Numerous cultivated and uncultivated viruses encode ribosomal proteins

Carolina M. Mizuno1, Charlotte Guyomar2, Simon Roux3, Régis Lavigne4, Francisco Rodriguez-Valera5, Matthew B. Sullivan6,7, Reynald Gillet2, Patrick Forterre1 & Mart Krupovic1

Viruses modulate ecosystems by directly altering host metabolisms through auxiliary metabolic genes. However, viral genomes are not known to encode the core components of translation machinery, such as ribosomal proteins (RPs). Here, using reference genomes and global-scale viral metagenomic datasets, we identify 14 different RPs across viral genomes arising from cultivated viral isolates and metagenome-assembled viruses. Viruses tend to encode dynamic RPs, easily exchangeable between ribosomes, suggesting these proteins can replace cellular versions in host ribosomes. Functional assays confirm that the two most common virus-encoded RPs, bS21 and bL12, are incorporated into 70S ribosomes when expressed in *Escherichia coli*. Ecological distribution of virus-encoded RPs suggests some level of ecosystem adaptations as aquatic viruses and viruses of animal-associated bacteria are enriched for different subsets of RPs. Finally, RP genes are under purifying selection and thus likely retained an important function after being horizontally transferred into virus genomes.
During billions of years of co-evolution with their hosts, viruses have evolved numerous strategies to directly modulate metabolic pathways and subvert key cellular biosynthetic machineries, which ensure their successful propagation. For example, ocean viruses that infect cyanobacteria (cyanophages) commonly encode core photosynthetic reaction center proteins, which serve to maintain the complex photosynthetic machinery during infection. These and other ocean viruses can similarly manipulate their host’s ability to alter central carbon metabolism, uptake phosphate, and cycle nitrogen, and sulfur—fundamental building blocks of life. Complementarily, viruses employ a diverse array of host take-over strategies to (i) fight off host defenses by encoding anti-restriction-modification or anti-CRISPR genes, (ii) control transcription by encoding sigma factors or polymerases themselves, and (iii) affect how proteins are translated. Indeed, many bacterial and some eukaryotic viruses with large double-stranded DNA genomes commonly encode a range of tRNA genes.

While these genes are presumed to boost the translational efficiency and virulence in diverse hosts, their importance during infection remains to be demonstrated experimentally. Giant mimiviruses, whose genomes approach the size of small bacterial genomes, carry many typically cellular genes including those for translation initiation, elongation, and termination, as well as a range of aminoacyl-tRNA synthetases. A single tRNA synthetase gene is also encoded in the genome of the largest known bacterial virus, Bacillus megaterium myxovirus, but its function lacks experimental confirmation. Some marine phages encode peptide deformylases, which are involved in post-translational modifications that, at least in cyanophages, may help preferentially produce the phage-encoded D1 photosystem protein. Finally, T7-like podoviruses encode serine/threonine kinases that have been shown to phosphorylate around 90 proteins, including several involved in protein translation, such as host-encoded ribosomal proteins bS1 and bS6, translation initiation factors IF1, IF2, and IF3, and elongation factors G and P.

It was suggested that phosphorylation of these proteins may stimulate translation of the phage late mRNAs. Although it is now clear that viruses have evolved different strategies to tinker with protein translation, the genes encoding proteins that directly participate in the formation of the ribosomes are not yet observed in the genomes of cultured viral isolates. In fact, this feature—ribosome-encoding or not—has been proposed to signify a major divide between cellular life forms and viruses. However, viral genome fragments assembled from environmental viral community sequence datasets (viral metagenomes), which vastly expand upon cultured sequence space, suggested that viruses might encode ribosomal proteins, specifically, bS1 and bS21. Though challenges insuring removal of contamination from cellular genomes and the lack of host context available warrants caution about such observations of “cellular features” in metagenome-only datasets, the findings are intriguing.

Here we leverage the greater genomic context now available from large-scale metagenomes and genomes to revisit the question of whether viral genomes encode ribosomal proteins (RPs). We identify 14 different RPs across viral genomes arising from cultivated viral isolates and metagenome-assembled viruses. We show that viruses tend to encode RPs known to be easily exchangeable between ribosomes, suggesting these proteins can replace cellular versions in host ribosomes, and confirm this experimentally for the two most common virus-encoded RPs, bS21 and bL12. Ecological distribution of virus-encoded RPs suggests certain level of ecosystem adaptations as aquatic viruses and viruses of animal-associated bacteria are enriched for different subsets of RPs. Overall, these results further blur the borders between viruses and cellular life forms.

**Results**

**Ribosomal proteins encoded in cultivated virus genomes.** To systematically investigate the presence of RP-encoding genes in viral genomes, we first searched available reference genomes of cultivated viruses. Of 106 RP domains, 18 were identified in at least one viral genome (Table 1). The genes were generally embedded within variable genomic contexts, even for homologous RP genes (Supplementary Fig. 1). Note that throughout this article we use the unified RP nomenclature, where capital letters “S” and “L”, respectively, indicate whether the protein is present in the small or large ribosome subunit, whereas the lowercase letters denote that the protein is specific to bacteria (b), eukaryotes/archaea (e), or are universal (u).

We first identified a ribosomal protein eS30 domain, a component of the small 40S ribosomal subunit in the Finkel–Biskis–Reilly murine sarcoma virus (FBR-MuSV), a member of the family Retroviridae. This domain was part of the *fau* gene fused to an N-terminal ubiquitin-like domain (Supplementary Fig. 2a). FBR-MuSV has acquired the cDNA copy of *fau* in inverse orientation, and production of the antisense RNA suppresses expression of endogenous *fau* mRNA, which leads to apoptosis inhibition and induces tumorigenesis. Although the viral protein is not translated, the antisense transcript affects the production of the cellular *fau* and thus might have an indirect effect on the ribosome biogenesis.

