

# Phage*Dive*: the comprehensive strain database of prokaryotic viral diversity

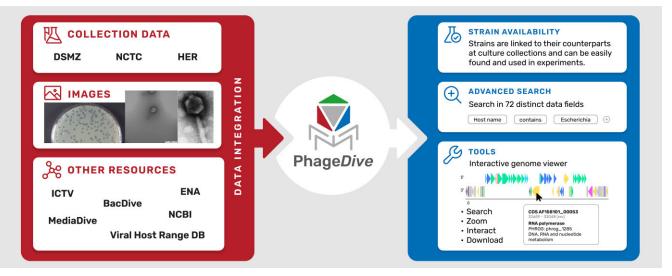
Clara Rolland, Johannes Wittmann<sup>®</sup>\*, Lorenz C. Reimer<sup>®</sup>\*, Joaquim Sardà Carbasse, Isabel Schober<sup>®</sup>, Christian-Alexander Dudek, Christian Ebeling, Julia Koblitz<sup>®</sup>, Boyke Bunk and Jörg Overmann

Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, 38124 Braunschweig, Germany \*To whom correspondence should be addressed. Tel: +49 5312616 214; Fax: +49531 2616 418; Email: johannes.wittmann@dsmz.de Correspondence may also be addressed to Lorenz C. Reimer. Tel: +49 5312616 311; Fax: +49531 2616-418; Email: lorenz.reimer@dsmz.de

# Abstract

Prokaryotic viruses represent the most diverse and abundant biological entities on Earth. So far, data on bacteriophages are not standardized, not readily available for comparative analyses and cannot be linked to the rapidly growing (meta)genomic data. We developed Phage*Dive* (https: //phagedive.dsmz.de), a comprehensive database for prokaryotic viruses gathering all existing data dispersed across multiple sources, like scientific publications, specialized databases or internal files of culture collections. Phage*Dive* allows to link own research data to the existing information through an easy and central access, providing fields for various experimental data (host range, genomic data, etc.) and available metadata (e.g. geographical origin, isolation source). An important feature is the link between experimental data, the culture collection number and the repository of the corresponding physical bioresource. To date, Phage*Dive* covers 1167 phages from three different world-renowned public collections (DSMZ, Félix d'Hérelle Reference Center for Bacterial Viruses and NCTC) and features an advanced search function using all data fields from the sections like taxonomy or morphology by controlled vocabulary and ontologies. Phage*Dive* is fully interoperable with other resources including NCBI, the Viral Host Range database (VHRdb) of Institute Pasteur or the Bac*Dive* and Media*Dive* databases of DSMZ.

# **Graphical abstract**



# Introduction

Since their discovery by Félix d'Hérelle and Fredrick Twort over a century ago >11 200 bacteriophage species have been characterized (1) which makes them the largest virus group known to date. The vast majority of the known bacteriophages are tailed and contain double-stranded DNA (2,3). Culture-independent molecular approaches have revealed that bacteriophages occur in a much larger diversity and in high abundance in essentially all environments on Earth, ranging from marine waters (4) and soils (5), to acidic hot volcanic springs (6) and the human gut (7). The crude estimates of total viral diversity amount to  $10^7-10^9$  virus species (8). At present, only a very limited fraction of this existing bacteriophage diversity is maintained in public collections worldwide, which provide the quality-controlled and documented cultures needed for scientific studies of bacteriophage

Received: August 14, 2024. Revised: September 18, 2024. Editorial Decision: September 21, 2024. Accepted: September 25, 2024

<sup>©</sup> The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

biology. Besides the isolation, characterization, biobanking, and genome analysis of bacteriophages, an easy access to structured (meta)data associated with the different phage species is essential for future studies, particularly for comparative analyses.

The current release on the update of virus taxonomy of the International Committee on Taxonomy of Viruses (ICTV) comprises 11 273 virus species in 2818 genera and 264 families (1). For these viruses, the information on the names, taxonomy, morphology, genome features, proteins, host range and origin can be found in the report chapters of the ICTV database (report on Virus Classification and Taxon Nomenclature, https://ictv.global/taxonomy/history, Retrieved 19 February 2024). However, this open access, public database is organized along taxonomic levels and does not use controlled vocabulary or ontologies. Most of the existing databases on bacteriophages focus on collecting and providing genomic data like genetic features or predicted lifestyles (e.g. PhageScope (9)), or on specific aspects of phage biology like virus-host interactions like MVP or VHRdb (10,11). The database PhagesDB collects and shares data exclusively for actinobacterial phages (12). The European Virus Archive Global (https://www.european-virusarchive.com/evag-portal) only covers human and animal pathogens. Hence, no database exists so far which can accommodate additional metadata on bacteriophages from literature or would enable specific searches for single features (e.g. a search term 'phage' to extract viruses that infect bacteria or archaea) or for specific combinations of characteristics (e.g. 'host + marine + Roseobacter', to recover all bacteriophages infecting bacteria of the genus Roseobacter; or 'DNA + host = *Staphylococcus*', to recover all DNA viruses infecting Staphylococcus).

We established the database PhageDive (https://phagedive. dsmz.de) that provides comprehensive information on bacteriophage biology, molecular mechanisms of host interaction, or phage ecology and links this information to established bacteriophage taxa.

# Materials and methods

#### Technical information

The Graphical User Interface (GUI) of Phage*Dive* was individually built with PHP 8.2. The frontend relies on JavaScript and JQuery. For better usability, the GUI only displays filled data fields and hides empty ones. Therefore, the Phage*Dive* GUI might appear differently from strain to strain. The relational database engine is MariaDB 10.8.3. The database scheme consists of 156 data fields stored in 17 tables. This technical framework ensures the flexibility to keep up with the growing database and allows to integrate new data types and to adapt the GUI according to user feedback and needs.

## Results

#### User interface

Currently, Phage*Dive* harbours data for bacteriophages and archaeal viruses which are physically available in public culture collections such as the Félix d'Hérelle Reference Center for Bacterial Viruses in Canada, the National Collection of Type Cultures (NCTC) in UK and DSMZ phage collections (Germany) amounting to a total of 1167 viruses. The decision to base Phage*Dive* on culture collections was taken because the access to the physical resources of viruses and their hosts enables future analysis, thereby ensuring reproducibility and improvement of the quality of the research data. All information collected through the culture collections and other various sources such as genomic resources or literature was standardized and classified into 99 data fields, grouped into 12 categories.

