RESEARCH COMMENTARY

Coevolution mechanisms that adapt viruses to genetic code variations implemented in their hosts

SUSHIL KUMAR[1,2]*, RENU KUMARI[2] and VISHAKHA SHARMA[1,2]

[1]*SKA Institution for Research, Education and Development, 4/11 SarvPriya Vihar, New Delhi 110016, India*
[2]*National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi 110067, India*

[Kumar S., Kumari R. and Sharma V. 2016 Coevolution mechanisms that adapt viruses to genetic code variations implemented in their hosts. *J. Genet.* **95**, 3–12]

Introduction

Viruses, the preponderant species, are the agents of horizontal gene transfer between cellular organisms, a major means for generation of genetic variability that drives evolution in varying environments. Recent work on virus × host interaction has led to revision of the conventional idea that the genetic code of the virus and host must be same so that the host translational system facilitates efficient, accurate and complete expression of the infecting viral genome. There is evidence now that differences between the genetic codes of viruses and their hosts are not an absolute barrier to virus multiplication. The recent work on mechanisms by which viruses overcome the mismatch in codon usage of host versus theirs is discussed here contextually. Examples of coevolution of viruses and their hosts, in terms of genetic code usage, discussed here agree with the concept that their evolution is reciprocally driven and therefore suggestive of a kind of long-term interdependent symbiotic relationship between them.

Standard genetic code

Evolution in species of living organisms occurs based on the origin of genetic variability in their genomes, that arises by occurrence of mutations to generate new alleles and recombination of variant alleles in the genes already present in their genome and by interspecies sexual hybridization, asexual cell fusions and plasmid, and virus-mediated horizontal gene transfers that expand their genomes. Intraspecies and interspecies exchange of genetic information requires that the genomes of concerned species use the same genetic code. Indeed, in a large majority of prokaryotic genomes (of eubacteria, archaea and plastids and/or mitochondria in eukaryotes (protists, fungi, plants and animals)) and nuclear genomes of eukaryotes, the standard/canonical genetic code is used (Crick *et al.* 1961; Matthaei *et al.* 1962; Salas *et al.* 1965; Khorana *et al.* 1966).

Normally, the living organisms use a set of 20 amino acids to make their proteins that perform most of their cellular functions. The genetic code is a map that relates the amino acids in proteins to the codons in the genomic DNA or messenger RNA (mRNA) products of the genes in genomic DNA. The standard genetic code consists of 64 codons of different sequences of three nucleotides. To achieve it, the nucleotides of four bases: thymine (T) in DNA or its counterpart uracil (U) in RNA and cytosine (C), adenine (A) and guanine (G), are arranged in all permutations of three nucleotides. During translation of each of mRNA into respective protein (polypeptide) on ribosomes, 61 codons are interpreted as 20 amino acids (table 1). Among these sense codons, AUG has dual function besides specifying Met, it also serves as the start signal. Three codons: UAA (ochre), UAG (amber) and UGA (opal): act as translation stop signals. In the translation process on ribosomes (Selmer *et al.* 2006; Alberts 2008; Korostelev *et al.* 2008; Root-Bernstein and Root-Bernstein 2015), transfer RNAs (tRNA) that carry anticodons and specific amino acids enable the sequence of codons in mRNA to get dictated into the sequence of amino acids in the polypeptide, via the codon–anticodon pairings (Kirchner and Ignatova 2015). Organisms contain upto 31 tRNAs that can read the 61 sense codons. The stop (nonsense) codons are directly read by release factor(s) (RF-protein) that mimic tRNA function and release the translated polypeptide products from ribosomes by interacting with them. There are two RFs in prokaryotes, RF1 reads UAA and UAG and RF2 reads UAA and UGA and the only release factor eRF1 reads all three stop codons in eukaryotes (Kanaya *et al.* 1999; Bertram *et al.* 2001; Kapp and Lorsch 2004; Gingold and Pilpel 2011). The genetic code has

*For correspondence. E-mail: sushil2000_01@yahoo.co.in.

Keywords. codon usage; eukaryotes; nonstandard genetic code; phages; prokaryotes; transfer RNA; translation factors; viruses.

Table 1. The standard genetic code (codon information shown in black colour) and changes from it observed in the alternate genetic codes of nuclear genomes of certain eukaryotes (information shown in blue colour), genomes of mitochondria of many diverse eukaryote species and certain bacterial and archaeal species (information shown in red colour) and genomes of a variety of bacteriophages (information shown in green colour). The word 'Start' bearing a black line below means that these codons serve as alternate start codons for one or more genes in certain organisms using standard genetic code. The codons dispensed from genomes of certain organism are identified by a '*' below them. The * is of orange colour for codons experimentally eliminated from certain strains of *Escherichia coli* and it is of brown colour for the naturally missing codons[a].

First \ Second	U	C	A	G	Third
U	UUU phenylalanine (Phe)	UCU Serine (Ser)	UAU tyrosine (Tyr)	UGU cysteine (Cys)	U
	UUC phe	UCC Ser *	UAC Tyr	UGC Cys	C
	UUA leucine (Leu) Stop Start	UCA Ser Stop	UAA stop Gln/Gln (Ochre) Ser/Tyr	UGA stop Trp/Trp/Trp/Cys (opal) Gly/Gly	A
	UUG Leu Start/Start[b]	UCG Ser	UAG stop Gln/Gln/Leu/ (Amber) * * Ala	UGG tryptophan (Trp)	G
C	CUU Leu * Thr	CCU proline (Pro)	CAU histidine (His)	CGU arginine (Arg)	U
	CUC Leu * Thr	CCC Pro *	CAC His	CGC Arg *	C
	CUA Leu * Thr	CCA Pro	CAA glutamine (Gln)	CGA Arg *	A
	CUG Leu Ser/Thr/Start	CCG Pro	CAG Gln Start	CGG Arg *	G
A	AUU isoleucine (Ile) Start	ACU theonine (Thr)	AAU aspergine (Asn)	AGU serine (Ser)	U
	AUC Ile Start	ACC Thr *	AAC Asn	AGC Ser	C
	AUA Ile * Met/Start	ACA Thr	AAA lysine (Lys) Asn	AGA Arg Ser/Gly/Stop *	A
	AUG methionine (Met) start	ACG Thr	AAG Lys	AGG Arg Ser/Gly/Lys/Stop * *	G
G	GUU valine (Val)	GCU alanine (Ala)	GAU asparticacid (Asp)	GGU glycine (Gly)	U
	GUC Val *	GCC Ala *	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA glutamic acid (Glu)	GGA Gly	A
	GUG Val Start/Start	GCG Ala	GAG Glu	GGG Gly	G

