomes come from a sorted single individual (virus or cell), genomic assembly and subsequent bioinformatic analyses are typically less complex than for metagenomics.

Figure 2. Single-virus genomics workflow. Viruses are fluorescently stained (for example with SYBR Gold for double stranded DNA viruses) and sorted one at a time from the natural sample using fluorescence-activated virus sorting or other tools such as microfluidics. Sorted single viral particles are deposited in multi-well plates and then the capsid is lysed, for example, by a combination of pH and temperature shock. Free viral DNA is then subjected to whole genome amplification (for example, real time multiple-displacement amplification). After amplification, enough DNA is available for DNA sequencing or gene target PCR screening .

Figure 3. Insights from single-virus genomics. Marine habitats are the first ecosystem in which single-virus genomics (SVG) has been applied[36,88,90,103]. From over 2,000 uncultured viral particles that were sorted, 44 novel viruses from the surface and the deep ocean were discovered[36]. These viruses, which had been overlooed by viromics, are highly abundant and widespread across all oceans. One of them, the viral single amplified genome (vSAG) 37-F6 potentially is the most abundant virus in the surface ocean and represents a new viral family infecting the candidate genus Pelagibacter[45]. SVG has been successfully applied to study marine eukaryotic nucleocytoplasmic large DNA viruses (NCLDVs), uncovering novel viruses, their ecology and interaction with their hosts[90,103]. The first application of SVG in human samples resulted in the discovery of *Streptococcus* spp. phages in saliva[91].

Figure 4. Present and future of single-virus technologies. Other single-virus technologies (Box 3) exist beyond single-virus genomics (SVG) and these methods have the potential to complement each other in the study of viruses and in other fields, for example, in vesicle research. Combination of SVG with fluorescence-based biorthogonal non-canonical amino acid tagging (BONCAT)[119] of new released viruses could help identify active uncultured viruses in complex natural communities. BONCAT fluorescently tagged viruses, either at the population or individual scale, from environmental samples could be sorted out by flow cytometry and analysed through the SVG pipeline to sequence and identify active viruses produced under certain environmental conditions. Other promising examples are based on advances in nanotechnology combined with mass spectrometry[120] or Raman spectroscopy[121] at the level of individual viral particles will support the analysis of the chemical structure of viruses.

Box 1 | Detecting viruses in (meta)genomes

Unequivocal identification of an assembled DNA fragment as a viral genome fragment can be challenging because of the lack of a universal gene marker for viruses[54]. There are two ways to address this task: manually, which is time-consuming, or automatically with bioinformatics programs, which use different technical approaches with their own limitations and biases. We would like to highlight a few of several available programs: Metavir[155], Virome[156], or MetaPhinder[157] are based on gene similarity searches and alignment of a query sequence to previously-known viral sequences. Next-generation bioinformatic tools, such as VirSorter[63], try to mine unknown viruses combining the identification of similar homologuous viral sequences and hallmark genes with other genetic `features´ common in many viruses, such as the enrichment of uncharacterized or unknown genes and/or the depletion of Pfam-affiliated genes. Recently, artificial intelligence has has been adapted to 'hunt' viruses. Recent examples are the machine learning applications VirMiner[158], MARVEL[62], or VirFinder[159], which search for *k*-mer frequency signatures from assembled metagenomic datasets and avoid gene-based similarity searches. Finally, PPR-Meta[160], and ViraMiner[161] are deep learning-algorithms based methods that have been tested with contigs from human metagenomes. Alternatively, we also recommend automatic annotation platforms. Probably the most comprehensive and robust one is the JGI-IMG, whichrelies on likely the most complete publicly viral databases to date, the Integrated Microbial Genome/Virus system v.2.0 (IMG/VR v.2.0)[19].