#### Individual strain page content

PhageDive offers entries for each individual virus strain, aggregating the available data and directly linking them to strain identifiers like official collection numbers of the original source, and catalogue entries of the respective collection, if available.

The individual pages for the viruses are structured by section as follows (Figure 1). A summary provides key data like name, morphotype, host species, as well as data on the morphology and pictures of phage plaques on a host culture, and electron photomicrographs, if available (Figure 1, section 1, label 1a). All sections displayed for a virus are listed in the right-hand upper corner (Figure 1, label 2). Basic data include strain identifiers directly linked to respective catalogues, the name of the strain and strain history (Figure 1, label 3). Data included in PhageDive cover basic information such as geographic origin, isolation sources, morphology, lifestyle, taxonomy (Figure 1, label 4), and associated literature. Particular attention was paid to virus morphology. Since current taxonomy focuses mainly on the genomic and proteomic data and phylogenetic relationships of viruses and in this course families reflecting morphological features of their members directly in their names like Myoviridae or Siphoviridae were recently abolished (13), this previous classification based on morphology is depicted in the morphotype section (e.g. siphovirus, myovirus, podovirus etc.) to maintain consistency of PhageDive with previous data. These morphotypes and the morphological description of plaques are supported by the transmission electron micrographs (TEM) and plaques pictures (Figure 1, label 1a). The latter can be enlarged interactively, providing good clarity. For detailed information on virus propagation, we also provide data on the recommended host strains and cultivation media through direct links to BacDive (14) and MediaDive (15), respectively. Information on virus-host interactions is based on internal host range data of the DSMZ as well as on the Viral Host Range Database hosted by the Pasteur Institute in Paris (11). In addition, general genomic information like accession number, genome length or GC content are directly displayed on the virus-specific webpage. If available, genome sequences with GenBank accessions were retrieved from NCBI, reannotated using Prokka version 1.14.6 (16) and the PHROG database (17) and implemented into a genome viewer tool.

As a strain page may aggregate data from different sources, each data entry of the Phage*Dive* database is labeled by a reference number to the left (Figure 1, label 3;). This number is linked to the reference section at the bottom of the virus description page. Through this extensive documentation users can easily retrace the origin of the data.

For a better usability of the interface, only section and data fields for which data is available are presented.

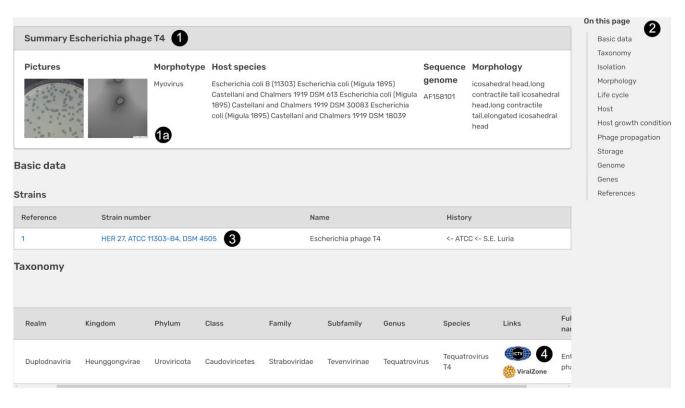


Figure 1. Example of an individual strain page (Escherichia phage T4, https://phagedive.dsmz.de/strain/160).

#### Linking and integrating available data sources

One of the main characteristics of PhageDive is to integrate and standardize the highly dispersed data on bacteriophages and to link them to the bioresource of origin. Depending on the source, data are either directly integrated into PhageDive and provided in a standardized format, or data are made accessible via direct links to the external sources to provide the user with the access to associated information. The former typically is the case for internal data from culture collections that are integrated in standardized data fields, which enables data comparison and systematic searches. For transparency, every single data point is associated with a reference id linking to the original source. Genome sequence data are reannotated and provided with the genome viewer. Other data types, in particular sequence data, bacterial strain data or cultivation media data, are in a highly specific, stable format and readily available through well-established, active databases. In these cases, data are directly linked to the respective web resource, allowing users to easily explore these data while ensuring access to the regularly updated information provided by the external sources. Important external sources directly linked in PhageDive are NCBI for sequence data, BacDive for bacterial strain data, MediaDive for cultivation media recipes, Viral Host Range database for virus host range data, ICTV for taxonomy, ViPTree for an overview on the phage phylogeny (18) and ViralZone (19) for general properties of taxonomic groups like genome features, further information on replication or putative receptors.

In turn, Phage*Dive* strains are indexed and linked by other web resources which increases visibility and findability. As a part of the newly established, overarching biodata infrastructure DSMZ Digital Diversity, viruses are linked to their respective host strain in Bac*Dive* and also indexed in the federated search of the DSMZ Digital Diversity Hub page.

#### Tools

Genome viewer. A genome browser is implemented directly on the individual webpage of each strain if the genome sequence is available (Figure 2). In the current state PhageDive contains 448 phage genomes that can be analysed with this browser tool. In order to standardize genome representation, the annotation was renewed using Prokka version 1.14.6 and the PHROG database (Prokaryotic Virus Remote Homologous Groups) (16), a constantly growing library of viral protein families generated by using a new clustering approach based on remote homology detection. Bacteriophage genomes generally harbour a lot of genes with unknown function. Using this library allows to add further annotations that are missed by standard annotation tools. The graphical display is based on the PHROG annotation, with genes individually coloured by PHROG category to give an initial overview of the general genome organization including gene clusters for structural genes, replication, lysis and others.