[a]The table has been compiled based on the following references: Elzanovski and Ostell (2013) and references given in their report on behalf of National Center for Biotechnology Information (NCBI), Bethesda. O'Flaherty *et al.* (2004); Naryshkina *et al.* (2006); Trotter *et al.* (2006); Shackelton and Holmes (2008); Lajoie *et al.* (2013); Taylor *et al.* (2013); and Ivanova *et al.* (2014).

[b]The viruses of archaea, bacteria, mitochondria and plastids have been treated as phages, since mitochondria and plastids are prokaryotes that entered into stable symbiotic association with eukaryotes by their propagation as organelles in cytoplasm of eukaryotic cells.

redundancies, such that except for two of the amino acids (Met and Trp), all the others are encoded by two to six (so called synonymous) codons. Likewise, there is redundancy of tRNAs for large majority of amino acids. These redundancies are responsible for bias in the use of standard genetic code in the organisms of its implementation, leading to rare, if at all, equal use of synonymous codons.

Nonrandom use of the synonymous codons in cellular organisms

Biased or nonrandom use of synonymous codons is a species-wise character in both prokaryotic and eukaryotic organisms (Ermolaeva 2001; Angellotti *et al.* 2007). It is an outcome of the interaction of genomes with environment in the course of evolvement of the present species and their parents (Francino and Ochman 1999; Lynn *et al.* 2002; Willenbrock *et al.* 2006; Lin and Forsdyke 2007; Behura and Severson 2013). The genome × environment interaction brings about several to many characteristics of individual species to determine the biased use of the standard genetic code. They include genome size and its base composition (G + C content; Lynn *et al.* 2002; Willenbrock *et al.* 2006; Agashe *et al.* 2013; Ma *et al.* 2015), structure, activity, stability and levels of the protein and RNA products of gene expression (dos Reis *et al.* 2004; Cooper and Brown 2008; Drummond and Wilke 2008; Angov 2011; Frenkel-Morgenstern *et al.* 2012; Hu *et al.* 2013; Pechmann *et al.* 2014; Dimitrieva and Anisimova 2014; Hockenberry *et al.* 2014; Ma *et al.* 2015), additional roles of various gene exons serving as binding sites for regulatory proteins (Stergachis *et al.* 2013), relative abundance of tRNAs and differential tightness of binding between different codons and corresponding anticodon synonyms borne on tRNAs (Ikemura 1985; Dong *et al.* 1996; Kramer and Farabaugh 2007; Stoletzki and Eyre-Walker 2007; Dana and Tuller 2014; Stoecklin and

Diederichs 2014; Xia 2015), and effectiveness of release factor(s) (Tate *et al.* 1995; Freistroffer *et al.* 2000).

The idea that the codon usage bias implemented in species is optimized for environmental challenges has found some experimental support. Genetically-engineered randomization in codon usage, in the cluster of genes concerned with nitrogenase function in *Klebsiella oxytoca*, resulted in lowering the ability of bacteria, possessing this gene cluster, to fix atmospheric nitrogen (Temme *et al.* 2012). Large malleability in the use of synonymous codons, that perhaps adapts the species to varying environments has been demonstrated in *Escherichia coli*. Thirteen identified codons for seven amino acids could be replaced by one or more of their synonyms in 42 essential genes in 80 strains. Such genome restructured strains, however, had altered fitness, in the form of enlarged generation time, even under highly favourable laboratory culture conditions (Gibson *et al.* 2010; Lajoie *et al.* 2013).

While meeting the demands of genome × environment interaction, allowance for interspecies genetic exchange for sharing improved adaptability traits, is a prerequisite in the implementation of genetic code in each species. In the *FAD7* gene of seven dicot species (*Arabidopsis thaliana*, *Betula pendula*, *Camellia sinensis*, *Glycine max*, *Helianthus annus*, *Olea europea* and *Solanum lycopersicon*) and three monocot species (*Musa basjoo*, *Oryza sativa* and *Zea mays*), the genetic code usage was largely similar among dicots on one hand and among monocots on the other hand, but the codon usage was significantly different between dicots and monocots (Ma *et al.* 2015). Thus, alleles of *FAD7* could be functional even if transferred, sexually or asexually, between phylogenetically distant genera among dicots or monocots.

Genome sequences of foreign origin, received horizontally borne on plasmids or of infective viruses, that do not conform to the genetic code bias of the recipient host, will not express efficiently unless they have evolutionarily adapted to the codon usage and tRNA redundancies of the translation system of their present host. In the organisms using alternate genetic codes, certain sense codon(s) and/or stop codon(s) are interpreted differently than in the organisms employing the standard genetic code. The altered genetic codes also suffer from codon biases. Thus, the translation systems of the cellular host organisms, that use alternate genetic codes and not the standard genetic code, become a formidable barrier against the horizontally transferred genes and viruses patterned after the standard genetic code or a differential form of alternate genetic code. Genetic code alterations in a cellular organism or its virus are a means to isolate a population of the species towards further speciation.

