Genomic and seasonal variations among aquatic phages infecting the Baltic Sea

Gammaproteobacteria Rheinheimera sp. BAL341

E. Nilsson, K. Li, J. Fridlund, S. Šulčius, C. Bunse, C.M.G Karlsson, M. Lindh, D. Lundin, J. Pinhassi and K. Holmfeldt

Linnaeus University, Faculty of Health and Life Sciences, Department of Biology and Environmental Science, Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), SE-39231 Kalmar, Sweden.

Current address: Nature Research Centre, Laboratory of Algology and Microbial Ecology, Akademijos str. 2, LT-08412 Vilnius, Lithuania

Current address: Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany

Current address: Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), Ammerländer Heerstrasse 231, D-26129 Oldenburg, Germany

Current address: Swedish Meteorological and Hydrological Institute (SMHI), Sven Källfelts gata 15, SE-42171 Göteborg, Sweden

#Corresponding author: karin.holmfeldt@lnu.se

Keywords: bacteriophage, Baltic Sea, genomics, temporal variation

Running title: Genomic and seasonal variation among aquatic myophages
Abstract

Knowledge in aquatic virology has been greatly improved by culture-independent methods, yet there is still a critical need for isolating novel phages to identify the large proportion of “unknowns” that dominate metagenomes and for detailed analyses of phage-host interactions. Here, 54 phages infecting *Rheinheimera* sp. strain BAL341 (*Gammaproteobacteria*) were isolated from Baltic Sea seawater and characterised through genome content analysis and comparative genomics. The phages showed a myovirus-like morphology and belonged to a novel genus, for which we propose the name *Barbavirus*. All phages had similar genome sizes and number of genes (80-84 kb; 134-145 genes), and based on average nucleotide identity and genome BLAST distance phylogeny the phages were divided into five species. The phages possessed several genes involved in metabolic processes and host signalling, such as ribonucleotide reductase, thymidylate synthase, *phoH*, and *mazG*. One species had additional metabolic genes involved in pyridine nucleotide salvage, possibly providing a fitness advantage by further increasing the phages’ replication efficiency. Recruitment of viral metagenomic reads (25 Baltic Sea viral metagenomes from 2012-2015) to the phage genomes showed pronounced seasonal variations, with increased relative abundance of barba-phages in August and September synchronised with peaks in host abundances as shown by 16S rRNA gene amplicon sequencing. Overall, this study provides detailed information regarding genetic diversity, phage-host interactions, and temporal dynamics of an ecologically important aquatic phage-host system.

Importance

Phages are important in aquatic ecosystems as they influence their microbial hosts through lysis, gene transfer, transcriptional regulation, and expression of phage metabolic genes. Still, there is limited knowledge of how phages interact with their hosts, especially at fine scales. This newly isolated *Rheinheimera* phage-host model system constitutes highly similar phages infecting one host strain.
This relatively limited diversity has previously only been seen when isolating smaller numbers of phages and points toward ecological constraints affecting the *Rheinheimera* phage diversity. The variation of metabolic genes among the species points towards varying fitness advantages, opening up for future hypothesis testing. Phage-host dynamics monitored over several years point towards recurring “Kill the Winner” oscillations and an ecological niche fulfilled by this system in the Baltic Sea. Identifying and quantifying ecological dynamics of such phage-host model systems *in situ* enable us to understand and study the influence of phages on aquatic ecosystems.

### Introduction

Viruses are the most abundant biological entities on the planet (1). With an approximate abundance of $10^7$ viral particles mL$^{-1}$ in surface waters of the global oceans they outnumber their microbial hosts approximately 10-fold (2). Through infection and lysis of 20% of the bacterial community on a daily basis, viruses that infect bacteria (bacteriophages or phages for short) are important factors influencing bacterial mortality and genetic diversity (3). Lately, high-throughput sequencing techniques have improved our knowledge regarding the viral communities, including viruses that infect so-far “uncultured” hosts. Metagenomic analyses have unravelled large-scale spatial variations within viral populations (4, 5), shed light on viral-host dynamics (6, 7), and predicted the functional gene potential of viral communities (8, 9). Moreover, single-cell genomics (10) of phage-host pairs from natural communities and viral tagging (11), where host-specific metagenomes are analysed, have uncovered viral diversity and host-interactions on finer scales. Since the discovery of auxiliary metabolic genes (AMGs) (12-14), which typically are metabolic genes transferred from the host’s genome to the phage genome that enable continued or increased phage replication during infection (15), the functional capacity of viral communities and the potential of these genes to influence biogeochemical cycles have gained increased interest.
The isolation of phage-host pairs is essential to elucidate fine-scale patterns of phage-host interactions. For example, different phages infecting the same bacterial species can display a wide range of infection abilities, both on a qualitative (who they infect) and quantitative (how well they infect) scale (16). In addition, micro diversity within viral species has the ability to affect not only host range but also replication efficiency, seen as e.g. burst size (17). Further, a viral strain can have different replication efficiencies on different host strains (17, 18) due to differences in the host transcriptional (19) and translational (20) response. Thus, the knowledge gained from phage isolation and experimental studies is imperative to fully understand phage-host systems and hypotheses that derive from sequencing methods such as metagenomics, particularly with regards to viral diversity and ecological importance. This is especially important for non-model systems since we do not know how they differ from model systems before they have been isolated.

Clinically important phages infecting different gammaproteobacterial species, including e.g. *Escherichia* spp. and *Pseudomonas* spp., are among the most well-studied phage-host systems in the world (21). However, in marine environments, phages infecting *Gammaproteobacteria* have received less attention, and research has mainly been focused on phages infecting *Pseudoalteromonas* (22-24) or *Vibrio* (13, 25, 26). These isolated *Pseudoalteromonas* (27, 28) and *Vibrio* (26) phages display a large diversity, with phages belonging to a variety of genera and families, which is similar to what has been seen for *Escherichia* (29) and *Pseudomonas* (30) phages. In addition, a podovirus and a myovirus, respectively, have been identified infecting the “uncultured” gammaproteobacterial SAR92 and SAR86 clades using single amplified genomics (10). However, information regarding phages infecting other *Gammaproteobacteria* species is sparse.
Members of the bacterial genus *Rheinheimera* exist in various environments and they have been found in soils (31) and aquatic habitats, including both freshwater (32) and marine (33) environments. They have also been shown to proliferate in more extreme environments, like iron backwash sludge (34), and are able to degrade hydrocarbons such as n-alkanes (35). Bacteria within this genus have been detected in the brackish Baltic Sea (36) and transplant experiments with shifting salinities have noted stabilising or increasing abundances as adjustment effects (37). Even though species in the *Rheinheimera* genus are widespread, to our knowledge, no phages infecting these species have been isolated.

The aim of this study was to isolate and characterise aquatic phages infecting environmentally important bacteria, to increase our knowledge about phage diversity and their ecological relevance within aquatic ecosystems. We present 54 genetically similar phages, that are distinct from previously described phages, infecting the Baltic Sea isolate *Rheinheimera* sp. strain BAL341. We propose that these phages are assigned to a novel viral genus, *Barbavirus*, consisting of five species, which contain genes potentially involved in various metabolic pathways. The phages’ prevalence is evident by metagenomic recruitment and their temporal abundance coincides with the host’s abundance in late-summer.

**Results**

**Genomic characteristics**

Fifty-four phage isolates, originating from individual plaques, infecting *Rheinheimera* sp. strain BAL341 were isolated from the long-term sampling station Linnaeus Microbial Observatory (LMO) in the Baltic Sea Proper (38, 39), where 31 isolates were obtained in August 2015 and 23 in September 2015. The proposed names of the isolates, following ICTV standards (40), are
Rheinheimera phage vB_RspM-barba followed by numbers indicating the order of isolation and A or S, representing the time of sampling (August or September), e.g. Rheinheimera phage vB_RspM-barba18A (supplemental Table S1). All 54 phage isolates were whole-genome sequenced and the genomes were assembled with a coverage ranging from 193x to 4,789x (supplemental Table S1). Identical k-mers at both ends of the genomes indicated that they were circular. The phage genome sizes varied between 80 and 84 kb, with between 134 and 145 predicted genes in each genome. All versus all genome comparison showed that the barba-phages shared more than 70% average nucleotide identity (ANI; Fig. 1, supplemental Table S2) across their entire genomes, indicating that they belong to the same genus (>50% ANI) (41).

Phylogenetic analysis using VICTOR (42) showed that the isolated phages formed a separate clade that was distinct from previously sequenced and characterised phages (Fig. 2, supplemental Table S3). Whereas the low similarity to the reference genomes prevents us from drawing firm conclusions between the novel phages and the reference genomes, the short branches within the clade containing the novel phages lets us propose that these phages should be assigned to a novel genus, for which the name Barbavirus (Baltic Sea Rheinheimer strain BAL341) is suggested. Within this novel genus, 48 of the phage genomes shared more than 95% ANI and therefore belong to one species (Fig. 1, supplemental Table S2). Using this species cut off, four additional species were distinguished, where two species were represented by two isolates and two species only by a single isolate each (Fig. 1, supplemental Table S2). This agreed with the VICTOR-analysis for which five species within this one genus was identified (Fig. 2, supplemental Table S3). The five species were named based on their suggested type phages: 1) Rheinheimera virus Barba18A, 2) Rheinheimera virus Barba21A, 3) Rheinheimera virus Barba5S, 4) Rheinheimera virus Barba8S, and 5) Rheinheimera virus Barba19A (Fig. 1). From here on, species are indicated with full species name, i.e. Rheinheimera virus Barba18A, while isolates are indicated with their
unique identifier, e.g. barba18A; also, we use the term “barba-phages” to denote all isolates collectively while *Barbavirus* indicates the species within the genus.

