Evolution of Codon Usage in the Smallest Photosynthetic Eukaryotes and Their Giant Viruses

Stephanie Michely[1,2,6], Eve Toulza[1,2], Lucie Subirana[1,2], Uwe John[3], Valérie Cognat[4], Laurence Maréchal-Drouard[4], Nigel Grimsley[1,2], Hervé Moreau[1,2], and Gwenaël Piganeau[5,*]

[1]UPMC Univ Paris 06, UMR7232, BIOM, Observatoire Océanologique, F-66650, Banyuls-sur-Mer, France

[2]CNRS, UMR7232, Observatoire océanologique, Banyuls-sur-Mer, France

[3]Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

[4]Institut de Biologie Moléculaire des Plantes, UPR2357 CNRS, associated with the University of Strasbourg, France

[5]School of Life Sciences, Sussex University, Falmer, United Kingdom

[6]Present address: INRA, UMR1319 Micalis, Jouy-en-Josas, France

*Corresponding author: E-mail: gwenael.piganeau@obs-banyuls.fr.

Accepted: March 29, 2013

Data deposition: The complete microarray data set has been submitted to the ArrayExpress public database at EBI (Parkinson et al. 2009) under the accession number: E-MEXP-3520. Table 1 summarizes the GenBank accession numbers of genomes analyzed in the article.

Abstract

Prasinoviruses are among the largest viruses (>200 kb) and encode several hundreds of protein coding genes, including most genes of the DNA replication machinery and several genes involved in transcription and translation, as well as transfer RNAs (tRNAs). They can infect and lyse small eukaryotic planktonic marine green algae, thereby affecting global algal population dynamics. Here, we investigate the causes of codon usage bias (CUB) in one prasinovirus, OtV5, and its host *Ostreococcus tauri,* during a viral infection using microarray expression data. We show that 1) CUB in the host and in the viral genes increases with expression levels and 2) optimal codons use those tRNAs encoded by the most abundant host tRNA genes, supporting the notion of translational optimization by natural selection. We find evidence that viral tRNA genes complement the host tRNA pool for those viral amino acids whose host tRNAs are in short supply. We further discuss the coevolution of CUB in hosts and prasinoviruses by comparing optimal codons in three evolutionary diverged host–virus-specific pairs whose complete genome sequences are known.

Key words: selection, codon usage bias, NCLDV, picoeukaryote, microarray, tRNA, *Ostreococcus*.

Introduction

Small phytoplanktonic eukaryotes (cell diameter < 2 μm) are important contributors to photosynthesis in many coastal areas, and seawater metagenomic screens have highlighted their high taxonomic diversity and ubiquity in the ocean's surface (Viprey et al. 2008; Massana 2011). Three genera of planktonic chlorophytes within the class Mamiellophyceae can be particularly prevalent in coastal surface waters: *Bathycoccus* (Monier et al. 2011; Vaulot et al. 2012), *Micromonas* (Worden et al. 2009), and *Ostreococcus* (Demir-Hilton et al. 2011). Genome comparisons between these organisms revealed unexpected ancient divergence and differences in gene content, reflecting potential niche adaptations (Jancek et al. 2008; Worden et al. 2009; Piganeau et al. 2011; Moreau et al. 2012). Recently, large nucleocytoplasmic double-stranded DNA lytic viruses infecting these algae have been found in seawater and may be one to two orders of magnitude more abundant than their hosts (Bellec et al. 2010). The genomes of seven prasinoviruses, lysing four different species of Mamiellophyceae, each contain approximately 200 genes, including genes involved not only in replication and transcription but also in translation, with the intriguing presence of 4–6 transfer RNAs (tRNAs) (Moreau et al. 2010).

Codon usage bias (CUB), the preferential use of a subset of the synonymous codons for a given amino acid in a protein

coding gene, is under a mutation–selection-drift balance (Bulmer 1991) that varies both between genes within a genome and between species (see Lynch [2007] Chapter 6 for a review). Selection may act to optimize the translation of individual genes or as a consequence of selection on nucleotide composition (Hershberg and Petrov 2010; Hildebrand et al. 2010). In many species, highly expressed genes have a higher CUB, and optimal codons correspond to the most abundant isoacceptor tRNAs (Ikemura 1985). This is as expected under selection for optimization of translation, as the use of optimal codons increases its accuracy (Stoletzki and Eyre-Walker 2007; Drummond and Wilke 2008; Zhou et al. 2009) and efficiency (Qian et al. 2012). Recently, Qian et al. (2012) provided evidence that expression is maximal for a balanced optimal codon frequency, proportional to the cellular cognate tRNA concentrations in yeast. Optimal codons in highly expressed genes are thus expected to coevolve with the most abundant cognate tRNAs (Qian et al. 2012). The null hypothesis of neutral evolution of CUB, as a consequence of mutational bias, does not predict a positive relationship between gene expression rates and CUB, unless the expression process itself is mutagenic. Analyses of CUB, and its relationship with gene expression rates and tRNA prevalence, enabled us to test the alternative hypothesis of selection for translational optimization. The understanding of the role of CUB in translational optimization of viral genes has important implications, as CUB may be a hallmark of host–virus specificity (Bahir et al. 2009) or virulence (Mueller et al. 2006). As these microalgae are of raising interest for their biotechnological potential, such as biopharmaceutical protein supplements or sustainable energy sources (Cadoret et al. 2012), CUB has also practical implications for transgene expression in the microalgae (van Ooijen et al. 2012).

Here, we first investigate whether we can find evidence of translational selection on CUB in *Ostreococcus tauri* and in its virus OtV5 (Derelle et al. 2008) by analyzing the relationship between CUB, expression rates, and tRNA gene prevalence both in the host and in the viral genome. Second, we test whether CUB of host and virus coevolve, using the available sequence data in *Bathycoccus prasinos*, *Micromonas pusilla*, *O. lucimarinus*, *O. tauri,* and the viruses infecting them, namely BpV1, MpV1, OlV1, and OtV5.

Materials and Methods

Ostreococcus tauri Growth and Lysis by OtV5 and RNA Extraction

Ostreococcus tauri (RCC745, first called OTTH0595; Courties et al. 1994) was cultured in continuous light (100 μmol·photon·m^{-2}·s^{-1}) at 20 °C on Keller's medium agitated by air bubbles. Three replicates of 2 l of *O. tauri* culture in exponential growth phase (5×10^7 cells/ml) were each inoculated with 2×10^{10} viral particles. Lysis was monitored by flow cytometry (Becton Dikinson, San Jose) every hour, by autofluorescence for the host and by Syber Green I (Molecular probes) staining for viruses (Marie et al. 1999; Brussaard et al. 2000).

Two hundred milliliters of each replicate was harvested by centrifugation at T0 (control before inoculation), T2, T5, and T10 (2, 5, and 10 h postinfection, respectively). Total RNA was extracted from each sample using RNeasy Plant Kit (Qiagen, MD) following manufacturer's instructions and after DNase I treatment (Qiagen).

