Single-virus genomics and beyond

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

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 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.

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

Metagenomics
The study of sequenced nucleic acids obtained from bulk environmental samples (enriched in cells or viruses).

Tara expedition
Oceanic 3-year-long expedition around the world to investigate planktonic and coral ecosystems and the impacts of global change. More than 150 international scientists have taken part.

[1]*Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA.*

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

✉*e-mail: m.martinez@ua.es*

https://doi.org/10.1038/s41579-020-00444-0

Most viruses are assumed to infect microorganisms, which are the most abundant cellular hosts on Earth, both free-living and those associated 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 (FIG. 1, top) 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 virus–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 represent only a small fraction of the total viral diversity[14–18]. The mismatch between viral diversity in the environment and the number of laboratory isolates is also evident from the number of sequenced viral genomes. As of February 2020, the US Department of Energy Joint Genome Institute IMG/VR database[19] contained 8,392 genomes of isolated viruses, three times less than the number of uncultured 'high-quality draft viral genomes' ($n = 25{,}659$) and almost 100 times less than the number of viral genomic fragments ($n = 735{,}112$) predicted from assembled shotgun metagenomes. Expanding the cultivable virus diversity is limited by the inability to grow most 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 double-stranded DNA (dsDNA) viruses with approximately 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] (FIG. 1, middle). 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 metagenomic 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 virus 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 more than 24,000 viral sequences and identified a suite of cosmopolitan virally encoded auxiliary metabolic genes that are potentially

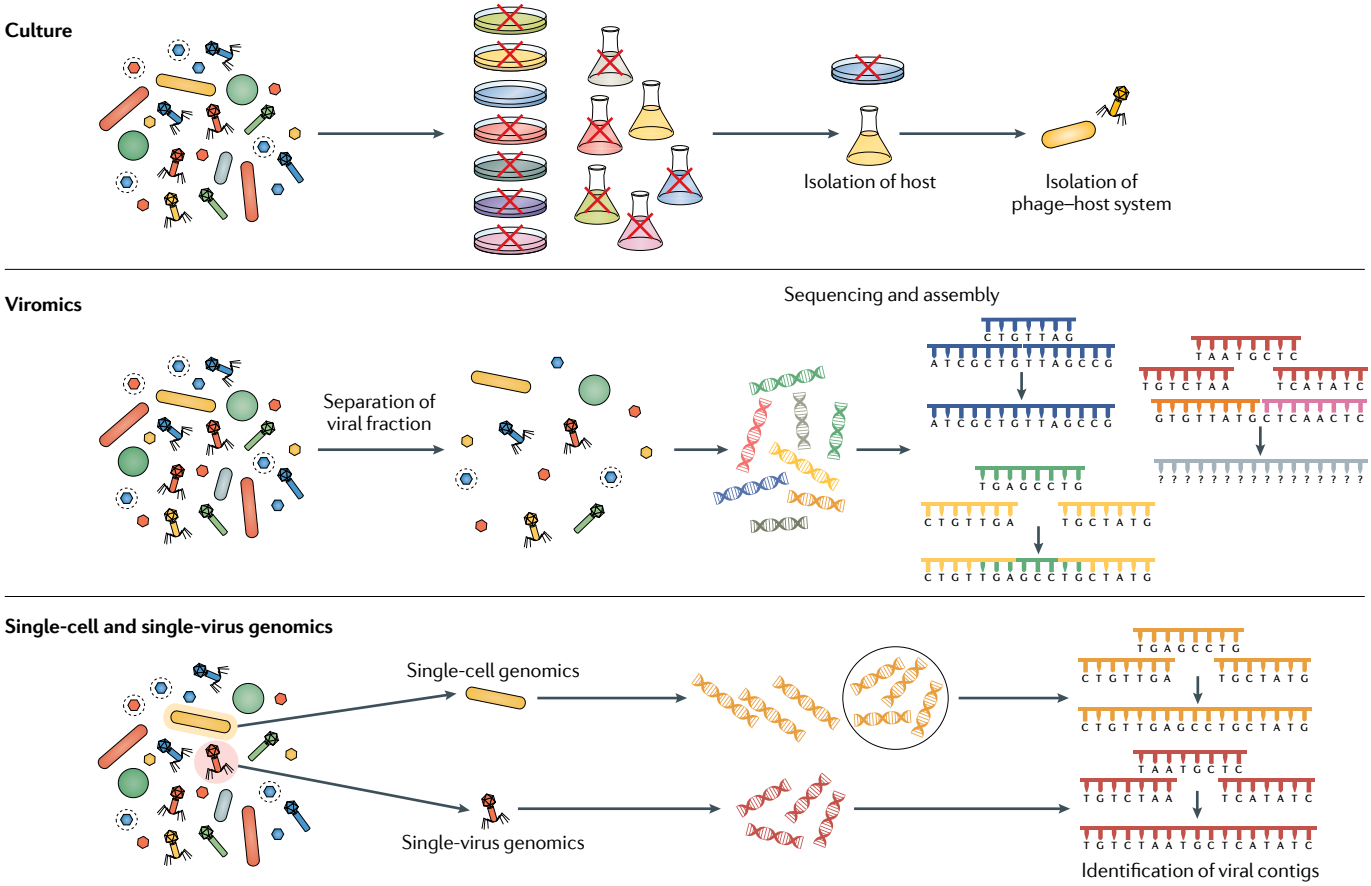

Fig. 1 | **Methods to study viruses.** Culturing has been the standard and most commonly used technique for many years to study viruses and their hosts[112,154] (top). Although undoubtedly this method is very powerful, it is restricted to the few virus–host pairs that can be cultured and maintained in the laboratory. Viromics[25,32,58] (middle), which usually requires high sample volumes, sequences nucleic acids obtained from bulk environmental samples (enriched in viruses or cells) to study uncultured microbial and viral genetic diversity. Metagenomic assembly of viruses is often 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.

