Single-stranded genomic architecture constrains optimal codon usage

Daniel J. Cardinale and Siobain Duffy*

Department of Ecology, Evolution and Natural Resources; School of Environmental and Biological Sciences; Rutgers; The State University of New Jersey; New Brunswick, NJ USA

Viral codon usage is shaped by the conflicting forces of mutational pressure and selection to match host patterns for optimal expression. We examined whether genomic architecture (single- or double-stranded DNA) influences the degree to which bacteriophage codon usage differ from their primary bacterial hosts and each other. While both correlated equally with their hosts' genomic nucleotide content, the coat genes of ssDNA phages were less well adapted than those of dsDNA phages to their hosts' codon usage profiles due to their preference for codons ending in thymine. No specific biases were detected in dsDNA phage genomes. In all nine of ten cases of codon redundancy in which a specific codon was overrepresented, ssDNA phages favored the NNT codon. A cytosine to thymine biased mutational pressure working in conjunction with strong selection against non-synonymous mutations appears be shaping codon usage bias in ssDNA viral genomes.

Introduction

Viruses usually exhibit genomic signatures that closely mimic those of their primary hosts,[1,2] in part to better evade innate and acquired immune responses.[3,4] However, the majority of the close adherence to host nucleotide usage is attributed to selection for improved translational speed and efficiency, which are correlates of viral fitness. Synonymous codons are used at different frequencies in virtually all organisms,[5,6] and the most frequently used codons correlate with the most abundant tRNAs within a cell.[7,8] These favored synonymous codons are therefore recognized[9] and translated[10-12] more rapidly. The most frequently expressed cellular genes within a given organism exhibit similar patterns of this codon usage bias (CUB) and are more biased than less frequently expressed genes.[11,13-16] For viruses, these factors should contribute to increased rate of replication when strictly adhering to host CUB. Therefore many viruses have been under selective pressure to match the CUB of their preferred hosts.[17] Despite increased attention to the genomic match between viruses and their hosts, there have been few studies examining how different viral genomic architectures facilitate or hinder adaptation to their hosts' genomes.

Phages are the optimal system in which to explore how genomic architecture affects viral molecular evolution. The codon bias expressed in prokaryotic hosts is constant for each host cell, unlike multi-cellular organisms, in which codon usage profiles are affected by tissue-specific gene expression.[18] Perhaps

due to this, phage are more strongly adapted to their primary hosts' CUB than eukaryotic viruses,[19] allowing the greatest potential to identify factors that diminish the match between virus and host genomes. Bacterial hosts also offer a wider range of genomic nucleotide content to examine compared with plant or mammalian hosts, and their CUB have been well-documented. Additionally, while phage host ranges are far from perfectly annotated, bacteriophage host ranges are usually quite narrow[20] and many of their host ranges have been better delineated than eukaryotic viruses, such as phytopathogens.[19]