The amount of total RNA was assessed using nanodrop ND-1000 spectrophotometer (PeqLab, Erlangen, Germany). RNA integrity was checked using Nano Chip Assay by Bioanalyzer 2100 (Agilent technologies, Böblingen, Germany). We kept only samples with $OD_{260/280} > 1.75$ and $OD_{260/230} \geq 1.75$.

Microarrays

The hybridization was performed on five 4X44K microarray slides where both the host *O. tauri* and the virus OtV5 genes were spotted. Four 60-mer oligonucleotide probes were designed for each gene on the microarray slide. A total of 33,093 probes were thus designed for 7,552 *O. tauri* genes and 1,056 probes for the 261 OtV5 genes. After 17 h at 65 °C of hybridization, microarrays were scanned on an Agilent G2505B scanner. Raw data were obtained by Agilent Feature Extraction Software version 9131 (FE), and the quality was monitored with Agilent QC Tool (v1.0) with the metric set GE2-v5_95. Five hundred nanograms of RNA was amplified and labeled by Low input Quick Amp Labeling kit (Agilent). For the first experiment, the control T0 cRNA was labeled with cyanine-3, and T2, T5, and T10 cRNA were labeled with dye cyanine-5. Total RNA from each postinfection sampling time was hybridized with the control (T0) for 17 h at 65 °C.

After microarray disassembly and a wash procedure according to the manufacturer's instructions (Agilent), microarrays were scanned with the Agilent G2505B scanner. Raw data processing was carried out with the Agilent Feature Extraction Software version 9.1.3.1 (FE), and the quality was monitored with Agilent QC Tool (v1.0) with the metric set GE2-v5_95.

The complete data set has been submitted to the ArrayExpress public database at EBI (Parkinson et al. 2009) under the accession number: E-MEXP-3520.

Expression Rate Estimation

To estimate the expression rate of genes in the host *O. tauri*, we used the expression signal obtained for the noninfected host cells. We first checked the correlation of the expression measures across the 12 replicates: All expression measures were highly correlated across replicates (Spearman $\rho > 0.91$, supplementary table S2, Supplementary Material online). We then computed the average expression over all probes for

each gene, transformed this into a rank (from 1 to 7,552), and summed the ranks across the 12 replicates for the 7,552 *O. tauri* genes.

To estimate the expression rate of genes in the virus OtV5, we used the expression signal obtained during infection at 2, 5, and 10 h postinoculation. All nine replicates were checked for consistency by computing correlations between expression measures between replicates (supplementary table S3, Supplementary Material online). We then computed the average expression rate for each of the 261 genes over probes and replicates.

Codon Usage Bias

CUB was estimated for each gene larger than 200 codons to reduce variance due to small samples. We used three measures of codon usage: 1) the relative synonymous codon usage (RSCU) value for each codon (Sharp and Li 1986), 2) the effective number of codons (Wright 1990), estimated by Ncp (Novembre 2002), and 3) we estimated the tRNA adaptation index (tAI) for each gene using the tAI.R code as described in dos Reis et al. (2004). tAI scores the optimality of a coding sequence with respect to a species' tRNA pools.

We defined optimal codons as codons whose frequency significantly increases with gene expression (Whittle et al. 2011) in *O. tauri* and OtV5.

Both host and viral complete genome sequences for five prasinoviruses, OtV5, OtV1, OlV1, BpV1, and BpV2 (table 1), were used. Another available sequenced prasinovirus, MpV1, was isolated from an as yet unsequenced *Micromonas* strain (RCC1109) (Moreau et al. 2010); we have, however, included the genomes of MpV1 and *Micromonas* RCC299 to analyze the general patterns of CUB in hosts and viruses. Two host–prasinovirus pairs came from samples taken at the North West Mediterranean Sea (Lion Gulf): *O. tauri* and OtV5, as well as *B. prasinos* RCC1105 and BpV1. One prasinovirus, OlV1, found in the Lion Gulf, infected a host strain, *O. lucimarinus,* isolated from the Pacific Coast near San Diego.

tRNA Data

The available annotated tRNAs did not allow the translation of all codons in both *Ostreococcus* genomes. We therefore used tRNA-scanSE (Lowe and Eddy 1997) with default parameters to identify tRNAs and completed these predictions with recent findings on permuted tRNAs in *Ostreococcus* and *Micromonas* (Maruyama et al. 2010). We curated these predictions manually to provide a complete annotation of 47 nuclear tRNA genes in *O. tauri* (supplementary table S4, Supplementary Material online). We used the software SPLITS (Sugahara et al. 2006) to detect additional permuted tRNA genes that cannot be detected with tRNA-scanSE in the six viral genomes and the recently sequenced *B. prasinos* genome.

Statistical Analysis

All statistical analyses were performed with the R software version 2.10.1 http://www.r-project.org (Ihaka and Gentleman 1996). We used ADE4 (Thioulouse and Dray 2007) and seqinR packages for within-group correspondence analysis (CA) as described by Charif et al. (2005).

The expression rates and CUB values deviate from a normal distribution (Shapiro-Wilk test, P value $< 10^{-10}$). Therefore, we used nonparametric statistics to test the significance of the correlation between expression rates and CUB. We used both the Spearman correlation coefficient on all genes with more than 200 codons (table 1) and the Kruskal–Wallis

Table 1
Genomes, Number of CDS, and tRNA Genes in Mamiellophyceae and Prasinoviruses

Genomes—Abbreviation	GenBank Accession	No. CDS	No. CDS > 200	No. tRNAs (Permuted tRNA)
Mamiellophyceae				
Ostreococcus tauri—Ota	CAID01000001-20	7,890	5,463	47 (6)
O. lucimarinus—Olu	NC_009355-75	7,603	5,806	44 (4)
Bathycoccus prasinos—Bpra	FO082260-78	8,747	7,388	33 (7)
Micromonas pusilla RCC299—Mpu	NC_013038-54	10,044	8,250	51 (3)
Prasinoviruses				
OtV5	NC_010191	264	107	5[a]
OtV1	FN386611	232	107	4[b]
OlV1	NC_014766	251	117	5[a]
BpV1	NC_014765	203	100	5 (1)[c]
BpV2	HM004430	210	98	5 (1)[c]
MpV1	NC_014767	244	109	6[d]

Note.—CDS, coding sequence.
[a]tRNA-Asn-GUU, Gln-UUG, Ile-UAU, Thr-AGU, and Tyr-GUA (Moreau et al. 2010).
[b]Same as (a) minus Tyr-GUA (Weynberg et al. 2009).
[c]tRNA-Asn-GUU, Ile-UAU, Leu-AAG, Tyr-GUA, and Thr-UGU (this study).
[d]Same as (a) + Leu-AAU.

analysis of variance test on gene expression classes, defined so that each class contains the same number of genes. We used the R pcor.test function from Kim and Yi (2006) to estimate partial correlation coefficients between expression rates, CUB, and gene length.