The primary view displays the organization of the entire virus genome (Figure 2A). To navigate through the annotated sequence different parameters are available. Zoom into the genome (Figure 2B) is possible via two alternatives. The first option is to enter the genomic coordinates of the area to be zoomed (Figure 2B, label 1). The second is to generate a zoom manually by framing the area of interest on the genomic sequence at the bottom (Figure 2B, label 2). Each gene has a 'mouse-over' window containing its main information, which appears when the mouse pointer is hovering over the gene in question (Figure 2B, label 3). Displayed are the gene name, position of the gene in the genome, function, its corresponding PHROG number and the PHROG assigned gene category. In addition, clicking on each gene also gives access to its nucleotide sequence that can be copied and pasted (Figure 2B,

	Nucleic	Secuence	Accession	Assembly		Assembly	Genome	GC	1
Reference	acid type	Sequence database	number	accession number	Sequence title	level	length	content	1
1	DNA, linear	NCBI	AF158101	GCA_000836945	Escherichia phage T4, complete genome	Complete	168903	35.30	
0	0		Search	genome					
5'	1								
3' <b>4</b>					( 📢				
Legend:	ation and excision	head and packaging	transcription regulation	n connector tail Ivsis Di	NA PNA and puclashida mataboliam	other unknown	function		
	netabolic gene and host		a anscription regulation	i i commector i can i visis i pr	na, nna ann nucleolule metabolism :		nandaon		
3257	24	1195	Search	genome 1					
5'					[				
3' pa	re 📢 📢				CDS AF1581		3 🤇 候	· • • • •	recA
3257		2				othetical protei			
	min 📫 🖬				PHROG: phro	SMZ:AF158101_ g_2024			
Legend: 🧧 integra	ation and excision 📃 I	nead and packaging 📃 t	ranscription regulation	n 🧧 connector 📕 tail 📕 lysis 🦳 DM	IA DALA was done a	gory: unknown f	function		
moron, auxiliary n	netabolic gene and host	takeover							
Forward posit	tion 17738 - 179	38 4							
(210)									
				acatticaat aattggilot igaalatoat iggoo				itattttgga aagtgagat	tt tga
				acatticaat aattggttet tgaatateat tggee aaatgaatta tacaattett catcatatgt atttg		gataa acctocacaa aata attaa tttaaaggag gata		statttigga aagtgagat aafggcant lyttaarge	tt tga
	ttattac ctgtatacgc ga							statttigga angtgagat malogeant fotoaloo stagatagee geatttag	tt tga se ato ge cat
1730) taataatgat tgtt 1700)	ttattac ctgtatacgc ga						atat <mark>atg</mark> gtacaaaaat t ↦	stattitgga angigagat aniggnadi tottanigo stagatagee geattitag	tt tgi to ati ge cal
17361 taataatgat tgt1 17661 ataagaaatg tata	ttattac ctgtatacgc ga			aaatgaatta tacsattett catoatatgt atttg getegaatte geboetyyig stankeltos syley			atat <mark>atg</mark> gtacaaaaat t ↦	itattitgga aagtgagat Anicoonal Iotianico itagatagee geattitag	tt tgi
1759) taataatgat tgtt 1769) ataagaaacg tate (799) tgggcaatat	ttattac ctgtatacgc ga						atat <mark>atg</mark> gtacaaaaat t ↦	itattitgga angigagat Antoponon tortkaroo Itagatagoo goattitag	tt tge
1796 tastaatgat tgtt 1619 1619 1619 1619 1619 1619 1619 16	ttattac cigiatacgo ga		ctcattt attgoagcaa a	aaatgaatta tacsattett catoatatgt atttg getegaatte geboetyyig stankeltos syley			atat <mark>atg</mark> gtacaaaaat t ↦	ttatttigga angigagat sotiopensi ligitAlarge itagatagee geattitag	co ata
17301 taataatgat tgt1 17001 alaagabeeg tata	ttattac cigiatacgo ga		ctcattt attgoagcaa a	aaatgaatta tacsattett catoatatgt atttg getegaatte geboetyyig stankeltos syley			atat <mark>atg</mark> gtacaaaaat t ↦	Galggeant by Haarge	co ata
0 5' 3'	ttattac cigiatacgo ga		ctcattt attgoagcaa a	aaatgaatta tacsattett catoatatgt atttg getegaatte geboetyyig stankeltos syley			atat <mark>atg</mark> gtacaaaaat t ↦	Galggeant by Haarge	ge cat
17301 tastaatgat tgtt 17000 11400 gaaacog 6426 17990 17900 1790 179	ttattac cigiatacgo ga		ctcattt attgoagcaa a	aaatgaatta tacsattett catoatatgt atttg getegaatte geboetyyig stankeltos syley			atat <mark>atg</mark> gtacaaaaat t ↦	Galggeant by Haarge	ge cat
0 5' 3'	ttattac cigiatacgo ga		ctcattt attgoagcaa a	aaatgaatta tacsattett catoatatgt atttg getegaatte geboetyyig stankeltos syley			atat <mark>atg</mark> gtacaaaaat t ↦		ge cat
C S C C C C C C C C C C C C C	ttattac cigiatacgo ga		ctcattt attgoagcaa a						ge cat
Conception of the second secon		Appetent agosttatat tot	CTCATTL ATTOCAGOAS A						ge cat
taataatgat tgtt Etgggcaatat 0 5 3 0 Legend: integra integra integra integra integra integra	tatta etgiatage ga	Appreciation agastitate tor Appreciation of the second of	interatit attigeageas a			attas ettasaggag gata tenes tetasaggag gata tenes tetasgataggata tenes tetasgataggata tenes tetasgataggataggata tenes tetasgataggataggataggataggataggataggatag	At a 'C' a constraint of the c		ego
Contract of the second	tatta etgiatage ga	AppleOpperture of a positive o	interest attgrages / head	anatypatta tacanticit catoriatyt attip	atagt astantaals optittigiga taaat associations as provide a start association and a start as a start as a start astanta and nucleotide metabolism roduct	atta tttasopp gat	tar to tageatat t → to ottageatat t ↓ → to ottage	PHROG cate	ego ing
Legend: Integra more, auxiliary n Start 1 15331 2 22049	tatta crejatage ga	Approximation of the second se	creatt attgrages / head	anarguatta tacanteet extendent attep	atagt astantaals optittigiga taaat associations as provide a start association and a start as a start as a start astanta and nucleotide metabolism roduct	atta tttasopp gat ttel attasopp	tate to tage at a tage at	PHROG catu d and packagi d and packagi	ego ing ing
Legend: 115331 115531 115551 1155551 1155551 1155551 1155551 1155555 115555555 1155555555	ttattat etgistange ga	Approver agreet table to the second s	Image: control of the second	connector tail yais DP virion structural protein head vertex assembly chai internal virion protein	atagt astantaals optittigiga taaat associations as provide a start association and a start as a start as a start astanta and nucleotide metabolism roduct	ettes ettessoppe gete other unknown other unknown phrog_ phrog_ phrog_	ta (27) → ctagestat t → ctagestat t -function trun	PHROG catta dand packagi dand packagi dand packagi	ego ing ing
1         1	ation and excision 1 15573 - 22393 - 655718 - 66415 -	Approximation of the second se	(head     (head     (head     (head     (head)     (head)	anatypatria tanattete esteatary attrip galaxies and a second sec	atagt astantaals optittigiga taaat associations as provide a start association and a start as a start as a start astanta and nucleotide metabolism roduct	atta ettaaogog gate dese etta	ter €2 constraint → constraints → constraints -	PHROG catu di and packagi di and packagi di and packagi	ego ing ing ing
1 1 1 1 1 1 1 1 1 1 1 1 1 1	trattac esplatance of a	AF15810 AF15810 AF15810 AF15810 AF15810	Image: control of the second	anatypatria tanaticit ostocatary attrip	atagt astantaals optittigiga taaat associations as provide a start association and a start as a start as a start astanta and nucleotide metabolism roduct	atta ettasoppo gata	ter to tragestate t → → → + + + + + + + + + +	PHROG cate PHROG cate di and packagi di and packagi di and packagi di and packagi di and packagi	ego ing ing ing
1       1	tatta etgiatage ga	exected agastists for association of the second event and packaging the takeover itrand L AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: Second	suspensa tensatient extendent attrap	atagt astantaals optittigiga taaat associations as provide a start association and a start as a start as a start astanta and nucleotide metabolism roduct	atta ettasoppo gata	ta (20) → ctageatat t → ctageatat t → ctageatat t (unction) IROG 1715 hea 7857 hea 7857 hea 2787	PHROG cate di and packagi di and packagi di and packagi mector	ego ing ing ing