The remainder of the virus-encoded ribosomal proteins—bS21, bL9, bL12, and ribosome hibernation promotion factor (HPF)—were found in bacterial viruses (bacteriophages) infecting proteobacteria (from three different classes) and mycobacteria (phylum Actinobacteria) (Table 1). Though RP-encoding bacteriophage genomes ranged in size between 44.5 and 358.6 kb, only large (>130 kb) genomes encoded bS21 and HPF, whereas only smaller (<80 kb) genomes encoded bL9 and bL12 (Table 1). However, many more phage genomes encoding RP would be needed to verify the significance of this observation.

The bS21 homolog was identified in pelagicphage HTVC008M, a myxovirus. bS21 is a conserved component of the bacterial 30S ribosomal subunit (Fig. 1a) required for the initiation of polypeptide synthesis and mediates the base-pairing reaction between mRNA and 16S rRNA. The viral protein was most similar (54% identity over the protein length) to the corresponding protein of its host, *Pelagibacter ubique* (Fig. 1b), an abundant member of the SAR11 clade (class Alphaproteobacteria), which is considered to represent one of the most numerous bacterial groups worldwide. Maximum likelihood phylogenetic analysis showed that bS21 homologs from different families of alphaproteobacteria cluster together and form a sister group to the mitochondrial homolog, consistent with the scenario under which mitochondria have evolved from an alphaproteobacterial ancestor. In this tree, all alphaproteobacterial sequences are basal to the viral protein, strongly suggesting that the phage gene was horizontally acquired from the *Pelagibacter* host (Fig. 1c).

Ribosomal protein bL9 was identified in *Mycobacterium* phage 32HC, a siphovirus. bL9 binds to the 23S rRNA and is a component of the large 50S ribosome subunit (Supplementary Fig. 3a). The protein is involved in translation fidelity and is required to suppress frameshifting, and stop codon “hopping”. bL9 has a highly conserved architecture consisting of two widely spaced globular RNA-binding domains connected by an elongated α-helix. While the C-terminal domain in the viral bL9 homolog has been apparently non-homologically replaced with a
sequence that lacks known function, the N-terminal RNA-binding domain and part of the α-helical spacer are preserved (Supplementary Fig. 2b), suggesting that the viral protein may bind to the 23S rRNA.

The next ribosomal protein encoded in sequenced viral genomes was bL12, which was found in 7 phages infecting proteobacteria from three different classes (Table 1). The bL12 proteins participate in the formation of the so-called bL12 stalk, a clearly defined morphological feature in the E. coli 50S ribosomal subunit, which besides bL12, contains ribosomal proteins uL10 and uL11 as well as the uL10- and uL11-binding region of the 23S rRNA (Supplementary Fig. 3a). The phage-encoded bL12 domains are similar (~30–40% identity) to bona fide cellular ribosomal homologs and contain conserved residues involved in interaction with uL11 and elongation factors EF-G and EF-Tu (Supplementary Fig. 4a). Although in some phages (e.g., Ralstonia phage RSB3), the bL12 domain spans the entire protein, it was more common to observe these domains variably positioned within much larger polypeptides (up to 724 aa-long; Supplementary Fig. 5). For example, proteins encoded by pedophage HTVC008M and uL11 as well as the uL10- and uL11-binding region of the 23S rRNA.

**Fig. 1** Virus-encoded ribosomal protein bS21. a Structure of the Escherichia coli 30S ribosomal subunit (PDB id: 4ADV). 16S ribosomal RNA is shown as blue ribbon. bS21 ribosomal protein is highlighted in pink. b Alignment of the ribosomal protein bS21 encoded by pelagiphage HTVC008M with homologs from representatives of distinct bacterial taxa and environmental sequences obtained from the Global Ocean Sampling (GOS) dataset. c Phylogenetic tree of ribosomal protein bS21. Taxonomic affiliations are represented by colored circles (see panel (b) legend).
Ribosomal proteins detected in viral metagenomes. To place these findings of cultivated virus-encoded RPs into a broader ecological context, we searched 424,225 viral contigs from two global metagenomic datasets8,38 for putative RPs using the same 106 sequence profiles (see Methods). Overall, 13 putative ribosomal protein genes were identified across 1403 contigs (Fig. 2, Supplementary Table 2, Supplementary Fig. 3). By matching CRISPR spacers, comparing k-mer nucleotide frequencies and performing BLASTn homology searches against reference cellular sequences, hosts could be predicted for 74 (5.3%) of the uncultivated viruses encoding 7 different RPs (Supplementary Data 1). Despite this relatively low fraction of predicted hosts, RP-encoding phages are already associated with 8 bacterial phyla (Proteobacteria, Firmicutes, Verrucomicrobia, Bacteroidetes, Deinococcus-Thermus, Thermotogae, Cyanobacteria, and Actinobacteria).

The bS21, bL12, and HPF, which were found in cultivated phages, were also detected in uncultivated phages, with bS21 homologs dominating (88%) the pool of RPs detected (Fig. 2, Supplementary Table 2). While found in only one cultivated phage (see above), maximum likelihood phylogeny and genome context comparison using these metagenomic data suggested that multiple virus-host exchanges of bS21 protein-coding genes have occurred, likely across various bacterial phyla (Fig. 3). Notably, bS21-encoding viruses were almost exclusively from aquatic samples (90% of bS21s detected). Such repeated transfers and enrichment in aquatic samples suggest that virus-encoded bS21 proteins likely can provide a direct fitness benefit to aquatic bacteriophages. By contrast, bL12 and HPF were found across a broad range of samples (Fig. 2 and Supplementary Figs 8 and 9), suggesting that their repeated acquisition could be beneficial in multiple types of conditions and hosts.