Alternate genetic codes

The standard genetic code for protein translation has been preserved in large majority of organisms, a small number of changes in codon assignments have got selected in the nuclear genomes of eukaryotes and prokaryotes and organelle (mitochondrium and chloroplast) genomes of a small number of cellular organisms and genomes of certain viruses. The alternate genetic codes have evolved independently some dozens of times. Consequently, the organelles of some clades of various classes of eukaryotes and members in the families of certain viruses, bacteria, fungi and invertebrates have come to possess alternate genetic codes of exclusive properties. Elzanovski and Ostell (2013) have compiled information on behalf of NCBI on the codon reassignments and affected organisms with respect to 17 different types of alternate genetic codes discovered among cellular organisms. Three alternate genetic codes are identified in viruses (table 2). Among the 17 nonstandard genetic codes of cellular organisms, 13 relate to mitochondria and two each to main genomes of eukaryotes and bacteria (including archaea and plastids). Each type of alternate genetic code occurs in one or more species. The species using the same alternate genetic code are closely related (for e.g., species in the genetic code types 6, 10, 12–14, 16, 21–25 of Elzanovski and Ostell (2013) or from very distant lineages (for e.g., species of bacteria, fungi, red algae, protozoa and coelentrates bearing the alternate genetic code of type 4 of Elzanovski and Ostell (2013). In the 20 types of nonstandard genetic codes, only 1–8 (mean = 2.6) codons have undergone change in their amino acids assignment, while the remaining of the codons continue to have normally assigned function in protein translation.

Table 1 summarizes codon usage in 21 genetic codes. As compared to the standard genetic code, 20 codons (three stop codons and 17 sense codons) in the 20 alternate genetic codes, have got reassigned or dispensed with from usage. Reassignments have resulted in increase of codons for the amino acids Trp, Met, Asn, Cys, Gln, Lys and Tyr that are encoded by one or two codons in the standard genetic code. They have also increased degeneracy in codons for the amino acids such as Ala, Gly, Leu, Ser and Thr for which the standard code has four or more codons. In different alternate genetic codes, one stop codon, one codon for Leu and three codons for Arg are not used, a codon each for Leu and Ser and two codons for Arg have assumed stop codon function and six codons serve as sense as will as start codons, in addition to three codons performing such function in the standard genetic code. The stop codons of the standard genetic code have been frequent target for reassignments in the alternate genetic codes, demonstrating that all three are not essential for usage in an organism. The significance of genetic code alterations in the biology of concerned organisms remains to be elucidated.

Viruses are parasites that depend on the translational machinery present in the organelles or cytoplasm of their hosts. Viruses overcome the genetic code biases and codon usage changes by a variety of means. It is seen in table 2 that some viruses with alternate genetic codes (mycoplasma and spiroplasma phages and mitoviruses) underwent the same kind of codon reassignments as in alternate genetic code bearing cellular (host) organisms. This means that these viruses have patterned their genetic code after that of their

Table 2. Examples of virus/phage responses to the emergence of alternate genetic codes in their host. On one hand viruses/phages have evolved their genomes to harmonize with the genetic code of the host and on the other hand viruses/phages adapt to the antagonism between their and host genetic codes.

Complexion of coimplementation of the genetic codes in virus and its host	Virus/phage[a]		Site of virus gene expression in the host cell	Host		Change(s)/reassignments from the standard code		Reference(s)
	Name	Genomic characteristics		Name	Habitat	Virus/phage genome has	Concerned host genome has	
Harmonius[b]	Phage K	Double-strand linear DNA	Cytoplasm	*Staphylococcus aureus*	Human nasal passage bacterium	UUG Leucine → start GUG Valine → start	UUG → start GUG → start	O'Flaherty *et al.* (2004)
	Phage 4268	As above	As above	*Lactococcus lactis*	Cheese bacterium	As above	As above	Trotter *et al.* (2006)
	Phage Phi YS40	As above	As above	*Thermus thermophilus*	Hot spring bacterium	As above	As above	Naryshkina *et al.*(2006)
	Phage MAV1	As above	As above	*Mycoplasma arthritidis*	Bacterium of arthritic rats	UGA stop → tryptophan	UGA stop → tryptophan	Shackelton and Holmes (2008)
	Phage SpV1-R8A2B	Single-strand DNA virus	As above	*Spiroplasma citri*	Bacterium that causes stubborn disease on citrus plants	As above	As above	as above
	Phage OnuMv1a and others	Double-strand RNA	Mitochondrium	*Ophiostoma novo-ulmi*	Ascomycete fungal pathogen of Dutch-elms	As above	As above	Cole *et al.* (2000) and as above
	Helicobasidium mompa mitovirus 1–18 (phages)	Single-strand RNA	As above	*Helicobasidium mompa*	Basidiomycete fungal pathogen of plants	As above	As above	Shackelton and Holmes (2008)
	SsVL	Double-strand RNA	Cytoplasm	*Scheffersomyces segobiensis*	Used in bioethanol production	Lost all the functionally relevant CUG codons	CUG leucine → serine	Taylor *et al.* (2013)
Antagonistic/ incompatible	Phage 2	Double-strand DNA	As above	NK	NK	UAG amber stop → glutamine in late expressed genes; additionally the gene (RF2) which is expressed for the release factor 2 early and for Gln-tRNA CUA and Ser tRNA CUA	UGA opal stop → glycine; retains RF1 but has lost RF2	Ivanova *et al.* (2014)

[a]Viruses of bacteria, archaea and of eukaryote mitochondria and plastids have been called phages here; [b]this class of coevolution has been called in the text as conventional for the phages K, 4268, PhiYS40, MAV1, SPV1-R8A2B, OnuMV1a and mitoviruses and complementary for the virus SsVL; NK, not known. As above indicate the information in the above row of the same column.

hosts, the translation processes of the viruses and their hosts are not discrepant. Contranstingly, another virus (phage 2) that differs in codon usage from that of its host, disrupts the translation of host's genetic code and thereby manipulates the host in favour of its multiplication, as discussed below.