Contraction of the second seco	tatita etgiatage ga 0 0 0 0 0 0 0 0 0 0 0 0 0	AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: Contract of a trigonage of a trige of a trige of a trigonage of a trigonage of a trigonage of a	substrate transition extended attract attraction of the second attracti	atagt astantaals optittigiga taaat associations as provide a start association and a start as a start as a start astanta and nucleotide metabolism roduct	ettes ettessegge gete ettes ettessegge gete ettessegge gettessegge gete ettessegge gete ettesse	ta (22) account of the second	PHROG cate dand packagi dand packagi dand packagi idand packagi inector	ego ing ing ing
Internet           0           5'           3'           0           5'           3'           0           Integrate           0           5'           3'           0           Integration           0           Integration           0           Start           1           15331           2           2           3           4           5           7           4           5           7           4           5           7           6           7           6           7           8           94555	ttattat etgistang ga particular and excision () attorn and excision () 15573 - 15573 - 22393 - 65718 - 66415 - 773878 - 77387 - 94553 + 95325 +	Approximation of the second se	Image: control of the second	connector tail by the protein internal virion protein internal virion protein internal virion protein head closure head-tail adaptor Ad2 head closure Hc2	atagt astantaals optittigiga taaat associations as provide a start association and a start as a start as a start astanta and nucleotide metabolism roduct	other unknown phrog_ phrog_ phrog_ phrog_	ter €2 constant → constant t → constant t	PHROG catte PHROG catte d and packagi d and packagi d and packagi id and packagi in an corr in a corr	ego ing ing ing
Control Contro	ation and excision	AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: control of the second	connector tail by tails of the protein internal virion protein internal virion protein internal virion protein head closure Hc2 terminase small subunit	atagt astantaals optittigiga taaat associations as provide a start association and a start as a start as a start astanta and nucleotide metabolism roduct	etter ettersoppi gete other unknown other unknown phrog phro	ter trajestat t → trajestat t → trajestat t → trajestat t + t	PHROG catu di and packagi di and packagi di and packagi inector inector inector inector inector inector inector inector	ego ing ing ing ing
Control Contro	trattac esplatance of a	AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: Control of the second	virion structural protein internal virion protein internal virion protein internal virion protein head closure head-tail adaptor Ad2 head closure Hc2 terminase small subunit terminase large subunit	atagt astartacis optitiopa taat according to the provide a table according to table accordi	etter ettersogie gete other unknown phrog p	ter trajectat t → trajectat t → trajectat t → trajectat t + + + + + + + + + + + + + + + + + + +	PHROG catu PHROG catu and packagi dand packagi	ego ing ing ing ing ing
Legend: Color C	Stop         Stop           15573         -           15573         -           22393         -           65718         -           73878         -           77337         -           94553         +           95688         +           98504         +           102781         +	AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: Second	virion structural protein internal virion protein internal virion protein internal virion protein head closure head-tail adaptor Ad2 head closure Hc2 terminase large subunit portal protein	atagt astartacis optitiopa taat according to the provide a table according to table accordi	ettes ettessoppe gete other unknown phrog p	ter to account of the second	PHROG cate dand packagi dand packagi	ego ing ing ing ing ing ing ing
Legend: Integr Tomoron. auditory in Start 1 15331 2 22049 3 65416 4 65834 5 73591 6 76885 7 93624 8 94555 9 96194 10 96672 11 001207 12 102781	Stop         Stop           15573         -           15573         -           15573         -           22393         -           65718         -           77337         -           94553         +           95325         +           96688         +           102781         +           103023         +	Image: constraint of the second sec	Image: Control of the second	supporter transference carectering attrapporter transference carectering attrapport transference carectering attrapport terminase large subunit terminase large subunit terminase large subunit portal protein prohead	atagt astartacis optitiopa taat according to the provide a table according to table accordi	other unknown phrog_ phrog_ phrog_ phrog_ phrog_ phrog_ phrog_ phrog_ phrog_ phrog_ phrog_ phrog_ phrog_	ter trajectat t → ctrajectat t → transition (function) +ROG 1715 hea 999 hea 7857 hea 5341 hea 4363 hea 1358 hea 218 con 211 con 679 con 735 hea 5989 hea 213 hea 1064 hea	PHROG cate dand packagi dand packagi	ego ing ing ing ing ing ing ing
Teachartgue type           1           1           1           0           3'           0           3'           0           1           1           1           1           2           2           2           3           4           5           3           4           5           3           4           5           7           3           6           7           3           4           5           7           3           5           7           3           4           5           7           5           9           9           9           9           9           9           9           9           9           9           9           9	Image: constraint of the second sec	AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: Control of the second	connector tail departed extended attrip connector tail by the protein head vertex assembly chain internal virion protein internal virion protein internal virion protein head closure head closure Hc2 terminase small subunit terminase large subunit protaid protein head scaffolding protein	And and nucleotide metabolism roduct aperone	other unknown other unknown phrog_	ter trajectat t → transformed to trajectat t → transformed to trajectat t → transformed to trajectat t → transformed to trajectat	PHROG catta PHROG catta PHROG catta and packagi d and packagi	ego ing ing ing ing ing ing ing
Control Contro	Stop         Stop           ation and excision         ation           ation and excision         ation           stop         ation           ation and excision         ation           ation and excision         ation           ation and excision         ation           75375         -           22393         -           66415         -           77337         -           95325         +           96688         +           102781         +           103428         +           103448         +           104086         +	AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: control of the second	connector tail by the protein head closure Hc2 terminase large subunit terminase large subunit protein protein head closure Hc2 terminase large subunit protein head closure Hc2 terminase large subunit terminal protein head maturation protease	And and nucleotide metabolism roduct aperone	etter etteraggi gete etter etteraggi gete etter etteraggi etter etteraggi etter etteraggi ette	ter CT according to the second secon	PHROG cature PHROG cature d and packagi d and packagi	ego ing ing ing ing ing ing ing ing ing