Another 10 RPs detected in uncultivated viruses were not previously identified in isolate genomes (Supplementary Table 2). Commonly (>10 viral contigs) detected among these are bL31 and bL33. Although the biological function of bL33 remains

### Table 1 Ribosomal protein domains found in cultivated viruses

| Domain          | Protein Name (family) | Genome length, kb | RP accession, length (aa) | Coverage, identity (%) | HHpred Probability (%) | E-value |
|-----------------|-----------------------|-------------------|---------------------------|------------------------|------------------------|---------|
| Ribosomal_S30  | eS30                  | 3,811             | NP_598374, 133            | 85, 86                 | 99.92                  | 2.2E-26 |
| Ribosomal_S21  | bS21                  | 147,284           | AGE60443, 67              | 59, 46                 | 99.81                  | 2.7E-19 |
| Ribosomal_L9_N | bL9                   | 50,781            | AHI86298, 86              | 33, 40                 | 98.32                  | 4.3E-07 |
| Ribosomal_L12  | bL12                  | 75,028            | AHI01035, 106             | 74, 32                 | 99.9                    | 1.0E-23 |
| Ribosomal_S30AE| HPF                   | 72,098            | AGF88198, 418             | 15, 36                 | 96.21                  | 2.6E-02 |
|                 |                       | 74,485            | AGH07436, 126             | 25, 47                 | 97.06                   | 1.8E-03 |
|                 |                       | 358,663           | AFC21633, 111             | 71, 34                 | 99.96                   | 4.2E-28 |
|                 |                       | 136,947           | AEZ65272, 105             | 74, 35                 | 99.92                   | 1.5E-23 |
|                 |                       | 137,947           | ABI79209, 105             | 74, 33                 | 99.96                   | 1.3E-27 |
|                 |                       | 136,910           | AGM12525, 105             | 74, 32                 | 99.96                   | 3.0E-28 |
|                 |                       | 131,952           | AKE47110, 105             | 74, 33                 | 99.96                   | 3.4E-28 |
|                 |                       | 139,020           | AEZ65272, 105             | 74, 35                 | 99.93                   | 7.9E-24 |

Retroviridae, M Myoviridae, S Siphoviridae, P Podoviridae, HPF ribosome hibernation promotion factor.
obscure\textsuperscript{39}, it appears to contact tRNAs in the ribosomal E(exit)-site\textsuperscript{40} (Supplementary Fig. 3b), whereas bL31, similar to HPF, plays a role in 100S formation, 70S association, and translation\textsuperscript{41}. As in the case of bS21, viral contigs encoding bL31 or bL33 were almost exclusively detected in aquatic environments (Fig. 2). Maximum likelihood phylogenies and genome context comparisons highlighted a consistent pattern of at least two independent events of virus-host transfers involving viruses infecting different bacterial phyla (Supplementary Figs 10 and 11).

Thus, at this point, there is an emerging picture that ribosomal protein genes acquired through random sampling of host DNA might, in some cases, become fixed in viral genomes. Most (>99%) of the viruses contained only a single ribosomal protein gene (exception: 8 uncultivated viral contigs that contained 2; Supplementary Fig. 12), which is clearly not enough for viruses to build functional ribosomes on their own. Presumably, these viruses are merely tweaking ribosomal functioning in their hosts—just as observed for auxiliary metabolic genes (AMGs) whereby viruses are merely tweaking ribosomal functioning in their hosts.

Virus-encoded RP genes are under strong purifying selection. Presence of ribosomal protein genes in viral genomes raises a question of what their functions in the course of the infection cycle might be and how do viruses benefit from carrying such genes. The eS30-encoding gene increases the transformation capacity of FBR-MuSV in vitro by twofold, providing clear fitness advantage to the virus\textsuperscript{30}. It is conceivable that homologs of other ribosomal proteins might be also beneficial for the bacteriophages that encode them. For instance, it is known that bS21 is necessary during translation initiation step and in the absence of bS21, ribosomes are incapable of binding natural mRNAs\textsuperscript{32}. Thus, phage-encoded bS21 might compete with and replace the cellular bS21, ensuring translation of viral transcripts. Similarly, viral bl12 domain proteins might provide interfaces for virus-specific translation factors. Protein bL9 is required for translational fidelity and is involved in suppression of frameshifting. In many members of Caudovirales, production of certain tail components is dependent on programmed translational frameshifting\textsuperscript{34} and a viral copy of bL9 might help to achieve optimal frameshifting in these genes. Finally, it has been demonstrated that stalling of phage protein synthesis is one of the major defense strategies in Bacteroidetes\textsuperscript{45}. Thus, viral homologs of HPF and bL31 might compete with the cellular homologs and prevent formation of ribosome dimers, thereby releasing translation inhibition and ensuring that phage transcripts are efficiently translated.

Given what seemed to be reasonable explanations for why viruses might benefit from encoding such genes, we next investigated whether virus-encoded RP genes appeared to be functional. To this end, we calculated the ratio of non-synonymous polymorphisms per non-synonymous site (pN) to the number of synonymous polymorphisms per synonymous site (pS). Here we used this metric to test if virus-encoded RPs were under purifying or positive selection, where the former (pN/pS <1) would indicate selection for a functional protein and the latter (pN/pS > 1) would indicate that the gene might be in the process of being phased out from the viral genome\textsuperscript{46}. We found that well-sampled viral-encoded RP genes (>10× coverage, and ≥1 single nucleotide polymorphism, or SNP) had an average pN/pS = 0.20 (Supplementary Table 2). This suggests that these genes are under strong purifying selection and thus likely retained an important function after being transferred into virus genomes.