Two distinct phage genomic architectures (single-stranded DNA, ssDNA and double-stranded DNA, dsDNA) have been amply sequenced; unfortunately, the small number of sequenced RNA phages precludes their close examination at this time. The two DNA-based architectures are subject to specific constraints: dsDNA phages can house the largest genomes, up to ~300 kb,[21,22] whereas even the largest ssDNA phages are smaller than 10 kb.[23] Many dsDNA phages encode their own tRNAs, (e.g., T4 encodes eight[24]), decreasing selection for adherence to host CUB, whereas none have been found in ssDNA phages. dsDNA phages have the lowest mutation rates among viruses, while ssDNA phage mutation rates are faster, approaching those of a dsRNA phage.[25,26] Eukaryotic viruses with the same ssDNA genomic architecture exhibit evolutionary rates orders of magnitude above those seen in eukaryotic dsDNA viruses.[27] Consequently, faster-evolving ssDNA phages might be better able to adapt to host-imposed genomic conditions. Conversely, the mutation frequency in
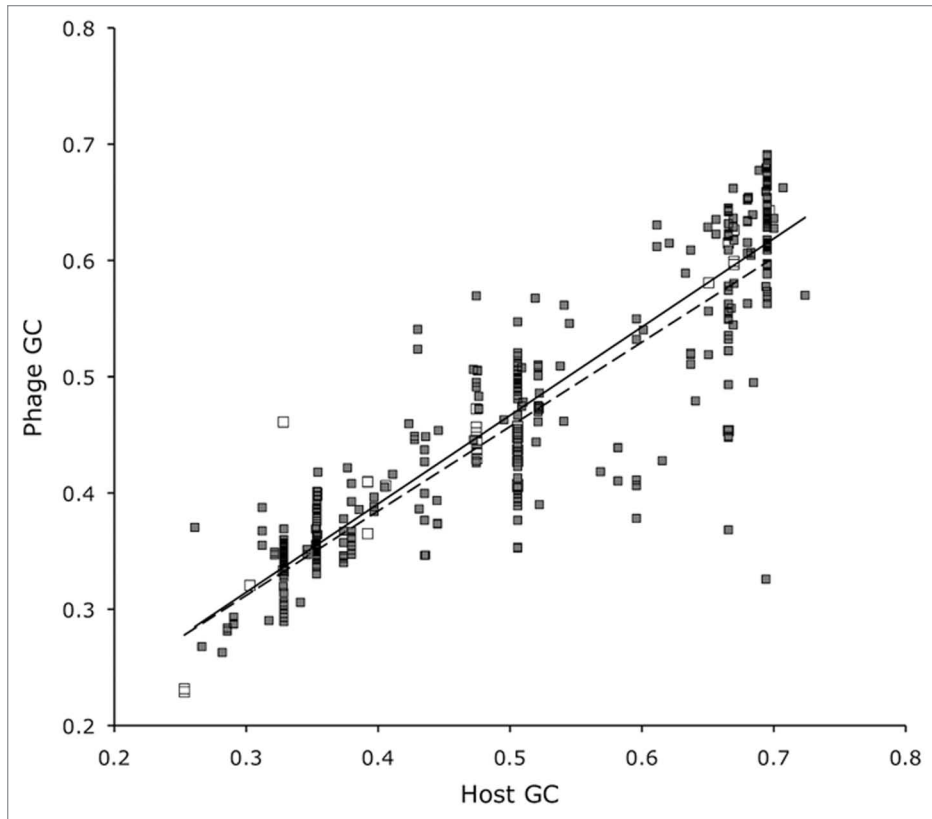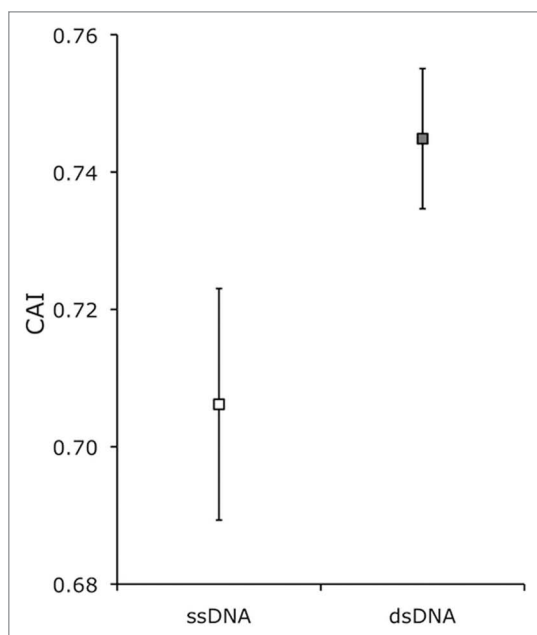
Figure 1. Correlation between host and phage genomic GC content. Grey squares indicate dsDNA phages, open squares ssDNA. Best-fit linear regression lines are solid for dsDNA (r^2 = 0.84) and dashed for ssDNA (r^2 = 0.82). There was no significant difference between the correlations (p = 0.72).

ssDNA phages may diminish their ability to conform to their host codon preferences.

Genomic GC content is a rough predictor of CUB, and many viruses match the GC content of their hosts.[28-32] Bacteriophage GC content, in particular, correlates strongly to that of their primary bacterial hosts.[33] We measured the similarity in GC content between each ssDNA and dsDNA GenBank phage reference genome and that of its primary host. We used the most numerous group of phages with a common host, *Escherichia coli*, to compare codon adaptation indices (CAI) and relative synonymous codon usage (RSCU) for a subset of highly expressed genes from dsDNA and ssDNA coliphages.[1] Our results show that genomic architecture correlates to statistically significant differences in nucleotide content and codon usage between ssDNA and dsDNA phages, and point to an enrichment of thymine as a cause.