involved in the metabolism of organic carbon in soil. Furthermore, metagenomics has provided insights into the viral communities in extreme environments (reviewed in REF.[33]). A recent study led by researchers at the US Department of Energy 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 microorganisms[26,34], they typically represent a low-abundance but diverse fraction of the community, which complicates the assembly of large contigs and the capture of 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 levels[36]. However, current metagenomic and bioinformatic 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[40–46] (further discussed in detail later). Most commonly, individual cells are sorted by flow cytometry and chemically lysed and their genomes are amplified (single amplified genomes (SAGs)) by multiple-displacement amplification before sequencing[47–50]. Although sorting is not targeted for infected cells, infected cells are often 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 (FIG. 1, bottom). The basic SVG workflow (FIG. 2) comprises sample collection (typically less than 1–5 ml of sample containing 10^4–10^6 virions per millilitre 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 virus particles

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

Nucleocytoplasmic large DNA viruses
(NCLDVs). A 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 cell nucleus and the host cell 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. The sample is often fluorescently stained with cell and/or virus markers.

Single amplified genomes
(SAGs). Genome sequences 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 done by Φ29 DNA polymerase.

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

(for example, by flow cytometry); lysis of virion capsids (for example, through a combination of temperature and chemical shock); whole-genome amplification (WGA; for example, multiple-displacement amplification); and sequencing and data analysis. For more details on SVG methods and protocols, see REF.[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 SVG 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 more than 1,000 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 DNA (ssDNA), dsDNA 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 is often enough[51], whereas metagenomics typically requires large volumes (litres). 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

