Chapter

Single-cell Genomics for Uncovering Relationships between Bacteriophages and their Hosts

Mária Džunková

Abstract

Microbial single-cell genomics represents an innovative approach to study microbial diversity and symbiosis. It allows us to recover genomes of microbes possessing specific features of our interest, or detect relationships between microbes found in close proximity to each other (one microbe inside of the other or microbes attached to each other). It can be used for linking phages with their bacterial hosts in different kinds of environmental samples, which often contain an enormous diversity of yet uncultured bacterial species and novel bacteriophages. In the typical microbial single-cell genomics workflow, fluorescence-activated cell sorting (FACS) is used to collect bacterial cells of interest, based on their cell size, internal granularity, or fluorescence. Femtograms of DNA from each sorted particle are then amplified up to the quantities required by the standard sequencing library preparation kits. Single-cell assemblies then reveal presence of phages in sorted bacterial cells. In case of highly abundant viral species, single-cell genomics can be coupled with metagenomics (shotgun sequencing of the total microbial community), which can provide insights into the bacteria-bacteriophage population fluctuations in time or space. In this chapter, we explain the details of uncovering relationships between bacteriophages and their hosts coming from so-called viral or bacterial dark matter.

Keywords: microbial single-cell genomics, single-amplified genomes, fluorescence activated cell sorting, bacteriophages, microbial dark matter

1. Introduction

Bacteriophages (the viruses of bacteria) influence biogeochemistry across all environments on Earth, and can also affect our health. They contribute to the bacterial evolution, impact ecosystems by killing their bacterial hosts, and have enormous industrial and pharmaceutical potential [1]. Thousands of novel phages are being discovered daily thanks to the recent advances in the sequence-based recovery of genomes of yet uncultured microbes in environmental samples [2]. In the typical workflow, the viral-like contigs are extracted from sequence assemblies of metagenomes, which contain DNA sequences from all microbes in a given sample, or from viromes, which are samples enriched for the particles smaller than 0.2 μm (**Figure 1**) [3]. The largest database of viral-like contigs derived from metagenomes, IMG/VR,

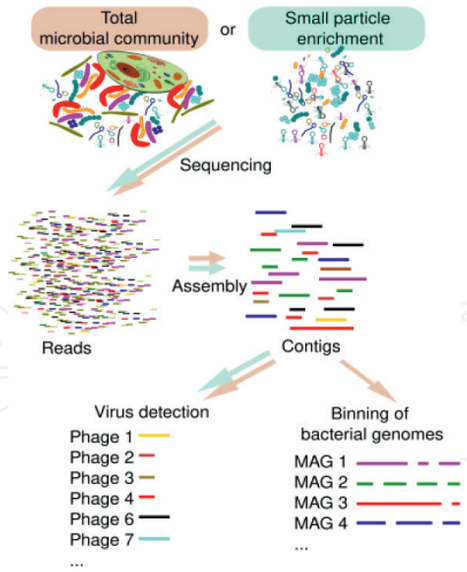

Figure 1.
Sequence-based recovery of microbial genomes.

currently contains nearly 3 million viral-like contigs and this number is increasing exponentially [4]. It is important to say that also the relatively well-studied environments formed by bacterial groups with cultured representatives, such as the human gut, harbor thousands of yet undiscovered phages [5].

For a long time, our knowledge of phages in the human gut has been limited to phages with easily culturable hosts. Then, scientists started to compare metagenomic samples from healthy volunteers with samples from patients suffering from different diseases, and many of these diseases resulted to be associated with novel uncultured phages targeting unknown hosts [6]. This suggests that the phages have an enormous potential to influence human health indirectly, by shaping the bacterial composition of the human microbiome. Therefore, the idea of employing phages in clinical practice is attracting a lot of scientific attention. Phages isolated by culture methods in the laboratory can be used for elimination of multidrug resistant pathogenic bacteria affecting organs with low number of commensal bacteria, such as the lungs or skin [7]. In the case of more complex microbiomes, such as the human gut and fecal microbiota transplant (FMT), is applied to aim to change the whole gut microbiome composition [8]. Each preparation of fecal material from a healthy donor is free of common pathogens but harbors an unknown diversity of bacteria and phages. Few clinical experimental studies showed that 0.2 μm filtered FMT preparation (containing only phages) can have the same beneficial effects as the traditional unfiltered FMT preparation containing bacterial cells, which suggests that phages play an important role in restoration of the healthy human gut, but the identity of their bacterial hosts remains unknown [9, 10].

It is intriguing that our knowledge of the phage biology is not catching up to speed with the sequence-based discoveries. For studying the biology of the novel phages, we first need to identify their bacterial hosts, which is traditionally done by plaque assays. Plaque assays are the most straightforward method for testing interaction of phages with culturable bacteria [11]. Bacterial culture is mixed with phage particles in agarose and distributed evenly on a standard agar plate, and after incubation, zones of clearing (plaques) appear in the bacterial lawn on the agarose overlay. Nevertheless, it is widely known that the plaque assays do not capture all viruses able

to infect the given bacterial strain [12]. The most important drawback of the plaque assays is that they cannot be used for uncultured bacteria, since in many environments, as much as 99.9% of bacterial species do not have any previously cultured representatives [13]. Even if a species of our interest has several easily culturable representatives, which is typical for the human gut bacteria, isolating new strains of the same species can be complicated by bacterial community complexity and strain-specific culture media requirements. Consequently, the culture-based approaches for studying phage-bacterial interactions cannot keep up with the rapidly increasing number of novel phages discovered by sequence-based methods.