Although these bioinformatic programs have been powerful and successful, they have some limitations, mainly owing to the viral database used to compare with or to perform the `training´ in the case of the machine-learning algorithms. For example, we have isolated and characterized some viruses from hypersaline environments that were not recognized as having a viral origin. On the other hand, other programs that assign viruses to hosts based on *k*-mer signatures have, in our example, assigned these hypersaline viruses that weal isolated from extremophile *Salinibacter ruber*, which inhabits environments of >15-35% of NaCl, to standard marine bacteria. Complementary analysis and manual in-house curation can help to correctly characterize and identify the obtained viral genome fragments. Careful review of the protein annotation of the supposedly viral assembled fragments to analyze the different gene functions in the same genome fragment and genetic context is fundamental. Complementary strategies might help to identify a putative virus, such as virome fragment recruitment, which support recognizing a viral contig when it shows a high recruitment rate when compared against reads obtained from viral metagenomes. Other tools such as the network-based application vConTACT 2.0 (ref. [162]) builds a viral network based on gene-sharing relationship among different viruses and help to visualize the location and relatedness of the recovered virus against other viruses from the same environment.

Box 2 | Non-genomic approaches to study single virus particles and hosts

In plaque-mass spectrometry imaging

In plaque-mass spectrometry imaging (MSI) visualizes temporal changes of the metabolome during viral infection[163]. Unlike studies with bulk liquid samples, in plaque-MSI maps metabolic states to infection states by analyzing individual plaques formed in host cultures grown on solid media[164]. A plaque originates from the infection of a single host cell by a single virus at the center of the plaque. As progeny virions infect adjacent cells, the plaque expands in

concentric rings, creating a temporal metabolomic record of the infection. This novel and powerful method provides a unique opportunity to further understand of infection dynamics and consequences. However, it is limited to hosts that are amenable to growing in the laboratory.

Viral-BONCAT

This method[119] is a modification of biorthogonal non-canonical amino acid tagging (BONCAT)[117] coupled to fluorophore addition through 'click chemistry'[118] to track host-virus interactions and to measure virus production. Viral BONCAT measures the transfer of L-homopropargylglycine (HPG), a methionine analogue, from HPG-labelled host cells to newly formed virions. The method has been tested in cultures and field seawater samples. Although epiflorescence was used to visualize labelled host cells and virus particles, flow cytometry could make this method compatible with SVG. Viral-BONCAT in combination with SVG may enable targeted genomic investigation of newly produced virions in environmental samples, that is, active, ecologically-relevant lytic viruses at the time of sampling.

Proteomics

Proteomic approaches and advances in mass spectrometry methods have revolutionized our ability to determine the composition and function of proteins in virions as well as protein interactions and cellular changes in infected host cells, advancing our understanding of viral diversity, infection mechanisms and pathogenesis[165,166]. Viruses are particularly suitable for genome-wide analyses due to their relatively small genomes. Sequenced viral genomes provide a blueprints for possible viral gene products and are powerful references for testing protein properties and functions by proteomics. Consequently, SVG complements proteomics through the expansion of the viral genome sequence space.

Optical tweezers

Optical tweezers use a laser to tether small particles (nanometer to micrometer size) in place under a microscope. Optical tweezers enable non-invasive manipulation of single cells and virions. Biotinylated single virions trapped to a bead by a DNA tether can be delivered to a live host cell to investigate the physical host-virus surface interactions and viral entry[167]. This method requires prior knowledge of host proteins that are incorporated into virions for biotinylation. Hence, it is limited to well-characterized cultured host-virus systems, but it is amenable to investigate RNA and ssDNA viruses, which are undersampled by fluorescence-based recognition techniques, including microscopy and flow cytometry[46,93].