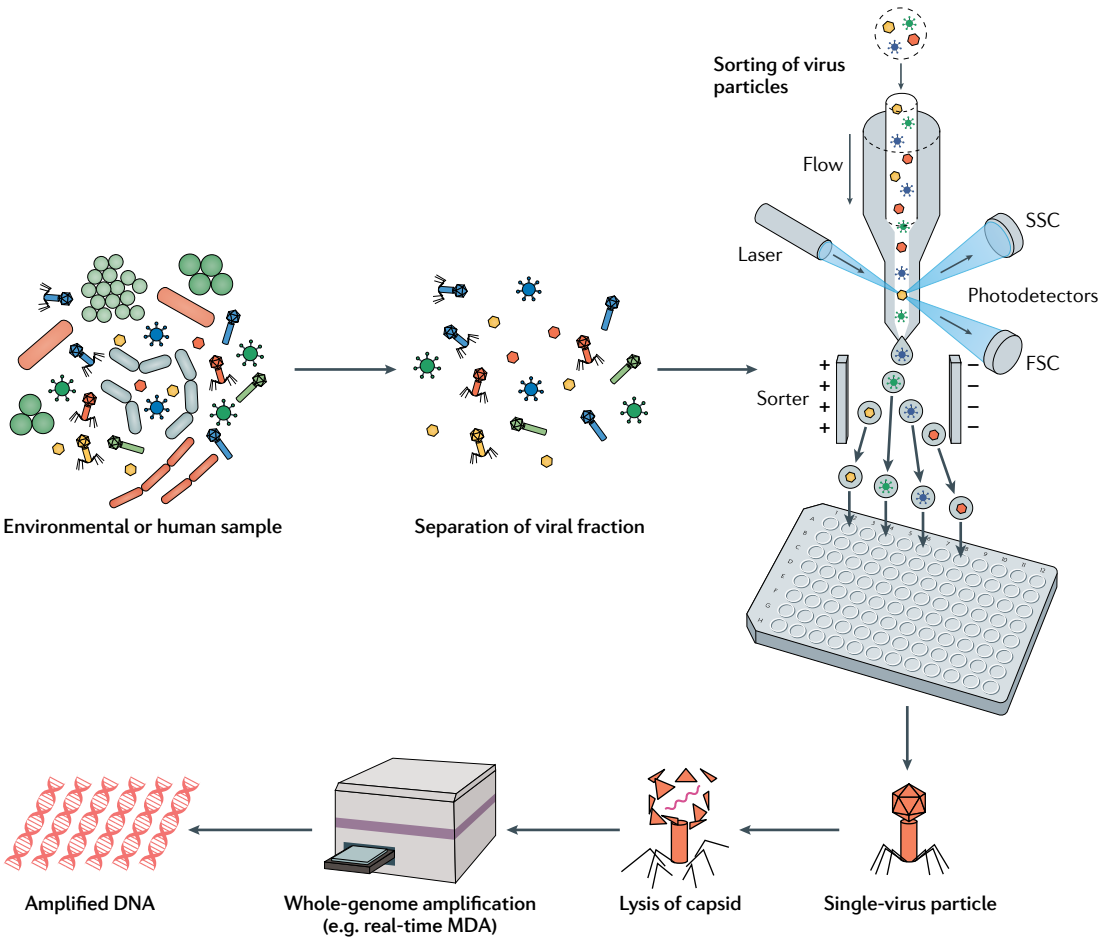

Fig. 2 | **Single-virus genomics workflow.** Viruses are fluorescently stained (for example, with SYBR Gold for double-stranded DNA viruses) and sorted one at a time from the natural sample by fluorescence-activated virus sorting or other tools such as microfluidics. Sorted single-virus particles are deposited in multiwell 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 (MDA)). After amplification, enough DNA is available for DNA sequencing or gene-target PCR screening. FSC, forward scatter; SSC, side scatter.

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 highlight a few available programs. Metavir[155], Virome[156] and MetaPhinder[157] are based on gene similarity searches and alignment of a query sequence with previously known viral sequences. Next-generation bioinformatics tools, such as VirSorter[63], try to mine unknown viruses, combining the identification of similar homologous 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 been used to 'hunt' viruses. Recent examples include the machine learning applications VirMiner[158], MARVEL[62] and 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 algorithm-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 US Department of Energy Joint Genome Institute IMG, which relies on likely the most comprehensive public viral database to date, the IMG/VR version 2.0 (REF.[19]).

Although these bioinformatics programs have been powerful and successful, they have some limitations, mainly owing to the viral database used for comparison 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 were isolated from the extremophile *Salinibacter ruber*, which inhabits environments of 15–35% NaCl, to standard marine bacteria.

Complementary analysis and manual in-house curation can help to correctly characterize and identify the viral genome fragments obtained. Careful review of the protein annotation of the supposedly viral assembled fragments to analyse 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 supports 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 the gene-sharing relationship among different viruses and helps to visualize the location and relatedness of the recovered virus against other viruses from the same environment.

~200 µl without introducing air bubbles in the fluidic lane of the instrument. Two previous studies analysed low-concentration, cryopreserved samples (~10^4 virions per millilitre) collected at 4,000-m depth in the Atlantic Ocean and Pacific Ocean[36,55], sorting 1 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 'viromic') studies include filtration through filters with 0.2–0.45-µm pore size to exclude cellular organisms[17,25,32,39,56–59] and for enrichment of virus 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 is 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 homologues 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 viral genome fragments (contigs) or, ideally, in complete viral 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 ultradeep 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 of more complete viral genomes and capturing of longer genomic islands than is 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 the sequencing of viruses with genomes larger than 50 kb (REFS[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 by SCG is a helpful cultivation-independent approach to discover new bacteria and archaea[84–89]. Following the success of 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

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

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

Ultradeep 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 a DNA insert of up to 40 kb in size.

from the staining of ssDNA and RNA virus genomes with commercially available dyes and their low side and forward scatter signals are below the 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 push the boundaries of SVG to capture hidden viral diversity. Other strategies for virus detection and sorting based on microfluidic nanodevices 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 Φ29 DNA polymerase and variations of it can amplify DNA, we are not aware of a commercially available enzyme with the required sensitivity for WGA of a single copy of an RNA virus genome. Also, no manufactured enzyme and molecular reagents are contaminant-free[99], which is crucial when one is 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 used 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 bioinformatic 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 Φ29 polymerase[101], which increased the speed, genome recovery and size of new assemblies, even for GC-rich genomes of bacteria, archaea, protists and virions[88], compared with the regular Φ29 polymerase[102]. Recent preliminary data from our laboratories indicate that this new Φ29 polymerase is an excellent choice for SVG experiments.

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 virus 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 metagenomic studies, in which hundreds or thousands of single-virus 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 virus 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. In 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 virus particles. However, we have not tested these sorters ourselves, and there are no published studies. 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 the WGA step, probably shows the best performance. With the current 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. On the basis of the recently proposed criteria for the minimum information about an uncultivated virus genome (MIUViG)[106], we conclude that genomes assembled from single-virus sequences are not finished genomes, but are rather 'genome fragments' (recovery of less than 90% of the expected genome length) or 'high-quality draft genomes' (90% or more 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' (less than 90% of the expected genome length).

Insights from single-virus genomics

Although only few SVG studies have been published, they provide important insights into viral diversity and ecology (FIG. 3). We hope the SVG case studies presented here 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 λ 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.