2. Methods for linking bacteriophages with their hosts

2.1 Computational methods

Computational biologists are currently trying to develop highly efficient methods for linking millions of recently discovered viral-like contigs with their bacterial hosts avoiding the need to culture them in the laboratory. The first step is the recovery of high-quality host genomes from metagenomes. If a microbial community is simple enough, it is possible to recover nearly full genomes of bacteria by assembling sequencing reads into longer contigs [14]. However, in environments with a more complex species composition, binning algorithms must be applied to organize the contigs into larger sets, which results in metagenome-assembled genomes – MAGs (**Figure 1**) [15]. Nowadays, articles published in high impact journals report hundreds of thousands of new MAGs at once, providing an enormous source of reference genomes of uncultured microbes for the whole scientific community [16–18]. There are initiatives for normalized taxonomic classification of these recently discovered bacterial genomes, for example, the genome taxonomy database project GTDB, which allows an objective assessment of novelty of a MAG based on a large set of single-copy marker genes [19]. Sequence databases currently contain thousands of uncultured bacterial species, and each species surely consists of thousands of strains, most of them are not yet been discovered. Each of these strains can be infected by several bacteriophages, which are also yet to be discovered.

The genomes of phages or their hosts contain genetic features revealing links between them. These signatures are acquired during the phage infection, which can occur in different ways (**Figure 2**). Phages use bacterial cell machinery for forming virions (viral particles) and releasing them from the cell. This can be done without destroying the host cells (chronic infections), or through host cell lysis (lytic infection). Chronic infections can impact the cell growth, but aggressive lytic infections can reduce the host population significantly. Some phages can integrate into the host genome and replicate along with the host chromosome as prophages without producing any virions, which is called lysogenic cycle. In general, phages can switch between lysogenic and lytic cycles depending on ecological factors influencing growth of their bacterial host (**Figure 2**) [20].

Integrated prophages can be easily detected in about one third of the genomes of uncultured bacteria, nonetheless, hosts of lytic phages cannot be identified by this method [21]. Lytic phages can be linked to their uncultured hosts by several genetic signatures found in the bacterial or phage genomes. The first type of these genetic signatures is the clustered regularly interspaced short palindromic repeats (CRISPR) arrays detectable in the bacterial genomes, which represent an evidence

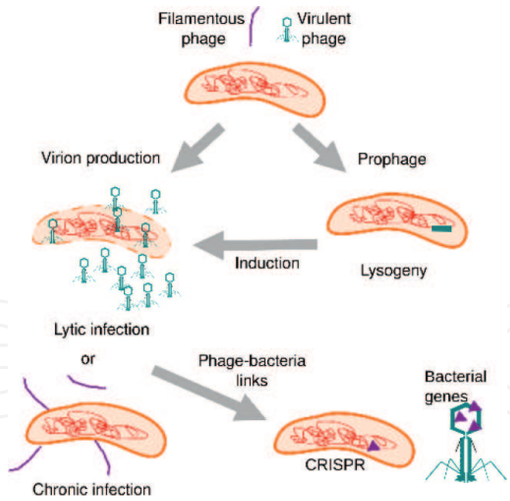

Figure 2.
Different types of phage infection.

of the previous phage infection. However, the percentage of bacterial genomes harboring a CRISPR array is quite low; the estimates vary from 30% in species from the human gut microbiome 5–10% of all bacterial lineages from different environments [22]. The second type of genetic signatures providing phages-hosts links is the genes acquired by the phages from bacteria during the past infections, for example, tRNA sequences [23], sequences of ribosomal proteins [24], or so-called auxiliary metabolic genes [25]. Nevertheless, these sequences do not provide strain-level host resolution and are present only in a small fraction of all viruses [4, 26]. As a consequence of this, the computational host-range predictions failed to reveal hosts for 85% of the millions of bacteriophages in the IMG/VR database [26]. In addition, these computational host-range predictions have another big disadvantage: they show links acquired during the past infections; however, we do not know how these phages adapted to constantly evolving bacterial hosts and changing environmental conditions [27, 28].

2.2 Experimental methods

Fortunately, a certain portion of cells collected directly from an environment contain viruses in the cell or attached to the cell, which provides evidence for their relationship. There are several experimental approaches for observing these interactions in their natural environment [29]. The first option is the digital PCR applied to emulsion droplets containing bacterial cells, in which the infection is detected by primers targeting a specific phage [30]. Digital PCR can be applied only to phages with previously known sequences, thus does not provide insights into the full diversity of phages able to attack one bacterial species. An example of a method, which is not limited to the previously known phages, is the proximity ligation MetaHiC [31]. It is a combination of experimental and computational approaches revealing which DNA molecules had physical contact in the cell, such as phages infecting the hosts. MetaHiC method is applied to the whole environmental sample and is limited to highly abundant bacteria and viruses, which are detectable by the traditional shotgun metagenomic sequencing.