To assess correspondence between CUB of hosts and viruses, we counted the number of identical optimal codons in the host and the virus genome. We then estimated the distribution of the number of expected identical optimal codons by randomly sampling *i* codons among the *n* amino acids that had an optimal codon for both the host and the virus. The sampling was performed with random generation for the binomial distribution with probability given by the frequency of each codon in the lowly expressed genes (supplementary table S6, Supplementary Material online).

Results

General Features of CUB in Mamiellophyceae and Prasinoviruses

We performed a correspondence analysis on codon usage in each genome of all available Mamiellophyceae host–virus pairs (fig. 1). Because of the heterogeneity in base composition along the genomes of *Bathycoccus, Micromonas,* and *Ostreococcus* (Piganeau et al. 2011), the codon usage of the host genomes was computed independently for the normal chromosomes, the small outlier chromosome and the big outlier chromosome, that have a lower GC composition. Because the variability in amino acid composition between proteins is a confounding factor when analyzing synonymous codon usage variability, we used total codon frequencies (Perriere and Thioulouse 2002). The first two eigenvectors from each analysis were used and incorporated most information from the data sets (89.4% of variance explained). The first axis (78.2% of variance) mainly reflects the GC content at the third codon position (Spearman correlation coefficient between first axis object coordinates and the GC3 $\rho = 0.86$, $P < 10^{-15}$). On the second axis, all prasinoviruses cluster together, whereas hosts formed independent, separate clusters, with the two *Ostreococcus* hosts grouped together. The general picture is that there is considerable variation in CUB between hosts and, as a consequence of the outlier chromosomes, between chromosomes within a host. Codon usage is G or C biased for all codons in *Micromonas,* for all codons but Arg2 in *O. tauri*, and

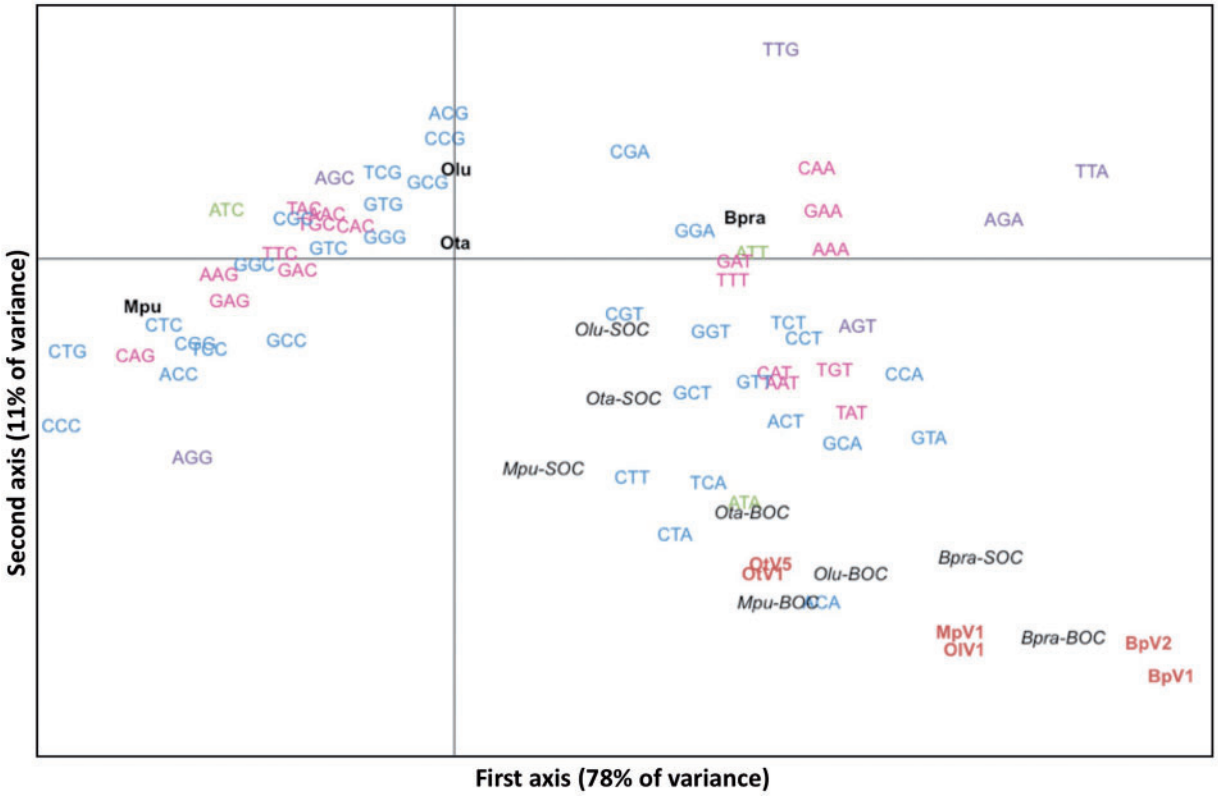

Fig. 1.—Principal component analysis of codon prevalence of the six prasinoviruses and the four Mamiellophyceae genomes, divided according to normal chromosomes (Ota, Olu, Bpra, and Mpu), low GC regions of the big outlier chromosome (Ota-BOC, Olu-BOC, Bpra-BOC, and Mpu-BOC), and small outlier chromosome (Ota-SOC, Olu-SOC, Bpra-SOC, and Mpu-SOC). Prasinoviruses are represented in red. We also projected codons on the two principal axes, 4-fold degenerate codons are blue, 3-fold degenerate codons are green, and 2-fold degenerate codons are pink.

for all codons but Arg2 and Gly in *O. lucimarinus*. In *B. prasinos,* seven amino acids (Arg2, Gln, Glu, Ile, Lys, Phe, and Arg4) have an A or T biased codon usage (supplementary table S1, Supplementary Material online).

The codon usage in prasinoviruses is A or T biased for most amino acids, the most extreme AT-biased genome, BpV1, using the doublet AT to end codons the most frequently. There is also considerable variation within viruses infecting different hosts, whereas CUB is nearly identical between OtV1 and OtV5, and between BpV1 and BpV2.