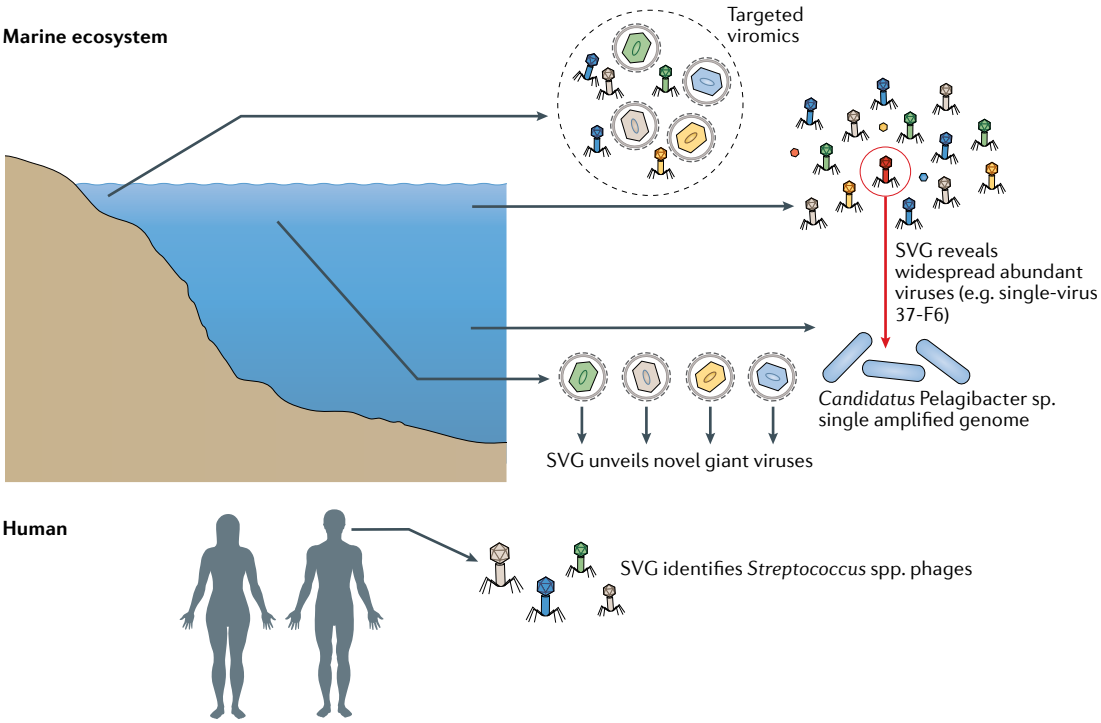

Fig. 3 | **Insights from SVG.** Marine habitats are the first ecosystem in which single-virus genomics (SVG) has been applied[36,88,90,103]. From more than 2,000 uncultured virus particles that were sorted, 44 novel viruses from the surface and the deep ocean were discovered[36]. These viruses, which had been overlooked by viromics, are highly abundant and widespread across all oceans. One of them, viral single amplified genome 37-F6, is potentially the most abundant virus in the surface ocean and represents a new virus family infecting the genus '*Candidatus* Pelagibacter'[45]. SVG has been successfully applied to study marine eukaryotic nucleocytoplasmic large DNA viruses, uncovering novel viruses, their ecology and their 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].

A few years later, a similar fluorescence-activated virus sorting and WGA strategy was used to study virus particles in a 1-ml surface seawater sample collected off the Patagonian Shelf[103]. However, this study did not investigate single-virus genomes but investigated a pool of ~5,000 virions. Flow cytometry resolved three distinct virus clusters, and they were sorted into three separate tubes before genome amplification. The study recovered sequences of uncultured giant viruses, validating the effectiveness of this approach to target specific virus 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. 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 of 2.4 Mb (REF.[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) (~20% of sorted viruses were successfully amplified) were randomly selected for Illumina sequencing. None of the 44 vSAGs matched known virus isolates or data in metagenomic 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 virus populations that dominate the oceans. Remarkably, the study also showed that the newly discovered virus vSAG 37-F6 is probably 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 SCG study demonstrated that several members of the 37-F6 viral population were present in uncultured sorted single cells belonging to the genus '*Candidatus* Pelagibacter' from different oceans[45]. Fine ecogenomic analyses indicated that virus 37-F6 was a distinct genetic population unrelated to previously

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

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

Viral shunt
Mechanism mediated by viral infection and consequently cell lysis that prevents (prokaryotic and eukaryotic) marine microbial particulate organic matter from migrating up trophic levels by recycling it into dissolved organic matter.

described pelagiphage isolates[45,112]. Now we know from SCG that 37-F6 is likely lytic and infects '*Candidatus* 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 millilitre are infected at any given time[113], with a potential carbon release from 124 fg to 4.9 pg per millilitre (assuming a 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 not been possible without looking at the most elemental component of viral communities, the single-virus 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 virus 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 individual giant ocean viruses directly recovered from a coastal seawater sample collected at Boothbay Harbour, Maine, USA. The 12 selected vSAGs that were sequenced revealed a wide viral diversity, 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 single-virus particles (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. Two recent SCG articles 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 an evolutionary history distinct from that of those that capture sunlight in cellular organisms[116]. As the giant viruses were 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 (REF.[90]), which had single-copy genomes as the starting material.

Although so far SVG has been used mainly 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 analysed by SVG combined with viromics. The 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.

Beyond single-virus genomics

Non-genomic approaches to study virus particles. Interrogating single-virus particles by SVG is only one of several techniques (BOX 2) that provide biologically meaningful information about single or quasi-individual virus 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 (FIG. 4). 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 virus particles of *E. coli* bacteriophage T5 (REF.[120]). Furthermore, Raman nanospectroscopy[121] can provide chemical and structural information at the single-virus particle level, and 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 of exploration (BOX 2).

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 microalga *Emiliania huxleyi* to promote infection[123].