The most advantageous approach for studying phage-bacteria interaction is single-cell genomics. In the typical single-cell genomics workflow,

fluorescence-activated cell sorting (FACS) is used to collect bacterial cells of interest, based on their cell size, internal granularity, or fluorescence, which is analyzed by the FACS instrument at a speed of several thousand of cells per second. The cells are sorted into 96 or 384 well plates and DNA is released from the cells by alkaline lysis. Afterward, a mixture of random hexamers and phi29 polymerase is applied to the single cells in an isothermal reaction of 4–12 hours to enrich the DNA by whole genome amplification (WGA). The femtograms of DNA from one cell are amplified up to the quantities required by the standard sequencing library preparation kits. Content of each single-cell is sequenced separately, resulting in so-called single amplified genomes SAGs (**Figure 3**) [32]. The SAG is then searched for viral-like contigs, which enables us to get links between previously unknown viruses and their uncultured hosts. Microbial single-cell genomics can be targeted toward minor bacterial groups by specific fluorescent probes, thus, it is not limited to highly abundant species as the shotgun metagenomics.

To avoid confusion, it is important to mention that microbial single-cell genomics largely differ from the single-cell genomics of human cells, which is widely used for characterization of gene expression in single cells from cancer tissues [33]. Expressed genes in eukaryotic organisms can be amplified through oligo(dT) primers; however, this is impossible in the case of bacteria, which lack the poly(A) tails in their transcripts. While the single-cell sequencing in cancer research is becoming more accessible for small labs and is moving into high-throughput scale [34], the single-cell genomics of bacterial cells is routinely managed only in few laboratories in the world, although has potential to be used more widely, especially if the requirements for sterile FACS sorting are fulfilled.

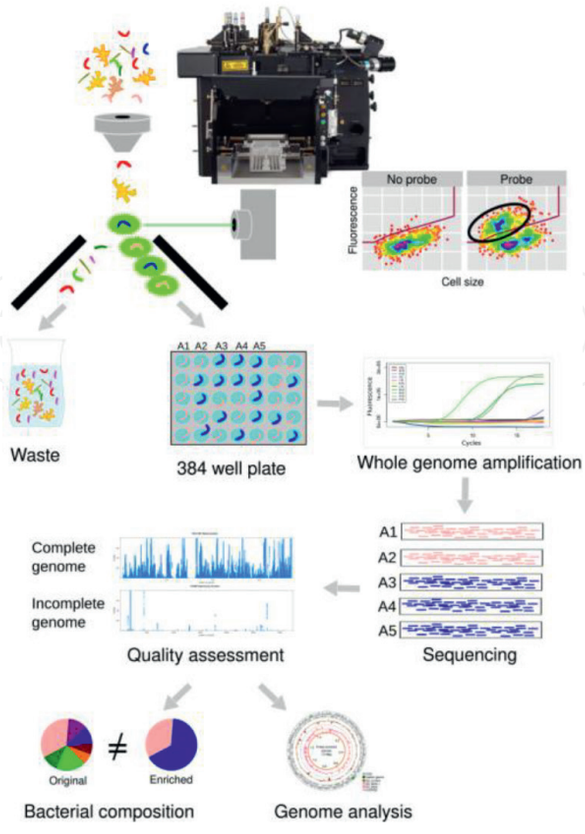

Figure 3.
Single-cell genomics workflow.

2.3 Assessing host-phage relationships from single-cell genomics data

Single-cell genomics in the context of studying viruses has been successfully applied to several habitats, ranging from seawater to hot springs [35, 36]. This approach has shed light on important aspects of viral biology, such as horizontal gene transfer, including the ability of viruses to reprogram their host's energy metabolism [37, 38].

If a phage-bacteria pair, which was detected on a single sorted particle, has a high abundance in the environmental sample, it is possible to assess the lifestyle of the phage by coupling the single-cell data with metagenomics performed on a longitudinal sample series or on samples collected in close proximity to each other, for example, different layers of a sediment or different soil layers. An example is a study on phage-host relationships in a hot spring microbial mat in California characterized by a layer-specific bacterial composition and high cellular density [39]. The single-cell genomics demonstrated that one quarter of microbial cells in this mat contained viral contigs. By mapping metagenomic reads from different mat layers to the sequences of virus-host pairs obtained by single-cell genomics, a low mobility of the viruses across the mat layers and a low copy number of viral genomes compared to their hosts were revealed (**Figure 4**). The stable host-phage ratio suggested that the lysogeny was the predominant lifestyle of these phages, or that these phages form only few virions during the infection, so they do not outnumber the host cells. If the phages replicate in an aggressive way, their genome coverage would be higher than the genome coverage of their host. The opposite situation, in which the host genome has higher coverage than the phage, would mean that the phage is infecting only a fraction of the total host population, or is specialized only to certain strains of the host species, which are not distinguishable by metagenomics.