Results and Discussion

GC content in ssDNA and dsDNA phages was highly correlated with host GC content (r^2 = 0.82 for ssDNA phages, 0.84 for dsDNA phages, equally correlated p = 0.72) across a very wide range of host GC content (~0.25 to ~0.72) (**Fig. 1**). A previous study found significant differences between ssDNA and dsDNA phage nucleotide correlation with their hosts,[33] but the additional 333 dsDNA and 13 ssDNA reference sequences added to GenBank since that analysis suggest there is no difference (**Table S1 and Fig. S1**). ssDNA phages exhibited a pronounced genomic thymine bias (average 0.30 T), but nonetheless infected hosts with a range of GC contents (0.25 to 0.70), as wide as that of dsDNA phages (0.26 to 0.72).

Correlated GC content was a poor predictor of strong CAI match between *E. coli* and the coat genes of its phages. The mean CAI of ssDNA coliphages was 0.706, while the dsDNA phages were significantly better matched to *E. coli* (0.744, p < 0.001, **Fig. 2**). This number includes eight dsDNA coliphage genomes for which tail protein encoding genes were used, rather that coat protein encoding genes, due to the absence of properly annotated coat genes. The inclusion of tail genes did not change the results of this analysis (p < 0.001 with and without the eight tail genes). The evidence of selection for translational efficiency is stronger for dsDNA phages.

Comparison of the GC content of the first two positions of each codon (GC1,2) and the third position (GC3) of these genes revealed an interesting pattern: for both ssDNA and dsDNA


Figure 2. Mean coat gene CAI with 95% confidence intervals of ssDNA (n = 11), dsDNA (n = 34) coliphages.

coliphages, the GC1,2 was restricted to a tight range between about 0.45 and 0.55. dsDNA GC3 varied along a wide range, from 0.26 to 0.69, but ssDNA GC3 occupied a narrower range, from 0.30 to 0.54 (**Fig. 3**). Furthermore, when plotted with a line representing a perfect correlation between GC1,2 and GC3, all but one of the ssDNA phages fell to the left of that line (**Fig. 3**), indicating a paucity of GC in the third codon position of their coat genes. Conversely, the dsDNA coat genes were GC3-rich or GC3-poor in approximately equal numbers. Past studies have indicated that strong mutational biases often occur with low levels of CUB,[34-36] possibly because a strong, non-specific mutational pressure would prevent any persistent, directional changes in the genome. The consistently lower GC3 content of the ssDNA genes suggests that a specific mutational pressure might be reducing GC3 content in a directional manner, which is disrupting the effects of selection for translational efficiency.

We further investigated the GC3-poor nature of ssDNA coliphage coat proteins with RSCU analysis. It revealed statistically significant variation in use for 15 of 59 codons between ssDNA and dsDNA phage ($p < 0.03$ for TTG, $p < 0.002$ for CTT and TCC, $p < 0.001$ for all other codons, **Fig. 4**). Notably, for four of the five codons more frequently used by ssDNA rather than dsDNA coliphages, thymine was in the third position. No codons enriched in dsDNA phage relative to ssDNA phage contained thymine in the third positions.

Calculation of RSCUs of coat genes in 28 ssDNA phages with a diverse host range confirmed this pattern: codons with thymine in the third position were extremely overrepresented ($p < 0.001$) for six amino acids (A, D, G, I, T, V), and were significantly favored ($p < 0.012$) in three more (H, P, S) (**Fig. 5**). Only one of the remaining nine degenerate amino acids had a statistically preferred codon in ssDNA phages (GAA for E, $p < 0.01$).

We subdivided our data set to separately examine the two morphologically distinct families of ssDNA phages, the *Inoviridae* and the *Microviridae*. Because inoviruses are frequently vertically transmitted and can productively infect their hosts without causing lysis, they might be under increased selective pressure to match the genomes of their more permanently associated hosts. RSCU comparisons revealed no consistent patterns associated with phage lifestyle. No difference in RSCU was evident for 11 of the 16 NNT codons in these groups (**Fig. S2**).