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

In plaque–MSI

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 analysing 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 centre 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 infection dynamics and consequences. However, it is limited to hosts that are amenable to being grown 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 epifluorescence was used to visualize labelled host cells and virus particles, flow cytometry could make this method compatible with single-virus genomics (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 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 (nanometre to micrometre 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 investigation of RNA and single-stranded DNA 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 and cryo-electron tomography have resulted in resolution limits similar to those of X-ray crystallography. These techniques enable the generation of images from individual symmetrical and asymmetrical virions under nearly native conditions without the need to crystallize the particles[168–171]. Nanoscale-infrared spectroscopic imaging was recently used to detect and quantify subtle chemical and structural changes in single enveloped virions before membrane fusion[172]. We suggest that flow cytometry could be used to sort individual virus particles from a discrete population, on the basis of 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 could be 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 different species with differing host ranges[173,174]. However, high-throughput analysis of individual sequenced and imaged particles would help inform patterns across taxonomic levels.

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 nanometres to more than 1 μm) and they have diverse contents (for example, ssDNA, dsDNA 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 60 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" (REF.[129]). Nevertheless, viruses are often seen as extracellular virions comprising genetic material enclosed in a capsid protein. As formidably stated by Patrick Forterre, individual "viral particles reveal their viral nature only if they encounter a host" (REF.[130]). In other words, a virus without its host is likely 'condemned' 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 quantifying the fitness of viruses in relation to the proliferation of viral genomes inside cells instead of enumerating free virus particles outside cells[132]. In summary, the study of infected cells is key for understanding virus–host dynamics, and dissociating viruses (at the community, population or individual level) from their host cells provides only 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 study authors sequenced three individual protist cells from seawater, which belonged to the clade Picobiliphytes, and sequences from one of these cells were dominated by reads assigned to a widespread ssDNA 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 stramenopile lineages that are widely distributed around the oceans identified 64 non-redundant viral contigs, only seven of which could be assigned to known virus families, including an endogenous Mavirus virophage[43]. Another SCG study[44] revealed two novel

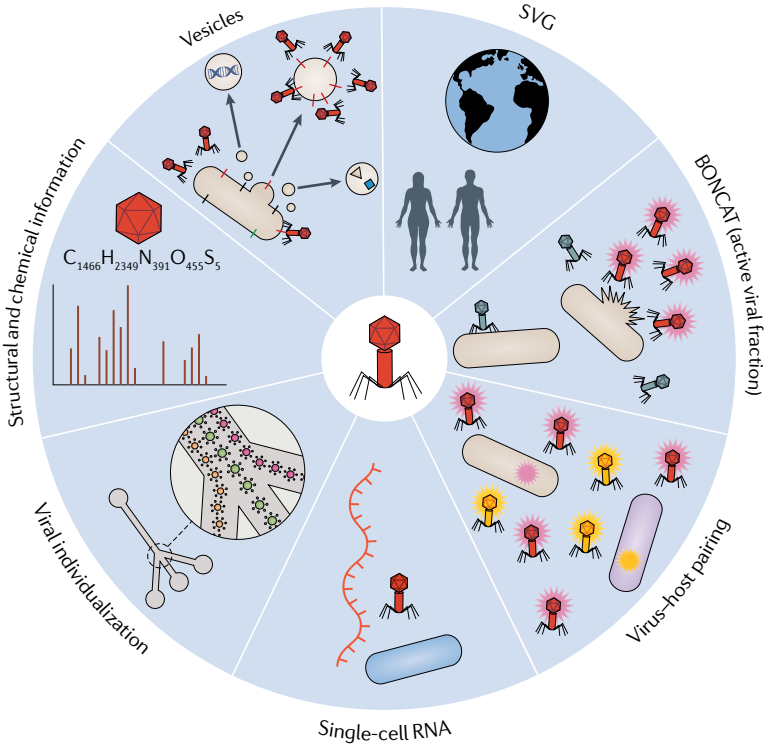

Fig. 4 | **Present and future of single-virus technologies.** Other single-virus technologies (BOX 2) exist beyond single-virus genomics (SVG), and these methods have the potential to complement each other in the study of viruses and in other fields (for example, in vesicle research). Combination of 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 scale or at the individual scale, from environmental samples could be sorted by flow cytometry and analysed through the SVG pipeline to sequence and identify active viruses produced under certain environmental conditions. Other promising examples based on advances in nanotechnology combined with mass spectrometry[120] or Raman spectroscopy[121] at the level of individual virus particles will support the analysis of the chemical structure of viruses.

circular DNA viruses in the genome of eukaryotic SAGs from the genera *Micromonas* and *Ostreococcus*.

There is a wealth of information on viruses and their microbial hosts from different environments, ranging from extreme ecosystems[134] to the human microbiota[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) by analysing 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 in nature than previously thought[136]. Recently, SCG combined with metagenomics was applied to unveil host–virus interactions in hot spring biofilms[137]. Around 26% of analysed single cells contained a viral contig, and the 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, WGA and sequencing. With this method, a study[135] identified 363 unique phage–host pairs in the human gut and demonstrated that most phages bind to only one bacterial species, limiting the risk of horizontal gene transfer between species (for example, of antibiotic resistance genes). In addition, a high level of cross-reactivity between phages and bacteria from different individuals was observed despite interindividual variability in the human microbiota. PhageFISH[139] or VirusFISH uses fluorescence in situ hybridization with specific probes targeting intracellular and extracellular viral DNA to monitor infection dynamics and virus production. Although host cells can be identified with use of ribosomal RNA probes, prior knowledge of virus genetic information is required to design target-specific DNA probes.

