The ISME Journal (2011), 1–10

www.nature.com/ismej

ORIGINAL ARTICLE

Cyanophage tRNAs may have a role in cross-infectivity of oceanic *Prochlorococcus* and *Synechococcus* hosts

Hagay Enav, Oded Béjà and Yael Mandel-Gutfreund
Faculty of Biology, Technion—Israel Institute of Technology, Haifa, Israel

Marine cyanobacteria of the genera *Prochlorococcus* and *Synechococcus* are the most abundant photosynthetic prokaryotes in oceanic environments, and are key contributors to global CO_2 fixation, chlorophyll biomass and primary production. Cyanophages, viruses infecting cyanobacteria, are a major force in the ecology of their hosts. These phages contribute greatly to cyanobacterial mortality, therefore acting as a powerful selective force upon their hosts. Phage reproduction is based on utilization of the host transcription and translation mechanisms; therefore, differences in the G + C genomic content between cyanophages and their hosts could be a limiting factor for the translation of cyanophage genes. On the basis of comprehensive genomic analyses conducted in this study, we suggest that cyanophages of the *Myoviridae* family, which can infect both *Prochlorococcus* and *Synechococcus*, overcome this limitation by carrying additional sets of tRNAs in their genomes accommodating AT-rich codons. Whereas the tRNA genes are less needed when infecting their *Prochlorococcus* hosts, which possess a similar G + C content to the cyanophage, the additional tRNAs may increase the overall translational efficiency of their genes when infecting a *Synechococcus* host (with high G + C content), therefore potentially enabling the infection of multiple hosts.
The ISME Journal advance online publication, 20 October 2011; doi:10.1038/ismej.2011.146
Subject Category: integrated genomics and post-genomics approaches in microbial ecology
Keywords: codon usage; cross-infectivity; marine cyanophages; *Prochlorococcus*; *Synechococcus*; tRNA

Introduction

Cyanobacteria of the *Synechococcus* and *Prochlorococcus* genera are important contributors to photosynthetic productivity in the open oceans (Li *et al.*, 1993; Liu *et al.*, 1997; Partensky *et al.*, 1999). Cyanophages, which are viruses infecting cyanobacteria, belong to three morphologically defined families: Podoviridae, Siphoviridae and Myoviridae (Suttle and Chan, 1993, 1994; Waterbury and Valois, 1993; Wilson *et al.*, 1993; Sullivan *et al.*, 2003). Among the cyanophages, podoviruses and siphoviruses tend to be host-specific, whereas myoviruses have a broader host range, even across genera (Sullivan *et al.*, 2003). Overall, myoviruses predominate over other phage groups in different oceanic regions (Angly *et al.*, 2006; DeLong *et al.*, 2006).

Phages rely on host cellular mechanisms to translate their proteins and reproduce. In order to take advantage of the host tRNA pool, it is expected that phage genes would be adapted to optimal bacterial codons. On the basis of this assumption, earlier studies have suggested that optimization of codon usage (CU) is a major force in phage–host co-evolution (Krakauer and Jansen, 2002). Further, in a comprehensive bioinformatics study, it has been shown that generally the CU of bacteriophages is strongly adapted to their specific host and differs from the CU of other bacterial hosts (Bahir *et al.*, 2009). Recently it has been reported that cyanophages, specifically myoviruses, carry up to 33 *bona fide* tRNA genes in their genomes (Sullivan *et al.*, 2010; Dreher *et al.*, 2011).

Bailly-Bechet *et al.* (2007) proposed that tRNAs carried in different phage genomes correspond to codons that are used highly by the phage genes but are rare in the host genome. In a recent study that tested CU adaptation in over 100 bacteriophages infecting 10 different bacterial hosts, it was shown that bacteriophage genomes are under codon-selective pressure imposed by the translational biases of their respective hosts (Carbone, 2008). On the other hand, Weigele *et al.* (2007) proposed that tRNA genes carried in viral genomes boost the expression of late phage genes encoding structural proteins. Interestingly, in the HIV-1 virus, which does not carry tRNA genes in its genome, it was recently

Correspondence: O Béjà or Y Mandel-Gutfreund, Faculty of Biology, Technion—Israel Institute of Technology, Haifa 32000, Israel.
E-mail: beja@tx.technion.ac.il or yaelmg@tx.technion.ac.il
Received 24 March 2011; revised 6 September 2011; accepted 7 September 2011

shown that the tRNA-encoding codons, which are highly used by the virus but avoided by its host, are overrepresented in its virions (van Weringh *et al.*, 2011).

An intriguing conjecture is that phages carry tRNA genes to enable cross-infectivity of hosts with different G+C contents. In order to study this question, we used the unique cyanobacterial–cyanophage system where myophages (~35–40% G+C) carry up to 33 different tRNA genes in their genomes (Sullivan *et al.*, 2010; Dreher *et al.*, 2011) and can, in some cases, cross-infect different hosts with very different %G+C contents (*Prochlorococcus* with ~30–40% G+C and marine *Synechococcus* with ~50–60% G+C) (Sullivan *et al.*, 2003).

Materials and methods

Data extraction and annotation
Annotated genomes were downloaded from the CAMERA website and the NCBI genome database. Information regarding cyanomyophages and their hosts was obtained from Sullivan *et al.* (2003) and G Sabehi and D Lindell (personal communication). Only phage–host pairs that had both the full genomic sequence available and were annotated were included in this work. A detailed list of all pairs of phages and hosts studied is given in Supplementary Table S1. tRNA genes were annotated using the tRNA-scanSE server (Lowe and Eddy, 1997) with parameters set to default.