Cytosines are comparatively unstable and readily undergo spontaneous deamination to uracil, resulting in C to T transitions after unrepaired replication.[37] This spontaneous deamination occurs 100 times more frequently in ssDNA than dsDNA, resulting in a higher mutation rate at cytosines[38] than at other bases in ssDNA phage.[39] ssDNA phage genomes appear to spend more time truly single-stranded, as they do not experience consistent intra-strand base pairing or regular secondary structure formation while encapsidated.[40-45]
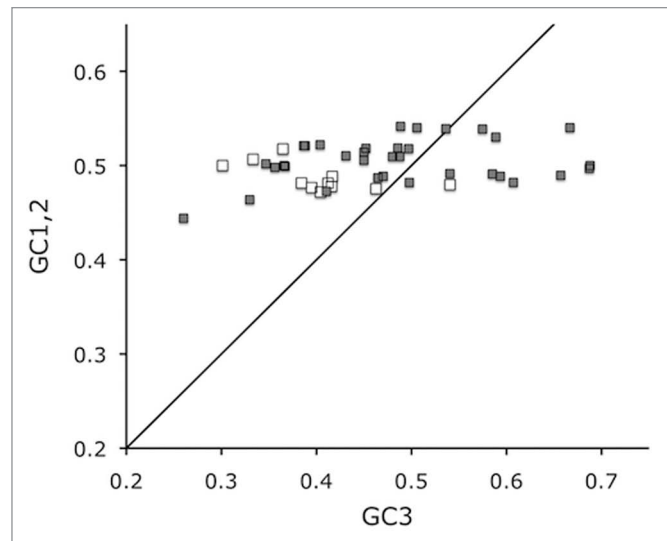

Figure 3. GC1,2/GC3 correlation for ssDNA (open squares) and dsDNA (gray squares) coliphage coat genes. Solid line indicates perfect correlation. Points above the line indicate genes deficient in GC3, points below denote genes enriched in GC3.

This causes ssDNA phages to more frequently have unpaired bases than ssRNA genomes, which are constrained by extensive stem-loop formation both in the cytosol and when encapsidated.[46]

Any thymine-increasing bias does not appear to have a discernible effect on genomic nucleotide content relative to the phages' primary hosts. Rather, it is likely that cytosine transitions in the first or second positions are subject to strong purifying selection relative to the wobble position,[47-49] and the signature of this mutational bias is only observed in the overabundance of thymine in the third position of synonymous ssDNA phage codons. The significant overrepresentation of NNT codons is
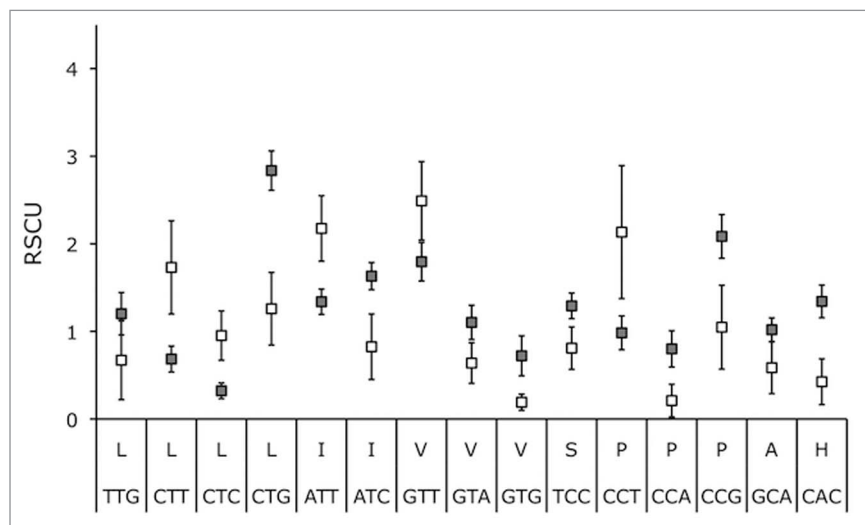
Figure 4. Mean RSCU values and 95% confidence intervals for individual codons with statistically significant differences in usage between ssDNA (open squares) and dsDNA (gray squares) coliphage coat or tail genes.
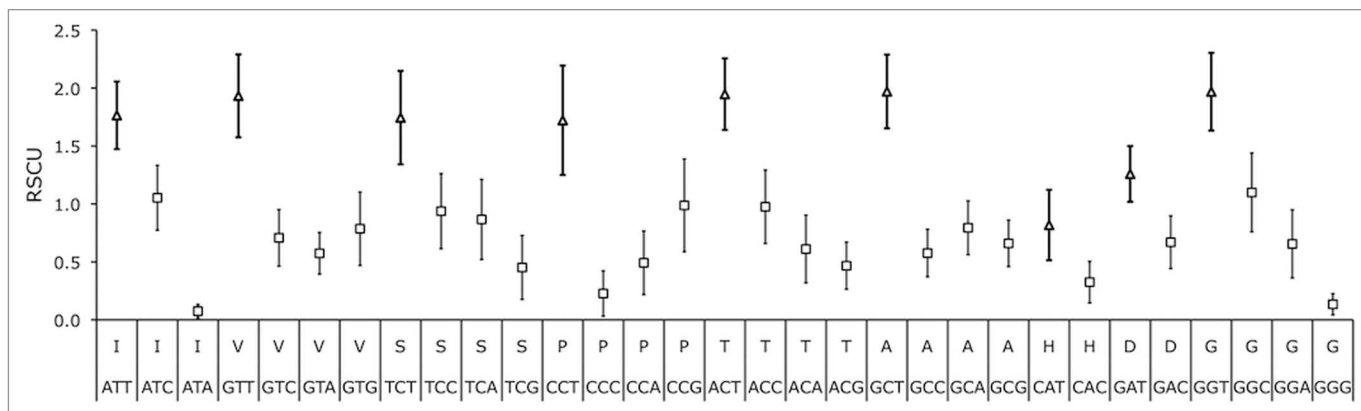
Figure 5. RSCU values and 95% confidence intervals for ssDNA phage coat gene codons that exhibited an NNT codon preference. Preferred NNT codons indicated by bold triangles, NNV codons indicated by squares.