High-resolution imaging

Imaging of single virus particles provides critical ultrastructural information. The combination of high-resolution imaging and SVG would greatly advance the understanding of uncultured viruses. X-ray crystallography has been the defining tool for structural biology to resolve the structure (and hence the function) of important biomolecules such as viral capsid and envelope proteins. Recently, improvements in X-ray free-electron lasers, Cryo-electron microscopy (Cryo-EM) and Cryo-electrom tomography (Cryo-ET) provide similar resolution limits as X-ray crystallography. These techniques enable the generation of images from individual symmetrical and asymmetrical virions under nearly native conditions without the need to crystalize the particles[168–171]. Nano-infrared spectroscopic imaging was recently implemented to detect and quantify subtle chemical and structural changes in single enveloped virions prior to membrane fusion[172]. We suggest that flow cytometry could be used to sort individual virus particles from a discrete population, based on fluorescence and light scattering signals, which likely would belong to the same viral species or genus. Some of the sorted particles could be sequenced and others imaged. Combining observations of structural details of the capsid and binding sites, with the genomic content of viruses from the same population would, for example, shed light on host specificity and infection mechanisms. A limitation of this approach would be that particles in a flow cytometrically resolved population might represent diverse strains within a species or even represent different species with differing host ranges[173,174]. However, high throughput analysis of individual sequenced and imaged particles would help informing patterns across taxonomic levels.

Glossary terms:

IMG/VR database: Integrated data management and analysis system for cultivated and environmental viral genomes that is publicly available for the scientific community

Metagenomics: the study of sequenced nucleic acids obtained from bulk environmental samples (enriched for cells or viruses).

***Tara* expedition:** Oceanic 3-years expedition around the world to investigate planktonic and coral ecosystems in the perspective of climate changes. More than 150 international scientists have taken part.

Auxiliary metabolic genes: Cellular host genes contained in the viral genome that modulate the cellular metabolism during infection to improve viral replication.

Nucleocytoplasmic large DNA viruses (NCLDVs): Group of large DNA viruses with genomes ranging from 150 kb to 1.2 MB classified within the phylum *Nucleocytoviricota*. These viruses are referred to as nucleocytoplasmic because they are often able to replicate in both the host's cell nucleus and cytoplasm.

Contigs: High confidence overlapped DNA sequenced reads that represent a consensus region of a genome.

Flow cytometry: Technique used to detect and measure some physical and chemical features of a population of cells, viruses or particles suspended in a fluid that flow one at a time through a laser beam, where the light scattered is detected along with other fluorescence features. Sample is often fluorescently stained with cell/virus markers.

Single-amplified genomes (SAGs): Genome sequence obtained from sequencing and assembly of the amplified genetic material from an individual sorted single cell.

Multiple displacement amplification: Common whole genome amplification technique used in Single-cell genomics to amplify minute amounts of DNA. DNA synthesis and amplification is carried out by Phi29 DNA polymerase.

Virions: Complete viral particles, in their extracellular phase, and able to carry out the infectious process. Typically, the viral genome is enclosed in a protein structure (capsid) and sometimes surrounded by a lipid membrane.

Gene-transfer agents: Phage-like entities that only contain a random piece of cellular genome, that is insufficient to encode its protein components.

Consensus sequences: is the calculated order of most frequent residues, either nucleotide or amino acid, found at each position in a sequence**.**

Ultra-deep sequencing: DNA sequencing performed at very high coverage. Deep sequencing refers to sequencing a genomic region multiple times, sometimes hundreds or even thousands of times

Fosmids: Clone system based on the bacterial F-plasmid usually in *Escherichia coli* that can hold DNA insert of up to 40 kb in size

Deep chlorophyll maximum: Region below the surface of water with the maximum concentration of chlorophyll.

Viral single-amplified genomes (vSAGs): Genome sequence obtained from sequencing and assembly of the amplified genetic material from an individual sorted single viral particle.

Viral shunt: mechanism mediated by virus infection and consequently cell lysis that prevents (prokaryotic and eukaryotic) marine microbial particulate organic matter (POM) from migrating up trophic levels by recycling them into dissolved organic matter (DOM)

ToC blurb

Viruses are extremely diverse and not all of this diversity has been captured so far. In this Review, Martínez Martínez, Martinez-Hernandez and Martinez-Garcia explore the potential and limitations of single-virus genomics and how this emerging technology can complement other methods.