	Stop         Stop           15573         -           22393         -           66415         -           73878         -           73878         -           73878         -           94553         +           96688         +           98504         +           102781         +           103448         +           104086         +           104286         +           104286         +           104286         +	AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: Control of the second	virion structural potein head vertex assembly cha internal virion protein internal virion protein head closure head-tail adaptor Ad2 head closure Hc2 terminase small subunit terminase large subunit prohead head scaffolding protein head scaffolding protein head scaffolding protein head scaffolding protein	And and nucleotide metabolism roduct aperone	etter etteragge gete deter etteragge gete deter etteragge phrog_ phrog phrog phrog phrog_{} phrog phrog_	ter tragestate t → tragestate t → tragestate t → tragestate t → tragestate t + + + + + + + + + + + + + + + + + + +	PHROG catu di and packagi di and packagi	ego ing ing ing ing ing ing ing ing ing ing
transmission         Tra	Stop         Stop           15573         -           22393         -           65718         -           73878         -           73878         -           73878         -           95325         +           95325         +           102781         +           10323         +           103448         +           104086         +           104926         +           104926         +           104510         +	AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: Control of the second	virion structural protein internal virion protein internal virion protein internal virion protein head closure head-tail adaptor Ad2 head closure Hc2 terminase small subunit terminase large subunit portal protein head acaffolding protein head acaffolding protein major head protein major head protein	And and nucleotide metabolism roduct aperone	etter ettersopge gete other unknown phrog_ phrog phrog phro	ter (22) actingentate ( → (14) actingentate ( → (14) actingentate ( ) (14) actingentate	PHROG cate and packagi d and packagi d and packagi d and packagi id and packagi	ego ing ing ing ing ing ing ing ing ing ing
Legend: integra 0 5' 3' 0 Legend: integra moren. audilary n 5 tart 1 total moren. audilary n 1 total 2 22049 3 65416 4 65834 5 73591 6 76855 7 93624 8 94555 9 96194 10 96672 11 101207 12 102781 13 103023 14 103448 15 104417 16 103445 17 107323	Stop         C           atten and excision         atten           atten and excision         atten           atten and excision         atten           stop         atten           f5573         atten           f5573         atten           f5573         atten           f5573         atten           f5573         atten           g5325         atten           g8504         atten           103023         atten           103048         atten           104926         atten           108606         atten	Lappoor apartial for Lappoor apartial for Lappoor and the second secon	Image: Second	virion structural protein head vertex assembly cha internal virion protein internal virion protein internal virion protein internal virion protein internal virion protein internal virion protein internal virion protein head closure head-tail adaptor Ad2 head closure head-tail adaptor Ad2 head closure Hc2 terminase large subunit terminase large subunit portal protein head maturation protease head scaffolding protein head maturation protease head scaffolding protein head maturation protease head scaffolding protein major head protein major head protein major head protein	And and nucleotide metabolism roduct aperone	other unknown other unknown phrog_	ter trajectat t → ctrajectat t → trunction + ROG 1715 hea 999 hea 7857 hea 5541 hea 4363 hea 138 hea 138 hea 138 hea	PHROG catta PHROG cata d and packagi d and packagi d and packagi d and packagi id and packagi	ego ing ing ing ing ing ing ing ing ing ing
Teatlastigit cipt           0           5'           3'           0           5'           3'           0           1           0           1           15:31           2           2           0           1           15:31           2           3           65:416           5           7:3591           6           6           7           9           9           9           9           9           9           9           9           9           9           9           9           9           10           10207           12           1020781           13           103023           14           103448           15           104945           17           107323           18           101087 <td>Image: constraint of the second sec</td> <td>AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810</td> <td>Image: Second Second</td> <td>connector tail departies extended at the second sec</td> <td>And and nucleotide metabolism roduct</td> <td>etter ettersopge gete other unknown phrog_ phrog phrog phro</td> <td>ter trajectat t → ctrajectat t → trunction + ROG 1715 hea 999 hea 7857 hea 5541 hea 4363 hea 138 hea 138 hea 138 hea</td> <td>PHROG cate and packagi d and packagi d and packagi d and packagi id and packagi</td> <td>ego ing ing ing ing ing ing ing ing ing ing</td>	Image: constraint of the second sec	AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810 AF15810	Image: Second	connector tail departies extended at the second sec	And and nucleotide metabolism roduct	etter ettersopge gete other unknown phrog_ phrog phrog phro	ter trajectat t → ctrajectat t → trunction + ROG 1715 hea 999 hea 7857 hea 5541 hea 4363 hea 138 hea 138 hea 138 hea	PHROG cate and packagi d and packagi d and packagi d and packagi id and packagi	ego ing ing ing ing ing ing ing ing ing ing
Statistics         Component of the state of the st	Stop         C           atten and excision         atten           atten and excision         atten           atten and excision         atten           stop         atten           f5573         atten           f5573         atten           f5573         atten           f5573         atten           f5573         atten           g5325         atten           g8504         atten           103023         atten           103048         atten           104926         atten           108606         atten	Lappoor apartial for Lappoor apartial for Lappoor and the second secon	Image: second	virion structural protein head vertex assembly cha internal virion protein internal virion protein internal virion protein internal virion protein internal virion protein internal virion protein internal virion protein head closure head-tail adaptor Ad2 head closure head-tail adaptor Ad2 head closure Hc2 terminase large subunit terminase large subunit portal protein head maturation protease head scaffolding protein head maturation protease head scaffolding protein head maturation protease head scaffolding protein major head protein major head protein major head protein	And and nucleotide metabolism roduct	other unknown other unknown phrog_	ter trajectat t → transition + transition	PHROG catta PHROG cata d and packagi d and packagi d and packagi d and packagi id and packagi	ego ing ing ing ing ing ing ing ing ing ing