strongly indicative of a biased mutational pressure acting in concert with strong selection against non-synonymous substitutions.

Genomic architecture (nucleic acid, segmentation, strandedness), while acknowledged as an important characteristic of virus taxonomy, is not typically included in broad-scale analyses of viral evolution. Instead, most comparisons focus within a single kind of virus,[50] and while many of these studies have provided insight into the codon usage biases of individual viruses, this is the first observation of a specific bias with a possible mechanistic explanation. Examining across two architectures, we saw strandedness play a critical role in the composition of phage genomes, and in determining the limits of ssDNA viral adaptation to their hosts.

Materials and Methods

All available ssDNA and dsDNA bacteriophage genome reference sequences were collected from GenBank on March 16, 2011. Reference sequences were used to avoid biasing our data sets toward any particular phage species, or highly studied phage, such as the model organisms PhiX174 or T7. These genomes were separated according to genomic architecture for further analysis. Initially collected were 41 ssDNA phages and 447 dsDNA phages (**Table S2**). For each phage having a known host with a sequenced genome (GenBank reference sequence), the relationship between the GC content of the phage and the host bacterium was examined. Because not every sequenced phage has an identified and sequenced host, not all phages were included in this analysis. Four ssDNA phages were excluded, as were 44 dsDNA phages (**Table S2**).

The codon usage biases of representative ssDNA and dsDNA phages were examined to gain a more complete picture of the CUB patterns in both architectures. Codon usage profiles were determined using major coat/capsid genes, or, in the eight cases for which coat genes were not available, tail gene sequences retrieved from GenBank reference genomes (**Table S3**). These structural proteins are highly expressed and exhibit the highest degrees of codon usage bias found in phage.[51,52] We compared codon usage between the two genomic architectures for phages infecting a single host: *Escherichia coli*. Coat or tail genes from 11

ssDNA and 34 dsDNA phages were used (**Table S3**). The online CAIcal tool[53] was used to calculate each phage's codon adaptation index (CAI), a measure of the degree to which one gene or set of genes adheres to the CUB of another gene or set of genes,[1] as implemented by Xia.[54] CAI ranges from zero to one; values closer to one indicate a strong correlation. The average CAI was calculated for both architectures.

Frequency of the first and second codon positions (GC1,2) and frequency of GC in the third position (GC3) were calculated for these genes using CAIcal and relationship between the two was analyzed. A plot of GC1,2 against GC3 is a common measure of the factors affecting CUB in a gene or set of genes; a strong correlation between the two implies that genome-wide mutational pressures are the driving force behind CUB, while a weaker correlation indicates that some force is unequally affecting the first two positions and the third position. Usually, this is interpreted as implying a selective force acting on CUB, as is expected to be the case for viruses under relatively strong selection for translational speed.