In some cases, phage lifestyle can be assessed directly by looking at the final completeness of the bacterial genome and the time when the fluorescence of WGA passes the critical point (Cp) detectable by the qPCR instrument. Bacterial genomes are not amplified uniformly due to the nature of the WGA reaction – some genomic regions will be over-amplified, while others will be absent in the final assembly. Nevertheless, this downside does not represent a big issue for the genome analysis – there are computational tools for estimating genome completeness, for example, CheckM, thus genomes with low completeness can be removed from the following analysis [40]. Normally, quickly amplifying wells (low Cp) result in bacterial genomes with a high genome completeness, while bacterial cells in wells with high Cp will have low genome completeness (**Figure 5**). If a well on the 384-well plate contains an easily accessible DNA fragment in several copies, such as phage replicating inside the bacterial cell, WGA reagents will preferentially enrich the phage genome rather than the bacterial genome, thus such a well will reach the Cp faster than the rest of the wells

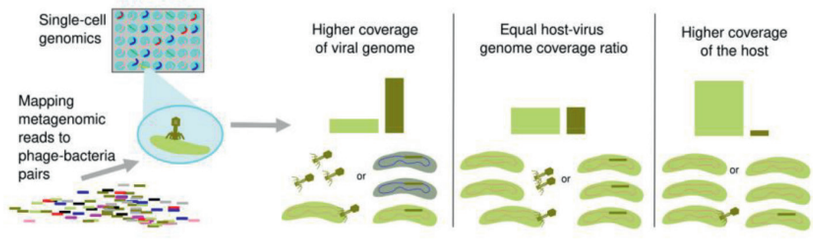

Figure 4.
Phage lifestyle assessed from single-cell genomics and metagenomics.

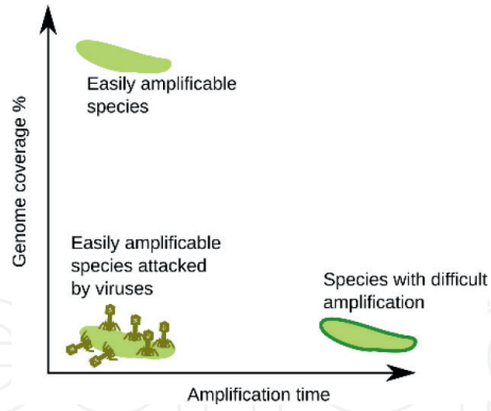

Figure 5.
Active phage infection captured by single-cell genomics.

on the plate, but it will have lower bacterial genome completeness (**Figure 5**). This phenomenon was observed in four out of 57 single cells from marine surface bacterioplankton, which is an environment with a high rate of lytic phage infections [35]. In contrast, it has not been detected in the single cells from the hot spring microbial mat, where lysogenic lifestyle prevails [39].

2.4 Viral tagging

The major advantage of the single-cell genomics is that it can be coupled with fluorescent probes targeting a specific subset of the total uncultured bacterial community, which possesses features of our interest (**Figure 3**) [32]. Fluorescent probes provide an experimental evidence of the tested feature, for example, the ability of the microbes to degrade cellulose [41] or to stimulate human immune system [42]. In addition, targeted single-cell genomics can lead to enrichment of low abundant bacteria with specific features of our interest, which would not be recovered by metagenome binning [43]. Targeted single-cell genomics can also employ phages as fluorescent tags. Phages stained with a generic nucleic stain can determine which bacteria are susceptible to phage attachment. This method is called viral tagging. The viral tagging represents a big advantage compared to detection of phages accidentally attached to the host cells in nature because the fluorescence provides evidence that the phages were present in the form of virions at the moment of the cell sorting, while detection of phages naturally occurring on single cells might be biased toward the prophages.

The first viral tagging experiments were performed in the 90's when infection of bacteria by stained phages was observed by a microscope [44]. A high throughput version of the viral tagging was developed two decades later and it involved FACS [45]. In a study published in Nature in 2014 [46], fluorescently labeled environmental phages were mixed with *Synechococcus*, a marine species cultured in a media containing isotopic nitrogen prior to the experiment. Phages, which were not attached, were removed prior to FACS by centrifugation, and some remaining phages were removed on FACS bi-plots by gating for particles of bacterial cell size. Only fluorescent bacterial cells were sorted using FACS in form of bulks of thousands of cells and the isotopic "heavy" host DNA was removed prior to sequencing to reduce sequencing efforts (**Figure 6**). By this way, a subset of 26 groups of environmental phages able to infect *Synechococcus* was revealed in a single FACS run, which saves a significant amount of work when compared with plaque assays.

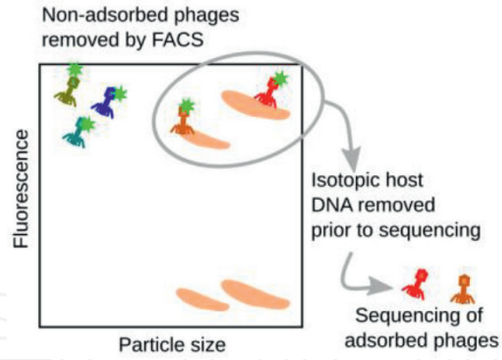

Figure 6.
Viral tagging.

In order to make viral tagging accessible to laboratories with no access to FACS instruments, a simplified adsorption assay has been recently developed, which does not involve fluorescent tagging. One of these methods is the removal of unattached phages by gel electrophoresis [47]. The principal advantage of the usage of the fluorescent viral tagging compared to the simplified methods is that correct phage attachment is confirmed by fluorescence emitted from the bacterial cells, while methods avoiding fluorescent tags can result in sequencing of cells containing no attached viruses, which results in wasting sequencing resources. Nevertheless, it is very convenient to apply the nonfluorescent separation techniques to communities with simple bacterial and viral compositions, which would easily detect the most active bacteriophages.