Modification of viruses to adapt to their host's genetic code

Viruses have been found to infect most of the cellular organisms that have been investigated for viral presence and it is believed that all cellular organisms may be sensitive to several to many species of virus species (say 10 virus species per cellular species on average basis) (Rohwer 2003; Koonin 2015a). It is important to note in this regard that 850 phages have been isolated for a strain of *Mycobacterium smegmatis* (Hatfull 2015). A large majority of earth's cellular species remain to be described (Suttle 2007; Costello *et al*. 2013; Brown *et al*. 2015). It is thought that although a large majority of terrestrial animal and plant species and cultivable species of terrestrial fungi, bacteria and archaea have been enumerated, however, as yet bulk of species of terrestrial microorganisms (inhabiting soil and water bodies) that are recalcitrant to cultivation (methods and media presently available) and prokaryotes and eukaryotes of all classes/groups inhabiting oceans remain to be enumerated (Rohwer and Thurber 2009; Mora *et al.* 2011; Kyrpides *et al.* 2014). Recent informed estimates suggest that earth's globe may be harbouring 8.74×10^6 of eukaryotes and at least 10^7 species of bacteria and archaea (Mora *et al.* 2011; Azvolinsky 2014). Thus, there may be more than 2×10^8 species of viruses/phages. The viromes of prokaryotes and eukaryotes comprise of +ve and −ve single strand (ss) and double strand (ds) classes of RNA viruses, ss and ds DNA viruses and reverse — transcribing viruses with +ve strand RNA genomes (Koonin *et al.* 2015a, b). The competence of viruses to parasitize all groups of cellular organisms, their diversity and the observation that viruses infect and multiply in certain cellular organisms that have undergone codon usage changes indicate that viruses are successful in surmounting the challenges posed to them by the genetic code biases and these together with alternate codes adopted by their hosts (tables 2 and 3).

Adaptation of viruses to the nonrandom codon usages in their hosts

Comparative analyses of the host range, life cycle, virion structure and genome, transcriptome and proteome of viruses belonging to various taxonomical classes have revealed that viruses have adopted in the main two routes for achieving translation efficiency of their RNA messages in the translational environments of varying codon usage of their hosts. (i) A large majority of nondsDNA viruses and some dsDNA viruses comprise a group which under the codon-selective pressure exerted by their hosts, have evolved the codon usage bias much like that of their hosts. This correspondence is much more for the viral proteins expressed in large quantities, such as virion structural proteins than for viral proteins expressed in small quantities, such as those concerned with infection specificity (Carbone 2008; Lucks *et al.* 2008; Bahir *et al.* 2009). Apparently, in the viruses of this groups, the codon usage has become correlated with tRNA pool composition in their hosts. (ii) The second group consists of dsDNA phages and viruses of large genomes (five examples are shown in table 3). The genomes of this group of viruses carry genes for tRNAs or tRNAs and amino-acyl-tRNA synthetases (aaRs). The virus tRNAs or tRNAs and aaRs complement the translation apparatus of its host such that requirements of the virus are met in using certain codons that are rarely used in its host (Bailly-Bechet *et al.* 2007; Dreher 2010). It is expected that coevolution (reciprocally-driven evolution, Ehrlich and Raven 1984) of virus and host would harmonize their codon usage, using the host translation apparatus. Presence of tRNA genes in genomes of some viruses indicates that the coevolution of virus and its host has not yet been completed. The virus in the interim has acquired and retained some tRNA genes from one of its hosts to meet the requirement of translating mRNAs from all its essential genes. This explanation requires to be tested by comparison of viruses of homologous properties.

Adaptation of viruses to hosts using nonstandard genetic codes

There are two views about the host × virus relationship. Since large majority of host infections are lytic, viruses are considered parasitic. Viruses transfer new genetic information and thereby promote ecological adaptation and evolution in their hosts. This property suggests host × virus interactions are symbiotic. The genetic code alterations in cellular organisms, a means of increasing their biodiversity, preclude infection by viruses that use the standard genetic code. Indeed, the range of viruses infecting the cellular organisms of altered genetic code has been observed to be small. A question has arisen whether genetic code alterations occur to resist virus parasitism. The observations that most viruses and hosts coexist and that cellular organisms with altered genetic code coevolve with viruses suggest that host × virus relationships are symbiotic and their coevolution is evolutionarily mutually beneficial. Discussed below are examples of three routes by which viruses adapt/coevolve in response to the use of nonstandard genetic code in their hosts. One of these provides strong support to the suggestion that the nature of host × virus relationships is symbiotic.

Conventional route of virus adaptation to genetic code alteration in host

Three examples of harmonius correspondence of virus and host genetic codes are shown in table 2 (rows 1–7). The bacterial species *Staphylococcus aureus*, *Lactococcus lactis* and

Table 3. Kinds of protein translation associated genes found in phages and viruses that complement the host translational system for viral gene expression, exemplified by a temperate phage and a nontemperate phage and an animal virus and two plant viruses.

		Bacteriophage		Virus		
Subject	Characteristic	D3	T4	OtV_5	PBCV1	Mimivirus APMV1
General						
	Nature of life cycle	Temperate	Nontemperate	Lytic	Lytic	Lytic
	Host	*Pseudomonas aeruginosa*	*Escherichia coli*	*Ostreococcus tauri*	*Chlorella variabilis*	*Acanthamoeba polyphage*
	Family	Sophoviridae	Myovirideae	Phycodnaviridae	Phycodnaviridae	Mimiviridae
	Genome size (bp)	56426	169000	186234	330740	1181404
Number of different translation related genes of varying functions carried in the virus genome						
	Transfer RNA (tRNA)	4	8	5	11	6
	Aminoacyl tRNA synthetase (aaRs)	0	0	0	0	4
	Translation elongation factor	0	0	0	1[a]	1[a]
	Translation release factor	0	0	0	0	1[a]
	Whether the above genes are expressed and function during viral growth cycle	0	0	0	Yes	Yes
	Whether they partially complement the host translational system for viral multiplication	Yes	Yes	Yes	Yes	Yes
	References	Kropinski (2000); Bailly-Bechet *et al.* (2007); Lucks *et al.* (2008).	Miller *et al.* (2003); Bailly-Bechet *et al.* (2007); Lucks *et al.* (2008); Dreher (2010).	Derelle *et al.* (2008); Michely *et al.* (2013).	Van Etten (2003); Da-Young *et al.* (2005); Yamada *et al.* (2006); Yanai-Balser *et al.* (2010).	Raoult *et al.* (2004); Suzan-Monti *et al.* (2006); Abergel *et al.* (2007); Jeudy *et al.* (2012); Colson *et al.* (2013); Abrahao *et al.* (2014).

[a]The roles of these virus-specified genes in the viral RNA translation process remains to be understood.