To examine the variation in codon usage that contributes to the differing CAI values and site-specific base compositions, relative synonymous codon usage (RSCU) values were calculated for the same sets of genes using CAIcal. RSCU is a measure of the relative codon usage for each individual degenerate amino acid compared with expected levels if synonymous codons were used with equal frequency. An RSCU of about one indicates that a codon is used as frequently as expected, while values above or below one indicate over or underuse of that synonymous codon, respectively. Mean dsDNA coliphage RSCUs were compared with ssDNA coliphage RSCU to determine the proximate cause of the observed variation in CAI. RSCU was also calculated for 17 additional sufficiently well-annotated genomes of ssDNA phages infecting a wide host range (primarily infecting Acholplasma, Bdellovibrio, Chlamydia, Escherichia, Propionibacteria, Pseudomonas, Ralstonia, Spiroplasma, Vibrio and Xanthomonas, **Table S4**), and the complete set of 28 ssDNA phage RSCUs was assessed for consistent CUB. For amino acids with 6-fold redundancy (L, R, S), RSCUs were calculated separately for the codon sets with 4-fold and 2-fold redundancy.

Significantly biased codon use was measured for each codon with one-tailed t-tests (Microsoft Excel) and Bonferroni correction for multiple comparisons (α = 0.017 for 4-fold, α = 0.025 for 3-fold).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Note

Supplemental material can be found at: www.landesbioscience.com/journals/bacteriophage/article/18496

References

1. Sharp PM, Li WH. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 1987; 15:1281-95; PMID:3547335; http://dx.doi.org/10.1093/nar/15.3.1281

2. Carbone A. Codon adaptation index as a measure of dominating codon bias. Bioinformatics 2003; 19:2005-15; PMID:14594704; http://dx.doi.org/10.1093/bioinformatics/btg272

3. Sharp PM. Molecular evolution of bacteriophages: evidence of selection against the recognition sites of host restriction enzymes. Mol Biol Evol 1986; 3:75-83; PMID:2832688

4. Wong EHM, Smith DK, Rabadan R, Peiris M, Poon LLM. Codon usage bias and the evolution of influenza A viruses. Codon Usage Biases of Influenza Virus. BMC Evol Biol 2010; 10:253; PMID:20723216; http://dx.doi.org/10.1186/1471-2148-10-253

5. Shields DC, Sharp PM, Higgins DG, Wright F. "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. Mol Biol Evol 1988; 5:704-16; PMID:3146682

6. Aota S, Ikemura T. Diversity in G+C content at the third position of codons in vertebrate genes and its cause. Nucleic Acids Res 1986; 14:6345-55; PMID:3748815; http://dx.doi.org/10.1093/nar/14.16.6345

7. Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J Mol Biol 1981; 151:389-409; PMID:6175758; http://dx.doi.org/10.1016/0022-2836(81)90003-6

8. Ikemura T. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. J Mol Biol 1982; 158:573-97; PMID:6750137; http://dx.doi.org/10.1016/0022-2836(82)90250-9

9. Curran JF, Yarus M. Rates of aa-tRNA selection at 29 sense codons in vivo. J Mol Biol 1989; 209:65-77; PMID:2478714; http://dx.doi.org/10.1016/0022-2836(89)90170-8

10. Pedersen S. *Escherichia coli* ribosomes translate in vivo with variable rate. EMBO J 1984; 3:2895-8; PMID:6396082

11. Bennetzen JL, Hall BD. Codon selection in yeast. J Biol Chem 1982; 257:3026-31; PMID:7037777

12. Robinson M, Lilley R, Little S, Emtage J, Yarranton G, Stephens P, et al. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. Nucleic Acids Res 1984; 12:6663-71; PMID:6091031; http://dx.doi.org/10.1093/nar/12.17.6663

13. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol 1986; 24:28-38; PMID:3104616; http://dx.doi.org/10.1007/BF02099948

14. Karlin S, Mrázek J, Campbell AM. Codon usages in different gene classes of the *Escherichia coli* genome. Mol Microbiol 1998; 29:1341-55; PMID:9781873; http://dx.doi.org/10.1046/j.1365-2958.1998.01008.x

15. Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet 2008; 42:287-99; PMID:18983258; http://dx.doi.org/10.1146/annurev.genet.42.110807.091442.