Viral tagging seems to be a very promising method for uncovering relationships between phages and their bacterial hosts. High specificity has been demonstrated by flow cytometry experiments and by mining phage-bacteria links detected on tagged single cells by computational methods [45, 46, 48]. Nevertheless, it can be argued that the adsorption of phages to the cell wall demonstrated by the acquisition of fluorescence does not always lead to infection. For example, there are several intracellular mechanisms, for example, CRISPR immunity, which can protect the bacterium from infection after the virus attachment to the cell wall. Nevertheless, the activity of the phage-bacteria links predicted by viral tagging can be assessed by metagenomic analysis of the same environmental samples, as explained above (**Figure 4**). In addition, in case of –culturable bacteria, the ability of attached phages to infect their hosts can be verified by following plaque assays [49].

Viral tagging has several advantages when compared to detection of phages naturally occurring on single cells. The principal advantage is that phages from one environment can be combined with bacteria from another environment. In a previous study, the viral tagging method was adjusted to the single-cell level and applied to the human gut microbiome. Viral tagging predicted phage-bacteria pairings, which could occur during a fecal microbiome transplant, in which viruses from a healthy individual are applied to restore altered gut microbiome composition [48].

2.5 Resolving phage-host relationships in complex communities, such as human gut

The human gut represents the most studied microbiome. However, while the majority of the most dominant bacterial groups in the human gut have some cultured representatives, the most of the human gut phages have not been cultured yet [50].

The results of the two recent studies on human gut phages clearly showed an enormous portion of novel phages. Nearly 190,000 phages clustered into 54,000 species-like phage groups were detected in a set of 11,000 human gut metagenomes, and 92% of them were not found in existing databases [5]. Another study reported 142,000 nonredundant gut phages from 28,000 human gut metagenomes [51]. Basically, we can say that in each new metagenomic sequencing run from human gut microbiome, some new phages are found.

Nevertheless, while the number of phages discovered by sequence-based method is increasing enormously, the transition from the phage genome sequence to its isolation and understanding of its biology can take years. In 2014, the first crAssphage was discovered in metagenomic sequences [52]. It was computationally associated with *Bacteroides* host, one of the most abundant bacterial species of the human gut. The crAssphages were found to form the most widespread phage group in the human gut. Isolation and replication of the first crAssphage representative were finally achieved in the laboratory 4 years later [53, 54]. Its isolation was not a simple task because its lifestyle differs from the typical lytic phages. It forms plaques but has a very small burst of progeny (2.5 plaque-forming units per infected cell), much lower than the burst size of the widely studied *E. coli* phages (burst size up to 300) [55]. The crAssphage does not exist in a form of a prophage, so it does not switch from lytic to lysogenic life cycle. This unusual lifestyle led to an equilibrated coexistence with its *Bacteroides* host, which might benefit from the reservoir of auxiliary metabolic genes harbored in the crAssphages [53, 54]. The high number of prophages found in the genomes of gut bacteria and the relatively stable composition of the human gut microbiome suggest that so-called "piggyback-the-winner" model of host-phage interactions is more likely in the gut than the "kill-the-winner" model, which is often observed in aquatic ecosystems and is characterized by significant fluctuations of the phage-bacteria populations [56].

The single-cell viral tagging technique applied to the human gut microbiome has demonstrated on a high-throughput scale that the "kill-the-winner" model is not widely spread in this environment [48]. The links obtained from analysis of the tagged single cells were used for mapping metagenomics reads from a temporal sample series obtained during 2 weeks from each bowel movement and samples collected with a time difference of 1 year. The host-phage genome coverage ratios were very similar for most of the phages during the 2 weeks' period, but many pairs experienced greater changes over the period of 1 year after. The results suggested that gut phages move between integrated and lytic states using small burst sizes to avoid overwhelming their hosts, which, so far, has been experimentally proven for the crAssphage group only. The equilibrated relationship between phages and bacteria in the human gut revealed by single-cell techniques is in accordance with previous bacterial composition reports showing that the human gut microbiome remains relatively stable for months; however, rare events, such as travel or enteric infection influence the community dynamics [57].

2.6 Novelty of the viruses recovered by single-cell genomics

Single-cell techniques often result in discovery of novel viruses, which are not captured by sequencing of viromes (metagenomes of environmental samples enriched for the particles smaller than 0.2 μm). For example, the previous studies focused on detection of viruses on sorted single-cell and recovered phages, which were under the detection limit in the viromes, while their bacterial hosts were

detectable [39, 48]. There are several reasons for this observation. The first reason is the randomness of the cell sorting – while metagenomics or viromics always recover the most abundant microbes, cell sorting can accidentally recover some microbes with low abundance, which would otherwise remain uncharacterized as part of the microbial dark matter [58]. The second reason might be the attachment ≠ infection argument claiming that not all attached phages are successful in the cell infection. Therefore, the phages recovered by the single-cell genomics might not be the most infectious ones [59]. The third reason is the possibility of recovering ssDNA viruses by whole genome amplification reaction, which are not detectable in traditional viromes, in which only dsDNA is extracted, thus ssDNA phages are missed [60]. The fourth reason might be ability of single-cell genomics to recover viruses, which are difficult to assemble if traditional viromes are sequenced. This was the case of a *Pelagibacter* phage, which is the most abundant virus in the world, but its microdiversity has been hindering the metagenomic assembly of its genome, despite large sequencing efforts of the global ocean microbiomes [61].