Evidence for Selection for Optimization of Translation in *O. tauri*

To estimate the expression level in *O. tauri* genes, we used the expression signal obtained for the noninfected host cells as a reference. We found a significant negative correlation between expression rate and the effective number of codons, Ncp, in *O. tauri* (fig. 2*A*, $n = 5{,}463$, Spearman $\rho = -0.22$, $P < 10^{-10}$), CUB increasing with gene expression. As CUB decreases with gene length in many eukaryotes (Duret and Mouchiroud 1999; Qiu et al. 2011; Whittle et al. 2011), gene length might be a confounding factor in the relationship between gene expression and CUB. Therefore, we reanalyzed the relationship between Ncp and expression rates taking gene length into account (partial correlation coefficient). The relationship between expression rates and gene lengths in our data set is tenuous (Spearman $\rho = 0.05$, $P < 10^{-3}$) and has no confounding effect on the relationship between Ncp and expression rates (Spearman partial correlation taking gene length into account $\rho = -0.22$, $P < 10^{-15}$).

tRNA Sharing Strategies in the Streamlined *O. tauri* Genome

If there is selection on CUB to optimize translation, we expect both the number of tRNA genes to correspond to the number of amino acids and the number of isoacceptor tRNAs to correspond to the optimal codons (Duret 2002; Qian et al. 2012). The number of tRNA copies in the genome correlates with cellular levels of tRNAs in *Escherichia coli* (Dong et al. 1996), *Bacillus subtilis* (Kanaya et al. 1999), *Chlamydomonas reinhardtii* (Cognat et al. 2008), and *Saccharomyces cerevisiae* (Percudani et al. 1997; Tuller et al. 2010). We therefore investigated whether optimal codons in highly expressed genes correspond to isoacceptor tRNA genes available in the genome.

We detected 47 tRNAs in *O. tauri* using tRNA-scan SE (Lowe and Eddy 1997) and SPLITS (Sugahara et al. 2006), which enabled the identification of six permuted tRNAs (supplementary table S4, Supplementary Material online). This is a modest number of tRNAs when compared with other eukaryotic green algae such as *Chlamydomonas* (259 tRNAs, [Cognat et al. 2008]) or in the yeast *S. cerevisiae* (275 tRNAs [Hani and Feldmann 1998]).

Because of this small number of tRNAs, tRNA sharing strategies must be considered in the coevolution of optimal codons and isoacceptor tRNAs to infer which tRNAs are dispensable (when only one tRNA is available for all codons corresponding to one amino acid, it cannot be used to characterize the tRNA-codon coevolution). Since the Wobble rule was first proposed (Crick 1966), allowing one tRNA to pair with several codons, many additional post-transcriptional tRNA modifications have been reported that allow one tRNA to pair with as many as four codons (see Agris et al. [2007] for a review). Four different

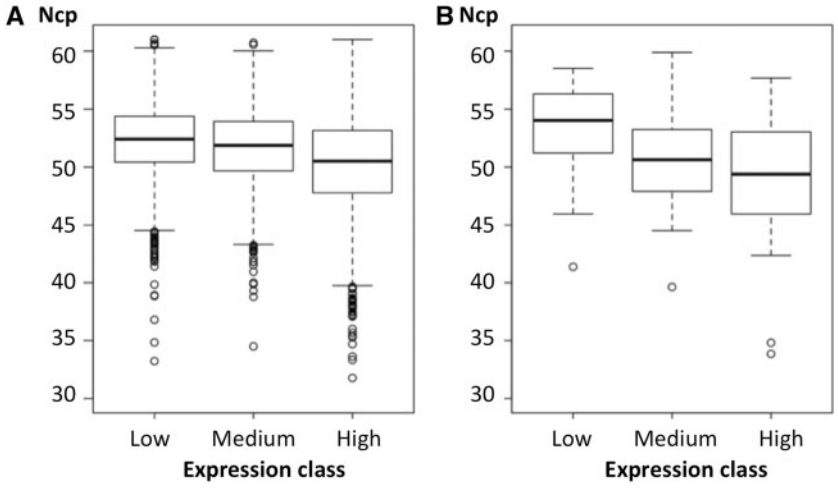

Fig. 2.—CUB, measured as Ncp, as a function of expression levels in *Ostreococcus tauri*, Kruskal–Wallis analysis of variance, $P < 10^{-10}$ (*A*) and OtV5, 10 h postinfection, $P < 0.01$ (*B*). Host and viral data include genes longer than 200 codons, and gene expression classes have been defined to have an equal number of genes between classes.

tRNA sharing strategies have been described, one of them allowing as few as 25 tRNAs to decode all 61 codons (Grosjean et al. 2010). The most common tRNA sharing strategy (pairing rule 1 in table 3) consists in the depletion of ANN-tRNAs (A at the first position of anticodon, corresponding to the third position of the codon) or GNN-tRNAs. A second pairing rule consists in the additional depletion of ANN or GNN and CNN-containing tRNAs. A third pairing rule has been reported in bacteria and involves the total depletion of ANN, GNN, and CNN tRNAs. A fourth pairing rule consists in the depletion of GNN and UNN-containing tRNAs. This last strategy has been reported for the arginine tRNA in most bacteria and for leucine and arginine in hemiascomycetous yeasts (Grosjean et al. 2010). Confronting these pairing rules with available tRNAs in *O. tauri,* we found that the combination of pairing rules 2 and 4 allows all codons to be decoded with the available tRNAs (table 3). Currently, pairing rule 2 has been observed in many eubacteria (such as *Bacillus halodurans*), a few Archea (such as *Methanopyrus kandleri*), but only exceptionally in Eukarya (Grosjean et al. 2010). This combination of pairing rules 2 and 4 implies that 14 tRNA genes are dispensable in *O. tauri*. Among these dispensable tRNAs, two correspond to nonoptimal codons, two correspond to codons whose frequency does not change with expression rates, and 10 correspond to optimal codons (table 3). Dispensable tRNAs are thus significantly more abundant for optimal codons in *O. tauri* (Binomial test, $P = 0.02$). Consistent with this excess of dispensable tRNA for optimal codons, the tRNA adaptation index (tAI) that estimates the correspondence of codon usage with available tRNAs for each gene (dos Reis et al. 2004) correlates positively with CUB, measured as Ncp (Spearman $\rho = -0.48$, $P < 10^{-15}$).

CUB Increases with Expression Rate in the Prasinovirus OtV5

We found that the CUB of the virus OtV5 genes, that infects *O. tauri*, also increases with expression levels for all times after infection in the experiment: 2 h after infection ($n = 107$, $\rho = -0.20$, $P < 0.05$), 5 h ($n = 107$, $\rho = -0.36$, $P < 10^{-3}$), and 10 h after infection (fig. 2*B*, $n = 107$, $\rho = -0.37$ $P < 10^{-4}$). There is no confounding effect of gene length on either expression rates or CUB in our viral gene data set. The positive relationship between CUB and expression rates of viral genes thus provides evidence for selection on translational optimization in the viral genome. Because the virus uses the translation machinery of its host, CUB of the viral genes is expected to converge toward the CUB of the host genes, which are GC rich in *O. tauri* (supplementary table S2, Supplementary Material online). Consistent with this, we observe a significant increase of the GC content at 4-fold degenerate sites in highly expressed viral genes (30% most expressed genes), GC3 = 52%, when compared with the other viral genes, GC3 = 47% (Wilcoxon test, $P < 0.02$).

How Can We Explain the Presence of Viral tRNAs?