16. Hiraoka Y, Kawamata K, Haraguchi T, Chikashige Y. Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*. Genes Cells 2009; 14:499-509; PMID:19335619; http://dx.doi.org/10.1111/j.1365-2443.2009.01284.x

17. Sharp PM, Rogers MS, Mcconnell DJ. Selection pressures on codon usage in the complete genome of bacteriophage T7. J Mol Evol 1984-1985; 21:150-60; PMID:6100189; http://dx.doi.org/10.1007/BF02100089

18. Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila and Arabidopsis. Proc Natl Acad Sci USA 1999; 96:4482-7; PMID:10200288; http://dx.doi.org/10.1073/pnas.96.8.4482

19. Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. Mol Syst Biol 2009; 5:311; PMID:19888206; http://dx.doi.org/10.1038/msb.2009.71

20. Hyman P, Abedon ST. Bacteriophage Host Range and Bacterial Resistance. In: Allen IL, Sima S, Geoffrey MG, Eds. Advances in Applied Microbiology: Academic Press 2010; 217-48

21. Serwer P, Hayes SJ, Zaman S, Lieman K, Rolando M, Hardies SC. Improved isolation of undersampled bacteriophages: finding of distant terminase genes. Virology 2004; 329:412-24; PMID:15518819

22. Thomas JA, Rolando MR, Carroll CA, Shen PS, Belnap DM, Weintraub ST, et al. Characterization of Pseudomonas chlororaphis myovirus 201 var-phi2-1 via genomic sequencing, mass spectrometry, and electron microscopy. Virology 2008; 376:330-8; PMID:18474389; http://dx.doi.org/10.1016/j.virol.2008.04.004

23. Kawasaki T, Nagata S, Fujiwara A, Satsuma H, Fujie M, Usami S, et al. Genomic Characterization of the Filamentous Integrative Bacteriophages RSS1 and RSM1, Which Infect *Ralstonia solanacearum*. J Bacteriol 2007; 189:5792-802; PMID:17557818; http://dx.doi.org/10.1128/JB.00540-07

24. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W. Bacteriophage T4 genome. Microbiol Mol Biol Rev 2003; 67:86-156; PMID:12626685; http://dx.doi.org/10.1128/MMBR.67.1.86-156.2003

25. Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA 1991; 88:7160-4; PMID:1831267; http://dx.doi.org/10.1073/pnas.88.16.7160

26. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol 2010; 84:9733-48; PMID:20660197; http://dx.doi.org/10.1128/JVI.00694-10

27. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 2008; 9:267-76; PMID:18319742; http://dx.doi.org/10.1038/nrg2323

28. Bernardi G, Bernardi G. Compositional constraints and genome evolution. J Mol Evol 1986; 24:1-11; PMID:3104608; http://dx.doi.org/10.1007/BF02099946

29. Karlin S, Mrázek J. What drives codon choices in human genes? J Mol Biol 1996; 262:459-72; PMID:8893856; http://dx.doi.org/10.1006/jmbi.1996.0528

30. Antezana MA, Kreitman M. The nonrandom location of synonymous codons suggests that reading frame-independent forces have patterned codon preferences. J Mol Evol 1999; 49:36-43; PMID:10368432; http://dx.doi.org/10.1007/PL00006532

31. Sueoka N, Kawanishi Y. DNA G+C content of the third codon position and codon usage biases of human genes. Gene 2000; 261:53-62; PMID:11164037; http://dx.doi.org/10.1016/S0378-1119(00)00480-7.

32. Adams MJ, Antoniw JF. Codon usage bias amongst plant viruses. Arch Virol 2004; 149:113-35; PMID:14689279

33. Xia X, Yuen KY. Differential selection and mutation between dsDNA and ssDNA phages shape the evolution of their genomic AT percentage. BMC Genet 2005; 6:20; PMID:15823202; http://dx.doi.org/10.1186/1471-2156-6-20

34. Ohkubo S. Muto A, Kawauchi Y, Yamao F, Osawa S. The ribosomal protein gene cluster of *Mycoplasma capricolum*. Mol Gen Genet 1987; 210:314-22; PMID:3481422; http://dx.doi.org/10.1007/BF00325700

35. Andersson SG, Sharp PM. Codon usage and base composition in *Rickettsia prowazekii*. J Mol Evol 1996; 42:525-36; PMID:8662004; http://dx.doi.org/10.1007/BF02352282

36. Ohama T, Muto A, Osawa S. Role of GC-biased mutation pressure on synonymous codon choice in *Micrococcus luteus*, a bacterium with a high genomic GC-content. Nucleic Acids Res 1990; 18:1565-9; PMID:2326195; http://dx.doi.org/10.1093/nar/18.6.1565

37. Poole A, Penny D, Sjöberg BM. Confounded cytosine! Tinkering and the evolution of DNA. Nat Rev Mol Cell Biol 2001; 2:147-51; PMID:11252956; http://dx.doi.org/10.1038/35052091