Thermus thermophilus use the standard genetic code except that, along with the standard AUG start codon, the GUG and UUG are also used by them as the nonstandard translation initiation codons. Phages of these bacteria, such as K (*S. aureus*), 4268 (*L. lactis*) and PhiYS40 (*T. thermophilus*) have also evolved to use GUG and UUG as the translation initiation sites in their open reading frames (ORF). The phages MAV1 and SpV1-RSA2B like their respective hosts *Mycoplasma arthritidis* and *Spiroplasma citri* use the opal translation terminator codon UGA of the standard genetic code as a reassigned codon for the Trp amino acid. Such harmonius genetic code coevolution has also occurred between the fungal mitochondria of ascomycete fungus *Ophiostoma novo-ulmi* and basidiomycete fungus *Helicobasidium mompa* and their respective phages OnuMv1a and Mitovirus1 and related phages are using the opal stop codon UGA as a codon for Trp.

Complementary type of host–virus coevolution

A unique host × virus genetic code coevolution is exemplified by the yeasts *Scheffersomyces segobiensis* and *S. stiptis* and the virus SsVL (table 2, row 8). In the yeasts, the CUG codon has been reassigned from Leu to Ser. In their totivirus species SsVL, all the CUG codons have been lost, except one in its gene for the capsid protein. The virus is unable to produce its own capsid protein. Interestingly, the genome of the yeast(s) has four tandem copies of totivirus capsid-like protein. The complementarity of the host in providing a functional protein to the virus which has adopted its genetic code signifies the importance of virus species over host species. This example is considered as strong evidence in favour of the suggestion that virus infections on their hosts, though apparently parasitic, are really symbiotic

Incompatible host and virus genetic codes favouring viral growth

Ivanova *et al.* (2014), in their study on stop codon reassignments in bacteria and phages of human oral cavity, came across T4 phage-like dsDNA caudovirale phages, exemplified by the phage 2, in whose genomes the amber UAG stop codon is reassigned for the amino acid Gln. These phages carried in their genome, the gene for nonstandard Gln $tRNA_{CUA}$ required for translation of reassigned amber codons and the gene for peptide chain release factor 2 (RF-2) which is required for termination of translation at opal UGA codons. Bacteria such as *Prevotella* of the human oral cavity did not carry amber stop codon reassignment like in phage 2 but were opal stop codon recoded and carried on their genome, the gene for peptide chain release factor 1 (RF-1) which terminates transcription at amber codons. Thus, the antagonism between the genetic codes of phage 2 and its putative host(s) was discovered. The mechanism which resolves the antagonism in favour of viral multiplication,

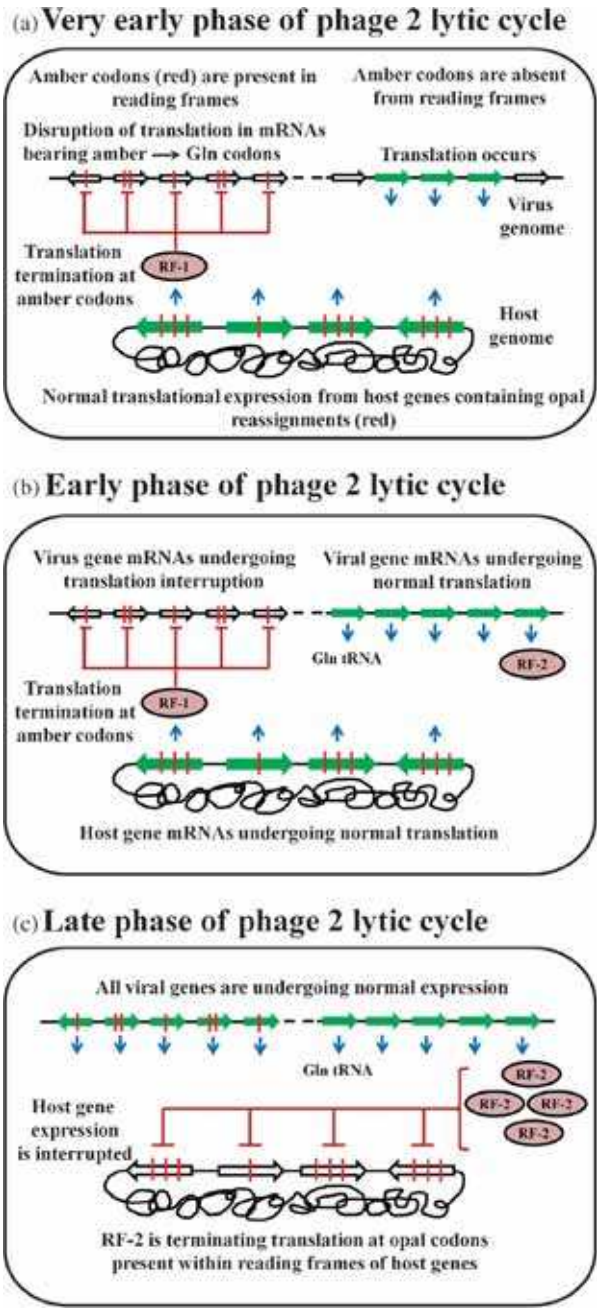

Figure 1. Diagrammatic representation of genetic interactions among phage 2, in which amber stop codon is reassigned for glutamine, and its host, in which opal stop codon is reassigned for glycine, leading to successful burst of phage progeny. (a) Very early phase in phage 2 life cycle: host expresses its genes, including that for the translation restriction factor 1 (RF-1) which terminates translation at amber codons. Early phage genes, that do not contain amber codons in reading frames, get expressed. But late phage genes, that contain amber codons in reading frames, do not express on account of the host encoded RF-1 action at the amber sites. (b) Mid-early phase of viral life cycle: viral gene for the translation restriction factor 2 (RF-2) begins to express. (c) Late phases of viral lytic cycle: phage produced RF-2 interrupts expression of host genes containing opal codon in their reading frames. Expression of RF-1 now does not occur. Consequently, all the genes of phage can express. Thus, despite the conflict between the genetic codes of host and phage, phage 2 is able to complete its life cycle.

became understood by further analysis of phage genome organization and the expression is shown in figure 1 and described below.