CU and nucleotide usage (NU) profiles
CU and NU profiles were calculated for each organism using all protein-coding genes in the genome. The CU for each codon was calculated as the frequency of the given codon in a window of 1000 codons. Euclidean Distances (ED) between the CU and NU profiles of all genomes studied were calculated as defined in Equation (1). ED values were normalized from 0 to 1, representing the highest to the lowest similarity, respectively. Hierarchical clustering was performed using MeV software (Saeed *et al.*, 2006).

$$\mathrm{ED} = \sqrt{\sum_{i=1}^{64} \sqrt{(p_i - h_i)^2}} \qquad (1)$$

where p_i is the frequency of codon i in the phage genome and h_i the frequency of codon i in the host genome.

Codon Adaptation Index (CAI) values were calculated for each bacterial gene using the CAIcal server (Puigbo *et al.*, 2008). CAI values for the genes in each genome were calculated relative to the CU table representing the codon frequencies of the genes encoding ribosomal proteins or the codes for related ribosomal function in the corresponding genome.

Relationship between phage genes and the host genome
The CU profiles were calculated for each phage gene, as described above. Each phage gene was characterized by a vector representing the codon frequencies observed for that gene. Cosine Similarity Distance and tRCI (tRNA Relative Contribution Index) were further calculated to compare the CU profile of each gene with its host, as described below:

Cosine similarity distance. Cosine similarity distance is calculated using the formula

$$\mathrm{Distance} = \frac{\sum_{i=1}^{64} p_i * h_i}{\sqrt{\sum_{i=1}^{64} p_i^2} * \sqrt{\sum_{i=1}^{64} h_i^2}} \qquad (2)$$

where p_i is the frequency of codon i in the phage gene and h_i the frequency of codon i in the host genome.

tRCI calculation
The tRCI was defined to estimate the potential gain in translational efficiency per phage gene when including phage tRNAs in the total tRNA pool. The tRCI is based on the comparison between frequencies of tRNA complementary codon (tCC) matching the phage tRNAs in the viral gene and the tCC in the host genome (see Equation (3)). Notably, the tRCI index is calculated uniquely for each gene relative to a specific host (namely, the same gene in a given genome might have different tRCI values when infecting different hosts).

$$\mathrm{tRCI} = \sum \log \frac{f_{\mathrm{tCC}}(\textit{phage gene})}{f_{\mathrm{tCC}}(\textit{host genome})} \qquad (3)$$

where f_{tcc} is the frequency of tCC.

Ranking the phage genes based on codon adaptation
Ranking the phage genes according to the cosine distance between the CU profile. For each host–phage interaction pair, the phage genes were sorted based on the cosine distance between the vectors representing their CU profile of the gene and the CU profile of the specific host. The analysis was repeated for all host–phage pairs.

Ranking the phage genes according to tRCI values. Following the calculation of tRCI of phage genes for each phage–host interaction, all genes in the phage genomes were sorted and ranked according to the tRCI calculated for a specific host.

Detecting gene enrichment among the sorted phage genes
Each sorted list (either sorted according to cosine distance or tRCI values) was divided into 10 subgroups (bins) with an equal number of genes

per bin. Bins were sorted according to the rank of the genes, that is, bin #1 had the highest ranked genes and bin #10 the lowest ranked genes. The content of each bin was analyzed to detect functional enrichment using the hypergeometric distribution probability (Equation (4)): $p(x)>X$ defining enrichment of a specific gene function within a given bin, and $p(x)<X$ representing under-representation of a gene function in a given bin. All analyses were conducted for all phage–host pairs, including (1) pairs involving *Synechococcus* hosts and (2) pairs involving *Prochlorococcus* hosts (see Supplementary Table S1).

$$p(x) = \frac{\text{choose}(m,\ x)\ \text{choose}(n,\ k-x)}{\text{choose}(m+n,\ k)} \qquad (4)$$

where m is the total number of genes tested in the genomes studied, N is the number of all genes within the genomes studied, K is the number of all genes in the bin and x is the number of genes with a specific function included in the bin.

Results and discussion

CU as an indication of the tRNA repertoire in cyanobacteria

One of the most important translational optimization mechanisms is the correlation between the cellular levels of tRNA molecules and the frequencies of their corresponding codons, resulting in higher translation rates (Ikemura, 1981, 1985; Varenne *et al.*, 1984) and accuracy (Akashi, 1994). This correlation is higher in fast-growing bacteria (Rocha, 2004) and in highly expressed genes (Ikemura, 1981; Ghaemmaghami *et al.*, 2003; Goetz and Fuglsang, 2005). In order to evaluate the adaptation of a gene to the cell tRNA pool, several indices have been proposed. One of the most frequently used indexes is the CAI, which calculates the codon bias of a gene relative to the bias of a set of highly expressed genes (Sharp and Li, 1987). In most organisms, the set of highly expressed genes is composed mostly of ribosomal protein-encoding genes. However, it is difficult to determine accurately an optimal gene set representing the codons that are favored in the selective process. Therefore, for each bacterium in our study, we examined both the correlation between the CU of the entire bacterial genome as well as the CU based solely on its ribosomal protein-encoding genes. Overall, we found that the two CU profiles were highly correlated, with an average correlation coefficient of 0.85 (± 0.066) (median = 0.875). We also calculated the CAI of all bacterial genes for each host, and found that the bacterial genomes included in our study showed a relatively uniform distribution of CAI (see Supplementary Figure 1). On the basis of these results, we decided to use the CU of the entire genome as a representative of the bacterial tRNA repertoire.