OtV5 uses the host translation machinery, the tRNA pool of the host, plus its five additional tRNAs to translate its gene pool. Under a purely neutral hypothesis, viral tRNAs are drawn at random from the host genome (Bailly-Bechet et al. 2007) and should thus correspond to the most abundant host tRNAs. This is not the case as the number of tRNAs present in OtV5 does not correspond to the most abundant tRNAs in the host (Wilcoxon test, $P = 0.32$). The alternative hypothesis is that these complementary tRNAs are maintained by selection. Under selection for translational optimization in the virus, we might hypothesize that 1) the additional tRNAs correspond to optimal codons in the viral genome and/or 2) the additional tRNAs correspond to the most abundant amino acids in the virus and/or 3) they complement the host tRNA pool to optimize viral protein translation. Only two of the five viral tRNA anticodons correspond to the optimal codon in the OtV5 viral genome (tRNA-TTC-Asn and tRNA-TAC-Tyr, table 1 and supplementary table S1, Supplementary Material online). One viral tRNA, tRNA-Gln-UUG, corresponds to an amino acid without optimal codon (i.e., neither UUG nor CUG frequencies vary with expression rates). This leaves us with two viral tRNAs, tRNA-ATA-Ile and tRNA-ACT-Thr, that correspond to the nonoptimal viral codon in both amino acids. There is thus no evidence for an excess of exact codon–anticodon matches in these viral tRNAs. The complementary viral tRNAs do not correspond to the most used amino acids in the viral genome (Wilcoxon test, $P = 0.60$) nor to the most used amino acids in the highly expressed viral genes (Wilcoxon test, $P = 0.55$).

To investigate whether these viral tRNAs complemented the host tRNA pool to render it better adapted to viral amino acid requirements, we compared the ratio of host tRNAs per viral amino acid 1) for the 15 amino acids that had no viral tRNA versus the 2) the five amino acids with one viral tRNA. We found that this ratio was significantly lower for the group of amino acids with one cognate viral tRNA in the *O. tauri*–OtV5 host prasinovirus pair (fig. 3, Wilcoxon $P = 0.02$).

We repeated this analysis in two other host–virus pairs: *O. lucimarinus* and OlV1 and *B. prasinos* and BpV1. When one tRNA was lacking in either *O. lucimarinus* and *B. prasinos* (table 1), we added tRNAs according to the minimal requirement of pairing rules 2 and 4 for these genomes. We observed that amino acids with cognate viral tRNA tend to have a shortage of corresponding host tRNAs for both OlV1–*O. lucimarinus* and BpV1–*B. prasinos*, but the difference is not significant (fig. 3). Overall, this trend suggests that the viral tRNAs optimize viral translation by increasing the available tRNAs for those viral amino acids where the host's tRNA shortage is more pronounced.

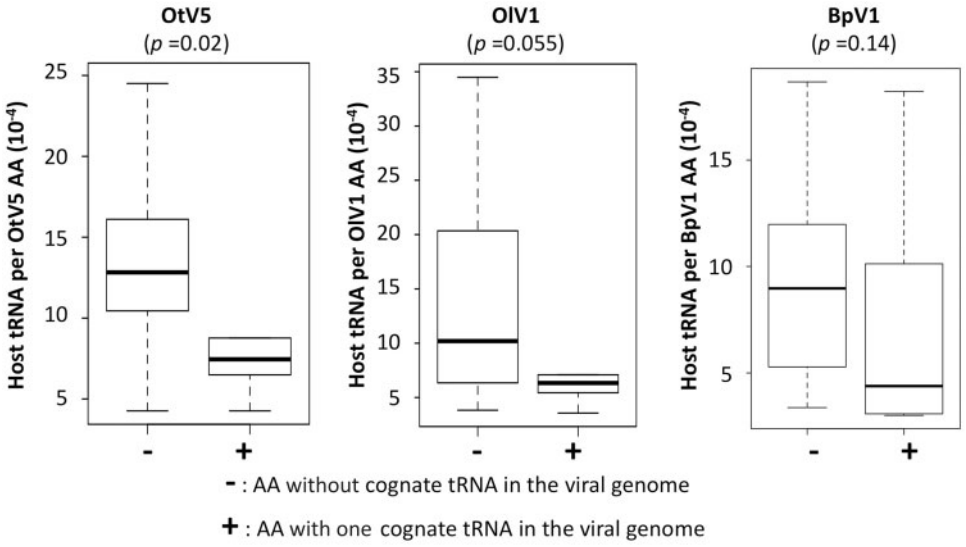

Fig. 3.—Proportion of host tRNA gene per viral AA, depending on whether the viral AA have a cognate viral tRNA (with tRNA) or not (without tRNA).

Does Host–Viral CUB Coevolution Reflect Viral Specificity?

If one virus and its host evolve for a sufficient amount of time, one may expect codon usage of a virus to be fine tuned to the translational machinery of its host and thus the CUB of its host (Carbone 2008). Under this assumption, the codon usage of the virus is expected to be more alike to codon usage of a host than to codon usage of a nonhost, despite different substitution patterns on their genomes reflected by average differences in GC content.

We investigated the correlation between host and virus CUB by randomly sampling RSCU from independent codons from host and viral genomes. However, except for the OtV5–*O. tauri* pair, we could not detect any significant positive relationship between viral and host RSCU (supplementary table S7, Supplementary Material online). As expected from the CA, CUB of host and prasinoviruses are more alike on low GC outlier chromosomes, but the higher correlation coefficients do not correspond to the specific host–virus pairs (supplementary table S7, Supplementary Material online).

We investigated the number of identical optimal codons between hosts and viruses. For *O. tauri* and OtV5, optimal codons can be inferred from the codons whose frequency increases significantly with expression. In the other host and viral genomes, we took Ncp as a proxy for expression rates to identify optimal codons, because of the positive correlation we found between Ncp and expression rates in *O. tauri* and OtV5. The expected number of identical codons between hosts and viruses can be inferred from the base composition of the viral genome alone (table 3, supplementary table S6, Supplementary Material online). The observed number of identical codons is always significantly higher than the expected number of identical codons for all host–virus pairs (table 3).