Like the model phage T_4, phage 2 has genes that in its lytic cycle are expressed early (genes concerning phage DNA replication and transcription) and others that are expressed late (genes concerning phage virion structure and host cell lysis). The very early expressed genes of phage 2 do not have amber codons in their reading frames. Therefore, expression of early phage 2 genes is not interrupted by the host-encoded RF-1. Phage 2 synthesizes Gln tRNA and RF-2 early, of which the latter blocks the synthesis of host proteins whose genes carry opal codons in their reading frames. Synthesis of RF-1 gets blocked and phage 2 late genes are expressed. Phage 2 is able to complete its lytic cycle, although its genetic code is different from the genetic code implemented in its host.

It appears that viruses can use their genetic code alterations to grow on hosts that implement genetic code like their own or of different kind(s). Viruses evolve genetic code modifications to counter any host genetic code changes which are inimical for their propagation.

Extension of host range in viruses possessing components of translation apparatus in their genomes

Another route to broader host range has been noted in mimivirus and megaviruses. In table 3, APMV mimivirus is shown to carry gene for translation-release factor. This gene also present in megaviridae viruses encodes a eukaryotic type of peptide chain-release factor. Certain eukaryotes, such as ciliates and others, are known to alter their genetic code such that the stop codon UGA is read as a sense codon (Salas-Marco *et al.* 2006; Cheng *et al.* 2009; Jeudy *et al.* 2012). It is likely that viruses possessing eRF1 in their genome can use such organisms possessing altered genetic code as their additional hosts.

Concluding remarks

Viruses, predominate living organisms, inhabiting terrestrial and oceanic environments of earth. Perhaps all organisms are infected by specific viruses; 850 viruses are known to infect a strain of *Mycobacterium smegmatis*. Functions in living organisms are performed by proteins, the information for which is encoded in genes borne on genomes of organisms. The main genomes of a large majority of prokaryotes (eubacteria and archea) and eukaryotes (fungi, animals and plants) and genomes of organelle (mitochondria and chloroplasts of eukaryotes) use the standard genetic code, in their protein-encoding genes. Among the 64 triplet codons of four nucleotides, abbreviated as A, T/U, G and C, 61 codons encode 20 amino acids and three function as stops in protein translation. All but two amino acids are encoded by 2–6 different codons, leading to degeneracy. There is species-specific codon usage preference. Genomes in some species have undergone alterations from the standard genetic code and their translation system is adapted to the genetic code change. Twenty altered genetic codes have been discovered. Viruses use the host translation apparatus to express their genes. Organelles and cells using nonstandard genetic codes become resistant to viruses using standard genetic code. Viruses in order to successfully propagate in their hosts must coevolve with hosts to match codon usage bias and genetic code implemented in hosts.

Viruses have taken two routes to meet their codon usage preference. Mostly, their coevolution with host have resulted in concordance of codon usage. Some viruses have recruited tRNA genes into their genomes to complement the translation apparatus of the host such that codons deficiently used in host but excessively used in virus that can be translated.

To be able to grow in hosts with altered genetic codes, mainly, viruses have evolved their genetic code to confirm with that of host. Where they happened to evolve a genetic code different from that of host's altered code, they have recruited tRNA gene(s) and translation release factor gene to overcome the incompatibility in their favour. In one instance, host is complementing an essential function in trans to virus that had incompletely adopted the genetic code of host. This is an example of gene flow from viruses to their hosts. Viruses acquire genetic information from their various hosts, and from each other, via nonhomologous recombination and thereby impact host evolution. Bacteria and archaea harbour phage genomes as episomes and eukaryotic genomes to the extent of 25–90% consists of various kinds of transposable elements derived from viruses (Katzourakis and Gifford 2010; Makałowski *et al.* 2012; Ayarpadikannan and Kim 2014). These features and virus × host adaptations altogether allow an inference that though their immediate relationship appears to be that of virus as parasite on its host, host × parasite relation is of coevolution and therefore, the true nature of host × parasite relationships in general can be considered as symbiotic.

Acknowledgements

Grateful thanks are due to Indian National Science Academy for the grant of an honorary-scientistship to SK and to Director of National Institute of Plant Genome Research for facilities.

References

Abergel C., Rudinger-Thirion J., Giege R. and Claverie J. M. 2007 Virus-encoded aminoacyl-tRNA synthetases: structural and functional characterization of mimivirus TyrRS and MetRS. *J. Virol.* **81**, 12406–12417.

Abrahao J. S., Dornas F. P., Silva L. C. F., Almeida G. M., Boratto P. V. M., Colson P. *et al.* 2014 *Acanthamoeba polyphaga* mimivirus and other giant viruses: an open field to outstanding discoveries. *Virol J.* **11**, 120.

Agashe D., Martinez-Gomez N. C., Drummond D. A. and Marx C. J. 2013 Good codons, bad transcript: large reductions in gene

expression and fitness arising from synonymous mutations in a key enzyme. *Mol. Biol. Evol.* **30**, 549–560.

Alberts B. 2008 *Molecular biology of the cell*. Garland Science Publishers, New York, USA.

Angellotti M. C., Bhuiyan S. B., Chen G. and Wan X. F. 2007 Codon O: codon usage bias analysis within and across genomes. *Nucleic Acids Res.* **35** (web server issue): W132–W136.

Angov E. 2011 Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol. J.* **6**, 650–659.

Ayarpadikannan S. and Kim H.-S. 2014 The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. *Genomics Inform.* **12**, 98–104.

Azvolinsky A 2014 Sequencing the tree of life. http://www.the-scientist.com/?articles.view/articleNo/39742/title/Sequencing-the-Tree-of-Life/.

Bahir I., Fromer M., Prat Y. and 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. and Rocha E. 2007 Causes for the intriguing presence of tRNAs in phages. *Genome Res.* **17**, 1486–1495.

Behura S. K. and Severson D. W. 2013 Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol. Rev.* **88**, 49–61.