CU profiles of cyanomyophages differ from their Synechococcus *hosts*

Previous studies have shown that the CU of bacteriophages is strongly adapted to their specific host but differs from the CU of other bacterial hosts (Bahir *et al.*, 2009). On the basis of these studies, our initial assumption was that cyanomyophages share a similar CU with the host they were isolated from (and their closely related hosts), as they both use the host tRNA pool. To examine this, we calculated the correlation between CU (Figure 1a) and NU profiles (Figure 1b) of bacteria and phages using the ED metric. As shown in Figure 1, the EDs calculated based on the CU and NU profiles showed very similar results. Notably, we identified clusters of organisms sharing similar CU: *Synechococcus* hosts

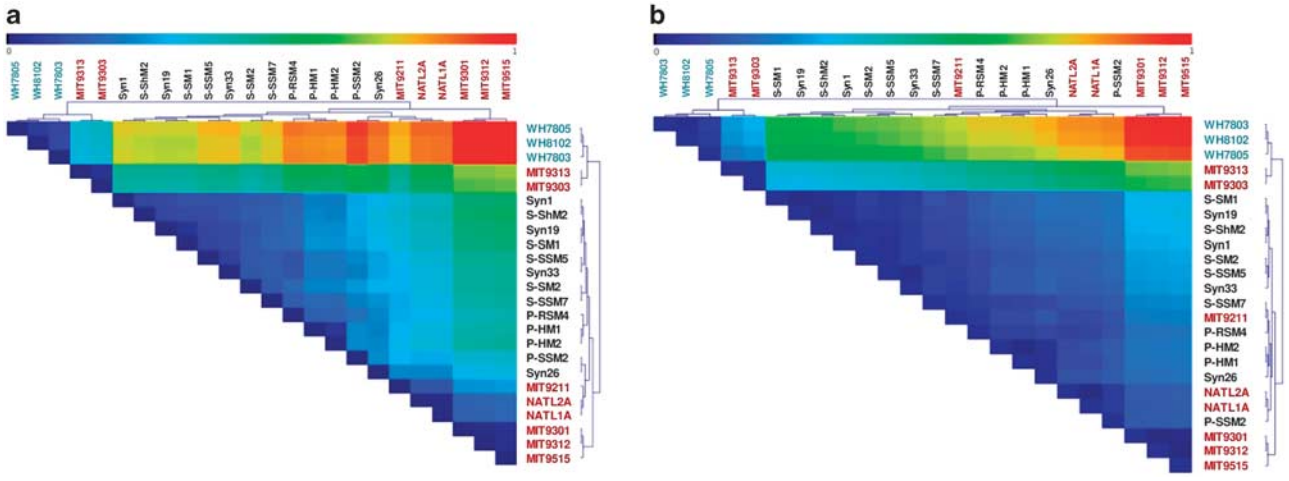

Figure 1 A clustered distance matrix based on CU (**a**) and NU (**b**) profiles. Colors range from blue, indicating close distances, and red, representing great distances. Color bar is given. Phage and host names are given: red labels indicate *Prochlorococcus,* blue labels indicate *Synechococcus* and black represents the phages.

comprise the most distinct cluster, phages tend to cluster together and the *Prochlorococcus* hosts cluster into two groups, one with a G+C content of ~50% and the other with a low G+C content of 35–40%. Surprisingly, all phages, including those isolated from a *Synechococcus* host, showed higher similarity (that is, closer distance) to *Prochlorococcus* than to *Synechococcus,* independent of their hosts.

Overall, the similar distance pattern between the different phages and hosts obtained when calculating either the CU or the NU implied that the CU profiles correlate with the nucleotide composition of the entire genomes. To further confirm this relationship, we plotted the normalized NU and CU distances of all phage–host pairs studied, each point representing the NU and CU distances of a single pair. Consistent with the ED analyses, we found a significant linear correlation between the CU- and NU-based distances (Supplementary Figure S2) with R^2 of 0.97, further confirming the close relationship between these two properties in cyanobacteria and their phages.

Codons complementary to phage tRNAs are used at high frequencies in phage genomes

From the above results, we may infer that the low G+C content of cyanomyophage genomes may shape the codon selection pattern of the phages. We further examined whether the anticodons of the tRNAs carried by the different phages are preferentially AT-rich, therefore enabling Watson–Crick pairing with the AT-rich codons of the phages. Indeed, we found that overall, the anticodons of phage tRNAs are AT-rich with an average G+C content of 31.5% (±5.2%) for all phages included in this study compared with 54.45% (±2.03%) in cyanobacterial anticodons. We also found that in 10 of the 12 viral tRNA genes, the anticodon can form Watson-Crick pairing with the lowest %GC codon coding for the given amino acid (AA) (Figure 2).

Further, we examined which codons are used at similar frequencies in the phage and host genomes, and which codons are preferred in one genome and avoided in the other. To answer this question, the codon frequencies of each phage genome were plotted against the codon frequencies of its host. This procedure was repeated for phages infecting multiple hosts, each time plotting the phage codon frequencies against the codon frequencies of one host. Overall, we found that for phages infecting a *Synechococcus* host, the plots are highly scattered (for example, Figures 3c and d), meaning that the phage and host genomes tend to use different codons. For phages infecting *Prochlorococcus* hosts, the degree of scattering varied, depending on the difference in G+C content between the two genomes (for example, Figures 3a and b). Notably, the frequencies of the codons that pair fully with the phage tRNAs (red triangles in Figure 3) are preferred in phages but are hardly used by the *Synechococcus*