Studying virus–host pairs goes beyond DNA; for example, single-cell RNA sequencing has revealed 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 sequencing of herpex simplex virus 1-infected cells has shed some light on the early stages of infection and connected metabolic activation of viral pathways with antiviral programmes[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 the 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 the screening of thousands of individual virus–cell interactions[144]. Studying free single-virus 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], such as microdiversity (as discussed herein), 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 to 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 the key to launching 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, immune system disorders[151] and cancer[152,153].

SVG has the potential to change our understanding of viral genetic diversity and viral infection in the environmental, agricultural and medical virology fields. Specifically, high-throughput SVG could speed up discovery by potentially delivering almost complete genomes of uncultured viruses; improving the coverage of viruses under-represented 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; and facilitating 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 questions about viral diversity (and microdiversity), evolution, adaptation, and ecology to be addressed in a more effective and comprehensive manner than previously possible.

Published online 6 October 2020

1. Koonin, E. V. The wonder world of microbial viruses. *Expert Rev. Anti Infect. Ther.* **8**, 1097–1099 (2010).
2. Yong, E. *I Contain Multitudes: The Microbes Within Us and A Grander View of Life* (Ecco, 2016).
3. Breitbart, M. & Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13**, 278–284 (2005).
4. Paez-Espino, D. et al. Uncovering Earth's virome. *Nature* **536**, 425–430 (2016).
This is a massive metagenomic study on global viral diversity and distribution and host specificity of viruses. A total of 125,000 partial DNA virus genomes are discovered.
5. Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nat. Rev. Microbiol.* **3**, 504–510 (2005).
6. Suttle, C. A. Marine viruses-major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
This is a fundamental must-read review of the general role of viruses in marine ecosystems.
7. Abedon, S. T. *Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses* (Cambridge Univ. 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. *Environ. Microbiol.* **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. *Environ. Microbiol.* **14**, 426–440 (2012).
12. Marston, M. F. et al. Rapid diversification of coevolving marine Synechococcus and a virus. *Proc. Natl Acad. Sci. USA* **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. *Nat. Commun.* **9**, 1–11 (2018).
14. Rappé, M. S. & Giovannoni, S. J. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**, 369–394 (2003).
This is a comprehensive review addressing a fundamental question in microbial ecology on the difficulty of culturing most microorganisms in the laboratory and how this bias impacts microbial discovery.
15. Pedrós-Alió, C. The rare bacterial biosphere. *Ann. Rev. Mar. Sci.* **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. *ISME J.* **7**, 1738–1751 (2013).
17. Brum, J. R. et al. Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498 (2015).
This is a pioneering, comprehensive metagenomic study on global marine viral diversity from hundreds of samples collected during the *Tara* expedition.
18. Trubl, G. et al. Soil viruses are underexplored players in ecosystem carbon processing. *mSystems* **3**, e00076-18 (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 Res.* **47**, D678–D686 (2019).
This article describes the most comprehensive genome database of uncultured viruses recovered by metagenomics from different ecosystems, including the human body, with more than 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. *Nat. Ecol. Evol.* **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? *Curr. Opin. Virol.* **3**, 566–571 (2013).
The authors address a hot topic in viral ecology (that is, how big the viral diversity in nature is) and estimate the total number of different viral proteins, which is a proxy for quantifying the 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. *Nat. Microbiol.* **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. *Nat. Commun.* **9**, 1–9 (2018).
The article reports the discovery of several relevant giant viruses, including one with a genome of 2.4 Mb, using metagenomics and a method that is similar to those used in SVG, but in this case targeting multiple sets of 100 viruses, instead of single-virus 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. *Nat. Commun.* **7**, 1–11 (2016).
29. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **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. *Proc. Natl Acad. Sci. USA* **99**, 14250–14255 (2002).
33. Dávila-Ramos, S. et al. A review on viral metagenomics in extreme environments. *Front. Microbiol.* **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. *Sci. Rep.* **9**, 1–9 (2019).
35. Simmonds, P. et al. Consensus statement: virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **15**, 161–168 (2017).
36. Martinez-Hernandez, F. et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat. Commun.* **8**, 1–13 (2017).
This is a pioneering reference high-throughput SVG study that unveils extremely abundant and ubiquitous 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 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 Genet.* **13**, e1007018 (2017).
40. Labonté, J. M. et al. Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. *ISME J.* **9**, 2386–2399 (2015).
The screening of sequencing data from hundreds of single cells obtained from seawater unveils virus–host interactions in different ecologically important bacterial and archaeal 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).