Figure 2. Screenshot of the genome viewer of *Escherichia* phage T4. (A) Primary view of the genome browser in the sequence genome section upon first access by the user. (B) Adapted browser view after zooming in a chosen genomic region (1) indicated by the red rectangular (2). Information on gene 33 is displayed in a mouse-over window (3), sequence information is provided separately below (4). (C) Results of a search for the term 'head' (5).

	itact			Search databas	e Q	
uery COD Copy query link						
Host species X contains - Staphylococcus au	Basic data		Taxonomy		Isolation	
AND	Strain number	Add Filter ①	Realm	Add Filter ①	Habitat	Add Filter ①
	strain synonym	Add Filter ①	Kingdom	Add Filter ①	Sampling date	Add Filter 🛈
Nucleic acid ty X contains V DNA	Designation	1	Phylum	Add Filter ①	Geographic location	Add Filter 🛈
AND 1	Ncbi tax id	Add Filter ①	Class	Add Filter ①	Country	Add Filter ①
Choose field X contains V insert * to show all (	link	Add Filter ①	Order	Add Filter ①	Coordinate	1
+ Add AND	Name	Add Filter ①	Family	Add Filter ①	Enriched sample	١
	Name short	Add Filter ①	Subfamily	Add Filter ①	Nagoya restrictions	Add Filter ①
kdd OR	History	Add Filter ①	Genus	Add Filter ①		
ıbmit			Species	٩		
			Nucleic acid type	0		
earch fields overview	Life evelo		Full scientific name	Add Filter ①	Morphology	
	Life cycle		Synonym	Ð	Category	Ð
	Type phage	Add Filter ①			Morphotype	Add Filter ①
	Type cycle	Add Filter ①			File	Add Filter ①
					Phage description	Add Filter 🛈
		h and then	Host		Phage size	Ð
	Host growt		Host species	Add Filter 🛈	Media type	0
	Medium name	Add Filter ①	Host range id	Ð	Plaque description	
		Add Filter <sup>(1)</sup>	Host range id Host range link	① Add Filter ①	Plaque description Caption	Add Filter 🛈
2					Caption Q	D
2	DSM medium nur			Add Filter ①	Caption Q	D
ery C-> Copy query link Host species X contains V Staphylococcus au	DSM medium nur			Add Filter ①	Caption Q	①
AND Nucleic acid ty X contains V DNA	DSM medium nur tact hits: 24 3 ID Name		Host range link Collection	Add Filter D	Caption e Q	① ownload table Nucle type
AND AND AND AND AND AND AND AND	DSM medium nur tact hits: 24 3 1D Name #554 Staphylor	nber Add Filter ①	Host range link Collection number	Add Filter ① Search databas Host species Staphylococcus	Caption e Q	O ownload table Nucle type DNA, I
ery Copy query link	DSM medium nur tact Ibits: 24 3 ID Name #558 Staphylou	nber Add Filter ①	Host range link Collection number NCTC 9858	Add Filter ① Search databas Host species Staphylococcus	Caption e Q c ureus PSX2	O     O
ery C=> Copy query link Host species X contains V Staphylococcus aul M AND Nucleic acid ty X contains V DNA M Add AND Add AND Add OR Ibmit	DSM medium nur tact hits: 24 3 ID Name #554 Staphylor #558 Staphylor #561 Staphylor	nber Add Filter ①	Host range link Host range link Collection number NCTC 9858 NCTC 9753	Add Filter ① Search databas Host species Staphylococcus	caption e Q Caption c c c c c c c c c c c c c c c c c c c	ownload table     Nucle     type     DNA, I     DNA, I
ery C=> Copy query link Host species X contains V Staphylococcus aul M AND Nucleic acid ty X contains V DNA M Add AND Add AND Add OR Ibmit	DSM medium nur tact hits: 24 3 1D Name #554 Staphylo #559 Staphylo #569 Staphylo	mber Add Filter ①	Host range link Collection number NCTC 9858 NCTC 9753 NCTC 9316	Add Filter ①	Caption e e Q C C C C C C C C C C C C C C C C C	Ownload table     Nucle     type     DNA, I     DNA, I     DNA, I
ery C=> Copy query link Host species X contains V Staphylococcus aul M AND Nucleic acid ty X contains V DNA M Add AND Add AND Add OR Ibmit	LDSM medium nur tact ID Name #554 Staphylor #558 Staphylor #569 Staphylor #569 Staphylor #569 Staphylor	mber Add Filter ①	Host range link Collection number NCTC 9858 NCTC 9753 NCTC 9316 NCTC 8429	Add Filter ① Search databas Host species Staphylococcus Staphylococcus Staphylococcus Staphylococcus	Caption e e Q C C C C C C C C C C C C C C C C C	O     O
ery C=> Copy query link Host species X contains V Staphylococcus aul M AND Nucleic acid ty X contains V DNA M Add AND Add AND Add OR Ibmit	DSM medium nur tact hits: 24 3 ID Name #554 Staphylor #559 Staphylor #561 Staphylor #569 Staphylor #570 Staphylor #577 Staphylor	mber Add Filter ①	Host range link Collection number NCTC 9858 NCTC 9753 NCTC 9316 NCTC 8429 NCTC 8428	Add Filter ① Search databas Host species Staphylococcus Staphylococcus Staphylococcus Staphylococcus	Caption  Caption Caption  Caption  Caption  Caption  Caption  Caption  Caption  Caption  Caption  Capt	ownload table     Nucle     Nucle     DNA, I     DNA, I     DNA, I     DNA, I     DNA, I
ery C=> Copy query link Host species X contains V Staphylococcus aui III AND Nucleic acid ty X contains V DNA IIII Add AND Add AND Add OR Abmit	DSM medium nur tatter hits: 24 3 10 Name #554 Staphylor #559 Staphylor #569 Staphylor #569 Staphylor #570 Staphylor #579 Staphylor	mber Add Filter ①	Host range link Collection number Collection NCTC 9858 NCTC 9753 NCTC 9316 NCTC 8429 NCTC 8428 NCTC 8420	Add Filter ①	Caption Captio	ownload table type DNA, ii DNA, ii DNA, ii DNA, ii DNA, ii
ery C-> Copy query link Host species X contains V Staphylococcus au	bits:24 bits:24 10 Name #554 Staphylou #558 Staphylou #569 Staphylou #569 Staphylou #570 Staphylou #570 Staphylou #577 Staphylou #579 Staphylou #579 Staphylou	mber Add Filter ①	Host range link Collection number NCTC 9858 NCTC 9753 NCTC 9316 NCTC 8429 NCTC 8428 NCTC 8428 NCTC 8420 NCTC 8418	Add Filter ① Search databas Host species Staphylococcus Staphylococcus Staphylococcus Staphylococcus Staphylococcus Staphylococcus	Caption  Cap	① ownload table Nuclei