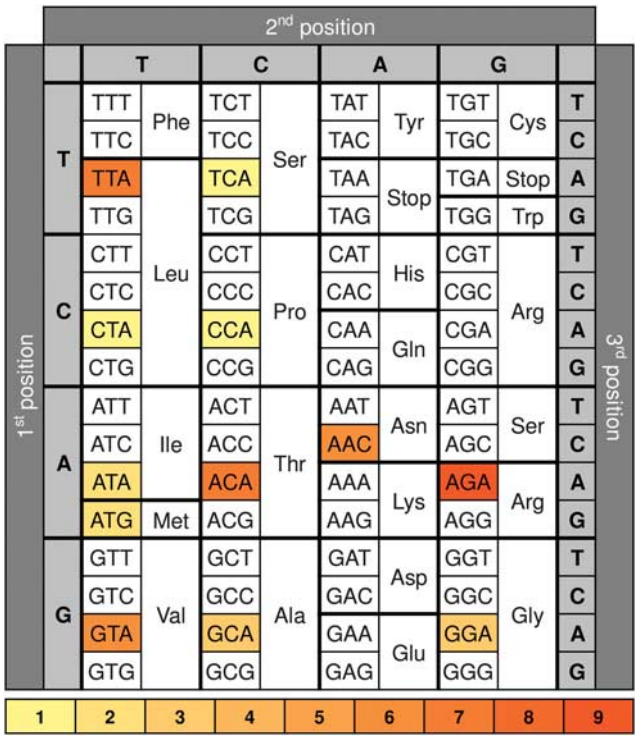

Figure 2 Distribution of phage tRNAs within cyanomyophage genomes. Genetic table presents the distribution of tRNA isoacceptors within the phage genomes. Phage tRNAs are denoted by their optimal complementary codon. Colors correspond to the number of appearances of each tRNA in the genomes studied.

hosts. In *Prochlorococcus* hosts (MIT9303 and MIT9313) having a G+C content of ~50%, we observed a similar, although slightly smaller, codon bias. In pairs associated with a *Prochlorococcus* host having a lower G+C content, the phage tRNAs optimal codons were used at similar frequencies in both the phage and the host genomes (Figure 3a).

AAs decoded by phage tRNAs are not preferred in phage genomes

On the basis of our initial results, we hypothesized that the presence of AT-rich tRNA genes in cyanomyophage genomes have an adaptive role by enhancing the translation of phage genes when infecting hosts with high G+C content. However, the presence of a specific set of tRNA genes in phages could also be interpreted as a selective pressure at the protein level, that is, when AAs encoded by phage tRNAs appear at higher frequencies within phage proteomes relative to their host proteome. To test which of the two conjectures could explain the existence of phage tRNAs, we compared the phage tCC preference (in phage and host genomes) with the preference for the AAs they code for. The frequency of each tCC within the phage genome was divided by the frequency of the tCC in the host genome. Consequently, the tCC ratio

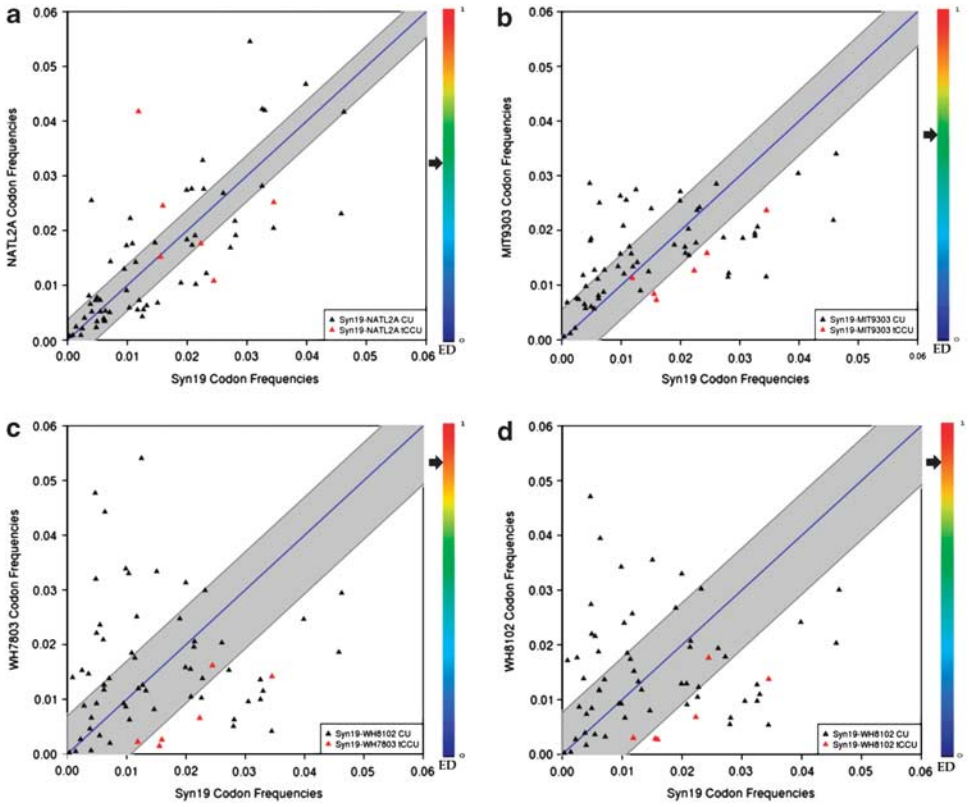

Figure 3 Phage codon frequencies compared with the codon frequencies of their hosts for four phage–host interactions: (**a**) Syn19 compared with NATL2A, a low %GC *Prochlorococcus* host; (**b**) Syn19 compared with MIT9303, a relatively high %GC *Prochlorococcus* host; (**c**) Syn19 compared with WH7803, a *Synechococcus* host; and (**d**) Syn19 compared with WH8102, a *Synechococcus* host. Each triangle stands for a specific codon; red triangles correspond to tCC. The regions between the lower and the upper quartile representing the distances of the triangle to the diagonal are shaded in gray.

was divided by the corresponding AA frequency. Our results clearly show that whereas phage tCCs are generally used at higher frequencies in phage genomes, the AAs they code for are not more commonly used in these phages. An example of the comparison between the phage tCC preference (in phage and host genomes) and the preference for AAs they code for is shown for cyanophage Syn19 in Supplementary Figure S3. These results refute the hypothesis that phage tRNA presence may be explained by the AA bias in the phage proteomes.