Bertram G., Innes S., Minella O., Richardson J. and Stansfield I. 2001 Endless possibilities: translation termination and stop codon recognition. *Microbiology* **147**, 255–269.

Brown C. T., Hug L. A., Thomas B. C., Sharon I., Castelle C. J., Singh A. *et al.* 2015 Unusual biology across a group comprising more than 15% of domain bacteria. *Nature* **523**, 208–211.

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

Cheng Z., Saito K., Pisarev A. V., Wada M., Pisareva V. P., Pestova T. V. *et al.* 2009 Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev.* **23**, 1106–1118.

Cole T. E., Hong Y. G., Brasier C. M. and Buck K.W. 2000 Detection of an RNA-dependent RNA polymerase in mitochondria from a mitovirus-infected isolate of the Dutch elm disease fungus, *Ophiostoma novo-ulmi*. *Virology* **268**, 239–243.

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

Cooper G. M. and Brown C. D. 2008 Qualifying the relationship between sequence conservation and molecular function. *Genome Res.* **18**, 201–205.

Costello M. J., May R. M. and Stork N. E. 2013 Can we name earth's species before they go extinct? *Science* **339**, 413–416.

Crick F. H., Barnett L., Brenner S. and Watts-Tobin R. J. 1961 General nature of the genetic code for proteins. *Nature* **192**, 1227–1232.

Dana A. and Tuller T. 2014 The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* **42**, 9171–9181.

Da-Young L., Michael V. G., Van Etten J. L. and Tae-Jin C. 2005 Functional implication of the tRNA genes encoded in the chlorella virus PBCV-l genome. *Plant Pathol. J.* **21** (doi: 10.5423/PPJ.2005.21.4.334).

Derelle E., Feraz C., Escande M. L., Eychenie S., Cooke R., Piganeau G. *et al.* 2008 Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green algae, *Ostreococcus tauri*. *PLoS One* **3**, e2250.

Dimitrieva S. and Anisimova M. 2014 Unraveling patterns of site-to-site synonymous rates variation and associated gene properties of protein domains and families. *PLoS One* **9**, e95034.

Dong H., Nilsson L. and Kurland C. G. F. 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. and Wernisch L. 2004 Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036–5044.

Dreher T. W. 2010 Viral tRNAs and tRNA-like structures. *Wiley Interdiscip. Rev. RNA* **1**, 402–414.

Drummond D. A. and Wilke C. O. 2008 Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352.

Ehrlich P. R. and Raven P. H. 1984 Butterflies and plants: a study in coevolution. *Evolution* **18**, 586–608.

Elzanovski A and Ostell J 2013 Genetic codes. (http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi*)*.

Ermolaeva M. D. 2001 Synonymous codon usage in bacteria. *Mol. Biol.* **3**, 91–97.

Francino M. P. and Ochman H. 1999 Isochores result from mutation not selection. *Nature* **400**, 30–31.

Freistroffer D. V., Kwiatkowski M., Buckingham R. H. and Ehrenberg M. 2000 The accuracy of codon recognition by polypeptide release factors. *Proc. Natl. Acad. Sci. USA* **97**, 2046–2051.

Frenkel-Morgenstern M., Danon T., Christian T., Igarashi T., Cohen L., Hou Y. M. and Jensen J. L. 2012 Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. *Mol. Syst. Biol.* **8**, 572.

Gibson D. G., Glass J. I., Lartigue C., Noskov V. N., Chuang R. Y., Algire M. A. *et al.* 2010 Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56.

Gingold H. and Pilpel Y. 2011 Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* **7**, 481.

Hatfull G. F. 2015 Dark matter of the biosphere: the amazing world of bacteriophage diversity. *J. Virol.* JV1–01340.

Hockenberry A. J., Sirer M. I., Amaral L. A. N. and Jewett M. C. 2014 Quantifying position-dependent codon usage bias. *Mol. Biol. Evol.* (doi: 10.1093/molbev/msu126).

Hu S., Wang M., Cai G. and He M. 2013 Genetic code-guided protein synthesis and folding in *Escherichia coli*. *J. Biol. Chem.* **288**, 30855–30861.

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

Ivanova N. N., Schwientek P., Tripp H. J., Rinke C., Pati A., Huntemann M. *et al.* 2014 Stop codon reassignments in the wild. *Science* **344**, 909–913.

Jeudy S., Abergel C., Claverie J. M. and Legendre M. 2012 Translation in giant viruses: a unique mixture of bacterial and eukaryotic termination schemes. *PLoS Genet.* **8**, e1003122.

Kanaya S., Yamada Y., Kudo Y. and 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.

Kapp L. D. and Lorsch J. R. 2004 The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem.* **73**, 657–704.

Katzourakis A. and Gifford R. J. 2010 Endogenous viral elements in animal genomes. *PLoS Genet.* **6**, e1001191.

Khorana H. G., Buchi H., Ghosh H., Gupta N., Jacob T. M., Kossel H. *et al.* 1966 Polynucleotide synthesis and the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* **31**, 39–49.

Kirchner S. and Ignatova Z. 2015 Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* **16**, 98–112.

Koonin E. V., Dolja V. V. and Krupovic M. 2015a Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* **479–480C**, 2–25.

Koonin E. V., Krupovic M. and Yutin N. 2015b Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. *Ann. N. Y. Acad. Sci.* **1341**, 10–24.

Korostelev A., Ermolenko D. N. and Noller H. F. 2008 Structural dynamics of the ribosome. *Curr. Opin. Chem. Biol.* **12**, 674–683.

Kramer E. B. and Farabaugh P. J. 2007 The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* **13**, 87–96.

Kropinski A. M. 2000 Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *J. Bacteriol.* **182**, 6066–6074.

Kyrpides N. C., Hugenholtz P., Eisen J. A., Woyke T., Goker M. *et al.* 2014 Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS Biol.* **12**, e1001920.

Lajoie M. J., Kosuri S., Mosberg J. A., Gregg C. J., Zhang D. and Church G. M. 2013 Probing the limits of genetic recoding in essential genes. *Science* **342**, 361–363.

Lin F. H. and Forsdyke D. R. 2007 Prokaryotes that grow optimally in acid have purine-poor codons in long open reading frames. *Extremophiles* **11**, 9–18.