Figure 3. Advanced search tool. (A) Query specification with choice of data fields (1) used in Phage*Dive* to characterize viruses. (B) Result of a 'Staphylococcus AND DNA' search. A query link (2), the number of hits (3) and a download option (4) are provided for traceability.

label 4). This tool also offers a function to search for specific genes by name or annotation category (Figure 2C, label 5). Results are displayed in a table containing the following information: start position, end position, DNA strand, locus tag and function (gene product), PHROG no. and category which can be copied and pasted. In addition, a list of these results can be seen on the genome overview table (Figure 2C).

Advanced search. The advanced search module is a key component for a refined search in PhageDive. This tool is accessible through the menu on the top of each page. To retrieve the specific data of interest, the dataset can be accessed through queries that combine several data fields and analyse all entries for the virus strains. The query builder is located on the left-hand side of the tool page (Figure 3). Users are able to combine up to five data fields in an 'AND' combined query and even up to 15 data fields by connecting up to three groups using 'AND' or 'OR' operators (Figure 3, label 1). For each of these data fields to be combined a user can choose

from a list of 72 options. Depending on the data field, input can be provided as text (e.g. plaque description), a numeric field (e.g. temperature) or from a selection field via a dropdown menu (e.g. biosafety level). For text fields, users are assisted by an autocomplete function providing suggestions of possible search terms. To start exploring data provided in the data fields, inserting the wildcard "\*" will allow users to retrieve all data available. As a result, a list of viruses corresponding to the request will be presented on the right (Figure 3B) with the number of hits noted above (Figure 3, label 3). The query can be bookmarked or shared using the 'copy query link' (Figure 3, label 2). The result list can be downloaded in a csv format table using the button on the upper right (Figure 3, label 4). In the same way the complete list of viruses included in PhageDive is also accessible. The viruses are listed in alphabetical order and by culture collection (starting with the Félix d'Hérelle collection). The retrieved data set is restricted to the data fields queried.

## Discussion

PhageDive is a new important resource in the domain of prokaryotic viruses designed to centralise all the data available for individual virus strain. Overall PhageDive currently provides access to 11 331 data points for 1167 phage strains. It is based on a solid foundation of data originating from culture collections with the latter offering direct physical access to a wide variety of viruses and host strains. At the moment PhageDive harbours viral data for members of at least 181 officially ICTV classified genera in 31 families (e.g. Ackermannviridae or Herelleviridae) in 6 classes over 4 realms (mostly Duplodnaviria) (Supplementary Figure S1). Additionally, the viruses in PhageDive cover host organisms from 7 different phyla (14 classes, 35 orders, 103 genera). A major strength of PhageDive is the breadth of the data space provided, notably the availability of plaque and/or TEM pictures, as well as multiple links to the various resources used (e.g. VHRdb for the host range, NCBI for genomes or ICTV for taxonomy). Though databases like e.g. PhageScope (9) might offer data from higher numbers of phages, this data is extracted (i) from sequences only and does not reflect any information from wet lab work like for instance propagation conditions (medium, temperature, need of bivalent cations) that might be valuable requirements for setting up experimental systems and (ii) the majority of analysed viruses is not available for practical experiments. PhageDive provides direct, easy access to standardized data. The consequent standardization of data allows the use of tools to reliably search and analyse the information on phage strains and conduct highly precise searches to find phages with specific properties. For instance, during the SARS-CoV-2 pandemic, researchers looked for model organisms that shared features with the virus like nucleic acid material or size and came up with *Pseudomonas* phage phi6 (20) which might also have been the result for a PhageDive search. Based on the information in PhageDive, candidates for phage mixtures for different experimental approaches can be identified that e.g. share same host strains or originate from the same habitat, but can be differentiated by other categories like morphotype or genomic properties. The implementation of the advanced search tool and an extended interlinkage network with the various data providers make PhageDive a highly robust source of information on prokaryotic viruses. With the integration of a genome viewer, genomes can be analysed more easily, providing a quick overview of their composition and organization. The integrated search function allows the fast and easy identification of specific genes of interest, e.g. endolysins or other enzymes, identifiable by standardized annotation. Various options for saving or downloading search results are provided. To take this a step further, an API will be implemented in the future to provide programmatic access to the continuous growing database. By applying the Creative Commons Attribution license (CC BY), an open access to the data is granted, ensuring the reuse of the data in the science community.

In the future, Phage*Dive* will be expanded in two directions. Firstly, the database will be enriched in parallel to the growth of the DSMZ phage collection and also through the integration of new datasets from external research groups and other phage banks. This will also include the targeted extension with phages for underrepresented organisms like anaerobic organisms or cyanobacteria. Secondly, new data fields will be added to complete the metadata sections, in particular by including transcriptomic and virus receptor data.