Discussion

General Features of Codon Usage in Mamiellophyceae and Their Giant Viruses

The general picture is that codon usage is biased toward GC in hosts, whereas codon usage is biased toward AT in viruses (fig. 1). However, slight differences in codon usage exist between hosts and between viral genomes, provided that they infect different host species. Previous analysis of base composition in 37 double-stranded DNA phages and their bacterial host provided evidence that base composition is on average 4% AT richer in the viral genome than in the host genome (Rocha and Danchin 2002). This trend might be the consequence of the lower cost of synthesis of A and T nucleotides when compared with G or C nucleotides. Thus, viral genomes with higher AT content would more efficiently exploit the cell resources and become selected (Rocha and Danchin 2002). Nucleocytoplasmic large double-stranded DNA viruses (NCLDV) consist of at least six families of viruses infecting a broad variety of eukaryotic hosts (Iyer et al. 2001; Yutin and Koonin 2012). Base composition analysis in 41 vertebrate-infecting NCLDVs suggest that CUB does reflect mutational pressures rather than translational selection, consistent with the lack of evidence of translational selection in their hosts (Shackelton et al. 2006). However, the life cycle of vertebrate NCLDVs, which can often persist in host cells over long periods, is quite different to those of the lytic prasinoviruses. Here, although regional mutational pressures are obvious given 1) the biased GC content of intergenic regions in

Table 2

Anticodon–Codon Correspondence in *Ostreococcus tauri*, According to Three Different tRNA Sparing Strategies

	Pairing Rule 1: Euk			Pairing Rule 2			Pairing Rule 3		
Amino Acid	**tRNA**	**Anticodon**	**Codon**	**tRNA**	**Anticodon**	**Codon**	**tRNA**	**Anticodon**	**Codon**
Arg2	1[a]	UCU →	AGA	1	UCU →	AGA	1		
Arg2	1	CCU →	**AGG**	1 (1o)	CCU	**AGG**	1 (1o)		
Gln	1	UUG →	CAA	1	UUG →	CAA	1		
Gln	nd	CUG	**CAG**		CUG	**CAG**			
Glu	3 (2u)	UUC →	GAA	3 (2u)	UUC →	GAA	3 (2n)	CfRule2	
Glu	nd	CUC	**GAG**		CUC	**GAG**			
Leu2	1[a]	UAA →	UUA	1	UAA →	UUA	1		
Leu2	1	CAA	**UUG**	1 (1o)	CAA	**UUG**	1 (1o)		
Lys	1	UUU →	AAA	1	UUU →	AAA	1		
Lys	1	CUU →	**AAG**	1 (1o)	CUU	**AAG**	1 (1o)		
Asn		AUU	AAU						
Asn	1, 1[a] (1o)	GUU	**AAC**	2 (1o)			2 (1o)		
Asp		AUC	GAU						
Asp	1	GUC	GAC	1			1		
Cys		ACA	UGU						
Cys	1[a]	GCA	**UGC**	1			1		
His		AUG	CAU						
His	1	GUG	**CAC**	1	CfRule1		1	CfRule1	
Ser2		ACU	AGU						
Ser2	1	GCU	**AGC**	1			1		
Phe		AAA	UUU						
Phe	1	GAA	**UUC**	1			1		
Tyr		AUA	UAU						
Tyr	1	GUA	**UAC**	1			1		
Ala	1	AGC →	GCU	1	AGC →	GCU	1 (1n)	AGC	GCU
Ala		GGC	GCC		GGC	GCC		GGC	GCC
Ala	1	UGC →	GCA	1	UGC →	GCA	1	UGC	GCA
Ala	1	CGC →	**GCG**	1 (1o)	CGC	**GCG**	1 (1o)	CGC	**GCG**
Arg4	1	ACG →	CGU	1	ACG →	CGU	1 (1n)	ACG	CGU
Arg4		GCG	**CGC**		GCG	**CGC**		GCG	**CGC**
Arg4	1	UCG →	CGA	1	UCG →	CGA	1	UCG	CGA
Arg4	nd	CCG	CGG		CCG	CGG		CCG	CGG
Gly	nd	ACC	GGU		ACC	GGU		ACC	GGU
Gly	1	GCC →	GGC	1	GCC →	GGC	1 (1n)	GCC	GGC
Gly	1	UCC →	GGA	1	UCC →	GGA	1	UCC	GGA
Gly	1	CCC →	**GGG**	1 (1o)	CCC	**GGG**	1 (1o)	CCC	**GGG**
Leu4	1	AAG →	CUU	1	AAG →	CUU	1 (1n)	AAG	CUU
Leu4		GAG	CUC		GAG	CUC		GAG	CUC
Leu4	nd	UAG	CUA		UAG	CUA	nd	UAG	CUA
Leu4	1	CAG →	**CUG**	1	CAG →	**CUG**	1 (1o)	CAG	**CUG**
Pro	1	AGG →	CCU	1	AGG →	CCU	1 (1n)	AGG	CCU
Pro		GGG	CCC		GGG	CCC		GGG	CCC
Pro	1	UGG →	CCA	1	UGG →	CCA	1	UGG	CCA
Pro	1	CGG →	**CCG**	1 (1o)	CGG	**CCG**	1 (1o)	CGG	**CCG**
Ser4	1	AGA →	UCU	1	AGA →	UCU	1 (1n)	AGA	UCU
Ser4		GGA	**UCC**		GGA	**UCC**		GGA	**UCC**
Ser4	1[a]	UGA →	UCA	1	UGA →	UCA	1	UGA	UCA
Ser4	2[a] (1n)	CGA →	UCG	2 (2n)	CGA	UCG	2 (2n)	CGA	UCG

(continued)

Table 2 Continued

	Pairing Rule 1: Euk			Pairing Rule 2			Pairing Rule 3		
Amino Acid	**tRNA**	**Anticodon**	**Codon**	**tRNA**	**Anticodon**	**Codon**	**tRNA**	**Anticodon**	**Codon**
Thr	1	AGU	ACU	1	AGU	ACU	1 (1n)	AGU	ACU
Thr		GGU	ACC		GGU	ACC		GGU	ACC
Thr	1	UGU	ACA	1	UGU	ACA	1	UGU	ACA
Thr	1	CGU	**ACG**	1 (1o)	CGU	**ACG**	1 (1o)	CGU	**ACG**
Val	1	AAC	GUU	1	AAC	GUU	1 (1u)	AAC	GUU
Val		GAC	GUC		GAC	GUC		GAC	GUC
Val	1	UAC	GUA	1	UAC	GUA	1	UAC	GUA
Val	2 (1o)	CAC	**GUG**	2 (2o)	CAC	**GUG**	2 (2o)	CAC	**GUG**
Ile	1	AAU	AUU	1	AAU	AUU	1 (1n)	AAU	AUU
Ile		GAU	**AUC**		GAU	**AUC**		GAU	**AUC**
Ile	1	UAU	AUA	1	UAU	AUA	1	UAU	AUA
Lacking tRNA	5			0			1		
Dispensable tRNA u:o (*p-value*)	2 : 2 (ns)			2 : 10 (0.02)			8 : 11 (ns)		

NOTE.—ns, nonsignificant. Pairing rule 4 for Leu4 is indicated by a gray arrow (Grosjean et al. 2010). For each dispensable tRNA, we indicate whether there is a corresponding optimal (o) codon, showing the best significant positive correlation to expression rate (codon in bold) or nonoptimal (u) codon, showing a significant negative correlation to expression rate. (n) indicates that there is no significant correlation between relative codon prevalence and expression for an amino acid.

[a]Permuted tRNAs.