Virus-encoded RPs are incorporated into ribosomes. Although encouraging, the results of these in silico functional assays did not exclude the possibility that the viral RPs function in a different framework compared to their bona fide cellular homologs. Thus, we next explored whether the viral proteins are incorporated into ribosomes, by focusing on 3 RPs encoded by cultivated phages and most frequently detected in uncultivated phage genomes (Fig. 2). These were pelagiphage-encoded bS21, bl12 from Salmonella phage FSL SP-076, and HPF from Escherichia coli phage rv5. Following moderate and controlled expression of the respective viral proteins, 70S ribosomes were isolated under high-stringency salt conditions (see Methods) to avoid unspecific association of viral proteins\textsuperscript{47}. Judging from the obtained ribosome profiles (Fig. 4a) and transmission electron microscopy (Supplementary Fig. 13), expression of the viral proteins did not affect the 70S stability. All examined
samples nearly exclusively contained 70S monoribosomes and a small portion of 100S particles (Fig. 4a). Subsequent mass spectrometry (MS) analysis of the 70S and 100S ribosomes purified on sucrose gradients unequivocally showed that bS21 and bL12 (Table 2, Supplementary Data 2 and 3), but not HPF (Supplementary Data 4), were stably incorporated into the ribosomes when expressed in E. coli. Consistently, there was no discernible effect on the formation of 100S particles in cells expressing the viral HPF homolog (Fig. 4a). Notably, HPF was detected using MS in the crude cell extracts (Table 2, Supplementary Data 5), indicating that lack of its incorporation into ribosomes is not due to poor protein expression, but may rather result from other factors, such as inadequate growth phase, genuine loss of ability to bind to ribosomes or dissociation due to stringent washes with salt during ribosome isolation.

Fig. 3 Ribosomal proteins bS21 identified in uncultivated viral genomes. Top: Phylogenetic tree of ribosomal protein bS21. Viral sequences are highlighted with blue branches. Bottom: Genome comparison of viral contigs encoding ribosomal protein bS21. Comparisons were done at the amino acid level, with the % identity displayed with a color scale. The predicted bS21-encoding genes are indicated in red. The position of these sequences in the tree (top panel) is indicated with colored circles.
In order to check whether the incorporation of virus-encoded ribosomal proteins alters bacterial translation, we next proceeded to purification of ribosomes carrying bS21 and bL12 viral proteins. Towards this aim, we overproduced the two streptavidin-tagged viral ribosomal proteins in *E. coli* and the ribosomes containing the tagged viral proteins were purified by affinity. Only ribosomes containing bS21 could be purified, suggesting that the tagged bL12 cannot be efficiently incorporated or that its interaction with the ribosome is not sufficiently strong to allow the purification. The presence of the viral bS21 within the ribosomes was confirmed by MS analysis. The activity of the purified ribosomes was then verified by an in vitro translation assay. The protein synthesis is severely impaired in the presence of viral bS21 (~5% compared to wild-type ribosomes; Fig. 4b). This impairment is not due to the presence of the affinity tag, because ribosomes containing the streptavidin-tagged *E. coli* bS21 were still active (~75% compared to wild-type ribosomes; Fig. 4b). The inactivity of ribosomes carrying the viral bS21 suggests that additional virus- or host-encoded factors may be needed for proper translation or that the RP from a phage infecting *Pelagibacter ubique* may not be fully compatible with *E. coli* ribosomes.

Regardless, these results indicate that following their transfer into viral genomes, bS21 and bL12 retained the ability to incorporate into ribosomes, successfully competing with the native cellular counterparts. Whether incorporation of these and other viral RPs modulates protein translation during phage infection remains to be demonstrated by further functional assays both with heterologously expressed RPs as well as in the framework of the infection with the corresponding phages.

**Discussion**

The evolution of life is a history of virus-host interactions, an incessant “arms race” between viruses and cells. To ensure their successful proliferation, both parties have evolved numerous molecular strategies which are continuously being uncovered. Among the most remarkable recent discoveries are various defense and counter-defense mechanisms, peptide-based communication strategies employed by bacterial viruses to alternate...
between the lytic and lysogenic life cycles, as well as virus-mediated manipulation of the behavior and reproductive strategies of the hosts. Horizontal gene transfer plays a key role in these processes, with many, if not all, molecular innovations being shuttled back and forth between viruses and cells, irrespective of who the original inventor was. Indeed, even such hallmark virus components as the major virion proteins have been recurrently exapted from cellular proteomes and occasionally reintroduced to perform diverse cellular functions. Similarly, numerous AMGs have been recruited by viruses from their hosts, enabling viral manipulation of various cellular pathways, and transforming the infected cell into a specialized virion factory. Genes for the core components of translation machinery, namely RPs, until now appeared as the last unbreached boundary between the cellular and viral kingdoms, despite the fact that certain viruses were known to tinker with protein translation by encoding tRNA genes, protein deformylases or tRNA synthetases. Our current work shows that diverse ribosomal proteins are in fact commonly encoded by numerous cultivated and uncultivated viruses with relatively small genomes and offers support for them having an evolutionary fitness advantage for viruses during infection. Notably, virus-encoded RPs appear to be differentially selected for across environments as aquatic viruses are enriched for bS6, uS9, uS15, and HPF (Fig. 2). Although ribosomes are highly stable macromolecular assemblies which retain most of their original components during cellular growth and division, some elements (proteins bS21, bL12, bL9, bL31, and bL33) are highly dynamic, solvent accessible, and among the few proteins that are loosely bound to the ribosome and can be exchanged in vivo between ribosomes. These dynamic ribosomal proteins are enriched in viruses, presumably because they are most suited to homologous replacement during infection. Just as in ocean virus AMGs, it appears that viruses co-opt and fix not all enzymes in a metabolic pathway, but instead only central regulators or enzymes for key rate-limiting steps in a pathway. More generally, such selective acquisition of key components of the multisubunit assemblies, such as ribosomes, or recruitment of central regulators of rate-limiting steps in metabolic pathways appears to be a general strategy employed by viruses to optimize the metabolic state of the infected cells and/or to achieve the takeover of the host. Together, these functional and computational findings highlight widespread means by which viruses might modulate translation upon infection and either call into question a commonly used definition for life—the presence of RPs—or further blur the borders between viruses and cellular life forms. Given that viral metagenomes becomes a major approach in virus discovery with new RNA and DNA virus genomes and large genome fragments being discovered by the thousands to tens of thousands, we are most certainly bound to discover many more new strategies of host take-over in the near future.