**Morphology and replication characteristics**

All phages displayed similar plaque morphology with round, clear plaques up to 4 mm in diameter, which appeared within 24-48 hours. Transmission electron microscopy of barba18A, the type phage of the genus, showed a myovirus-morphology (Fig. 3a), with a capsid diameter of 72.1 (standard deviation ±2.7) nm, a tail length of 88.7 (±2.2) nm, and a tail width of 18.8 (±1.5) nm (n=50). When grown in MLB medium and replicating on *Rheinheimera* sp. strain BAL341, barba18A had a burst size of 162 (±20) phages and a latent period of 50 (±8.7) minutes (n=3; Fig. 3b). Barba19A, the type phage of the most divergent species (Fig. 1 and 2), had a significantly smaller burst size of 101 phages (±17; n=3; Welch two sample t-test, df=3.88, p=0.017) and longer latent period of 80 minutes (±0; n=3; Welch two sample t-test, df=2, p=0.027; Fig. 3b).

We found three bacterial strains in our in-house bacterial collection (namely *Rheinheimera* sp. strain BAL331, *Rheinheimera* sp. strain BAL335, and *Rheinheimera* sp. strain BAL336) that had 97-99% 16S rRNA gene sequence identity to other cultured *Rheinheimera* bacterial strains in NCBI (supplemental Table S4). Compared to the model bacterium *Rheinheimera* sp. strain BAL341 they had 95-98% sequence identity on the sequenced parts of the 16S rRNA gene (supplemental Table S4). Neither barba18A nor barba19A could infect any of these other *Rheinheimera* strains.

**Gene functionality**
The core genome shared among the 54 novel phage genomes consisted of 97 genes with at least 70% amino acid identity, while the flexible genome consisted of 99 additional genes, most of them found in *Rheinheimera virus Barba19A*. Thus, the pan-genome of these phages was made up of a total of 196 genes (supplemental Table S5). Of the genes in the pan-genome, 81 genes had significant (e-value <0.001) sequence identity to proteins in NCBI nr, where 81% of the matches were to either *Gammaproteobacteria* or phages infecting bacteria within this class (supplemental Table S5). Combined with matches to the pfam database (adding function to 16 additional genes), 43 genes could be assigned to a predicted function (Table 1, supplemental Table S5). No tRNAs were detected in the genomes.

The genomes of the five species were similarly arranged and consisted of two larger modules, one for structural genes and one for host interaction genes (Fig. 4). Within the structural module, genes involved in the structure of the phage (e.g. capsid proteins and tail elements), packaging of DNA (e.g. large terminase subunit), and peptidases were detected (Table 1, supplemental Table S5). The host interaction module consisted of genes involved in for example DNA processing, replication, and recombination (e.g. DNA polymerase and primase), host signalling (nucleoside triphosphate pyrophosphohydrolase (*mazG*)), and various metabolic processes (e.g. ribonucleotide reductase (RNR), thymidylate synthase (*thyl*), and protein PhoH (*phoH*) (Table 1, supplemental Table S5). Of the metabolic genes, two were part of the core genome, the RNR genes *nrdA* and *nrdB*. These genes had 99.9-100% amino acid sequence identity in all barba-phage isolates. The other metabolic genes found in the barba-phages, glutaredoxin (*glrx*) and *thyl* involved in nucleotide metabolism and *phoH* potentially involved in phosphate metabolism, were part of the flexible genome. For these genes, two different variants were detected among the barba-phages (Table 2), where the common version was shared among all phages except the
phages within one particular species, either *Rheinheimer virus Barba19A* or *Rheinheimer virus Barba5S* (Table 2). The genes annotated as *glrx* in barba18A (the common one) and in barba19A (the rare) shared 43% amino acid identity across 92% of the genes (Table 2). Similarly, the common and rare (in barba5S) versions of the genes annotated as *phoH* shared 31% amino acid similarity across 94% of the genes. However, the two versions of the thymidylate synthase gene did not show significant sequence identity (blastp, no e-value < 0.001) between the two different phage variants (Table 2). In addition, the *Barba19A* isolates contained two metabolic genes that the other isolates lacked, a ribose-phosphate pyrophosphokinase (*prs*) and a nicotinate phosphoribosyltransferase (*pncB*) that are involved in pyridine nucleotide salvage. These phage genes, potentially involved in different metabolic processes, either showed low sequence similarity matches (<40% amino acid identity if e-value <0.001) or did not provide any significant matches (no e-value <0.001) to genes with similar functions within the host genome (Table 2).

**Temporal variations**

The barba-phages were isolated at two different time points, August 19\textsuperscript{th} 2015 and September 16\textsuperscript{th} 2015. The majority of the isolates belonged to *Rheinheimer virus Barba18A* and were isolated from both time points (Fig. 1). In addition, the two *Rheinheimer virus Barba5S* phages were isolated from both August and September (Fig. 1). Phages belonging to the other species were either isolated in August (*Rheinheimer virus Barba21A* and *Rheinheimer virus Barba19A*: one and two isolates respectively) or September (*Rheinheimer virus Barba8S*: one isolate).
The seasonal dynamics of *Barbavirus* were investigated by competitive recruitment of viral metagenomic reads from the LMO station against the genomes of the five representative phages (barba18A, barba21A, barba5S, barba8S, and barba19A; supplemental Table S6). In order for a phage to be counted as present in a metagenomic dataset, at least 75% of a phage genome should be covered by viral metagenomic reads with 90% nucleotide identity (5, 43). According to this definition and using the competitive recruitment, barba19A was detected at one time point while the other four species did not have enough coverage at any time points (supplemental Table S6). Due to the high sequence similarity among the barba-phages, competitive recruitment was unable to distinguish the true source of reads that belong to the identical parts of the genomes and thereby be unable to convey the prevalence of the phages *in situ*. Therefore, barba19A was selected to represent the barba-phages and was used for individual recruitment analysis, which resulted in detection of barba19A (<75% coverage) in August and September samples, and one July sample (Fig. 5). The seasonal dynamics of host abundance in the Baltic Sea was determined by quantification of a 16S rRNA gene amplicon sequence variant (ASV) with 100% sequence identity to *Rheinheimera* sp. strain BAL341. These results showed increasing relative host abundances in June to August between 2012-2015, which largely mirrored or slightly preceded the phage dynamics (Fig. 5).

**Discussion**

In this study we describe a previously unknown phage genus, whose members all infected the genome-sequenced bacterial isolate *Rheinheimera* sp. strain BAL341. Genetically, the phage genes showed highest similarity to different *Gammaproteobacteria* or phages infecting *Gammaproteobacteria*. Several of the best viral matches, in particular among genes within the structural module, were to myoviruses (supplemental Table S5), which, together with the
morphological characterisation, placed this genus within the family of *Myoviridae* (Fig. 3a). The head and tail width were average sized compared to other myoviruses, whereas the tail length was relatively short (44, 45). Replication studies of the type phage of the genus, barba18A, and the type phage of the most divergent species, barba19A, showed a burst size (162 and 101 phages, respectively) and latent period (50 and 80 minutes, respectively) within the range that is commonly seen among marine phages (17, 25, 46-49). These myoviruses do not fall into any previously known genera and do not share enough sequence similarity to the most similar phages to be able to determine which is their closest evolutionary relatives (Fig. 2). Therefore, we propose that this novel phage genus should be named *Barbavirus*.

All 54 barba-phages belonged to the same genus and thus their diversity appeared limited compared to the genetic diversity seen among phages infecting e.g. human associated *Escherichia coli* and *Pseudomonas aeruginosa* as well as the Baltic Sea bacterium *Cellulophaga baltica*, where phages belonging to multiple genera and even different families have been isolated (50). Potentially, the limited diversity within the barba-phages could be due to the use of a single host strain for isolation, compared to e.g. 20 host strains for the *C. baltica* system (51). This limited diversity is also known from other phage-host systems where single host strains have been used for isolation of phages, e.g. among phages infecting *Roseobacter SIO67* (52) and *Bacillus thuringiensis* (53). However, it should be taken into consideration that a smaller number of phages were isolated for these systems compared to the relatively high number of barba-phages. On the other hand, multiple phage genera and even phages of different families have been isolated for other model systems using an individual host strain (47, 54-56), also when few phages were isolated (47). Thus, the use of one individual host strain for isolation cannot be the sole explanation for the limited diversity noted among the barba-phages. Instead, this isolation of
only highly similar phages might suggest that the barba-phages are the dominant viral predators for *Rheinheimera* sp. strain BAL341, as observed during the late summer 2015. Further, the high nucleotide similarity among the barba-phages might be a result of ecological constraints, such as the reliance on host genes, imposed on the phages while infecting *Rheinheimera* sp. strain BAL341.

The majority (78%) of the pan-genome of the barba-phages could not be given a predicted function, but the genes for which a functional annotation could be provided showed a genome organisation within the barba-phages that is typical among phages (22). One module contained genes involved in host interactions and one module with genes involved in phage structure and packaging. The host interaction module contained a few genes involved in DNA replication, for example DNA primase, DNA polymerase, and DNA binding protein. Such genes are commonly found in phages (52, 55) and are involved in the replication of the phage DNA. However, several functions needed for DNA replication were not detected through amino acid similarity searches and, potentially, the phage could use host-genes for those functions (57, 58) or these functions are among the large number of genes without predicted function.