38. Frederico LA, Ta K, Shaw BR. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry 1990; 29:2532-7; PMID:2185829; http://dx.doi.org/10.1021/bi00462a015

39. Cuevas JM, Duffy S, Sanjuán R. Point mutation rate of bacteriophage PhiX174. Genetics 2009; 183:747-9; PMID:19652180; http://dx.doi.org/10.1534/genetics.109.106005

40. Shen CK, Ikoku A, Hearst JE. A specific DNA orientation in the filamentous bacteriophage fd as probed by psoralen crosslinking and electron microscopy. J Mol Biol 1979; 127:163-75; PMID:430560; http://dx.doi.org/10.1016/0022-2836(79)90237-7

41. Incardona NL, Prescott B, Sargent D, Lamba OP, Thomas GJ. Phage phi X174 probed by laser Raman spectroscopy: evidence for capsid-imposed constraint on DNA secondary structure. Biochemistry 1987; 26:1532-8; PMID:2954582; http://dx.doi.org/10.1021/bi00380a006

42. Benevides JM, Stow PL, Ilag LL, Incardona NL, Thomas GJ. Differences in secondary structure between packaged and unpackaged single-stranded DNA of bacteriophage phi X174 determined by Raman spectroscopy: a model for phi X174 DNA packaging. Biochemistry 1991; 30:4855-63; PMID:1827990; http://dx.doi.org/10.1021/bi00234a004

43. Welsh LC, Marvin DA, Perham RN. Analysis of X-ray diffraction from fibres of Pf1 Inovirus (filamentous bacteriophage) shows that the DNA in the virion is not highly ordered. J Mol Biol 1998; 284:1265-71; PMID:9878347; http://dx.doi.org/10.1006/jmbi.1998.2275

44. Wen ZQ, Armstrong A, Thomas GJ Jr. Demonstration by ultraviolet resonance Raman spectroscopy of differences in DNA organization and interactions in filamentous viruses Pf1 and fd. Biochemistry 1999; 38:3148-56; PMID:10074370; http://dx.doi.org/10.1021/bi981965m

45. Tsuboi M, Tsunoda M, Overman SA, Benevides JM, Thomas GJ. A Structural Model for the Single-Stranded DNA Genome of Filamentous Bacteriophage Pf1. Biochemistry 2010; 49:1737-43; PMID:20078135; http://dx.doi.org/10.1021/bi901323a

46. Thurner C. Conserved RNA secondary structures in Flaviviridae genomes. J Gen Virol 2004; 85:1113-24; PMID:15105528; http://dx.doi.org/10.1099/vir.0.19462-0

47. Crick FH. Codon-anticodon pairing: the wobble hypothesis. J Mol Biol 1966; 19:548-55; PMID:5969078; http://dx.doi.org/10.1016/S0022-2836(66)80022-0

48. Boyer SH, Scott AF, Kunkel LM, Smith KD. The proportion of all point mutations which are unacceptable: and estimate based on hemoglobin amino acid and nucleotide sequences. Can J Genet Cytol 1978; 20:111-37; PMID:350360

49. Holmes EC. Patterns of Intra- and Interhost Nonsynonymous Variation Reveal Strong Purifying Selection in Dengue Virus. J Virol 2003; 77:11296-8; PMID:14512579; http://dx.doi.org/10.1128/JVI.77.20.11296-8.2003

50. Jenkins GM, Holmes EC. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res 2003; 92:1-7; PMID:12606071; http://dx.doi.org/10.1016/S0168-1702(02)00309-X

51. Carbone A. Codon bias is a major factor explaining phage evolution in translationally biased hosts. J Mol Evol 2008; 66:210-23; PMID:18286220; http://dx.doi.org/10.1007/s00239-008-9068-6

52. Lucks JB, Nelson DR, Kudla GR, Plotkin JB. Genome landscapes and bacteriophage codon usage. PLOS Comput Biol 2008; 4:1000001; PMID:18463708; http://dx.doi.org/10.1371/journal.pcbi.1000001

53. Puigbò P, Bravo IG, Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation. Biol Direct 2008; 3:38; PMID:18796141; http://dx.doi.org/10.1186/1745-6150-3-38

54. Xia X. An Improved Implementation of Codon Adaptation Index. Evolutionary Bioinformatics 2007:53-8