Methods
Sequencing and Analyses. All viral genomes were downloaded from viral RefSeq database (ftp://ftp.ncbi.nih.gov/refseq/release/viral/). A hidden Markov model (HMM) profile was downloaded from the PFAM database (http://pfam.xfam.org/) for each domain listed in Supplementary Table 1. In total, 106 sequence profiles corresponding to distinct ribosomal protein domains were used as seeds to search the proteomes of viruses infecting hosts from the three cellular domains, as well as proteins predicted on viral contigs from two previously published global metagenomic datasets, Global Ocean Virome, and Earth genomic datasets, Global Ocean Virome, and Earth. The domains were identified by HHSearch with E-value of 1e-5. For isolates, the identified hits were then manually inspected using HHpred. All alignments were constructed using PROMALS3D. Maximum likelihood phylogenetic trees were constructed using PhyML with a WAG substitution model and the proportion of invariable sites estimated from the data. For metagenomic predicted proteins, multiple alignments were built with Muscle and maximum likelihood phylogenetic trees were constructed with FastTree. Genomic comparisons were performed using BLAST with the BLOSUM45 matrix. The ribosomal structure was downloaded from PDB database and visualized using Chimera. Selective temperate phages were identified by searching for the following PFAM domains in ribosomal protein-encoding viral contigs: Mu-transpos_C, Phage_int SAM_5, and Phage_intermase (hmmssearch, threshold of 30 on bit score). Host predictions for viral metagenomic contig were obtained from the original studies (i.e., ref. and and, in the case of the Earth’s virome, complemented with a k-mer based prediction computed with WshH (p-value threshold 5.3e-001).

To further confirm the functionality of RPs encoded on uncultivated viral genomes, selective constraint on these AMGs was evaluated through pN/pS calculation, as in ref. Briefly, synonymous and non-synonymous SNPs were observed in each ribosomal protein gene covered ±10 k, and compared to expected ratios of synonymous and non-synonymous SNPs under a neutral evolution model if at least 1 SNP was identified. The interpretation of pN/pS is similar as for dN/dS analyses, with the operation of purifying selection leading to pN/pS values <1.

Genetic Constructions. The genes encoding for bS21 protein from Pelagibacter phage HTVC008M (ACE604433), HPF protein from Escherichia coli bacteriophage rv5 and bL12 protein from Salmonella phage FSP-SP076 (ACE88397) were synthesized by Eurofins Genomics (Ebersberg, Germany). bS21 and HPF genes were cloned into pEX-A2 plasmid and bL12 gene into pEX-K4 plasmid. The gene corresponding to HPF viral protein was digested by BsaI and HindIII and inserted into a pBAD24 vector between NcoI and HindIII restriction sites. The genes corresponding to bS21 and bL12 viral proteins were cloned into the same vector, using EcoRI and HindIII restriction sites. The C-terminal streptavidin-tagged versions of bS21 from Pelagibacter phage and E. coli were also synthesized and cloned into pBAD24 plasmids. The pBAD24 plasmid harbors an arabinose dependent promoter, a pBAD32 origin and the ampicillin resistance coding sequence.

Protein expression and cell lysis. Escherichia coli strain NM2522 was used for expression of viral bS21, HPF, and bL12 proteins. The same strain harboring empty pBAD24 was used as a negative control. Overnight pre-cultures were grown in the presence of 1 mM of L-arabinose and 100 µg/mL of ampicillin. Then the expression was maintained in the cell culture until the end of exponential phase. Once the cultures reached an OD600 of 1, the cells were centrifuged at 8000 g for 7 minutes at 4 °C. The cell pellet was then washed into saline water at a concentration of 9 g/L of NaCl. A second centrifugation was made and the bacterial pellet was frozen at −80 °C.

70S Ribosome purification. The E. coli cells were resuspended in Buffer 1 (Tris-HCl pH7.5 20 mM, MgOAc 50 mM, NaCl 100 mM, EDTA 0.5 mM and DTT 1 mM) and finally lysed using the French Press. The lysate was centrifuged and the supernatant was put above the same volume of high-salt sucrose buffer (Tris-HCl pH7.5 20 mM, MgCl2 10 mM, NaCl 100 mM, EDTA 0.5 mM and DTT 1 mM) and finally centrifuged at 90,000 g for 20 h at 4 °C using Type 70 Ti rotor (BECKMAN L-90 ultracentrifuge), the ribosomes form a translucent pellet. The ribosome pellet was washed several times to remove membranes and then resuspended in Buffer 2 (Tris-HCl pH7.5 10 mM, MgCl2 10 mM, NaCl 50 mM, EDTA 0.5 mM and DTT 1 mM) on ice. An equivalent of 2000 OD600 units of ribosomes were loaded on top of a 10–50% sucrose gradient into polycarbonate tubes. The ultracentrifugation was performed at 95,000 g for 18 h at 4 °C using SW28 rotor (BECKMAN L-90 ultracentrifuge). The gradient was then fractioned into 500 µL aliquots. The OD260 values were determined for each fraction to locate the 70S absorbance peak. The corresponding fractions were pooled in one volume of buffer 2 and centrifuged at 95,000 g for 20 h at 4 °C using Type 70 Ti rotor in order to remove sucrose. The pellet was recovered in buffer 2 after dialysis, the ribosomes were ready for mass spectrometry analysis or purification using StreptTactin HP columns (GE healthcare) for in vitro translation assays.