This is the first report of SCG 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. *Mol. Ecol.* **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. *Phil. Trans. R. Soc. B* **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. *ISME J.* **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. *Proc. Natl Acad. Sci. USA* **99**, 5261–5266 (2002).
48. Raghunathan, A. et al. Genomic DNA amplification from a single bacterium. *Appl. Environ. Microbiol.* **71**, 3342–3347 (2005).
49. Stepanauskas, R. & Sieracki, M. E. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Natl Acad. Sci. USA* **104**, 9052–9057 (2007).
50. Rinke, C. et al. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* **9**, 1038–1048 (2014).
51. Martinez-Garcia, M., Martinez-Hernandez, F. & Martínez Martínez, J. Single-virus genomics: studying uncultured viruses, one at a time. *Ref. Module Life Sci.* https://doi.org/10.1016/b978-0-12-809633-8.21497-0 (2020).
The authors provide methodological details and protocols for implementing SVG to complement other existing methods in viral ecology.
52. Lindell, D. et al. Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc. Natl Acad. Sci. USA* **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. *Nat. Rev. Microbiol.* **13**, 147–159 (2015).
This review is recommended for readers who would like an introduction to 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. *Front. Microbiol.* **10**, 1801 (2019).
56. Aylward, F. O. et al. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc. Natl Acad. Sci. USA* **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**, e01903-17 (2017).
58. Angly, F. E. et al. The marine viromes of four oceanic regions. *PLoS Biol.* **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 Biol.* **3**, 130160 (2013).
61. Zolfo, M. et al. Detecting contamination in viromes using ViromeQC. *Nat. Biotechnol.* **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. *Front. Genet.* **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. *Front. Microbiol.* **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. *Environ. Microbiol.* **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 pioneering study successfully combines long-read and short-read sequencing data to improve viral metagenomic assemblies and shows the potential of Nanopore sequencing data to advance virus discovery.
72. Beaulaurier, J. et al. Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Res.* **30**, 437–446 (2020).
73. Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E. & Ghai, R. Expanding the marine virosphere using metagenomics. *PLoS Genet.* **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. *Front. Microbiol.* **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–e00816 (2016).
77. Martinez-Garcia, M. et al. High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *ISME J.* **6**, 113–123 (2012).
78. Stepanauskas, R. Single cell genomics: an individual look at microbes. *Curr. Opin. Microbiol.* **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. *Sci. Rep.* **9**, 1–11 (2019).
80. Lasken, R. S. Genomic sequencing of uncultured microorganisms from single cells. *Nat. Rev. Microbiol.* **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. *Sci. Rep.* **7**, 1–14 (2017).
82. Mangot, J. F. et al. Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. *Sci. Rep.* **7**, 1–12 (2017).
83. Seeleuthner, Y. et al. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nat. Commun.* **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 article is an excellent example of the power of single-cell technologies to provide biological insights into uncultured microorganisms.
85. Swan, B. K. et al. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc. Natl Acad. Sci. USA* **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. *ISME J.* **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. *Nat. Commun.* **8**, 1–10 (2017).
The authors use flow cytometry to sort uncultured single viruses and they amplify their genomes with a new variant of an efficient Φ29 enzyme, which is commonly used in SCG and SVG. This study is another SVG 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 J.* **8**, 2503–2516 (2014).
90. Wilson, W. H. et al. Genomic exploration of individual giant ocean viruses. *ISME J.* **11**, 1736–1745 (2017).
This reference SVG study targets for the first time uncultured giant viruses in nature, which are commonly 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).
This is the first study on SVG applied to the human virome. The authors implement this novel technology, combined with metagenomics, in salivary human samples and discover important, abundant phages.
92. Allen, L. Z. et al. Single virus genomics: a new tool for virus discovery. *PLoS ONE* **6**, e17722 (2011).
This is the first report showing the feasibility of SVG as a new tool for virus discovery. The authors successfully use this technology to sequence several single sorted virus particles of viral isolates T4 and λ of *E. 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. *Appl. Environ. Microbiol.* **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. *J. Extracell. Vesicles* **4**, 25530 (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 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., Payet, J. P., Winter, C. & Weinbauer, M. G. Quantification of aquatic viruses by flow cytometry. *Man. Aquat. Viral Ecol.* **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 Microbiol. Ecol.* **89**, 495–515 (2014).
99. Blainey, P. C. & Quake, S. R. Digital MDA for enumeration of total nucleic acid contamination. *Nucleic Acids Res.* **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 Eng. Des. Sel.* **29**, 617–628 (2016).
102. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**, 175–188 (2016).
This is one of the most comprehensive technical and scientific reviews of SCG technologies of unicellular and multicellular organisms, and discusses how these technologies have enabled new discoveries in multiple fields from microbiology to cancer or immunology.