The barba-phage genomes contained several genes potentially involved in metabolic processes, which are of interest since they can increase the amount of phage progeny (59). Genes involved in nucleotide metabolism, e.g. RNR and thymidylate synthase, are commonly detected within phage genomes and viral metagenomes (60-62). This includes aquatic myoviruses, for example T4-like cyanophages (61, 63) and *Vibrio* phage KVP40 (13). All barba-phages have a complete class I RNR, consisting of the *nrdA* and *nrdB* genes, a glutaredoxin gene, and thymidylate synthase, which are all clustered close together in the genome (Fig. 4). The proteins encoded by
the RNR genes *nrdA* and *nrdB* form an enzyme which reduces ribonucleotides to deoxyribonucleotides (64). The class I enzyme is thereafter reduced to its active form by either glutaredoxin or thioredoxin (64), which for the barba-phages would suggest the use of the phage encoded glutaredoxin. Another key part of nucleotide metabolism is thymidylate synthase, the enzyme that converts deoxyuridine 5'-monophosphate to deoxythymidine 5'-monophosphate (65), and is thus an important function for the phages to be self-sustaining on the nucleotide dTTP. The barba-phages gene is most similar to *thy1/X*, which has the greatest efficiency in the presence of FAD (65). Thymidylate synthase is commonly found among other marine phages, in particular in relatively large (>70 kb) T4-like cyanomyophages (14), N4-like *Roseobacter* phages (66) and large sipho- and podophages infecting *C. baltica* (50). These genes, thymidylate synthase, RNRs and glutaredoxin, all potentially involved in different parts of the nucleotide metabolism, would provide the barba-phages with the potential to utilize their own genes to acquire the metabolites they need for DNA replication, if host nucleotide metabolism is not sufficient or is down regulated during phage infection.

The *Rheinheimera virus Barba19A* isolates also contained two additional genes potentially involved in pyridine nucleotide salvage, a ribose-phosphate pyrophosphokinase and a nicotinate phosphoribosyltransferase (NAPRTase) (67). Ribose-phosphate pyrophosphokinase is an enzyme that transforms ribose 5-phosphate and ATP to phosphoribosyl pyrophosphate (PRPP) and AMP (68). PRPP is a substrate in several enzymatic processes in bacteria (69), and is a precursor of purine and pyrimidine nucleotides, as well as the pyridine nucleotide NAD+ (70). Purine and pyrimidine are used by RNR to produce the nucleotides that are needed during the replication of the phage. Whereas within the pyridine nucleotide salvage pathway, PRPP and nicotinic acid is converted by NAPRTase to nicotinate mononucleotide and pyrophosphate, most efficiently by...
hydrolysis of ATP (67). Nicotinate mononucleotide can then be converted to NAD+, which is involved in amino acid catabolism, DNA ligase reactions and is a key coenzyme in redox reactions where its reduced form (NADH) transfers electrons to the electron transport chain (71). However, the function within phages has not been clarified. A complete pyridine nucleotide salvage pathway, with nadV and natV, is functional in Vibrio phage KVP40, where nadV has functional similarity to NAPRTase (72). The nadV gene has also been found in other phages, including another Vibrio phage (73) and a marine Pseudoalteromonas phage (74). The reaction of nadV is followed by the conversion of nicotinamide mononucleotide to NAD+ by natV in the presence of ATP (72), for which a homologue is missing in Rheinheimera virus Barba19A. While the particular functionality of the ribose-phosphate pyroposphokinase and the NAPRTase in the Rheinheimera virus Barba19A phages have not been verified, these extra metabolic genes could be hypothesised to provide this species a fitness advantage compared to the other barba-species by increasing the efficiency of the RNRs through redox reactions. However, this is not evident for replication on Rheinheimera sp. strain BAL341, where barba19A had a significantly smaller burst size and longer latent period compared to barba18A. It should be taken into consideration that during the replication experiment, the nutrient supply was high and did not correspond to in situ conditions, rendering it hard to detect the increased metabolic potential of barba19A that might be evident as a fitness advantage in the Baltic Sea.

All barba-phage genomes contained a mazG homologue, even though the Rheinheimera virus Barba19A gene showed no sequence similarity compared to the gene in the other phages (Table 1 and 2). MazG is predicted to halt self-programmed cell death (75) and has been hypothesised to be used by the phage to keep the host alive during phage propagation (76). A functional MazG acts by decreasing the cellular pool of guanosine 3’,5’-bispyrophosphate (ppGpp), which changes
the metabolism of the cell as a stress response, that in turn will halt the toxic effects of MazF by re-synthesising MazE (75). Bioinformatic analysis has suggested that MazG is overrepresented among marine phages (22), such as the marine *Pseudoalteromonas* H105/1 (22), *Roseophage* SIOI (52), and T4-like cyanophages (77). Therefore, MazG has been suggested to have an important role in marine phage systems (22). Also, *phoH* was detected in all barba-phages, even though the *phoH* gene in *Rheinheimera virus Barba5S* differed from the version found in the other barba-phages. In *E. coli*, *phoH* is part of the Pho regulon, where it is expressed during phosphate limitation, potentially to transport or use phosphate (78). However, a bioinformatic approach to investigate the function of PhoH suggests that the protein could also be involved in phospholipid metabolism and RNA modification or fatty acid beta-oxidation (79). PhoH has been shown to be upregulated during late infection of the cyanobacterium *Prochlorococcus* MED4 by cyanophage P-SSP7 (80). This upregulation is speculated to be a reaction to stress in the host due to phage infection (80), or a consequence of P-limitation caused by phage production (27), where an increased up-take of phosphorus would increase phage replication success. Even though the exact mechanism of the *phoH* gene is not fully determined, the gene appear to be more important for marine phages than phages from other environments; the *phoH* gene is present in 40% of sequenced marine phages compared to 4% of phages from other environments (81). The presence of both *mazG* and *phoH* in the phage genome might be indicative of adaptations to counter the host response to stressors caused by the phage-infection and might increase the quantity of newly created phages.

Phage metabolic genes can be used to increase replication success and might originate from the host’s genome. However, regarding the barba-phages, their metabolic genes were distantly related to genes of similar function occurring in the host *Rheinheimera* sp. strain BAL341 (Table 15
For example, the RNR genes within the barba-phages shared low sequence similarity to the host’s RNR genes (*nrdA* 23% and *nrdB* 22%; Table 2). In general, phages have the same RNR as their host, but the opposite has also been noted in several instances (60, 82) and has been suggested as a sign of either horizontal gene transfer within phages with extended host range or phages with shifted host range over longer time periods (83), which potentially could be the case for these genes in the barba-phages.

The phages isolated in August and September were highly similar both at the nucleotide and amino acid level and there were no obvious indications that this phage population had changed genetically between August and September. Within the temporal viral metagenomes from 25 different samplings between 2012 and 2015, the barba-phages, represented by barba19A, were detected from July until September, when their host was either abundant or had just declined (Fig. 5). The relative abundance of barba19A within this viral metagenomic dataset was low compared to investigation of uncultured phages within marine metagenomes (5), but the recurring presence of barba19A during the entire investigated time-frame indicate that the barba-phages form a stable part of the Baltic Sea phage community in late summer (Fig. 5).

While barba19A was the species that was covered by the largest number of viral metagenomic reads during a competitive recruitment (supplemental Table S6) and therefore represented the abundance of the barba-phages in the temporal data set (Fig. 5), *Rheinheimera virus Barba18A* had most associated isolates (Fig. 1). This could be due to a shift in the phage community where *Rheinheimera virus Barba19A* had dominated the barba-phage population during the earlier years, as seen as coverage by viral metagenome reads, while *Rheinheimera virus Barba18A* was the dominant species at the time-points of isolation, hence the large number of associated...
isolates. On the other hand, *Rheinheimera* sp. strain BAL341 may not be the optimal or preferred host for the wild type *Rheinheimera virus Barba19A* phages, therefore not retrieving as many plaques as *Rheinheimera virus Barba18A*. This could explain the reduced replication success for barba19A when replicating on BAL341 compared to barba18A (Fig. 3b). Similar behaviour is known from aquatic *Cellulophaga* phages replicating on non-optimal hosts (16, 17). Efficiency of infection can be another indication of host suitability, but based on the one-step growth curves (difference between total and free phages before the burst, Fig. 3b), it was similar for barba18A and barba19A. Yet, efficiency of infection is highly linked to the host bacterium that was latest infected (16) and thus a high efficiency of infection of our isolated barba19A, which has been passed through *Rheinheimera* sp. strain BAL341 multiple times, might not be representative for the efficiency of infection of the wild type barba19A. However, no alternative host could be detected for barba18A and barba19A within our bacterial culture collection, but it should be noted that the range of suitable alternative hosts was limited. Still, *Rheinheimera virus Barba19A* is the species that could be detected in the Baltic Sea during several consecutive years (Fig. 5, supplemental Table S6) and potentially their additional metabolic genes provide them with the fitness advantage to make them the dominant member of the *Barbavirus* population.