The StreptTactin Sepharose column (StreptTact HP columns, GE Healthcare Life Sciences) was equilibrated with the buffer 3 (MgOAc 9 mM, NaCl 10 mM, KCl 50 mM, HEPEs–KOH pH7, 5 mM and DTT 1 mM). After injection of the purified ribosomes, the column was washed with 100 mL of buffer 3 before elution with 100 mL of d-Destabilin (GE Healthcare) for in vitro translation assays. Fractions containing ribosomes were finally concentrated (Amicon 30 kDa) and reequilibrated in the buffer 3 for in vitro translation assays.
In vitro translation assays. In vitro translation was performed using PURExpress® 
ARibosome Kit (New England Biolabs). Wild-type ribosomes (controls) or 
edult ribosomes carrying E. coli or virus-encoded tagged S21 were tested (final 
centrification: 10 pmol). A green fluorescent protein was translated by using a PCR 
product containing 77 promoter to rapidly evaluate the translation rates.

Fluorescence analysis. After 3 h incubation at 37°C, the reaction volume was 
adjusted to 125 μL and distributed in cuvettes for Fluorescence measurement with 
LS 55 Fluorescence Spectrometer (PerkinElmer). Fluorescence intensity of trans-
lated GFP was determined using the FinLab software with following conditions: 
λ excitation 485 nm/λ emission 775 nm 510 nm.

Negative staining. Following ribosome separation, we diluted samples 10 times in 
Buffer 2 and applied them to freshly glow-discharged 300-mesh carbon-coated grids. After 
three washes in this buffer, grids were stained with 2% uranyl acetate on FormVAR. 
They were then observed with the Tecnai G2 Spirit electron micrograph operating at 200 kV. Images were recorded with a 
4000 x 4000 Gatan Ultrascan 4000 CCD camera at a nominal magnification of 
×50,000.

Liﬂid digestion of ribosomal samples. 25 μg of ribosomes were digested 
according to the following protocol: first, 53.5 μL of 50 mM ammonium bicarbonate buffer 
(pH 7.8) was added to the sample to 65 μL total volume. After warming 1 minute, 
tubes were incubated 10 minutes at 80°C and then sonicated for two 
minutes. Reduction of disul ﬂide bonds step was processed by adding 12.5 μL of 65 
mm DTT to the sample and was incubated 15 minutes at 37°C after agitation 1 minute. Alkylation of reduced disul ﬂide bonds was realized by adding 135 mM 
iodoacetamide. Microube was then incubated 15 minutes in the dark at room 
temperature, under agitation. Finally, proteins were digested overnight at 37°C 
with 10 μL of either modiﬁed endoproteiﬁnase glu-c (0.1 μg/μL), Promega, Madison, 
WI) in 50 mM ammonium bicarbonate buffer for bS21 (due to high lysine and 
arginine content in bS21) or with modiﬁed Trypsine (0.1 μg/μL), Promega, Madison, 
WI) in 50 mM ammonium bicarbonate buffer for HPF, bl2 and control.

Protein prefractionation and digestion. Twenty ﬁve micrograms of soluble crude 
protein extracts of E. coli were boiled for 10 min with 5 mL of LDS Sample buffer 4X 
and 2 μL of reducing agent (DTT 10 × (500 mM)). They were then separated on a 
same size in the region of 7 kDa–14 kDa. The slices were 
then visualized using the FinLab software with following conditions: 
λ excitation 485 nm/λ emission 775 nm 510 nm.

LC-MS/MS analysis. Shotgun analyses were conducted on a LTQ-Orbitrap XL 
(Thermo Fisher Scientific) mass spectrometer. The MS measurements were done with 
a high-performance liquid chromatography (HPLC) system (Dionex, LC Packings Ultimate 3000) connected to a hybrid LTQ-Orbitrap XL (Thermo 
Fisher Scientific) equipped with a nanoelectrospray ion source (New Objective). 
The HPLC system consisted of a solvent degasser nano 
C18, 5 μm, Dionex, LC Packings). An aliquot of 10 μL of prepared peptide mixture was loaded onto a trapping pre-
column (5 mm × 300 μm i.d., 300 Å pore size, Pepmap C18, 5 μm) over 3 min in 2% 
buffer B at a flow rate of 25 μL/min. This step was followed by reverse-phase separations at a flow rate of 0.250 μL/min using an analytical column (15 cm × 300 
μm i.d., 300 Å pore size, Pepmap C18, 5 μm, Dionex, LC Packings). We ran a 
gradient from 2 – 35% buffer B for the first 60 min, 35 – 60% buffer B from minutes 
60 – 85, and 60 – 90% buffer B from minutes 85–105. Finally, the column was 
declamped at 0% buffer B for 15 min and then washed with 2% buffer B for 15 min before the 
next sample was loaded. The peptides were detected by directly eluting them from the HPLC column into the electrospray ion source of the mass spectrometer. An 
electrospray ionization (ESI) voltage of 1.6 kV was applied to the HPLC buffer 
using the liquid junction provided by the nanoelectrospray ion source, and the ion 
carrier gas was nitrogen at a pressure of 0.6 MPa, the spray temperature set to 200°C. The LTQ-Orbitrap XL instrument was 
operated in its data-dependent mode by automatically switching between full 
survey scan MS and consecutive MS/MS acquisitions. Survey full scan MS spectra 
(mass range 400–2000) were acquired in the Orbitrap section of the instrument with 
a resolution of r = 60 000 at m/z 400; ion injection times were calculated for each 
chromatogram to allow for accumulation of 106 ions in the Orbitrap. The ten most 
in intense peptide ions in each survey scan with an intensity above 2000 counts (to 
avoid triggering fragmentation too early during the peptide elution profile) and a 
charge state ≥2 were sequentially isolated at a target value of 10,000 and frag-
mented in the linear ion trap by collision-induced dissociation. Normalized collid-
ence energy was set to 35% with an activation time of 30 ms. Peaks selected for 
fragmentation were automatically put on a dynamic exclusion list for 30s with a 
mass tolerance of ± 10 ppm to avoid selecting the same ion for fragmentation more 
than once. The following parameters were used: the repeat count was set to 1, the 
exclusion list size limit was 500, singly charged precursors were rejected, and the 
mixture injection time was set at 300 and 300 ms for full MS and MS/MS scan 
events, respectively. For an optional duty cycle, the fragment ion spectra were 
recorded in the LTQ mass spectrometer in parallel with the Orbitrap full scan 
detection.