The possible role of viral tRNAs in viral gene translation

The differences observed between the overall CU profiles of phages and their host genomes motivated us to examine the differences at the gene level. On the basis of our working hypothesis that the presence of tRNA genes in phage genomes has an adaptive role enabling the phage to increase its fitness in a given environment, we wished to learn which of the phage genes might benefit from the presence of tRNA genes. This information may shed light on which phage genes are most critical in adapting to a given environment or host. In order to study the differences at the gene level, we examined two parameters: a) the distance between the CU of each phage gene and host genome; and b) the potential contribution of the viral tRNA genes to the translation efficiency of each gene in the phage genome.

Initially, we wished to examine which of the phage genes have a similar CU to the host of the phage it is associated with, thus enabling its efficient translation without relying on the expression of viral tRNAs, compared with those that benefit most from phage tRNAs for their efficient translation. To examine this, we calculated the cosine distances between the vectors of the CU of the individual phage genes and the overall CU of their host genomes, and ranked the phage genes in each genome according to the distance value (see Materials and methods). Notably, the distances calculated per phage gene strongly depend on the CU of the host, thus a specific phage gene will be ranked differently for different phage–host interactions. Interestingly, when annotating the ranked genes, we noticed a significantly high proportion of hypothetical genes (genes that have no detectable functionally characterized homolog in the non-redundant database) among the genes, which showed the highest discrepancy in the CU between the phage and the host, compared with the proportion

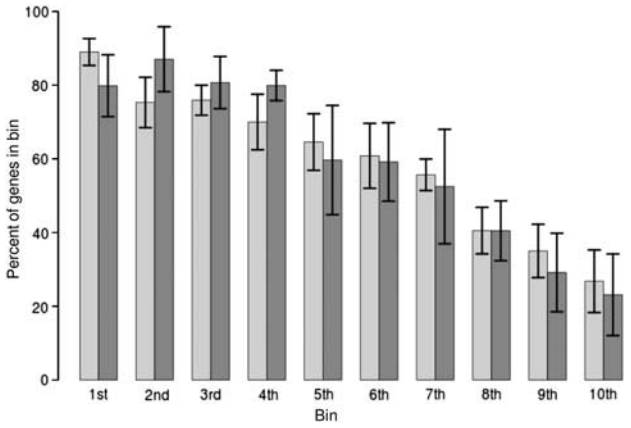

Figure 4 Distribution of genes encoding for hypothetical proteins in phage genomes sorted according to the distance between the phage gene and the host CU grouped into 10 groups (bins). Bins are marked from 1 to 10 corresponding to groups of genes having the highest to the lowest distance from the host CU, respectively. Each bin contains an equal number of genes. The height of the bar denotes the average number of hypothetical genes in each bin calculated from all studied genomes; error bars correspond to s.d.'s. Cyan bars represent the results obtained from phages associated with *Synechococcus* hosts, whereas red bars illustrate the results obtained from phages associated with *Prochlorococcus* hosts. Probabilities of the gene enrichment calculated based on the hypergeometric distribution test are detailed in Supplementary Table S2. The colour reproduction of this figure is available at the *ISME journal* online.

of hypothetical genes among genes that had a similar CU to the CU of the host. Notably, 52% of these hypothetical genes had no apparent homolog in the database (based on an E-value cut off of 10^{-5}). Strikingly, we found that in all subgroups of genes containing the 10% most distinct genes in all phage–host pairs, the proportion of hypothetical genes was extremely high ($>90\%$), whereas in the subgroup containing the 10% most similar genes in the sorted lists, the proportion of hypothetical genes was $<25\%$ (Figure 4). Overall, as shown in Figure 4, we noticed a clear tendency of hypothetical proteins to be enriched in the bins with higher ranked genes (showing the highest discrepancy between the phage and host CU) compared with the bins with lower ranked genes (showing the highest similarity between the phage and host CU). Interestingly, this phenomenon was observed when the CU was calculated both for phage—*Synechococcus* pairs and for phage—*Prochlorococcus* pairs. To evaluate whether these results are statistically significant, we further tested for each phage–host pair the probability of enrichment/underrepresentation of hypothetical genes in each bin using the hypergeometric distribution test. As shown in Supplementary Table S2, consistent with the tendency mentioned above, we detected a statistically significant enrichment of hypothetical genes in the bins, including the genes that showed the highest discrepancy between the gene CU and the host CU (ranked highest when sorting the genes according to the cosine difference between the CU vectors). However, in the bins with the lower ranked genes (that is, lowest cosine differences), the hypothetical genes were significantly underrepresented. On the basis of these results, we postulate that the genes benefiting most from the presence of tRNA genes carried by the phages may be unique genes that have an important role in adapting the phage to its specific host.

Specific viral genes tend to benefit from phage tRNA genes

To evaluate the contribution of viral tRNAs to the translation efficiency of phage genes, we defined a new term, the tRCI (see Materials and methods). As previously mentioned, the tRCI is calculated relative to the host and is unique for each phage gene, yielding different tRCI values for each gene in cases in which the phage is capable of infecting more than one host (Figure 5). Subsequently, as conducted for the CU differences, we sorted all genes in the phage genome according to their tRCI values and divided them into 10 subgroups (bins) with equal numbers of genes in each subgroup. Further, we calculated the gene enrichment in the top-ranked subgroups (including genes with the highest tRCI values). Among the gene families that were consistently enriched in the top-ranked subgroups in all phage–host interacting pairs, we found genes belonging to the high-light-induced (*hli*) gene family and the *cpeT* gene family (Figure 6).

hli *genes*

hli genes are widely distributed among cyanophages (Lindell *et al.*, 2005; Mann, 2005) and are apparently responsible for proper replication of the phage in high-light stressed cells whereas protecting the photosynthetic mechanims from light damage by dissipating excess light energy (Havaux *et al.*, 2003; Lindell *et al.*, 2005). In our analysis, *hli* genes were detected in different ranked groups, however, they were significantly more abundant at the top of the lists where tRCI values are higher (Figure 6 and Supplementary Table S3). Viral *hli* genes were previously shown to be unregulated immediately upon phage infection (Lindell *et al.*, 2007). Although these were tested only in a system using a podophage, which does not contain any tRNA genes, and in a *Prochlorococcus* host, on the basis of our results we postulate that *hli* genes may also be expressed immediately after a myophage infects a *Synechococcus* host.