Novel types of phages with atypical head or tail structures, unusual nucleic acids (such as RNA phages), or unique lifestyles are constantly being discovered [62]. The phage research community involves scientists from all around the world and there are many efforts to improve our ability to characterize novel phages. For example, the completeness of a viral genome recovered from metagenomes, viromes, or single-cell assemblies can be assessed by computational tools, such as CheckV [63]. There are also several computational tools for taxonomic classification of phages [64], and also, a possibility of a normalized taxonomic classification based on single-copy marker genes has been explored [65]. The International Committee on Taxonomy of Viruses (ICTV) is constantly updating the taxonomy system to accommodate the large number of novel viruses [66]. It is possible that if metagenomes or single cells sequence in the past are reanalyzed with new phage-mining computational tools, much more novel phages will be discovered.

3. Conclusion

Identification of the bacterial hosts of phages is important for elucidating their impact on the bacterial community in an environment. Traditional plaque assays provide this information for culturable bacteria only, and computational methods detect genome signatures reflecting past infections in a limited portion of bacterial or viral species. A certain portion of bacterial cells in each environment contains bacteriophages in their interior or attached to their surface. Microbial single-cell genomics is a high-throughput technique for capturing links between novel phages and their uncultured hosts if they are colocated on the same single particle. Lifestyle of uncultured phages in their natural environment can be revealed from whole-genome amplification curves of the collected single-particles or from their genome assemblies. Metagenomics of the same environmental samples, from which single-cells have been collected, provide additional information on the phage lifestyle. Phages from one sample can be combined with bacteria from another sample and their compatibility via phage adsorption can be tested by viral tagging. In summary, microbial single-cell genomics is a useful tool for obtaining important ecological data on bacterial and viral dark matter, which can influence biogeochemistry across all environments on Earth, and also affect our health.

Acknowledgements

MD is supported by Generalitat Valenciana program GenT grant number CDEIGENT/2021/008.

Author details

Mária Džunková
Institute for Integrative Systems Biology (UV/CSIC), Paterna, Valencia, Spain

*Address all correspondence to: maria.dzunkova@uv.es

References

[1] Chevallereau A, Pons BJ, van Houte S, Westra ER. Interactions between bacterial and phage communities in natural environments. Nature Reviews. Microbiology. 2022;**20**:49-62. DOI: 10.1038/s41579-021-00602-y

[2] Ofir G, Sorek R. Contemporary phage biology: From classic models to new insights. Cell. 2017;**172**:1260-1270. DOI: 10.1016/j.cell.2017.10.045

[3] Nami Y, Imeni N, Panahi B. Application of machine learning in bacteriophage research. BMC Microbiology. 2021;**21**:193. DOI: 10.1186/s12866-021-02256-5

[4] Roux S, Páez-Espino D, Chen IA, Palaniappan K, Ratner A, Chu K, et al. IMG/VR v3: An integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. Nucleic Acids Research. 2020;**49**:D1. DOI: 10.1093/nar/gkaa946

[5] Nayfach S, Páez-Espino D, Call L, Low SJ, Sberro H, Ivanova NN, et al. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. Nature Microbiology. 2021;**6**:960-970. DOI: 10.1038/s41564-021-00928-6

[6] Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, et al. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. Cell Host & Microbe. 2019;**26**:764-778.e5. DOI: 10.1016/j.chom.2019.10.009

[7] Furfaro LL, Payne MS, Chang BJ. Bacteriophage therapy: Clinical trials and regulatory hurdles. Frontiers in Cellular and Infection Microbiology. 2018;**8**:376. DOI: 10.3389/fcimb.2018.00376

[8] Giles EM, D'Adamo GL, Forster SC. The future of faecal transplants. Nature Reviews. Microbiology. 2019;**17**:719. DOI: 10.1038/s41579-019-0271-9

[9] Ott SJ, Waetzig GH, Rehman A, Moltzau-Anderson J, Bharti R, Grasis JA, et al. Efficacy of sterile fecal filtrate transfer for treating patients with clostridium difficile infection. Gastroenterology. 2017;**152**:799-811.e7. DOI: 10.1053/j.gastro.2016.11.010

[10] Zuo T, Wong SH, Lam K, Lui R, Cheung K, Tang W, et al. Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome. Gut. 2018;**67**:634-643. DOI: 10.1136/gutjnl-2017-313952

[11] Kropinski AM, Mazzocco A, Waddell TE, Lingohr E, Johnson RP. Enumeration of bacteriophages by double agar overlay plaque assay. Methods in Molecular Biology. 2009;**501**:69-76. DOI: 10.1007/978-1-60327-164-16_7

[12] Kaletta J, Pickl C, Griebler C, Klingl A, Kurmayer R, Deng L. A rigorous assessment and comparison of enumeration methods for environmental viruses. Scientific Reports. 2020;**10**:18625. DOI: 10.1038/s41598-020-75490-y

[13] Steen AD, Crits-Christoph A, Carini P, DeAngelis KM, Fierer N, Lloyd KG, et al. High proportions of bacteria and archaea across most biomes remain uncultured. The ISME Journal. 2019;**13**:3126-3130. DOI: 10.1038/s41396-019-0484-y

[14] Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. Community structure and metabolism

through reconstruction of microbial genomes from the environment. Nature. 2004;**428**:37-43. DOI: 10.1038/nature02340