Protein orbitrap measurements, an external calibration was used before each 
injection series ensuring an overall error mass accuracy below 5 ppm for the 
detected peptides. MS data were saved in RAW file format (Thermo Fisher 
Scientific) using XCalibur 2.0.7 with tune 2.4. The data analysis was performed with 
Proline software 1.4 supported by Mascot Distiller and Mascot server 
(v2.5.1; http://www.matrixscience.com) database search engine for peptide and 
protein identification using its automatic decoy database search to calculate a 
false discovery rate (FDR) of 1% at the peptide level. MS/MS spectra were compared to the Escherichia coli Reference proteome set database containing the 
phage ribosomal proteins (UniProt release 2017_01, 18 January 2017, 
23022 sequences, 7070297 residues). Mass tolerance for MS and MS/MS was 
set at 5 ppm and 0.5 Da, respectively. The enzyme selectivity was set to full 
trypsin with one misscavage allowed for samples HPF and bl2 and the 
enzyme selectivity was set to full V8-DE with one mislavage allowed for sample 
bS21.

Protein modifications were fixed carbamidomethylation of cysteines, variable 
oxidation of methionine, variable acetylation of lysine, and variable acetylation of 
N-terminal residues.

Reporting summary. Further information on experimental design is available in the 
Nature Research Reporting Summary linked to this article.

Data availability

The authors declare that the data supporting the findings of this study are available 
within the paper and its supplementary information files. Viral contigs analyzed in this 
study were from two previously published global metagenomes datasets (Global Ocean 
Virome and Earth’s Virome) that are available at https://img.jgi.doe.gov/cgi-bin/vr/main.cgi and http://datacommons.cyverse.org/browse/iplant/home/shared/Virus/GOV.

Received: 22 September 2017 Accepted: 18 January 2019 
Published online: 14 February 2019

References
1. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. 
Photosynthesis genes in marine viruses yield proteins during host infection. 
NATURE 438, 86–89 (2005).
2. Fridman, S. et al. A myovirus encoding both photosystem I and II proteins 
enhances cyclic electron flow in infected Prochlorococcus cells. Nat. 
Microbiol. 2, 1350–1357 (2017).
3. Hurwitz, B. L., Hallam, S. J. & Sullivan, M. B. Metabolic reprogramming by 
viruses in the sunlit and dark ocean. Genome Biol. 14, R123 (2013).
4. Sullivan, M. B. et al. Genomic analysis of oceanic cyanobacterial myoviruses 
compared with T4-like myoviruses from diverse hosts and environments. 
Environ. Microbiol. 12, 3035–3056 (2010).
5. Hurwitz, B. L., Brum, J. R. & Sullivan, M. B. Depth-stratiﬁed functional and 
taxonomic niche specialization in the ‘core’ and ‘flexible’ Pacific Ocean Virome. 
ISME J. 9, 472–484 (2015).
6. Thomson, L. R. et al. Phage auxiliary metabolic genes and the redirection of 
cyanobacterial host carbon metabolism. Proc. Natl. Acad. Sci. USA 108, 
6757–6764 (2011).
7. Anantharaman, K. et al. Sulfur oxidation genes in diverse deep-sea viruses. 
Science 344, 757–760 (2014).
8. Roux, S. et al. Ecolgenomics and potential biogeochemical impacts of globally 
abundant ocean viruses. Nature 537, 689–693 (2016).

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-08672-6 | www.nature.com/naturecommunications 9
Sontheimer, E. J. & Davidson, A. R. Inhibition of CRISPR-Cas systems by mobile genetic elements. Curr. Opin. Microbiol. 37, 120–127 (2017).

Stanley, S. Y., & Maxwell, K. L. Phage-Encoded Anti-CRISPR Defenses. Annu. Rev. Genet. 23, 445–464 (2018).

Nechaev, S. & Severinov, K. Bacteriophage-induced modifications of host RNA polymerase. Annu. Rev. Microbiol. 57, 301–322 (2003).

Colson, P., Fournous, G., Diene, S. M. & Raoult, D. Codon usage, amino acid usage, transfer RNA and amino-acyl-tRNA synthetases in Mimiviruses. Intervirology 56, 364–375 (2013).

Miller, E. S. et al. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67, 86–156 (2003).

Sencik, A. et al. Snapshot of haloarchaeal tailed virus genomes. RNA Biol. 10, 383–386 (2013).

Bailly-Bechet, M., Vergassola, M. & Rocha, E. Causes for the intriguing spatial and temporal patterns in the marine environment. Curr. Opin. Microbiol. 6, 619–628 (2012).