cpeT *genes*

cpeT genes have a regulatory effect on the biosynthesis of the light-harvesting protein pigment phycoerythrin (Cobley *et al.*, 2002; Shen *et al.*, 2006), and are therefore predicted to affect the photosynthetic

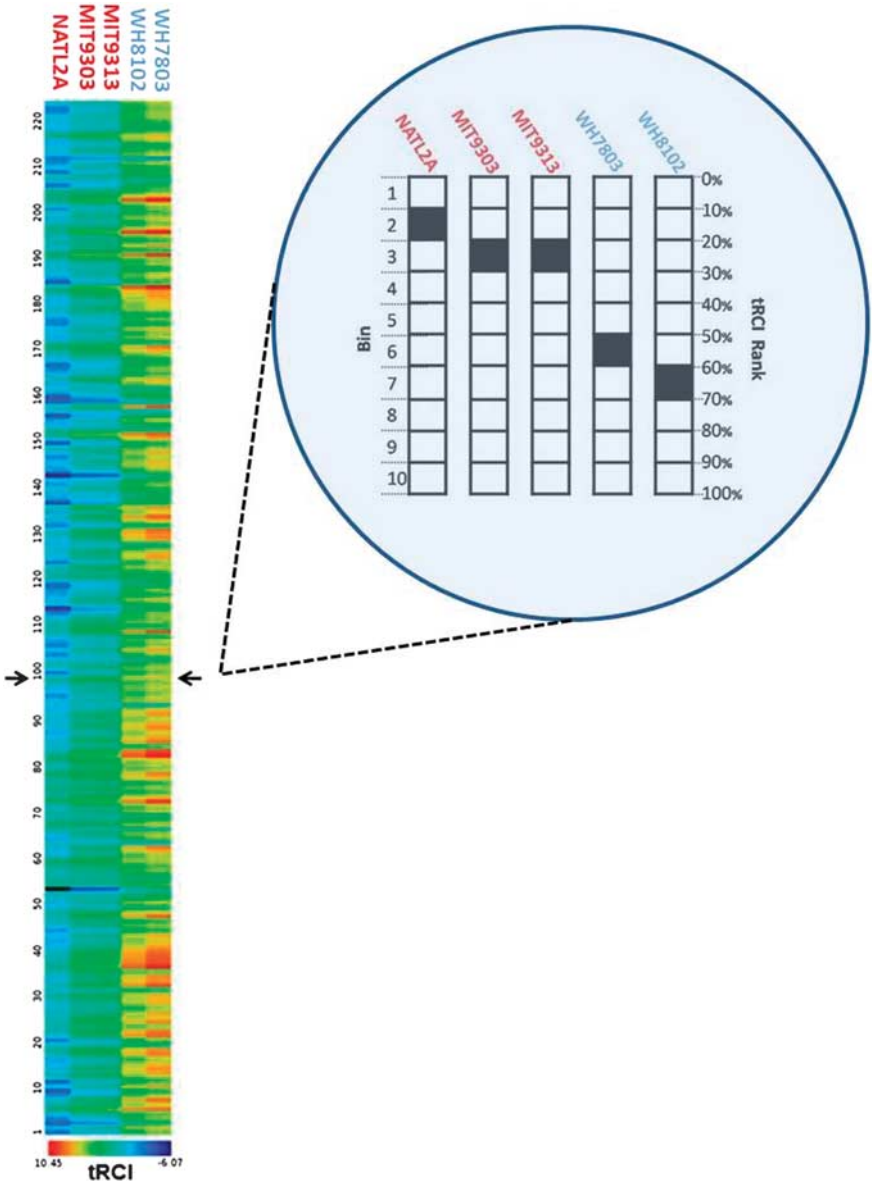

Figure 5 (Left) Heat map representing the tRCI values calculated for all ORFs in the Syn19 genome. Color bar is shown on the left of the heat map. The tRCI values are calculated independently for each of the Syn19 known hosts. Vertical lines represent the tRCI values of the gene, calculated relative to the specific host, listed on the right. (right) Example showing the ranking of the tRCI of the Syn19 gp16 gene (pointed by arrows in the heat map) when calculated for the five different hosts. tRCI values were sorted and ranked for each host from the highest tRCI (100th percentile) to the lowest (1st percentile) and grouped into 10 bins. As demonstrated, the same gene can obtain a different tRCI value that can be ranked differently and grouped in a different bin depending on the phage–host interactions.

properties of the infected cyanobacterial cells. We searched for *cpeT* genes following the procedure described above. We found that 93% of these genes were found in the top five bins and 55% in the top two bins with the highest ranked genes, according to the tRCI values. These results suggest that the expression of these genes is enhanced when phage tRNAs are expressed (Figure 6). Furthermore, taken together the enrichment of *hli* genes and cpeT genes among the genes that are predicted to benefit the most from the tRNA gene pool of the phage strongly points to the role played by the tRNA phage genes in phage adaptation.

The role of tRNA in cross-infectivity

According to our hypothesis, phage tRNA genes may have a different role in different hosts, increasing the expression efficiency of some genes when infecting one host and other genes when infecting a different host. This is reflected by the different tRCI values calculated for one host gene depending on the host the phage infects. As shown in the example given in Figure 5, a single phage gene can be ranked differently based on its tRCI when it is associated with different hosts.