Table 3

Number of Observed: Expected Identical Optimal Codons between Host and Viral Genomes

	OtV5	OlV1	BpV1	MpV1
Ostreococcus tauri	8:4** (14)	10:5** (14)	5:2** (10)	7:4* (12)
O. lucimarinus	12:4* (15)	12:7** (14)	8:3*** (10)	9:4*** (12)
Bathycoccus prasinos	11:4*** (13)	10:6* (12)	7:2*** (8)	7:4 (10)
Micromonas pusilla RCC299	14:6*** (15)	11:6* (14)	7:2*** (9)	12:4*** (12)

NOTE.—ns, nonsignificant. The expected number of matches is estimated from average base frequencies at 4-fold and 2-fold synonymous positions in the viral genes (supplementary table S6, Supplementary Material online). *P* values are computed by random sampling the distribution of expected identical codons. The number of amino acids for each comparison is indicated in brackets and corresponds to amino acids where both host and virus have one optimal codon.

*P value < 0.05.

**P value < 0.01.

***P value < 0.001.

Ostreococcus (Piganeau et al. 2009) and 2) the significant correlation in GC frequency at first, second, and third codon positions of genes in Mamiellophycean hosts (Piganeau et al. 2011) and 3) GC content at first and third positions in OtV5 genes (Spearman $\rho = 0.37$, $P < 10^{-4}$), we provide evidence of translational selection on codon usage in the host and viral genomes.

Translational Optimization in *O. tauri* and Its Virus OtV5

There are two hallmarks of selection for translational optimization in *O. tauri* and its virus OtV5. First, there is a positive relationship between CUB and expression rates. This phenomenon is well known in diverse organisms such as *E. coli*, *S. cerevisiae, Neurospora tetrasperma, N. discreta*, *Caenorhabditis elegans*, *Arabidopsis thaliana, Drosophila Melanogaster,* and *Silene latifolia* (Gouy and Gautier 1982; Ikemura 1985; Duret and Mouchiroud 1999; Coghlan and Wolfe 2000; dos Reis et al. 2003; Qiu et al. 2011; Whittle et al. 2011).

An alternative hypothesis to explain the positive relationship between expression rate and CUB in organisms with an excess of GC optimal codons could be a transcription-coupled AT→GC mutation rate (see Steele [2009] for a review on transcription coupled mutation processes). Because all but two optimal codons end with G or C in *O. tauri* (table 2), we performed additional analyses to test this hypothesis. Under this alternative scenario, the nucleotides that are subject to the transcription coupled mutation process should always increase with expression (i.e., GC content should increase with expression if the transcription-induced mutation bias goes from AT→GC). However, this is not the case; the Proline codon ending in C is nonoptimal, and its frequency decreases with expression (Spearman $\rho = -0.10$ and $P < 10^{-3}$). For another amino acid, glycine, one of the optimal

codons ends with T and the correlation between optimal codon frequency and expression is positive, though tenuous (Spearman $\rho = 0.07$ and $P < 10^{-3}$). These two observations infirm the hypothesis of a transcription-induced GC-biased mutation process, and we are therefore confident that the observed correlation between optimal codon usage and expression rate is the consequence of selection for translational optimization.

Second, there is an excess of dispensable tRNAs for optimal codons in *O. tauri* (table 3). A similar correspondence with isoacceptor tRNA availability has been observed in *E. coli, S. cerevisiae,* and *C. elegans* (Duret 2002). Also, the number of tRNA genes per amino acid increases with amino acid frequency in *O. tauri* (Spearman $\rho = 0.75$, $P < 10^{-4}$), whereas the viral tRNA genes complement the host tRNA pool for viral amino acids with less corresponding host–tRNAs (fig. 3).

The number of tRNA gene copies is positively correlated to growth rates in bacterial species (Rocha 2004). Consistent with the low number of tRNA genes in *O. tauri* (47, table 3), this phytoplanktonic cell has a slow doubling time of 8 h in optimal growth conditions (Farinas et al. 2006), which is over four times the doubling time of *S. cerevisiae*, which has a similar genome size.

Serendipitously, our tRNA annotation revealed a permuted tRNA in BpV1 (table 1 and supplementary table S5, Supplementary Material online). Permuted tRNAs are a recently discovered shared characteristic between archaeal and eukaryotic species (Maruyama et al. 2010), and here, we show that this characteristic can now be extended to double-stranded DNA prasinoviruses.

CUB Reflects Host–Viral Coevolution Not Host Specificity

Although whole CUB is not a good indicator of coevolution in prasinoviruses and their hosts (fig. 1), we found a significant excess of identical optimal codons between host and viruses (table 3). This is a consequence of translational selection acting on the host and on the viral codon usage, viral genes that mimic their host CUB will benefit from the same translation efficiency and accuracy as host genes. This suggests that host and viral genomes have coevolved to have a convergent CUB in highly expressed genes and that host switching should have a significant cost on viral gene expression. This being said, the number of identical optimal codons cannot be used as a proxy for host specificity, because, for example, BpV1 shares a high number of identical preferred codons with *O. lucimarinus*, a nonhost species.

In conclusion, we found evidence for optimization of translation in the host and prasinovirus genomes for highly expressed genes, which preferentially use codons with a higher number of genomic copies of host tRNAs. Our analyses provide evidence for coevolution of optimal codons between OtV5 and *O. tauri* and more generally between all prasinoviruses and their Mamiellophycean hosts. Careful annotation of tRNA genes, which may contain permuted tRNA genes in hosts and prasinoviruses, enabled us to infer tRNA sharing rules in *O. tauri*. The presence of tRNAs in prasinoviruses is biased toward viral amino acid with fewer tRNA genes in the host genome, suggesting selection on viral tRNA content for translational optimization of the viral genes. The identification of optimal codons might be useful to consider for fine tuning transgene expression both in the microalgae (van Ooijen et al. 2012) and in the prasinovirus for future research, including biotechnological applications.

Supplementary Material

Supplementary tables S1–S7 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank two anonymous referees for constructive comments and the Genomics of phytoplankton team for stimulating discussions, especially Romain Blanc-Mathieu and Rozenn Thomas. They also thank Adam Eyre-Walker for insightful suggestions on a previous version of the manuscript. This work was funded by Agence Nationale de la Recherche grants PHYTADAPT no. NT09_567009, REVIREC no. 12-BSV7-0006-01, and European Community 7th Framework Program FP7 under grant agreement no. 254619.

Literature Cited

Agris PF, Vendeix APF, Graham WD. 2007. tRNA's Wobble decoding of the genome: 40 years of modification. J Mol Biol. 366:1–13.

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.

Bellec L, Grimsley N, Desdevises Y. 2010. Isolation of prasinoviruses of the green unicellular algae *Ostreococcus* spp. on a worldwide geographical scale. Appl Environ Microbiol. 76:96–101.

Brussaard CP, Marie D, Bratbak G. 2000. Flow cytometric detection of viruses. J Virol Methods. 85:175–182.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897–907.

Cadoret J, Garnier M, Saint-Jean B. 2012. Microalgae, functional genomics and biotechnology. Adv Bot Res. 64:285–341.

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

Charif D, Thioulouse J, Lobry JR, Perriere G. 2005. Online synonymous codon usage analyses with the ade4 and seqinR packages. Bioinformatics 21:545–547.