Whether or not barba18A or barba19A was the dominant barba-phage species during 2015, these replication differences and the temporal dynamics of phages and their host seen over multiple years are indicative of a within-genus phage variation, where different phage species will have varied impacts on their host strain. The pattern of increased phage abundance that coincides with the increase of host abundance, measured as relative abundance of 16S rRNA gene amplicon reads (Fig. 5), is representative of the Lotka-Volterra-like phage-host oscillations suggested in the “kill the winner”-hypothesis (84). The yearly, reoccurring increases of *Rheinheimera* sp. strain 17
BAL341 and its phages after the summer cyanobacteria bloom (38) points towards an ecological importance of this phage-host system connected to bloom degradation or the exudates excreted during bloom declines. Given its ecological relevance and the high number of metabolic genes, the Rheinheimera phage-host system represents an important novel model system for hypothesis driven research, such as describing different genes of unknown function, how the expression of metabolic genes during infection at various growth conditions influence phage progeny, and how the phage replication success is affected by different environmental factors, such as temperature or nutrient concentrations. These are important aspects to consider for an increased understanding regarding phage-host interactions and the ecological implications of these microbial players in aquatic environments.
Material and methods

Bacterial isolation

Water for bacterial isolation was collected at 2 meters depth at LMO (N 56° 55.8540’, E 17° 3.6420), situated 10 km east off the coast of Öland, Sweden, with a Ruttner sampler on 12th of July 2012.

Bacterial colonies were grown on Zobell agar plates (1 g yeast extract (Becton, Dickinson and Company (BD)), 5 g Bacto-peptone (BD), and 15 g of Bacto-agar (BD) in 800 mL filtered Baltic Sea water and 200 mL MQ water) at room temperature for three to four days and each isolate was clean-streaked three times. Bacterial DNA for sequencing of the 16S rRNA gene (BAL331, BAL335, BAL336, and BAL341) and for whole genome sequencing (BAL341) was isolated using the E.Z.N.A. Tissue DNA kit (Omega bio-tek) according to the manufacturer’s instructions. Briefly, cells were lysed, protease treated, and bound to a HiBind DNA Mini column. While bound to the column, the DNA was washed with DNA Wash Buffer diluted with 100% ethanol and then eluted in 10 mM Tris. The 16S rRNA gene was PCR amplified with the primers 27F and 1492R and sequenced using Sanger dideoxy sequencing at Macrogen Europe, Amsterdam, Netherlands. The whole genome DNA was verified with regards to quality and quantity with Nanodrop (Thermo Scientific) and Qubit (high sensitivity DNA kit; Invitrogen, Life technologies), respectively, and sequenced with Illumina HiSeq at SciLife/NGI (see section Sequencing of whole genomes).

Viral isolation

Water for viral isolation was collected on 19th of August and 16th of September 2015 at LMO. At each time point, 10 L of water was prefiltered through 0.22 µM sterivex filters (Millipore) and the water was concentrated using a 30 kDa tangential flow filter (Millipore). The resulting 100-150 mL was further concentrated through a 50 kDa amicon ultra centrifugal-filters (Millipore) to 13-30 mL. In order to isolate phages, plaque assays (85) with the two viral concentrates were conducted using
Rheinheimera sp. strain BAL341 as host. Here, bacteria grown overnight in liquid Zobell medium (1 g yeast extract (BD) and 5 g Bacto-peptone (BD) in 800 mL filtered Baltic Sea water and 200 mL MQ water) were mixed with the viral concentrates and molten, 32°C top-agar (marine sodium magnesium (MSM) buffer (450 mM NaCl (Sigma), 50 mM MgSO₄·7H₂O (Fisher), and 50 mM Trizma base (Sigma), pH 8) with 0.5% low melting point agarose (Thermo Fisher Scientific)). The mixture of bacteria, viral concentrate, and molten agar was spread evenly onto Zobell agar plates. The plates were incubated at room temperature (RT) for 2 days and plaque formation was monitored on a daily basis. To obtain pure phage isolates, individual plaques (54 in total) were picked with a sterile 100 µL pipette tip, dispersed in MSM, and re-plated three times as described above, but now with phages suspended in MSM and not viral concentrates. Pure isolates were harvested by adding 5 mL MSM to fully lysed plates, the top agar-layer was shredded with an inoculation loop, and the plates were incubated on a shaking table (40 rpm) for at least 30 minutes. The MSM with suspended phages was collected into a falcon tube, centrifuged for 10 minutes at 10,000 x g, filtered through 0.2 µM syringe filter (BD) and stored at 4°C.

**Transmission Electron Microscopy**

Phage particles were purified through caesium chloride (CsCl₂) density gradient centrifugation (described by (86), with modifications as described below) and analysed by transmission electron microscopy (TEM). Briefly, a high titre of phage lysate (~10¹¹ mL⁻¹) was applied to a CsCl₂ step gradient (densities: 1.1, 0.9, 0.7, and 0.5 g mL⁻¹) and centrifuged in a Spinco SW39 rotor for 3 h at 24000 rpm (280,000 x g), 4°C. The band with the highest opalescence was collected with a syringe and dialysed three times with SM buffer (100 mM NaCl, 8 mM MgSO₄, 50 mM Tris-HCl, pH 7.5) in dialysis tubes (0.020 mm; Viskase, USA) for 24 h at 4 °C. The suspension was applied to carbon-
coated nitrocellulose grids (Agar Scientific Elektron Technology, UK), stained with two successive drops of 2% uranyl acetate (pH 4.5), dried, and examined using a Morgagni™ 268(D) transmission electron microscope (FEI, USA). In total, 50 phage particles were measured for determination of capsid size and tail length using the Morgagni™ integrated image acquisition and analysis software.

Host range

Host range analysis based on efficiency of plating was performed with two species type-phages from this study, barba18A and barba19A, on the in-house Rheinheimera bacterial strains, Rheinheimera sp. BAL341, BAL331, BAL335, and BAL336, isolated from LMO. Plaque assays as described previously were performed in duplicate for each phage-host pair with phage dilutions of $10^2$ to $10^4$-fold, except for Rheinheimera sp. strain BAL341 where the phages were diluted from $10^2$ to $10^{-9}$, and plates were monitored daily for five days to detect plaque formation.

One step growth curve

To determine the exponential growth phase of Rheinheimera sp. strain BAL341, the bacterium was grown in 10 mL Zobell medium in glass tubes at RT in triplicates and optical density at 600 nm with a CO8000 Cell Density Meter (WPA, Cambridge, UK) was measured every hour for 14 hours. One-step growth curves, providing phage replication characteristics, were performed as described in Holmfeldt et al. (16), with minor modifications. Briefly, exponentially growing bacteria were infected with barba18A or barba19A at a MOI of 0.1 for five minutes, after which the solution was diluted 1000-fold in order to prevent new infections. Free and total number of phages was enumerated using plaque assay (described above) every 15 minutes for two hours for barba18A and every 20 min for two hours followed by every 30 minutes for an additional hour for barba19A. To enumerate free (extracellular) phages, subsamples of the diluted infection was filtered through a 0.2 µM syringe filter
(BD) and the flow-through was used for the plaque assay. For total phages, including both extracellular and phages attached to or replicating inside bacteria, subsamples were retrieved directly from the incubation and used for plaque assay. Plaque formation was examined within 48 hours.

Statistical analysis and plotting were done with R (version 3.5.1) (87) through RStudio (version 1.1.383) (88) using packages plyr (version 1.8.4) (89) and ggplot2 (version 3.1.0) (90).

Viral DNA extraction
DNA extraction from the 54 isolated phages were performed with the Wizard PCR DNA Purification Resin and Minicolumns (Promega). First, 1 mL resin was added to 1 mL of the newly harvested, high titre viral sample, thoroughly mixed, and then pushed through a minicolumn. The column was washed twice using 1 mL 80% isopropanol and residues of isopropanol was removed by centrifuging the column for 2 minutes at 10,000 x g. The DNA was then eluted with pre-heated (80°C) 10 mM Tris (Trizma base (Sigma), pH 8) by first vortexing the column and then centrifuging at 10,000 x g for 30 seconds. Quality and quantity of the DNA was verified with Nanodrop (Thermo Scientific) and Qubit (high sensitivity DNA kit; Invitrogen, Life technologies), respectively.

Sequencing of whole genomes
Library preparation of DNA from whole-genome extraction of Rheinheimera sp. BAL341 and the 54 viral isolates was done with Nextera XT (Illumina Inc., San Diego, CA, USA) and sequencing was performed by SciLife/NGI (Solna, Sweden) on a HiSeq 2500 (Illumina Inc.), which resulted in paired-end 125 bp reads (average 1.2 million reads per sample).