To comprehensively test our hypothesis and further examine whether the same genes are

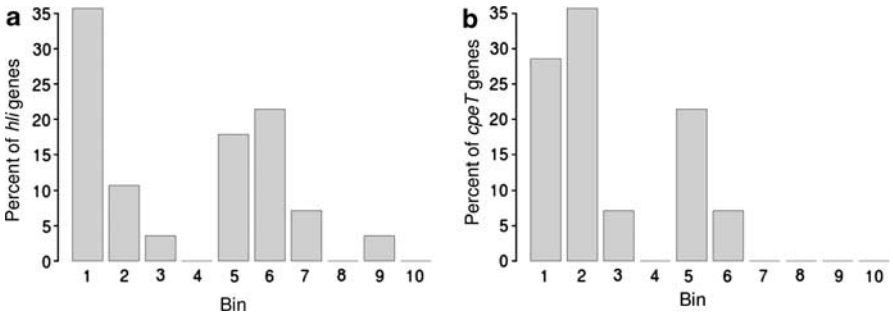

Figure 6 *hli* (**a**) and *cpeT* (**b**) enrichment in genomes sorted according to tRCI values and grouped into 10 bins from the most distant tRCI values (bin 1) to the least distant (bin 10). *P*-values for the hypergeometric distribution test are detailed in Supplementary Table S3. The colour reproduction of this figure is available at the *ISME journal* online.

promoted when phages infect different hosts or different gene groups are promoted in hosts belonging to different ecotypes, we used the Syn19 phage as a test case. Syn19 is known to infect both *Synechococcus* and *Prochlorococcus* hosts (Sullivan *et al.*, 2003). We compared the ranks of each gene of the Syn19 phage when it is associated with different hosts using the Pearson correlation coefficient. As expected, for closely related hosts, the ranks of the phage genes were found to be highly correlated (Pearson coefficient $\sim$1), whereas when comparing the ranks of the genes when the phage is associated with more distantly related hosts, very weak correlations were observed (Supplementary Figure S4). Whereas these results are generally expected based on the similarities and differences in CU profiles of the different hosts, they are generally consistent with the notion that the tRNA presumably supports increased translation resulting from higher levels of gene expression of certain phage genes in each type of host they infect.

Overall, cyanomyophage genomes tend to be AT-rich. It has been postulated that the lower G+C content may increase phage fitness as AT-rich sequences require less energy in order to melt the DNA strands, and enable faster replication of the phage genome (Miller *et al.*, 2003). AT-rich sequences also result in a slightly modified structure of the double helix, allowing better access of DNA-binding proteins, such as the complexes involved in replication and transcription (Leslie *et al.*, 1980; Calladine and Drew, 1996; El Hassan and Calladine, 1996). However, the lower G+C content of cyanomyophages compared with some of their hosts prevents the phage genomes from effectively using the tRNA pool of their hosts. We suggest that this bias may explain the presence of tRNA genes within phage genomes. Previous studies suggested that host tRNA genes may be integrated randomly in phage genomes to be lost later (Bailly-Bechet *et al.*, 2007). We postulate that tRNA genes having anticodons with low G+C content, which complement codons that are rare in some host genomes and abundant in phage genomes, have been selected to improve the expression of low GC phage genomes, increasing their fitness and fixation within phage genomes.

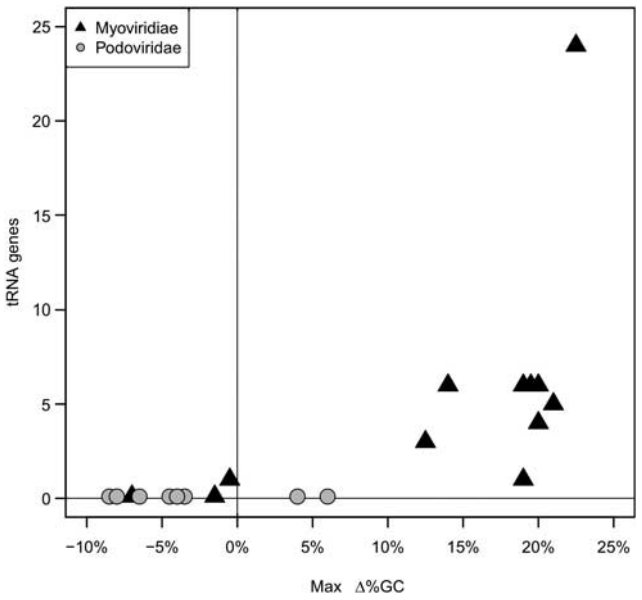

Figure 7 Phage tRNA gene copy number plotted against the difference in G+C content between the phage and its most distant host (in terms of NU). Black triangles stand for cyanomyophages (wide host range), gray circles represent cyanopodophages (narrow host range).

Some marine cyanomyophages have a broad host range, with some infecting both low G+C content hosts (*Prochlorococcus*) and high G+C content hosts (*Synechococcus*) as compared with the narrow host range of cyanopodophages (Sullivan *et al.*, 2003). Until now, no tRNA genes were reported in genomes of cyanopodophages, and all reports are from cyanomyophages (Sullivan *et al.*, 2010). Taken together, from the information available on the presence of tRNAs in marine cyanomyophages, it is clearly evident that cyanomyophages that are known to infect only *Prochlorococcus* tend to have no or a low number of tRNA in their genomes, whereas those infecting both *Prochlorococcus* and *Synechococcus* or just *Synechococcus* have a higher number of tRNA genes (see Figure 7). We therefore suggest that cyanomyophages use the strategy of carrying tRNA genes in order to be able to expand their repertoire of potential hosts while not changing the G+C content of their genomes. This

is a different strategy compared with cyanopodophages, which maintain a genome with G+C content similar to their hosts (for example, cyanopodophages P60 and Syn5 that infect *Synechococcus* and have GC content above 53% compared with P-SSP7 and P-SSP3 that infect *Prochlorococcus* and have a GC content below 40%).

The ability of cyanomyophages to infect a broad range of hosts may explain their predominance over other phages in different oceanic regions. Moreover, this was previously suggested to be a driving force in shaping photosynthetic gene diversity by facilitating the exchange of genetic materials between *Prochlorococcus* and *Synechococcus* via their shared phages (Zeidner *et al.*, 2005; Sullivan *et al.*, 2006). Therefore, the use of tRNA genes carried by these phages may have a global-scale effect on primary production in surface waters.