Lucks J. B., Nelson D. R., Kudla G. R. and Plotkin J. B. 2008 Genome landscapes and bacteriophage codon usage. *PLoS Comp. Biol.* **42**, e1000001.

Lynn D. J., Singer G. A. and Hickey D. A. 2002 Synonymous codon usage is subject to selection in thermophilic bacteria. *Nucleic Acids Res.* **30**, 4272–4277.

Ma X.-X., Feng Y.-P., Bai J.-L., Zhang D.-R., Lin X.-S. and Ma Z.-R. 2015 Nucleotide composition bias and codon usage trends of gene. *J. Genet.* **94**, 251–260.

Makałowski W., Pande A., Gotea V. and Makałowska I. 2012 Transposable elements and their identification. *Methods Mol. Biol.* **855**, 337–359.

Matthaei J. H., Jones O. W., Martin R. G. and Nirenberg M. W. 1962 Characteristics and composition of RNA coding units. *Proc. Natl. Acad. Sci. USA* **48**, 666–677.

Michely S., Toulza E., Subirana L., John U., Cognat V., Marechal-Drouard L. *et al.* 2013 Evolution of codon usage in the smallest photosynthetic eukaryotes and their giant viruses. *Genome Biol. Evol.* **5**, 848–859.

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

Mora C., Tittensor D. P., Adl S., Simpson A. G. and Worm B. 2011 How many species are there on Earth and in the ocean? *PLoS Biol.* **9**, e1001127.

Naryshkina T., Liu J., Florens L., Swanson S. K., Pavlov A. R. *et al.* 2006 *Thermus thermophilus* bacteriophage phiYS40 genome and proteomic characterization of virions. *J. Mol. Biol.* **364**, 667–677.

O'Flaherty S., Coffey A., Edwards R., Meaney W., Fitzgerald G. F. and Ross R. P. 2004 Genome of staphylococcal phage K: a new lineage of *Myoviridae* infecting gram-positive bacteria with a low G+C content. *J. Bacteriol.* **186**, 2862–2871.

Pechmann S., Chartron J. W. and Frydman J. 2014 Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP *in vivo*. *Nat. Struct. Mol. Biol.* **21**, 1100–1105.

Raoult D., Audic S., Robert C., Abergel C., Renesto P., Ogata H. *et al.* 2004 The 1.2-megabase genome sequence of mimivirus. *Science* **306**, 1344–1350.

Rohwer F. 2003 Global phage diversity. *Cell* **113**, 141.

Rohwer F. and Thurber R. V. 2009 Viruses manipulate the marine environment. *Nature* **459**, 207–212.

Root-Bernstein M. and Root-Bernstein R. 2015 The ribosome as a missing link in the evolution of life. *J. Theor. Biol.* **367**, 130–158.

Salas M., Smith M. A., Stanley W. M., Wahba A. J. and Ochoa S. 1965 Direction of reading of the genetic message. *J. Biol. Chem.* **240**, 3988–3995.

Salas-Marco J., Fan-Minogue H., Kallmeyer A. K., Klobutcher L. A., Farabaugh P. J. and Bedwell D. M. 2006 Distinct paths to stop codon reassignment by the variant-code organisms Tetrahymena and Euplotes. *Mol. Cell Biol.* **26**, 438–447.

Selmer M., Dunham C. M., Murphy F. V., Weixlbaumer A., Petry S., Kelley A. C. *et al.* 2006 Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942.

Shackelton L. A. and Holmes E. C. 2008 The role of alternative genetic codes in viral evolution and emergence. *J. Theor. Biol.* **254**, 128–134.

Stergachis A. B., Haugen E., Shafer A., Fu W., Vernot B., Reynolds A. *et al.* 2013 Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* **342**, 1367–1372.

Stoecklin G. and Diederichs S. 2014 tRNAs: new tricks from old dogs. *EMBO J.* **33**, 1981–1983.

Stoletzki N. and Eyre-Walker A. 2007 Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol. Biol. Evol.* **24**, 374–381.

Suttle C. A. 2007 Marine viruses major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812.

Suzan-Monti M., La Scola B. and Raoult D. 2006 Genomic and evolutionary aspects of mimivirus. *Virus Res.* **117**, 145–155.

Tate W. P., Poole E. S., Horsfield J. A., Mannering S. A., Brown C. M., Moffat J. G. *et al.* 1995 Translational termination efficiency in both bacteria and mammals is regulated by the base following the stop codon. *Biochem. Cell Biol.* **73**, 1095–1103.

Taylor D. J., Ballinger M. J., Bowman S. M. and Bruenn J. 2013 Virus–host co-evolution under a modified nuclear genetic code. *Peer J.* **1**, e50.

Temme K., Zhao D. and Voigt C. A. 2012 Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc. Natl. Acad. Sci. USA* **109**, 7085–7090.

Trotter M., McAuliffe O., Callanan M. *et al.* 2006 Genome analysis of the obligately lytic bacteriophage 4268 of *Lactococcus lactis* provides insight into its adaptable nature. *Gene* **366**, 189–199.

Van Etten J. L. 2003 Unusual life style of giant chlorella viruses. *Annu. Rev. Genet.* **37**, 153–195.

Willenbrock H., Friis C., Juncker A. S. and Ussery D. W. 2006 An environmental signature for 323 microbial genomes based on codon adaptation indices. *Genome Biol.* **7**, R114.

Xia X. 2015 A major controversy in codon–anticodon adaptation resolved by a new codon usage index. *Genetics* **199**, 573–579.

Yamada T., Onimatsu H. and Van Etten J. L. 2006 Chlorella viruses. *Adv. Virus Res.* **66**, 293–336.

Yanai-Balser G. M., Duncan G. A., Eudy J. D., Wang D., Li X., Agarkova I. V. *et al.* 2010 Microarray analysis of *Paramecium bursaria* chlorella virus 1 transcription. *J. Virol.* **84**, 532–542.

Received 14 July 2015; accepted 4 August 2015
Unedited version published online: 7 August 2015
Final version published online: 29 February 2016