Bioinformatics analysis
Reads in fastq-format were trimmed to remove adapters and poor quality reads (trimmomatic v0.30,
settings: -PE -threads 2 -phred33 ILLUMINACLIP:nextera_linkers.txt:2:30:10 LEADING:3

TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30) (91), and quality was then evaluated with

FastQC (92). Reads were assembled with Spades (v3.6.0, settings: --careful --t 8 --pe1-1 --pe1-2 -o)

(93) and Abyss (v1.3.5, settings: -np 16 --k61 --q3 --coverage hist --s -o) (94), and the assembler

providing the longest contig was chosen. The bacterial genome was annotated with the RAST online

service (95, 96). The phage genomes were manually inspected and gene prediction was performed

with GeneMark (version 3.26, gene code 11, heuristic parameters as 2010) (97) and manually curated

on completed genomes. Predicted genes were then annotated with Diamond (v0.8.26, settings: blastp

--p 16 --q 500 --min-score 30 --sensitive --tmpdir -f 6 -o) (98) against the NCBI non-redundant

database (February 2018) and hmmsearch (v3.1b2, settings: --tblout -E 1e-3 --cpu 4) (99) against the

Pfam database (100)(February 2018, v31.0), and an alignment was considered significant if e-value

was less than 0.001. Also, to identify the phages with highest similarity to the barba-phages, a

Diamond search with the same parameters as described above was run against a custom viral database

(viraldb) consisting of viral genomes from Refseq (downloaded 14th of September 2018). The results

from the search against the viraldb were used to define reference genomes for the VICTOR analyses

described below. Searches for tRNA were performed with tRNAscan-SE (settings: -qQ --detail -o# -
m# -f# -l# -c tRNAscan-SE.conf -s# -B) (101). Core and pan-genome were calculated with Roary

(102), with settings not to split paralogs and to group genes with more than 70% amino acid similarity

(-s -i 70). Whole genome comparisons on nucleotide level was done with gegenees (version 3.0.0,

settings: fragment size = 500, step size = 500, blastn) (103) that fragments the genomes and performs

all against all blastn alignments before calculating average nucleotide identity across the entire

genomes. The gegenees results were plotted with reshape2 (version 1.4.3) (104) and Tidyverse

(version 1.2.1) (105) in R (version 3.5.1) through RStudio (version 1.1.383). Comparison of gene

distribution between a select set of genomes was conducted using EasyFig (106).
Phylogenetic analysis

To distinguish taxonomic classification of the barba-phages compared to other viruses, all 54 barba-phage isolates and selected reference genomes, based on alignments to NCBI nr and viraldb, were analysed with VICTOR, a genome-to-genome distance calculator (42). Pairwise comparisons of nucleotide sequences were done by using the Genome-BLAST Distance Phylogeny (GBDP) method (107) under settings for the D0 formula (settings: word length = 11, e-value filter = 1.0, algorithm = greedy-with-trimming) recommended for prokaryotic viruses (42). The distances were then used to infer a balanced minimum evolution tree with branch support (SPR branch swapping, 100 pseudo-bootstrap replicates) via FastME (108). Also, the genomes were clustered at species, genus and family levels with the OPTSIL program (109) with recommended thresholds (42) and an F value (fraction of links required for cluster fusion) of 0.5 (110).

Viral metagenome sampling and recruitment to the phage genomes

At 25 time-points between 2012 and 2015 (for dates see supplemental Table S7), water was collected at the LMO in the same manner as described above. The water was prefiltered through an 0.22 µm filter (Sterivex cartridge filter, Millipore) to remove larger organisms. Viruses were precipitated using the iron-chloride method (111), where the flow-through was treated with FeCl₃ (final concentration 1 mg L⁻¹, Sigma) to aggregate the viral particles, which was collected on a 1.0 µM filter (polycarbonate, Maine manufacturing LLC) and stored in falcon tubes at +4°C until DNA extraction. For DNA extraction, the viral particles were resuspended in ascorbate solution (20 mL 0.5 M EDTA pH 8, 12.5 mL 1 M Tris pH 8, 1.9 g MgCl₂, 3.52 g Ascorbic acid, ~4 mL NaOH, and MQ water up to 100 mL) within 1 – 18 months since sampling and concentrated using amicon spin-filters (Millipore). The viral concentrates were thereafter treated with DNase I (Invitrogen) according to the manufacturers
recommendations and inactivated with EDTA (100 mM; Sigma) by incubation at 65° for 15 min. The viral particles were treated with pre–heated (37°C for 30 min) proteinase K (20 mg/mL; Fisher) at 37°C for 12 hours and the DNA was then extracted with the Wizard PCR DNA Purification kit as described above. The quality and quantity of DNA was assessed with Qubit (Invitrogen) and Nanodrop (Thermo Scientific), respectively, and sequenced at SciLifeLab/NGI (Solna, Sweden). Nine of the viral metagenomes were sequenced during 2014 and the last 16 samples during 2016. Libraries for the first set was prepared using Illumina TruSeq DNA, clustered with ‘cbot’, and sequenced with a 2x101 setup. For the second set, the libraries were prepared using Illumina TruSeq Nano, clustered with ‘cbot’, and sequenced with a 2x126 setup (supplemental Table S7). Both sets were sequenced on Illumina HiSeq 2500 (Illumina Inc., San Diego, CA, USA). Reads from the 25 viral metagenomes were recruited against barba18A, barba21A, barba5S, barba8S, and barba19A competitively, and against barba19A exclusively using Bowtie2 (v2.3.3.1, default settings) (112). To calculate coverage and depth, samtools (v 1.9, settings: -F4) (113) was used to filter out alignments with low quality (< 90% identity to the reference) and bedtools (v 2.27.1) (114) was used to perform calculations of coverage and depth. Read depth was normalised based on the length of the genome in Kb and the size of the viral metagenome in Mb as previously done for other viral metagenome analyses (5).

**Amplicon sequencing and host identification**

Bacterial 16S rRNA gene amplicon sequence data for monitoring host abundance were obtained according to Boström et al. (115), as modified by Bunse et al. (116). Briefly, water samples from LMO (2012-2015) were collected and either sequentially filtered through 3.0 µm pore size, 48 mm diameter, polycarbonate filters (Pall life sciences) and then 0.2 µm Sterivex cartridge filters (Millipore) or directly through Sterivex cartridge filters (Millipore). DNA was extracted using the phenol/chloroform method (117) from all filters and the V3-V4 region of the 16S rRNA gene.
was amplified in PCR with the primer pair 341f-805r (118, 119). Illumina adapters were attached
during a second PCR and the samples were sequenced (300 bp, pair-end) on the Illumina MiSeq
platform (Illumina Inc., USA) at SciLifeLab (Stockholm, Sweden). Raw reads were subsequently
processed with the DADA2 pipeline (120) to produce amplicon sequencing variants (ASVs;
filterAndTrim: --trimLeft=8,8 --truncLen=290,210, learnErrors: --randomize=TRUE, --
samples=24, mergePairs: --minOverlap=10, isBimeraDenovo: minParentAbundance=8,
minFoldParentAbundance=4). Using RStudio (version 1.1.383)(88) and the Tidyverse package
(version 1.2.1)(105), the ASV sequence matching Rheinheimera sp. strain BAL341 (100% nt
identity, blastn) was extracted from the dataset from all three filter fractions (3.0 µm, 0.2 µm and
0.2 µm after 3.0 filtration) and plotted as relative abundances (proportions) over time, when
detected, together with the phage abundances using ggplot2 (90).

Accession numbers
16S rRNA gene accession numbers in NCBI: Rheinheimera sp. BAL341: KM586890,
BAL331: KM586876, BAL335: KM586880, BAL336: KM586917.
Rheinheimera sp. strain BAL341 in European nucleotide archive, raw reads: bioproject
accession number: PRJEB29737; whole genome and annotation: CAAJGR01000000
Barba-phage genome accession numbers in NCBI: MK719701-MK719754
Viral metagenome bioproject accession number in NCBI: PRJNA474405
Acknowledgements

We would like to acknowledge Anders Månsson, Kristofer Bergström, Emil Fridolfsson, and the others in the LMO sampling crew, and Ecochange; Sabina Arnautovic who has helped with bacterial isolation, DNA extraction and sequencing; Stina Israelsson who helped with viral metagenome sampling; Diego Brambilla who has helped with genome submission. The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and the Swedish National Infrastructure for Computing (SNIC) at Uppmax for helping in massive parallel sequencing and resources for computation. The project was funded by a grant from the Swedish research council (2013-4554) and the Crafoord foundation (20140681) to KH.

Author contribution

KH and JP conceived the study. EN, KL, and SS isolated and processed viral isolates and carried out viral experiments. EN and KH prepared and processed viral metagenome samples. ML isolated bacteria, and CB and ML undertook LMO sampling efforts and processed 16S-sequencing data. EN, JF, CB, ML, and DL performed bioinformatic analyses. EN, KH, CB, CMGK, DL, and JP analysed and interpreted the data. EN and KH wrote the initial draft of the manuscript with the help of SS, CMGK, and CB. All authors commented and approved the final manuscript.

Conflict of interest

The authors declare no conflict of interest.
References

1. Suttle CA. 2007. Marine viruses - major players in the global ecosystem. Nat Rev Microbiol 5:801-812.

2. Bergh O, Borsheim KY, Bratbak G, Heldal M. 1989. High abundance of viruses found in aquatic environments. Nature 340:467-468.

3. Suttle CA. 1994. The significance of viruses to mortality in aquatic microbial communities. Microb Ecol 28:237-243.

4. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F. 2006. The marine viromes of four oceanic regions. PLoS Biol 4:e368.

5. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Tara Oceans Coordinators, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. 2015. Ocean plankton. Patterns and ecological drivers of ocean viral communities. Science 348:1261498.

6. Roux S, Hallam SJ, Woyte K, Sullivan MB. 2015. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. eLife 4:e08490.

7. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol Rev 40:258-272.

8. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brule JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White...
BA, Rohwer F. 2008. Functional metagenomic profiling of nine biomes. Nature 452:629-646.

Hurwitz BL, Hallam SJ, Sullivan MB. 2013. Metabolic reprogramming by viruses in the sunlit and dark ocean. Genome Biol 14:R123.

Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, Sullivan MB, Woyke T, Wommack KE, Stepanauskas R. 2015. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. ISME J 9:2386-2399.

Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, Sullivan MB. 2014. Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. Nature 513:242-245.

Mann NH, Cook A, Millard A, Bailey S, Clokie M. 2003. Marine ecosystems: bacterial photosynthesis genes in a virus. Nature 424:741.

Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, Feldblyum TV, White O, Paulsen IT, Nierman WC, Lee J, Szczypinski B, Fraser CM. 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J Bacteriol 185:5220-5233.

Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. 2005. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. PLoS Biol 3:e144.

Breithart M, Thompson LR, Suttle CA, Sullivan MB. 2007. Exploring the vast diversity of marine viruses. Oceanography 20:135-139.

Holmfeldt K, Solonenko N, Howard-Varona C, Moreno M, Malmstrom RR, Blow MJ, Sullivan MB. 2016. Large-scale maps of variable infection efficiencies in aquatic Bacteroidetes phage-host model systems. Environ Microbiol 18:3949-3961.
17. Holmfeldt K, Howard-Varona C, Solonenko N, Sullivan MB. 2014. Contrasting genomic patterns and infection strategies of two co-existing Bacteroidetes podovirus genera. Environ Microbiol 16:2501-2513.

18. Dang VT, Howard-Varona C, Schwenck S, Sullivan MB. 2015. Variably lytic infection dynamics of large Bacteroidetes podovirus phi38:1 against two Cellulophaga baltica host strains. Environ Microbiol 17:4659-4671.

19. Howard-Varona C, Roux S, Dore H, Solonenko NE, Holmfeldt K, Markillie LM, Orr G, Sullivan MB. 2017. Regulation of infection efficiency in a globally abundant marine Bacteroidetes virus. ISME J 11:284-295.

20. Howard-Varona C, Hargreaves KR, Solonenko NE, Markillie LM, White RA, Brewer HM, Ansong C, Orr G, Adkins JN, Sullivan MB. 2018. Multiple mechanisms drive phage infection efficiency in nearly identical hosts. ISME J 12:1605-1618.

21. Rodrigues RAL, Andrade A, Boratto PVM, Trindade GS, Kroon EG, Abrahao JS. 2017. An anthropocentric view of the virosphere-host relationship. Front Microbiol 8:1673.

22. Duhaime MB, Wichels A, Waldmann J, Teeling H, Glöckner FO. 2011. Ecogenomics and genome landscapes of marine Pseudoalteromonas phage H105/1. ISME J 5:107-121.

23. Lara E, Holmfeldt K, Solonenko N, Sa EL, Ignacio-Espinoza JC, Cornejo-Castillo FM, Verberkmoes NC, Vaque D, Sullivan MB, Acinas SG. 2015. Life-style and genome structure of marine Pseudoalteromonas siphovirus B8b isolated from the northwestern Mediterranean Sea. PLoS One 10:e0114829.

24. Wichels A, Gerdts G, Schüt C. 2002. Pseudoalteromonas spp. phages, a significant group of marine bacteriophages in the North Sea. Aquat microb ecol 27:233-239.

25. Baudoux AC, Hendrix RW, Lander GC, Bailly X, Podell S, Paillard C, Johnson JE, Potter CS, Carragher B, Azam F. 2012. Genomic and functional analysis of Vibrio phage SIO-2
reveals novel insights into ecology and evolution of marine siphoviruses. Environ Microbiol 14:2071-2086.

26. Comeau AM, Chan AM, Suttle CA. 2006. Genetic richness of vibriophages isolated in a coastal environment. Environ Microbiol 8:1164-1176.

27. Duhaime MB, Solonenko N, Roux S, Verberkmoes NC, Wichels A, Sullivan MB. 2017. Comparative omics and trait analyses of marine Pseudoalteromonas phages advance the phage OTU concept. Front Microbiol 8:1241.

28. Yu ZC, Chen XL, Shen QT, Zhao DL, Tang BL, Su HN, Wu ZY, Qin QL, Xie BB, Zhang XY, Yu Y, Zhou BC, Chen B, Zhang YZ. 2015. Filamentous phages prevalent in Pseudoalteromonas spp. confer properties advantageous to host survival in Arctic sea ice. ISME J 9:871-881.

29. Grose JH, Casjens SR. 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology 468-470:421-443.

30. Kwan T, Liu J, Dubow M, Gros P, Pelletier J. 2006. Comparative genomic analysis of 18 Pseudomonas aeruginosa bacteriophages. J Bacteriol 188:1184-1187.

31. Ryu SH, Chung BS, Park M, Lee SS, Lee SS, Jeon CO. 2008. Rheinheimera soli sp. nov., a gammaproteobacterium isolated from soil in Korea. Int J Syst Evol Micr 58:2271-2274.

32. Chen WM, Lin CY, Young CC, Sheu SY. 2010. Rheinheimera aquatica sp nov., antimicrobial activity-producing bacterium isolated from freshwater culture pond. J Microbiol Biotechn 20:1386-1392.

33. Yoon JH, Bae SE, Kang SJ, Oh TK. 2007. Rheinheimera aquimaris sp. nov., isolated from seawater of the East Sea in Korea. Int J Syst Evol Micr 57:1386-1390.
34. Schröder J, Braun B, Liere K, Szewzyk U. 2016. Draft genome sequence of Rheinheimera sp. strain SA_1 isolated from iron backwash sludge in Germany. Genome Announc 4:e00853-16.

35. Karlsson CM, Cerro-Gálvez E, Lundin D, Karlsson C, Vila-Costa M, Pinhassi J. Accepted May 15th 2019. Direct effects of organic pollutants on the growth and gene expression of the Baltic Sea model bacterium Rheinheimera sp. BAL341. Microb Biotechnol doi:10.1111/1751-7915.13441.

36. Brettar I, Christen R, Höfle MG. 2002. Rheinheimera baltica gen. nov., sp. nov., a blue-coloured bacterium isolated from the central Baltic Sea. Int J Syst Evol Micr 52:1851-1857.

37. Lindh MV, Figueroa D, Sjöstedt J, Baltar F, Lundin D, Andersson A, Legrand C, Pinhassi J. 2015. Transplant experiments uncover Baltic Sea basin-specific responses in bacterioplankton community composition and metabolic activities. Front Microbiol 6:223.

38. Bunse C, Israelsson S, Baltar F, Bertos-Fortis M, Fridolfsson E, Legrand C, Lindehoff E, Lindh MV, Martinez-Garcia S, Pinhassi J. 2019. High frequency multi-year variability in Baltic Sea microbial plankton stocks and activities. Front Microbiol 9:3296.

39. Lindh MV, Sjöstedt J, Andersson AF, Baltar F, Hugerth LW, Lundin D, Muthusamy S, Legrand C, Pinhassi J. 2015. Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. Environ Microbiol 17:2459-2476.

40. Kropinski AM, Prangishvili D, Lavigne R. 2009. Position paper: The creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. Environ Microbiol 11:2775-2777.

41. Adriaenssens EM, Brister JR. 2017. How to name and classify your phage: an informal guide. Viruses-Basel 9:70.
42. Meier-Kolthoff JP, Göker M. 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33:3396-3404.

43. Roux S, Emerson JB, Elof-Fadrosch EA, Sullivan MB. 2017. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ 5:e3817.

44. King AM, Lefkowitz E, Adams MJ, Carstens EB. 2011. Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Elsevier.

45. Brum JR, Schenck RO, Sullivan MB. 2013. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. ISME J 7:1738-1751.

46. Jiang SC, Kellogg CA, Paul JH. 1998. Characterization of marine temperate phage-host systems isolated from Mamala Bay, Oahu, Hawaii. Appl Environ Microbiol 64:535-542.

47. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, Ellisman M, Deerinck T, Sullivan MB, Giovannoni SJ. 2013. Abundant SAR11 viruses in the ocean. Nature 494:357-360.

48. Suttle CA, Chan AM. 1993. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. Mar Ecol Prog Ser 92:99-109.

49. Hadass H, Einav M, Fishov I, Zaritsky A. 1997. Bacteriophage T4 development depends on the physiology of its host *Escherichia coli*. Microbiology 143:179-185.

50. Holmfeldt K, Solonenko N, Shah M, Corrier K, Riemann L, Verberkmoes NC, Sullivan MB. 2013. Twelve previously unknown phage genera are ubiquitous in global oceans. Proc Natl Acad Sci USA 110:12798-12803.
51. Holmfeldt K, Middelboe M, Nybroe O, Riemann L. 2007. Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobacterium* hosts. Appl Environ Microbiol 73:6730-6739.

52. Angly F, Youle M, Nosrat B, Srinagesh S, Rodriguez-Brito B, McNairnie P, Deyanat-Yazdi G, Breitbart M, Rohwer F. 2009. Genomic analysis of multiple Roseophage SIO1 strains. Environ Microbiol 11:2863-2873.

53. Sauder AB, Quinn MR, Brouillette A, Caruso S, Cresawn S, Erill I, Lewis L, Loesser-Casey K, Pate M, Scott C, Stockwell S, Temple L. 2016. Genomic characterization and comparison of seven *Myoviridae* bacteriophage infecting *Bacillus thuringiensis*. Virology 489:243-251.

54. Gregory AC, Solonenko SA, Ignacio-Espinoza JC, LaButti K, Copeland A, Sudek S, Maitland A, Chittick L, Dos Santos F, Weitz JS, Worden AZ, Woyke T, Sullivan MB. 2016. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. BMC Genomics 17:930.

55. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoye NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O’Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW. 2010. Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. J Mol Biol 397:119-143.

56. Pope WH, Jacobs-Sera D, Russell DA, Peebles CL, Al-Atrache Z, Alcoser TA, Alexander LM, Alfano MB, Alford ST, Amy NE. 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. PLoS One 6:e16329.
57. Weigel C, Seitz H. 2006. Bacteriophage replication modules. FEMS Microbiol Rev 30:321-381.

58. Seco EM, Zinder JC, Manhart CM, Lo Piano A, McHenry CS, Ayora S. 2013. Bacteriophage SPP1 DNA replication strategies promote viral and disable host replication in vitro. Nucleic Acids Res 41:1711-1721.

59. Breitbart M, Middelboe M, Rohwer F. 2008. Marine viruses: community dynamics, diversity and impact on microbial processes, p 443-479. In Kirchman DL (ed), Microbial ecology of the oceans, 2nd ed. John Wiley & Sons, Hoboken, NJ.

60. Dwivedi B, Xue B, Lundin D, Edwards RA, Breitbart M. 2013. A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. BMC Evol Biol 13:33.

61. Ignacio-Espinoza JC, Sullivan MB. 2012. Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. Environ Microbiol 14:2113-2126.

62. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Tara Oceans Coordinators, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature 537:689-693.

63. Mann NH, Clokie MR, Millard A, Cook A, Wilson WH, Wheatley PJ, Letarov A, Krisch HM. 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine Synechococcus strains. J Bacteriol 187:3188-3200.

64. Nordlund P, Reichard P. 2006. Ribonucleotide reductases. Annu Rev Biochem 75:681-706.
65. Myllykallio H, Lipowski G, Leduc D, Filee J, Forterre P, Liebl U. 2002. An alternative flavin-dependent mechanism for thymidylate synthesis. Science 297:105-107.

66. Zhao Y, Wang K, Jiao N, Chen F. 2009. Genome sequences of two novel phages infecting marine roseobacters. Environ Microbiol 11:2055-2064.

67. Gross JW, Rajavel M, Grubmeyer C. 1998. Kinetic mechanism of nicotinic acid phosphoribosyltransferase: implications for energy coupling. Biochem 37:4189-4199.

68. Miller Jr GA, Rosenzweig S, Switzer RL. 1975. Oxygen-18 studies of the mechanism of pyrophosphoryl group transfer catalyzed by phosphoribosylpyrophosphate synthetase. Arch biochem biophys 171:732-736.

69. Hove-Jensen B. 1988. Mutation in the phosphoribosylpyrophosphate synthetase gene (prs) that results in simultaneous requirements for purine and pyrimidine nucleosides, nicotinamide nucleotide, histidine, and tryptophan in Escherichia coli. J Bacteriol 170:1148-1152.

70. Jensen KF. 1983. Metabolism of 5-phosphoribosyl 1-pyrophosphate (PRPP) in Escherichia coli and Salmonella typhimurium, p 1-25. In Munch-Petersen A (ed), Metabolism of nucleotides, nucleosides and nucleobases in microorganisms. Academic Press, London.

71. Horton H, Moran L, Ochs R, Rawn D, Scrimgeour K. 1996. Principles of biochemistry, 2nd edition ed. Prentice Hall, Upper Saddle River, NJ.

72. Lee JY, Li ZQ, Miller ES. 2017. Vibrio phage KVP40 encodes a functional NAD(+) salvage pathway. J Bacteriol 199:e00855-16.

73. Seed KD, Bodi KL, Kropinski AM, Ackermann HW, Calderwood SB, Qadri F, Camilli A. 2011. Evidence of a dominant lineage of Vibrio cholerae-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. mBio 2:e00334-10.
74. Wang DB, Sun MQ, Shao HB, Li Y, Meng X, Liu ZY, Wang M. 2015. Characterization and genome sequencing of a novel bacteriophage PH101 infecting *Pseudoalteromonas marina* BH101 from the Yellow Sea of China. Curr Microbiol 71:594-600.

75. Gross M, Marianovsky I, Glaser G. 2006. MazG - a regulator of programmed cell death in *Escherichia coli*. Mol Microbiol 59:590-601.

76. Clokie MR, Mann NH. 2006. Marine cyanophages and light. Environ Microbiol 8:2074-2082.

77. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osbourne MS, Henn MR, Chisholm SW. 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environ Microbiol 12:3035-3056.

78. Kim SK, Makino K, Amemura M, Shinagawa H, Nakata A. 1993. Molecular analysis of the *phoH* gene, belonging to the phosphate regulon in *Escherichia coli*. J Bacteriol 175:1316-1324.

79. Kazakov AE, Vassieva O, Gelfand MS, Osterman A, Overbeek R. 2003. Bioinformatics classification and functional analysis of PhoH homologs. In Silico Biol 3:3-15.

80. Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, Kettler G, Sullivan MB, Steen R, Hess WR, Church GM, Chisholm SW. 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature 449:83-86.

81. Goldsmith DB, Crosti G, Dwivedi B, McDaniel LD, Varsani A, Suttle CA, Weinbauer MG, Sandaa RA, Breitbart M. 2011. Development of *phoH* as a novel signature gene for assessing marine phage diversity. Appl Environ Microbiol 77:7730-7739.
Sakowski EG, Munsell EV, Hyatt M, Kress W, Williamson SJ, Nasko DJ, Polson SW, Wommack KE. 2014. Ribonucleotide reductases reveal novel viral diversity and predict biological and ecological features of unknown marine viruses. Proc Natl Acad Sci USA 111:15786-15791.

Loderer C, Holmfeldt K, Lundin D. 2019. Non-host class II ribonucleotide reductase in Thermus viruses: sequence adaptation and host interaction. PeerJ 7:e6700.

Thingstad TF. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr 45:1320-1328.

Sambrook J, Russel DW. 2001. Molecular cloning: a laboratory manual (3-volume set), 3rd ed. Cold spring harbor laboratory press, Cold Spring Harbor, NY.

Carlson K, Miller E. 1994. Experiments in T4 genetics, vol 4. American Society for Microbiology, Washington DC.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

RStudio Team. 2015. RStudio: integrated development for R. RStudio, Inc., Boston, MA.

Wickham H. 2011. The split-apply-combine strategy for data analysis. J Stat Softw 40:1-29.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer, New York.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,
Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-477.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABYSS: A parallel assembler for short read sequence data. Genome Res 19:1117-1123.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid annotations using subsystems technology. BMC Genomics 9:75.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia FF, Stevens R. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 42:D206-D214.

Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451-W454.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59-60.

Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279-85.

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54-57.
Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691-3693.

Ågren J, Sundström A, Håfström T, Segerman B. 2012. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. PLoS One 7:e39107.

Wickham H. 2007. Reshaping data with the reshape package. J Stat Softw 21:1-20.

Wickham H. 2017. Tidyverse: Easily install and load 'tidyverse' packages.

Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27:1009-1010.

Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60.

Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol 32:2798-2800.

Göker M, García-Blázquez G, Voglmayr H, Tellería MT, Martín MP. 2009. Molecular taxonomy of phytopathogenic fungi: a case study in Peronospora. PloS One 4:e6319.

Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C, Rohde M, Fartmann B, Goodwin LA, Chertkov O, Reddy TBK, Pati A, Ivanova NN, Markowitz V, Kyrpides NC, Woyke T, Göker M, Klenk HP. 2014. Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci 9:2.
111. John SG, Mendez CB, Deng L, Poulos B, Kauffman AK, Kern S, Brum J, Polz MF, Boyle EA, Sullivan MB. 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environ Microbiol Rep 3:195-202.

112. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359.

113. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

114. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

115. Boström KH, Simu K, Hagström Å, Riemann L. 2004. Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. Limnol Oceanogr Meth 2:365-373.

116. Bunse C, Bertos-Fortis M, Sassenhagen I, Sildever S, Sjöqvist C, Godhe A, Gross S, Kremm A, Lips I, Lundholm N. 2016. Spatio-temporal interdependence of bacteria and phytoplankton during a Baltic Sea spring bloom. Front Microbiol 7:517.

117. Riemann L, Steward GF, Azam F. 2000. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. Appl Environ Microbiol 66:578-587.

118. Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J 5:1571-1579.

119. Hugerth LW, Wefer HA, Lundin S, Jakobsson HE, Lindberg M, Rodin S, Engstrand L, Andersson AF. 2014. DegePrime, a program for degenerate primer design for broad-
taxonomic-range PCR in microbial ecology studies. Appl Environ Microbiol 80:5116-5123.

120. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581-583.
Table 1. Predicted genes with given function and the functional group they belong to, where “Structural and packaging” indicate that they belong to the structural module and the other functional groups belong to the host interaction module. For details see supplemental Table S5.

| Gene name           | Length of protein (aa) | Functional group                                    | Function                              |
|---------------------|------------------------|-----------------------------------------------------|---------------------------------------|
| Barba18A_gp001      | 334                    | Nucleotide metabolism and recycling                 | ribonucleotide reductase              |
| Barba18A_gp002      | 951                    | Nucleotide metabolism and recycling                 | ribonucleotide reductase              |
| Barba18A_gp004      | 223                    | Nucleotide metabolism and recycling                 | thymidylate synthase                 |
| Barba18A_gp005      | 85                     | Recycling                                           | glutaredoxin                         |
| Barba18A_gp006      | 119                    | Host signalling                                     | mazG                                  |
| Barba18A_gp008      | 404                    | DNA processing, replication and recombination       | metallo-dependent phosphatase         |
| Barba18A_gp012      | 202                    | DNA processing, replication and recombination       | homing endonuclease HNH              |
| Barba18A_gp013      | 350                    | DNA processing, replication and recombination       | putative exodeoxyribonuclease         |
| Barba18A_gp022      | 730                    | DNA processing, replication and recombination       | DNA polymerase                       |
| Barba18A_gp023      | 631                    | DNA processing, replication and recombination       | DNA primase                          |
| Barba18A_gp030      | 164                    | Host signalling                                     | macro domain protein                 |
| Barba18A_gp072      | 228                    | DNA processing, replication and recombination       | CLP_protease                         |
| Barba18A_gp073      | 262                    | DNA processing, replication and recombination       | phoH                                  |
| Barba18A_gp071      | 271                    | DNA adenine methylase                               | DNA-binding motif containing protein  |
| Barba18A_gp073      | 262                    | Phosphate metabolism                                | putative DNA binding protein         |
| Barba18A_gp082      | 102                    | Structural and packaging                            | terminase large subunit              |
| Barba18A_gp090      | 205                    | Structural and packaging                            | putative portal protein              |
| Barba18A_gp095      | 455                    | Structural and packaging                            |                                       |
| Barba18A_gp096      | 491                    | Structural and packaging                            |                                       |
| Accession | Start Position | Category                        | Description                                            |
|-----------|----------------|---------------------------------|--------------------------------------------------------|
| Barba18A_gp097 | 469            | Peptidase                       | peptidase                                              |
| Barba18A_gp099 | 352            | Structural and packaging         | major capsid protein                                    |
| Barba18A_gp104 | 359            | Structural and packaging         | tail sheath protein                                     |
| Barba18A_gp105 | 154            | Structural and packaging         | putative structural protein                             |
| Barba18A_gp108 | 213            | Structural and packaging         | putative DNA binding protein                            |
| Barba18A_gp109 | 566            | Structural and packaging         | tail length tape-measure protein                        |
| Barba18A_gp113 | 222            | Structural and packaging         | putative baseplate                                      |
| Barba18A_gp116 | 504            | Structural and packaging         | assembly protein                                         |
| Barba18A_gp118 | 220            | Structural and packaging         | baseplate                                               |
| Barba18A_gp119 | 414            | Structural and packaging         | tail fiber                                              |
| Barba18A_gp123 | 121            | Peptidase                       | peptidase M15                                           |
| Barba19A_gp006 | 82             | Nucleotide metabolism and recycling | glutaredoxin                                           |
| Barba19A_gp007 | 190            | Host signalling                 | mazG                                                   |
| Barba19A_gp009 | 209            | DNA processing, replication and recombination | putative DNA binding protein                           |
| Barba19A_gp010 | 338            | DNA processing, replication and recombination | metallo-dependent phosphatase                         |
| Barba19A_gp033 | 496            | Pyridine nucleotide salvage     | nicotinate phosphoribosyltransferase                    |
| Barba19A_gp034 | 268            | Pyridine nucleotide salvage     | ribose-phosphate                                        |
| Barba19A_gp035 | 167            | Host signalling                 | pyrophosphokinase                                       |
| Barba19A_gp061 | 322            | DNA processing, replication and recombination | macro domain protein                                   |
| Barba19A_gp063 | 419            | DNA processing, replication and recombination | nucleotidyltransferase                                 |
| Barba19A_gp098 | 219            | Structural and packaging         | tRNA                                                   |
| Barba19A_gp103 | 427            | Structural and packaging         | nucleotidyltransferase                                   |
| Barba23A_gp115 | 374            | DNA processing, replication and recombination | putative DNA binding protein                           |
| Barba5S_gp006  | 303            | Nucleotide metabolism and recycling | thymidylate synthase                                    |
| Barba5S_gp080  | 254            | Phosphate metabolism            | phoH                                                   |
Table 2. Alignment of genes with metabolic functions within the pan-genome of the barba-phages with blastp, barba18A was used as subject in phage-phage alignments where the rare gene variant was compared to the common, while the host was used as query for the phage-host alignment. Only significant alignments (e-value <0.001) are included.

| Function                        | Common gene no. of isolates | gene name     | Rare gene no. of isolates | gene name     | e-value | % id | % cov. | Best matching host gene, blastp against first gene | e-value | % id | % cov. | Best matching host gene, blastp against additional gene | e-value | % id | % cov. |
|---------------------------------|----------------------------|---------------|---------------------------|---------------|---------|------|-------|-----------------------------------------------|---------|------|-------|------------------------------------------------|---------|------|-------|
| thymidylate synthase            | 52                         | Barba18A_gp004| 2                         | Barba5S_gp006| -       | -    | -     | fhjl7557.7.peg.3030                          | 6.00E-04| -    | -     | fhjl7557.7.peg.2541                                 | -       | -    | -     |
| glutaredoxin                    | 52                         | Barba18A_gp005| 2                         | Barba19A_gp006| 5.00E-20| 43   | 92    | fhjl7557.7.peg.5555                          | 5.00E-27| 88   | 28    | fhjl7557.7.peg.1895                                 | 4.00E-05| 25   | 96    |
| protein phosphatase             | 52                         | Barba18A_gp073| 2                         | Barba5S_gp080| 42      | 31   | 94    | fhjl7557.7.peg.2145                          | 4.00E-06| 54   | -     | fhjl7557.7.peg.2145                                 | 4.00E-08| -    | -     |
| ribonucleotide reductase nudF   | 54                         | Barba18A_gp001| -                         | -             | -       | -    | -     | fhjl7557.7.peg.2913                          | 8.00E-10| 23   | 29    | -                                               | -       | -    | -     |
| ribonucleotide reductase nudA   | 54                         | Barba18A_gp002| -                         | -             | -       | -    | -     | fhjl7557.7.peg.2912                          | 8.00E-12| 27   | 79    | -                                               | -       | -    | -     |
| nicotinate phosphoribosyltransferase | 2                      | Barba18A_gp033| b                        | -             | -       | -    | -     | fhjl7557.7.peg.3024                          | 2.00E-08| 27   | 79    | -                                               | -       | -    | -     |
| ribose-phosphate pyrophosphokinase | 2                        | Barba19A_gp034| b                        | -             | -       | -    | -     | fhjl7557.7.peg.1160                          | 2.00E-08| 27   | 79    | -                                               | -       | -    | -     |

*All phages have the same gene

*No other phages had this gene
Figures

Fig 1. Pair-wise blastn comparison where average nucleotide identity across the entire genomes were calculated using Gegenees (version 3.0.0; settings: fragment size = 500, step size = 500, blastn) (103). The isolates belonging to the different species are boxed and the names are color-coded accordingly: isolates belonging to i) Rheinheimera virus Barba18A (species 1) are written in black, ii) Rheinheimera virus Barba21A (species 2) are written in blue, iii) Rheinheimera virus Barba5S (species 3) are written in pink, iv) Rheinheimera virus Barba8S (species 4) are written in green, and v) Rheinheimera virus Barba19A (species 5) are written in orange. The type phages are underlined on the left axis.

Fig 2. Genome BLAST distance phylogeny (GBDP) tree created with FastME (SPR branch swapping), as a part of VICTOR (42). The numbers above branches are GBDP pseudo-bootstrap support values from 100 replications (values >50 are reported), and the branch lengths are scaled in terms of the D0 distance formula (settings: word length = 11, e-value filter = 1.0, algorithm = greedy-with-trimming). Isolates belonging to Barbavirus are coloured based on which species they belong to as in Fig. 1.

Fig 3. (a) TEM of barba18A shows a myovirus morphology with a capsid that was measured to 72.1 nm (standard deviation ±2.7 nm) and the tail length to 88.7 nm (±2.2 nm) and width to 18.8 nm (±1.5 nm). Scale bar is 100 nm. (b) One step growth curve where barba18A and barba19A infected Rheinheimera sp. strain BAL341. “All phages” includes phages within and attached to cells as well as unattached phages, while “free phages” only
includes unattached phages. The values are normalised based on the concentration of “all phages” at time zero for each replicate. Error bars indicate standard deviation (n=3) but is not shown for barba18A at 60 minutes for free phages due to too large deviation (±7.18).

Fig 4. Comparison of genomes from the type-phages of the five species within the novel genus Barbavirus. Arrows indicate predicted genes, coloured by functionality and the shade of grey between the genomes indicate percent nucleotide identity (blastn). The metabolic genes and mazG discussed in the text, are indicated with numbers (1-6) or letters (a-b) if they exist in all species or only one, respectively. NrdA = ribonucleotide reductase alpha subunit, nrdB = ribonucleotide reductase beta subunit, thyl = thymidylate synthase, glrx = glutaredoxin, mazG = nucleoside triphosphate pyrophosphohydrolase, phoH = protein phoH, prs = ribose-phosphate pyrophosphokinase, pncB = nicotinate phosphoribosyltransferase. Created with EasyFig (106).

Fig 5. Temporal variation of barba19A and relative abundances of 16S rRNA gene amplicon sequences similar to the bacterial host Rheinheimera sp. strain BAL341 (triangles). The differently shaded circles indicate the coverage of that phage within the sample.