[15] Yue Y, Huang H, Qi Z, Dou HM, Liu XY, Han TF, et al. Evaluating metagenomics tools for genome binning with real metagenomic datasets and CAMI datasets. BMC Bioinformatics. 2020;**21**:334. DOI: 10.1186/s12859-020-03667-3

[16] Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nature Biotechnology. 2021;**39**:105-114. DOI: 10.1038/s41587-020-0603-3

[17] Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nature Microbiology. 2017;**2**:1533-1542. DOI: 10.1038/s41564-017-0012-7

[18] Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nature Biotechnology. 2021;**39**:499-509. DOI: 10.1038/s41587-020-0718-6

[19] Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nature Biotechnology. 2018;**36**:996-1004. DOI: 10.1038/nbt.4229

[20] Correa AMS, Howard-Varona C, Coy SR, Buchan A, Sullivan MB, Weitz JS. Revisiting the rules of life for viruses of microorganisms. Nature Reviews. Microbiology. 2021;**19**:501-513. DOI: 10.1038/s41579-021-00530-x

[21] Akhter S, Aziz RK, Edwards RA. PhiSpy: A novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. Nucleic Acids Research. 2012;**40**:e126. DOI: 10.1093/nar/gks406

[22] Sorek R, Kunin V, Hugenholtz P. CRISPR–A widespread system that provides acquired resistance against phages in bacteria and archaea. Nature Reviews. Microbiology. 2008;**6**:181-186. DOI: 10.1038/nrmicro1793

[23] Bailly-Bechet M, Vergassola M, Rocha E. Causes for the intriguing presence of tRNAs in phages. Genome Research. 2007;**17**:1486-1495. DOI: 10.1101/gr.6649807

[24] Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, et al. Numerous cultivated and uncultivated viruses encode ribosomal proteins. Nature Communications. 2019;**10**:752. DOI: 10.1038/s41467-019-08672-6

[25] Kieft K, Breister AM, Huss P, Linz AM, Zanetakos E, Zhou Z, et al. Virus-associated organosulfur metabolism in human and environmental systems. Cell Reports. 2021;**36**:109471. DOI: 10.1016/j.celrep.2021.109471

[26] Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. Nature. 2016;**536**:425-430. DOI: 10.1038/nature19094

[27] Koskella B. Bacteria-phage interactions across time and space: Merging local adaptation and time-shift experiments to understand phage evolution. The American Naturalist. 2014;**184**:S9-S21. DOI: 10.1086/676888

[28] Sant DG, Woods LC, Barr JJ, McDonald MJ. Host diversity slows

bacteriophage adaptation by selecting generalists over specialists. Nature Ecology & Evolution. 2021;**5**:350-359. DOI: 10.1038/s41559-020-01364-1

[29] Dang VT, Sullivan MB. Emerging methods to study bacteriophage infection at the single-cell level. Frontiers in Microbiology. 2014;**5**:724. DOI: 10.3389/fmicb.2014.00724

[30] Lim SW, Lance ST, Stedman KM, Abate AR. PCR-activated cell sorting as a general, cultivation-free method for high-throughput identification and enrichment of virus hosts. Journal of Virological Methods. 2017;**242**:14-21. DOI: 10.1016/j.jviromet.2016.12.009

[31] Marbouty M, Thierry A, Millot GA, Koszul R. MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut. eLife. 2021;**26**:10. DOI: 10.7554/eLife.60608

[32] Woyke T, Doud DFR, Schulz F. The trajectory of microbial single-cell sequencing. Nature Methods. 2017;**14**:1045-1054. DOI: 10.1038/nmeth.4469

[33] Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. Nature Medicine. 2018;**24**:1277-1289. DOI: 10.1038/s41591-018-0096-5

[34] Qiu Q, Hu P, Qiu X, Govek KW, Cámara PG, Wu H. Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. Nature Methods. 2020;**17**:991-1001. DOI: 10.1038/s41592-020-0935-4

[35] Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, et al. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. The ISME Journal. 2015;**9**:2386-2399. DOI: 10.1038/ismej.2015.48

[36] Munson-McGee JH, Peng S, Dewerff S, Stepanauskas R, Whitaker RJ, Weitz JS, et al. A virus or more in (nearly) every cell: Ubiquitous networks of virus-host interactions in extreme environments. The ISME Journal. 2018;**12**:1706-1714. DOI: 10.1038/s41396-018-0071-7

[37] Labonté JM, Field EK, Lau M, Chivian D, Van Heerden E, Wommack KE, et al. Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. Frontiers in Microbiology. 2015;**6**:349. DOI: 10.3389/fmicb.2015.00349

[38] Roux S, Hawley AK, Torres Beltran M, Scofield M, Schwientek P, Stepanauskas R, et al. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. eLife. 2014;**3**:e03125. DOI: 10.7554/eLife.03125

[39] Jarett JK, Džunková M, Schulz F, Roux S, Paez-Espino D, Eloe-Fadrosh E, et al. Insights into the dynamics between viruses and their hosts in a hot spring microbial mat. The ISME Journal. 2020;**14**:2427-2541. DOI: 10.1038/s41396-020-0705-4

[40] Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research. 2015;**25**:1043-1055. DOI: 10.1101/gr.186072.114

[41] Doud DFR, Bowers RM, Schulz F, De Raad M, Deng K, Tarver A, et al. Function-driven single-cell genomics uncovers cellulose-degrading bacteria

from the rare biosphere. The ISME Journal. 2020;**14**:659-675. DOI: 10.1038/s41396-019-0557-y