# Data availability

All the data are publicly available at https://phagedive.dsmz. de licensed under CC BY 4.0. All data can be downloaded freely without restrictions, the only restriction being that the origin of the data must be properly cited when it is used in other works.

## Supplementary data

Supplementary Data are available at NAR Online.

## Acknowledgements

We cordially thank Prof. Sylvain Moineau and Denise Tremblay (Félix d'Hérelle Reference Center for Bacterial Viruses, Canada) and Dr Jake D. Turnbull (NCTC, UK) for sharing their data. We thank Andrew Millard for help in implementing the PHROG database in Prokka annotation. We kindly thank Laurent Debarbieux (Institut Pasteur) for sharing data from the Virus Host Range Database (VHRdb).

Author contributions: Clara Rolland: Data curation, Investigation, Writing – original draft. Johannes Wittmann: Conceptualization, Data curation, Funding acquisition, Writing – review & editing. Lorenz C. Reimer: Conceptualization, Writing – review & editing. Joaquim Sardà Carbasse: Software. Isabel Schober: Formal analysis, Software. Christian-Alexander Dudek: Software. Christian Ebeling: Software. Julia Koblitz: Visualization, Software. Boyke Bunk: Conceptualization. Jörg Overmann: Conceptualization, Funding acquisition, Writing – review & editing.

# Funding

Deutsche Forschungsgemeinschaft [SPP 2330, OV 20/29-1, WI 5350/1-1 project 465136285]. Funding for open access charge: Deutsche Forschungsgemeinschaft [OV 20/29-1, WI 5350/1-1 project 465136285].

# **Conflict of interest statement**

None declared.

#### References

- Adriaenssens, E.M., Roux, S., Brister, J.R., Karsch-Mizrachi, I., Kuhn, J.H., Varsani, A., Yigang, T., Reyes, A., Lood, C., Lefkowitz, E.J., *et al.* (2023) Guidelines for public database submission of uncultivated virus genome sequences for taxonomic classification. *Nat. Biotechnol.*, 41, 898–902.
- 2. Ackermann,H.-W. (2011) Bacteriophage taxonomy. *Microbiol. Austr.*, **32**, 90–94.
- 3. Dion, M.B., Oechslin, F. and Moineau, S. (2020) Phage diversity, genomics and phylogeny. *Nat. Rev. Micro.*, **18**, 125–138.
- Roux,S., Brum,J.R., Dutilh,B.E., Sunagawa,S., Duhaime,M.B., Loy,A., Poulos,B.T., Solonenko,N., Lara,E., Poulain,J., *et al.* (2016) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*, 537, 689–693.
- Williamson,K.E., Fuhrmann,J.J., Wommack,K.E. and Radosevich,M. (2017) Viruses in soil ecosystems: an unknown quantity within an unexplored territory. *Annu Rev Virol*, 4, 201–219.
- Bolduc,B., Shaughnessy,D.P., Wolf,Y.I., Koonin,E.V., Roberto,F.F. and Young,M. (2012) Identification of novel positive-strand RNA

viruses by metagenomic analysis of archaea-dominated Yellowstone Hot Springs. J. Virol., 86, 5562.

- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P. and Rohwer, F. (2003) Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriology*, 185, 6220–6223.
- Koonin, E.V., Krupovic, M. and Dolja, V.V. (2023) The global virome: how much diversity and how many independent origins? *Environ. Microbiol.*, 25, 40–44.
- 9. Wang,R.H., Yang,S., Liu,Z., Zhang,Y., Wang,X., Xu,Z., Wang,J. and Li,S.C. (2024) PhageScope: a well-annotated bacteriophage database with automatic analyses and visualizations. *Nucleic Acids Res.*, **52**, D756–D761.
- Gao,N.L., Zhang,C., Zhang,Z., Hu,S., Lercher,M.J., Zhao,X.M., Bork,P., Liu,Z. and Chen,W.H. (2018) MVP: a microbe-phage interaction database. *Nucleic Acids Res.*, 46, D700–D707.
- Lamy-Besnier, Q., Brancotte, B., Ménager, H. and Debarbieux, L. (2021) Viral Host Range database, an online tool for recording, analyzing and disseminating virus-host interactions. *Bioinformatics*, 37, 2798–2801.
- 12. Russell,D.A. and Hatfull,G.F. (2017) PhagesDB: the actinobacteriophage database. *Bioinformatics*, 33, 784–786.
- Turner,D., Shkoporov,A.N., Lood,C., Millard,A.D., Dutilh,B.E., Alfenas-Zerbini,P., van Zyl,L.J., Aziz,R.K., Oksanen,H.M., Poranen,M.M., *et al.* (2023) Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. *Arch. Virol.*, 168, 74.

- Reimer,L.C., Sardà Carbasse,J., Koblitz,J., Ebeling,C., Podstawka,A. and Overmann,J. (2022) BacDive in 2022: the knowledge base for standardized bacterial and archaeal data. *Nucleic Acids Res.*,50, D741–D746.
- Koblitz, J., Halama, P., Spring, S., Thiel, V., Baschien, C., Hahnke, R.L., Pester, M., Overmann, J. and Reimer, L.C. (2023) MediaDive: the expert-curated cultivation media database. *Nucleic Acids Res.*, 51, D1531–D1538.
- Seemann,T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30, 2068–2069.
- 17. Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Pérez Bucio, R.E., Mom, R., Toussaint, A., Petit, M.A. and Enault, F. (2021) PHROG: families of prokaryotic virus proteins clustered using remote homology. NAR Genom. Bioinform., 3, lqab067.
- Nishimura,Y., Yoshida,T., Kuronishi,M., Uehara,H., Ogata,H. and Goto,S. (2017) ViPTree: the viral proteomic tree server. *Bioinformatics*, 33, 2379–2380.
- Hulo,C., de Castro,E., Masson,P., Bougueleret,L., Bairoch,A., Xenarios,I. and Le Mercier,P. (2011) ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.*, 39(Database issue), D576–D582.
- 20. Gomes, M., Bartolomeu, M., Vieira, C., Gomes, A.T.P.C., Faustino, M.A.F., Neves, M.G.P.M.S. and Almeida, A. (2022) Photoinactivation of Phage Phi6 as a SARS-CoV-2 model in wastewater: evidence of efficacy and safety. *Microorganisms*, 10, 659.