Coghlan A, Wolfe KH. 2000. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. Yeast 16:1131–1145.

Cognat V, et al. 2008. On the evolution and expression of *Chlamydomonas reinhardtii* nucleus-encoded transfer RNA genes. Genetics 179:113–123.

Courties C, et al. 1994. Smallest eukaryotic organism. Nature 370: 255–255.

Crick FH. 1966. Codon–anticodon pairing: the wobble hypothesis. J Mol Biol. 19:548–555.

Demir-Hilton E, et al. 2011. Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. ISME J. 5: 1095–1107.
Derelle E, et al. 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. PLoS One 3:e2250.
Dong H, Nilsson L, Kurland CG. 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. J Mol Biol. 260:649–663.
dos Reis M, Savva R, Wernisch L. 2004. Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res. 32: 5036–5044.
dos Reis M, Wernisch L, Savva R. 2003. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. Nucleic Acids Res. 31: 6976–6985.
Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134: 341–352.
Duret L. 2002. Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev. 12:640–649.
Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc Natl Acad Sci U S A. 96:4482–4487.
Farinas B, et al. 2006. Natural synchronisation for the study of cell division in the green unicellular alga *Ostreococcus tauri*. Plant Mol Biol. 60: 277–292.
Gouy M, Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10:7055–7074.
Grosjean H, de Crècy-Lagard V, Marck C. 2010. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. FEBS Lett. 584: 252–264.
Hani J, Feldmann H. 1998. tRNA genes and retroelements in the yeast genome. Nucleic Acids Res. 26:689–696.
Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet. 6(9):e1001115.
Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-content in bacteria. PLoS Genet. 6(9):e1001107.
Ihaka R, Gentleman R. 1996. R: a language for data analysis and graphics. J Comput Graph Stat. 5:299–314.
Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol. 2:13–34.
Iyer LM, Aravind L, Koonin EV. 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J Virol. 75:11720–11734.
Jancek S, Gourbiere S, Moreau H, Piganeau G. 2008. Clues about the genetic basis of adaptation emerge from comparing the proteomes of two *Ostreococcus* ecotypes (Chlorophyta, Prasinophyceae). Mol Biol Evol. 25:2293–2300.
Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene 238: 143–155.
Kim SH, Yi SV. 2006. Correlated asymmetry of sequence and functional divergence between duplicate proteins of *Saccharomyces cerevisiae*. Mol Biol Evol. 23:1068–1075.
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.
Lynch M. 2007. The Origins of Genome Architecture. Sinauer Associates.
Marie D, Brussaard CPD, Thyrhaug R, Bratbak G, Vaulot D. 1999. Enumeration of marine viruses in culture and natural samples by flow cytometry. Appl Environ Microbiol. 65:45–52.
Maruyama S, Sugahara J, Kanai A, Nozaki H. 2010. Permuted tRNA genes in the nuclear and nucleomorph genomes of photosynthetic eukaryotes. Mol Biol Evol. 27:1070–1076.
Massana R. 2011. Eukaryotic picoplankton in surface oceans. Annu Rev Microbiol. 65:91–110.
Monier A, et al. 2011. Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. Environ Microbiol. 14:162–176.
Moreau H, et al. 2010. Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. J Virol. 84:12555–12563.
Moreau H, et al. 2012. Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. Genome Biol. 13:R74.
Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. 2006. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol. 80:9687–9696.
Novembre JA. 2002. Accounting for background nucleotide composition when measuring codon usage bias. Mol Biol Evol. 19: 1390–1394.
Parkinson H, et al. 2009. ArrayExpress update-from an archive of functional genomics experiments to the atlas of gene expression. Nucleic Acids Res. 37:D868–D872.
Percudani R, Pavesi A, Ottonello S. 1997. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. J Mol Biol. 268: 322–330.
Perriere G, Thioulouse J. 2002. Use and misuse of correspondence analysis in codon usage studies. Nucleic Acids Res. 30:4548–4555.
Piganeau G, Grimsley N, Moreau H. 2011. Genome diversity in the smallest marine photosynthetic eukaryotes. Res Microbiol. 162: 570–577.
Piganeau G, Vandepoele K, Gourbiere S, Van de Peer Y, Moreau H. 2009. Unravelling cis-regulatory elements in the genome of the smallest photosynthetic eukaryote: phylogenetic footprinting in *Ostreococcus*. J Mol Evol. 69:249–259.
Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. 2012. Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet. 8: e1002603.
Qiu S, Bergero R, Zeng K, Charlesworth D. 2011. Patterns of codon usage bias in *Silene latifolia*. Mol Biol Evol. 28:771–780.
Rocha E, Danchin A. 2002. Base composition bias might result from competition for metabolic resources. Trends Genet. 18: 291–294.
Rocha EP. 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. 14:2279–2286.
Shackelton LA, Parrish CR, Holmes EC. 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J Mol Evol. 62:551–563.
Sharp PM, Li WH. 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. Nucleic Acids Res. 14: 7737–7749.
Steele EJ. 2009. Mechanism of somatic hypermutation: critical analysis of strand biased mutation signatures at A:T and G:C base pairs. Mol Immunol. 46:305–320.
Stoletzki N, Eyre-Walker A. 2007. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. Mol Biol Evol. 24:374–381.
Sugahara J, et al. 2006. SPLITS: a new program for predicting split and intron-containing tRNA genes at the genome level. In Silico Biol. 6: 411–418.
Thioulouse J, Dray S. 2007. Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages. J Stat Softw. 22:1–14.

Tuller T, et al. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141:344–354.
van Ooijen G, Knox K, Kis K, Bouget FY, Millar AJ. 2012. Genomic transformation of the picoeukaryote *Ostreococcus tauri*. J Vis Exp. e4074.
Vaulot D, et al. 2012. Metagenomes of the picoalga *Bathycoccus* from the Chile coastal upwelling. PLoS One 7:e39648.
Viprey M, Guillou L, Ferreol M, Vaulot D. 2008. Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. Environ Microbiol. 10:1804–1822.
Weynberg KD, Allen MJ, Ashelford K, Scanlan DJ, Wilson WH. 2009. From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. Environ Microbiol. 11: 2821–2839.
Whittle CA, Sun Y, Johannesson H. 2011. Evolution of synonymous codon usage in *Neurospora tetrasperma* and *Neurospora discreta*. Genome Biol Evol. 3:332–343.
Worden AZ, et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. Science 324:268–272.
Wright F. 1990. The "effective number of codons" used in a gene. Gene 87:23–29.
Yutin N, Koonin EV. 2012. Hidden evolutionary complexity of nucleo-cytoplasmic large DNA viruses of eukaryotes. Virol J. 9:161.
Zhou T, Weems M, Wilke CO. 2009. Translationally optimal codons associate with structurally sensitive sites in proteins. Mol Biol Evol. 26: 1571–1580.

Associate editor: Purificación López-García