[42] Yang C, Mogno I, Contijoch EJ, Borgerding JN, Aggarwala V, Li Z, et al. Fecal IgA levels are determined by strain-level differences in bacteroides ovatus and are modifiable by gut microbiota manipulation. Cell Host & Microbe. 2020;**27**:467-475.e6. DOI: 10.1016/j.chom.2020.01.016

[43] Grieb A, Bowers RM, Oggerin M, Goudeau D, Lee J, Malmstrom RR, et al. A pipeline for targeted metagenomics of environmental bacteria. BMC Microbiome. 2020;**8**:21. DOI: 10.1186/s40168-020-0790-7

[44] Hennes KP, Suttle CA, Chan AM. Fluorescently labeled virus probes show that natural virus populations can control the structure of marine microbial communities. Applied and Environmental Microbiology. 1995;**61**:3623-3627. DOI: 10.1128/aem.61.10.3623-3627.1995

[45] Deng L, Gregory A, Yilmaz S, Poulos BT, Hugenholtz P, Sullivan MB. Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. MBio. 2012;**3**:e00373-e00312. DOI: 10.1128/mBio.00373-12

[46] Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, et al. Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. Nature. 2014;**513**:242-245. DOI: 10.1038/nature13459

[47] de Jonge PA, von Meijenfeldt FAB, Costa AR, Nobrega FL, Brouns SJJ, Dutilh BE. Adsorption sequencing as a rapid method to link environmental bacteriophages to hosts. iScience. 2020;**23**:101439. DOI: 10.1016/j.isci.2020.101439

[48] Džunková M, Low SJ, Daly JN, Deng L, Rinke C, Hugenholtz P. Defining the human gut host-phage network through single-cell viral tagging. Nature Microbiology. 2019;**4**:2192-2203. DOI: 10.1038/s41564-019-0526-2

[49] Jang HB, Chittick L, Li YF, Zablocki O, Sanderson CM, Carrillo A, et al. Viral tag and grow: A scalable approach to capture and characterize infectious virus–host pairs. ISME Communications. 2022;**2**:12. DOI: 10.1038/s43705-022-00093-9

[50] Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature. 2016;**533**:543-546. DOI: 10.1038/nature17645

[51] Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. Cell. 2021;**184**:1098-1109.e9. DOI: 10.1016/j.cell.2021.01.029

[52] Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nature Communications. 2014;**5**:4498. DOI: 10.1038/ncomms5498

[53] Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Nature Communications. 2018;**9**:4781. DOI: 10.1038/s41467-018-07225-7

[54] Guerin E, Shkoporov A, Stockdale SR, Clooney AG, Ryan FJ, Sutton TDS, et al. Biology and taxonomy of crass-like bacteriophages, the most abundant virus in the human gut. Cell Host & Microbe. 2018;**24**:653-664.e6. DOI: 10.1016/j.chom.2018.10.002

[55] Hyman P, Abedon ST. Practical methods for determining phage growth parameters. Methods in Molecular Biology. 2009;**501**:175-202. DOI: 10.1007/978-1-60327-164-6_18

[56] Silveira C, Rohwer F. Piggyback-the-winner in host-associated microbial communities. NPJ Biofilms Microbiomes. 2016;**2**:16010. DOI: 10.1038/npjbiofilms.2016.10

[57] David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. Genome Biology. 2014;**15**:R89. DOI: 10.1186/gb-2014-15-7-r89

[58] Balachandran M, Cross KL, Podar M. Single-Cell Genomics and the Oral Microbiome. Journal of Dental Research. 2020;**99**:613-620. DOI: 10.1177/0022034520907380

[59] Thanki AM, Taylor-Joyce G, Dowah A, Nale JY, Malik D, Clokie MRJ. Unravelling the links between phage adsorption and successful infection in clostridium difficile. Viruses. 2018;**10**:411. DOI: 10.3390/v10080411

[60] Székely AJ, Breitbart M. Single-stranded DNA phages: from early molecular biology tools to recent revolutions in environmental microbiology. FEMS Microbiology Letters. 2016;**363**:fnw027. DOI: 10.1093/femsle/fnw027

[61] Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez Martínez J, et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. Nature Communications. 2017;**8**:15892. DOI: 10.1038/ncomms15892

[62] Sausset R, Petit MA, Gaboriau-Routhiau V, De Paepe M. New insights into intestinal phages. Mucosal Immunology. 2020;**13**:205-215. DOI: 10.1038/s41385-019-0250-5

[63] Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nature Biotechnology. 2021;**39**:578-585. DOI: 10.1038/s41587-020-00774-7

[64] Turner D, Kropinski AM, Adriaenssens EM. A roadmap for genome-based phage taxonomy. Viruses. 2021;**13**:506. DOI: 10.3390/v13030506

[65] Low SJ, Džunková M, Chaumeil PA, Parks DH, Hugenholtz P. Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. Nature Microbiology. 2019;**4**:1306-1315. DOI: 10.1038/s41564-019-0448-z

[66] Krupovic M, Turner D, Morozova V, Dyall-Smith M, Oksanen HM, Edwards R, et al. Bacterial viruses subcommittee and archaeal viruses subcommittee of the ICTV: Update of taxonomy changes in 2021. Archives of Virology. 2021;**166**:3239-3244. DOI: 10.1007/s00705-021-05205-9