103. Martínez Martínez, J., Swan, B. K. & Wilson, W. H. Marine viruses, a genetic reservoir revealed by targeted viromics. *ISME J.* **8**, 1079–1088 (2014). **This study uses technologies similar to those used in SVG 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. *J. Comput. Biol.* **19**, 455–477 (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). *Nat. Biotechnol.* **37**, 29–37 (2019).
107. Hercher, M., Mueller, W. & Shapiro, H. M. Detection and discrimination of individual viruses by flow cytometry. *J. Histochem. Cytochem.* **27**, 350–352 (1979).
108. Lippé, R. Flow virometry: a powerful tool to functionally characterize viruses. *J. Virol.* **92**, e01765-17 (2017).
109. Koonin, E. V. & Yutin, N. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Adv. Virus Res.* **103**, 167–202 (2019).
110. Brum, J. R. et al. Illuminating structural proteins in viral 'dark matter' with metaproteomics. *Proc. Natl Acad. Sci. USA* **113**, 2436–2441 (2016).
111. Alonso-Sáez, L., Morán, X. A. G. & Clokie, M. R. Low activity of lytic pelagiphages in coastal marine waters. *ISME J.* **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. *Environ. Microbiol. Rep.* **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. *Appl. Environ. Microbiol.* **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. *Phil. Trans. R. Soc. B* **374**, 20190086 (2019).
116. Needham, D. M. et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl Acad. Sci. USA* **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). *Proc. Natl Acad. Sci. USA* **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. *Environ. Microbiol.* **16**, 2568–2590 (2014).
119. Pasulka, A. L. et al. Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. *Environ. Microbiol.* **20**, 671–692 (2018).
120. Dominguez-Medina, S. et al. Neutral mass spectrometry of virus capsids above 100 megadaltons with nano-mechanical 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. *J. Virol.* **88**, 8504–8513 (2014).
123. Schatz, D. et al. Communication via extracellular vesicles enhances viral infection of a cosmopolitan alga. *Nat. Microbiol.* **2**, 1485–1492 (2017).
124. Berleman, J. & Auer, M. The role of bacterial outer membrane vesicles for intra- and interspecies delivery. *Environ. Microbiol.* **15**, 347–354 (2013).
125. Van Niel, G., D'Angelo, G. & Raposo, G. Shedding light on the cell biology of extracellular vesicles. *Nat. Rev. Mol. Cell Biol.* **19**, 213–228 (2018).
126. Machtinger, R., Laurent, L. C. & Baccarelli, A. A. Extracellular vesicles: roles in gamete maturation, fertilization and embryo implantation. *Hum. Reprod. 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 J.* **11**, 394–404 (2017).
128. Kulp, A. & Kuehn, M. J. Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annu. Rev. Microbiol.* **64**, 163–184 (2010).
129. Jacob, F. & Wollman, E. L. Viruses and genes. *Sci. Am.* **204**, 93–107 (1961).
130. Forterre, P. The virocell concept and environmental microbiology. *ISME J.* **7**, 233–236 (2013).
131. Forterre, P. Manipulation of cellular syntheses and the nature of viruses: the virocell concept. *C. R. Chim.* **14**, 392–399 (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 Evol.* **5**, vez006 (2019).
133. Martinez-Garcia, M. et al. Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. *ISME J.* **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'. *Nat. Commun.* **5**, 1–8 (2014).
135. Džunková, M. et al. Defining the human gut host–phage network through single-cell viral tagging. *Nat. Microbiol.* **4**, 2192–2203 (2019). **This is probably one of the most comprehensive SCG studies within the context of the human gut microbiota, and unveils a total of 363 unique host–phage pairings, expanding the known host–phage network of the gut microbiota**.
136. Munson-Mcgee, J. H. et al. A virus or more in (nearly) every cell: Ubiquitous networks of virus-host interactions in extreme environments. *ISME J.* **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. *ISME J.* **14**, 2527–2541 (2020).
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. *Environ. Microbiol.* **15**, 2306–2318 (2013).
140. Zanini, F. et al. Virus-inclusive single-cell RNA sequencing reveals the molecular signature of progression to severe dengue. *Proc. Natl Acad. Sci. USA* **115**, E12363–E12369 (2018).
141. Steuerman, Y. et al. Dissection of influenza infection in vivo by single-cell RNA sequencing. *Cell Syst.* **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. *Nat. Commun.* **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. *Sci. Rep.* **7**, 1–11 (2017).
144. Liu, W. et al. More than efficacy revealed by single-cell analysis of antiviral therapeutics. *Sci. Adv.* **5**, eaax4761 (2019).
145. Lasken, R. S. Single-cell genomic sequencing using multiple displacement amplification. *Curr. Opin. Microbiol.* **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. *Proc. Natl Acad. Sci. USA* **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. *Nat. Microbiol.* **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. *Nat. Neurosci.* **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. *Nat. 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* **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. *Nat. Biotechnol.* **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. *Nat. Microbiol.* **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. *Microbiol. Mol. Biol. Rev.* **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 Rev. Proteom.* **13**, 325–340 (2016).
167. Cheng, W. & Schimert, K. A method for tethering single viral particles for virus-cell interaction studies with optical tweezers. *Proc. SPIE* **10723**, 107233B (2018).
168. Ekeberg, T. et al. Three-dimensional reconstruction of the giant mimivirus particle with an X-ray free-electron laser. *Phys. Rev. Lett.* **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. *J. Biol. Chem.* **294**, 5181–5197 (2019).
171. Subramaniam, S., Bartesaghi, A., Liu, J., Bennett, A. E. & Sougrat, R. Electron tomography of viruses. *Curr. Opin. Struct. Biol.* **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. *Appl. Environ. Microbiol.* **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. *Aquat. Microb. Ecol.* **74**, 17–28 (2015).

Acknowledgements

This work was supported by the Gordon and Betty Moore Foundation (grant 5334), the US National Science Foundation (NSF-OPP 1644155, NSF-OCE 1933289), the Spanish Ministry of Economy and Competitiveness (CGL2013-40564-R, RTI2018-094248-B-I00 and SAF2013-49267-EXP) and Generalitat Valenciana (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 Microbiology thanks Julio Cesar Ignacio Espinoza and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.