Robertson, E. S. & Nicholson, A. W. Phosphorylation of Escherichia coli elongation factor G and ribosomal protein S6 in bacteriophage T7-infected cells. Mol. Microbiol. 10, 1057 (1994).

Korepanov, A. P., Korobeinikova, A. V., Shestakov, S. A., Garber, M. B. & Gongadze, G. M. Protein L15 is crucial for in vivo assembly of the bacteriophage 50S ribosomal subunit central proteolipid. Biotechnol. Bioeng. 100, 9533–9539 (2012).

Ueta, M., Wada, C., Bessho, Y., Maeda, M. & Wada, A. Ribosomal protein L31 in Escherichia coli contributes to ribosome subunit association and translation, whereas short L31 cleaved by protease 7 reduces both activities. Genes Cells 22, 452–471 (2017).

Van Duin, J. & Wijnands, R. The function of ribosomal protein S21 in protein synthesis. FEBS Lett. 407, 240–249 (2009).

Korean, S. O., Sjöberg, P., Nilsson, M. & Söding, J. WIsH: who is the puppet master? Replication of a parasitic virus in its bacterial cellular ancestors. Proc. Natl. Acad. Sci. USA 114, E4201–E4210 (2017).

Koonin, E. V. & Dolja, V. V. A virocentric perspective on the evolution of life. Curr. Opin. Virol. 3, 546–557 (2013).

Erez, Z. et al. Communication between viruses guides lysis-lysogenic decisions. Nature 541, 488–493 (2017).

Mehta, P. & Sicheri, F. X-ray crystal structure and functional analysis of mimic TyrRS and MetRS. Nature 240, 157–159 (2002).

Robertson, E. S. & Nicholson, A. W. Phosphorylation of Escherichia coli RNA polymerase. Annu. Rev. Genet. 40, 229–250 (2006).

Ban, N. et al. A new system for naming ribosomal proteins. Biochemistry 31, 4822–4827 (1992).

Sharon, I. et al. Comparative metagenomics of microbial traits within oceanic viral communities. ISME J. 5, 1178–1190 (2011).

Frank, J. A. et al. Structure and function of a cyanobacterial encoded peptide defensyme. ISME J. 7, 1150–1160 (2013).

Robertson, E. S., Aggison, L. A. & Nicholson, A. W. Phosphorylation of elongation factor G and ribosomal protein S6 in bacteriophage T7-infected Escherichia coli. Mol. Microbiol. 11, 1045–1057 (1994).

Robertson, E. S. & Nicholson, A. W. Phosphorylation of Escherichia coli translation initiation factors by the bacteriophage T7 protein kinase. Biochemistry 31, 4822–4827 (1992).

Ferreira, P., Krupovic, M. & Prangishvili, D. Cellular domains and viral lineages. Trends Microbiol. 22, 554–558 (2014).

Raoult, D. & Forterre, P. Redefining viruses: lessons from Mimivirus. Nat. Rev. Microbiol. 6, 315–319 (2008).

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

Ban, N. et al. A new system for naming ribosomal proteins. Curr. Opin. Struct. Biol. 24, 165–169 (2014).

Michiels, L., Van der Rauwelaert, E., Van Hasselt, F., Kas, K. & Merregaert, J. fau cDNA encodes a ubiquitin-like-S30 fusion protein and is expressed as an antisense sequence in the Finkel-Biskis-Reilly murine sarcoma virus. Onconase 8, 2537–2546 (1996).

Mouradta-Maarabouni, M., Kirkham, L., Farzaneh, F. & Williams, G. T. Regulation of apoptosis by fau revealed by functional expression cloning and antisense expression. Oncogene 23, 9419–9426 (2004).

Van Duin, J. & Wijnands, R. The function of ribosomal protein S21 in protein synthesis. FEBS J. 118, 615–619 (1981).

Morris, R. M. et al. SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420, 806–810 (2002).

Atkins, J. F. & Bjork, G. R. A gripping tale of ribosomal frameshifting: extragenic suppressors of frameshift mutations spotlight P-site realignment. Microbiol. Mol. Biol. Rev. 73, 178–210 (2009).

Hoffman, D. W., Cameron, C. S., Davies, C., White, S. W. & Ramakrishnan, V. Ribosomal protein L19: a structure determination by the combined use of X-ray crystallography and NMR spectroscopy. J. Mol. Biol. 264, 1058–1071 (1996).

Gudkov, A. The L7/L12 ribosomal domain of the ribosome: structural and functional studies. FEBS Lett. 407, 253–256 (1997).

Trauner, A., Looseheid, K. E., Bennett, M. H., Hingley-Wilson, S. M. & Williams, H. D. The dormancy regulator DsrR controls ribosome stability in hypoxic mycobacteria. J. Biol. Chem. 287, 24053–24063 (2012).

Paz-Espino, D. et al. Uncovering Earth’s virome. Nature 536, 425–430 (2016).

Müller, B. A. & Wild, D. G. The roles of proteins L28 and L33 in the assembly and function of Escherichia coli ribosomes in vivo. Mol. Microbiol. 23, 237–245 (1997).

Acknowledgements

This work was supported by grant ERC UE 34040 to PF; Agence Nationale pour la Recherche grants to M.K. (ANR-17-CE15-0005-01) and R.G. (Direction Générale de la Recherche scientifique).
M.K. conceived the study. C.M.M. and S.R. performed sequence analyses. C.M.M., S.R., M.B.S. and M.K. interpreted the results. C.G. and R.G. planned and performed functional assays. R.L. performed LC-MS/MS analysis. F.R.-V. and P.F. contributed resources. C.M.M. and M.K. wrote the first draft of the manuscript, with substantial contribution from S.R., M.B.S. and R.G. All authors edited and approved the final version of the manuscript.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08672-6.

Competing Interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.