Acknowledgements

We thank D Lindell for her helpful discussions and I Paz for her graphical assistance. This work was supported in part by grant 580/10 from the Israel Science Foundation (OB).

References

Akashi H. (1994). Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy. *Genetics* **136**: 927–935.

Angly F, Felts B, Breitbart M, Salamon P, Edwards R, Carlson C *et al.* (2006). The marine viromes of four oceanic regions. *PLoS Biol* **4**: e368.

Bahir I, Fromer M, Prat Y, Linial M. (2009). Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol* **5**: 311.

Bailly-Bechet M, Vergassola M, Rocha E. (2007). Causes for the intriguing presence of tRNAs in phages. *Genome Res* **17**: 1486–1495.

Calladine CR, Drew HR. (1996). A useful role for 'static' models in elucidating the behaviour of DNA in solution. *J Mol Biol* **257**: 479–485.

Carbone A. (2008). Codon bias is a major factor explaining phage evolution in translationally biased hosts. *J Mol Evol* **66**: 210–223.

Cobley JG, Clark AC, Weerasurya S, Queseda FA, Xiao JY, Bandrapali N *et al.* (2002). CpeR is an activator required for expression of the phycoerythrin operon (*cpeBA*) in the cyanobacterium *Fremyella diplosiphon* and is encoded in the phycoerythrin linker-polypeptide operon (*cpeCDESTR*). *Mol Microbiol* **44**: 1517–1531.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.

Dreher TW, Brown N, Bozarth CS, Schwartz AD, Riscoe E, Thrash C *et al.* (2011). A freshwater cyanophage whose genome indicates close relationships to photosynthetic marine cyanomyophages. *Environ Microbiol* **13**: 1858–1874.

El Hassan MA, Calladine CR. (1996). Structural mechanics of bent DNA. *Endeavour* **20**: 61–67.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N *et al.* (2003). Global analysis of protein expression in yeast. *Nature* **425**: 737–741.

Goetz RM, Fuglsang A. (2005). Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from *Escherichia coli*. *Biochem Biophys Res Commun* **327**: 4–7.

Havaux M, Guedeney G, He Q, Grossman AR. (2003). Elimination of high-light-inducible polypeptides related to eukaryotic chlorophyll *a/b*-binding proteins results in aberrant photoacclimation in *Synechocystis* PCC6803. *Biochim Biophys Acta* **1557**: 21–33.

Ikemura T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol* **146**: 1–21.

Ikemura T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* **2**: 13–34.

Krakauer DC, Jansen VA. (2002). Red queen dynamics of protein translation. *J Theor Biol* **218**: 97–109.

Leslie AG, Arnott S, Chandrasekaran R, Ratliff RL. (1980). Polymorphism of DNA double helices. *J Mol Biol* **143**: 49–72.

Li WKW, Zohary T, Yacobi YZ, Wood AM. (1993). Ultraphytoplankton in the eastern Mediterranean Sea—towards deriving phytoplankton biomass from flow cytometric measurements of abundance, fluorescence and light scatter. *Mar Ecol Prog Ser* **102**: 79–87.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86–89.

Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T *et al.* (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**: 83–86.

Liu H, Nolla HA, Campbell L. (1997). *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquat Microb Ecol* **12**: 39–47.

Lowe TM, Eddy SR. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955–964.

Mann NH. (2005). The third age of phage. *PLoS Biol* **3**: e182.

Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W. (2003). Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**: 86–156.

Partensky F, Hess WR, Vaulot D. (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**: 106–127.

Puigbo P, Bravo IG, Garcia-Vallve S. (2008). CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct* **3**: 38.

Rocha EPC. (2004). Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res* **14**: 2279–2286.

Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA *et al.* (2006). TM4 microarray software suite. *Methods Enzymol* **411**: 134–193.

Sharp PM, Li WH. (1987). The codon Adaptation Index- a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**: 1281–1295.

Shen G, Saunee NA, Williams SR, Gallo EF, Schluchter WM, Bryant DA. (2006). Identification and characterization of a new class of bilin lyase: the *cpcT* gene encodes a bilin lyase responsible for attachment of phycocyanobilin to Cys-153 on the beta-subunit of phycocyanin in *Synechococcus* sp. PCC 7002. *J Biol Chem* **281**: 17768–17778.

Sullivan MB, Waterbury JB, Chisholm SW. (2003). Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**: 1047–1051.

Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. (2006). Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4**: e234.

Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR *et al.* (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* **12**: 3035–3056.

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.

Suttle CA, Chan AM. (1994). Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl Environ Microbiol* **60**: 3167–3174.

van Weringh A, Ragonnet-Cronin M, Pranckeviciene E, Pavon-Eternod M, Kleiman L, Xia X. (2011). HIV-1 modulates the tRNA pool to improve translation efficiency. *Mol Biol Evol* **28**: 1827–1834.

Varenne S, Buc J, Lloubes R, Lazdunski C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol* **180**: 549–576.

Waterbury JB, Valois FW. (1993). Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophage abundant in seawater. *Appl Environ Microbiol* **59**: 3393–3399.

Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF *et al.* (2007). Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* **9**: 1675–1695.

Wilson WH, Joint IR, Carr NG, Mann NH. (1993). Isolation and molecular characterization of five marine cyanophages propogated on *Synechococcus* sp. strain WH 7803. *Appl Environ Microbiol* **59**: 3736–3743.

Zeidner G, Bielawski JP, Shmoish M, Scanlan DJ, Sabehi G, Béjà O. (2005). Potential photosynthesis gene recombination between *Prochlorococcus* & *Synechococcus* via viral intermediates. *Environ Microbiol* **7**: 1505–1513.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)