The Fitness Effects of Synonymous Mutations in DNA and RNA Viruses

José M. Cuevas,[1] Pilar Domingo-Calap,[1] and Rafael Sanjuán*,[1,2,3]

[1]Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, València, Spain

[2]Departament de Genètica, Universitat de València, València, Spain

[3]Unidad Mixta de Investigación en Genómica y Salud, Centro Superior de Investigación en Salud Pública (CSISP), València, Spain

*Corresponding author: E-mail: rafael.sanjuan@uv.es.

Associate editor: Oliver Pybus

Abstract

Despite being silent with respect to protein sequence, synonymous nucleotide substitutions can be targeted by natural selection directly at the DNA or RNA level. However, there has been no systematic assessment of how frequent this type of selection is. Here, we have constructed 53 single random synonymous substitution mutants of the bacteriophages Qβ and ΦX174 by site-directed mutagenesis and assayed their fitness. Analysis of this mutant collection and of previous studies undertaken with a variety of single-stranded (ss) viruses demonstrates that selection at synonymous sites is stronger in RNA viruses than in DNA viruses. We estimate that this type of selection contributes approximately 18% of the overall mutational fitness effects in ssRNA viruses under our assay conditions and that random synonymous substitutions have a 5% chance of being lethal to the virus, whereas in ssDNA viruses, these figures drop to 1.4% and 0%, respectively. In contrast, the effects of nonsynonymous substitutions appear to be similar in ssRNA and ssDNA viruses.

Key words: synonymous substitutions, nonsynonymous substitutions, silent substitutions, RNA structure, codon usage bias, RNA virus, DNA virus, selection, evolution, site-directed mutagenesis, mutational robustness, mutational fitness effects.

Letter

Although synonymous nucleotide substitutions have been traditionally considered as selectively neutral or subject to minor constraints, they are important for several processes. For instance, synonymous codons can have different translation efficiencies because transfer RNAs vary in concentration within the cell, typically leading to codon usage biases (Shields et al. 1988). The thermodynamic stability of the mRNA is also under selection (Chamary and Hurst 2005), and its secondary structure influences gene expression by 5′ untranslated region–mediated initiation of translation (Kudla et al. 2009), riboswitches (Tucker and Breaker 2005), recognition by the double-stranded RNA-activated protein kinase (PKR) (Kaempfer 2003), or activation of posttranscriptional silencing (Voinnet 2009). The structure of the mRNA can also modulate the folding of the nascent amino acid chain by ribosome pausing (Watts et al. 2009) and, in some cases, alter the protein sequence by inducing ribosomal slippage (Xu et al. 2001).

Selection acting on synonymous substitutions should be most relevant in viruses with highly compact genomes because their protein-coding sequences are often involved in gene regulation, cell trafficking, or virus encapsidation. Particularly, it has been shown that some of the secondary structures adopted by RNA virus genomes are under strong selective pressure because they regulate translation (Groeneveld et al. 1995; Olsthoorn and van Duin 1996; Klovins, Tsareva, et al. 1997; Watts et al. 2009), control replication initiation (Klovins et al. 1998), or are targeted by cellular RNases (Klovins, van

Duin, et al. 1997). These selective pressures have been demonstrated through the analysis of synonymous genetic variation (Le et al. 1988, 1989; Rodríguez-Alvarado and Roossinck 1997; Yoshida et al. 1997; Simmonds and Smith 1999; Tuplin et al. 2002), and in experimentally evolving single-stranded RNA (ssRNA) viruses or ssDNA viruses, the same synonymous substitutions are sometimes fixed repeatedly in independent lines, suggesting positive selection at these sites (de la Torre et al. 1992; Bull et al. 1997; Cuevas et al. 2002; Novella et al. 2004).

Although there is currently little doubt that synonymous substitutions need not to be neutral, there has been no systematic quantitative assessment of their fitness effects. Site-directed mutagenesis offers us a powerful tool for achieving this goal. In previous studies, we and others have created series of single-nucleotide mutants for several ssRNA and ssDNA viruses (reviewed in Sanjuán 2010), but these studies did not focus specifically on synonymous substitutions. Here, we have used this technique to introduce 53 synonymous, but otherwise random single-nucleotide substitutions in the genomes of the bacteriophages Qβ and ΦX174 and have determined their fitness effects (fig. 1). We created 27 mutants for Qβ and 26 for ΦX174, all of which were viable. In Qβ, the average fitness effect was −0.0295 with variance 0.0036, and 5/27 mutations (all mapping to the maturase gene) were significantly deleterious: C628U ($s = -0.087 \pm 0.007$), C1048U ($s = -0.056 \pm 0.005$), C1118A ($s = -0.178 \pm 0.080$), U1186G ($s = -0.040 \pm 0.009$), and C1303A ($s = -0.202 \pm 0.028$; mutations refer
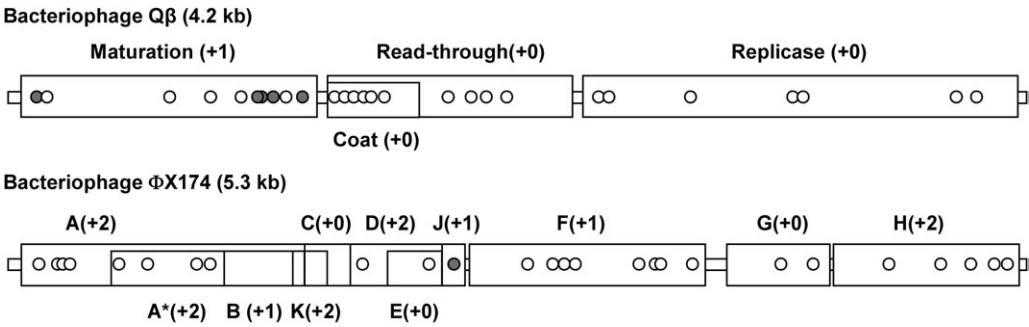
FIG. 1. Genome structure of bacteriophages Qβ and ΦX174 and location of the random synonymous substitutions introduced by site-directed mutagenesis. Significantly deleterious mutations are shown in gray. Genome sizes are 4.2 and 5.4 kb, respectively, of which >90% are protein-coding regions. Each gene is indicated with a box, and inner boxes indicate overlapping genes. The reading frame of each gene is indicated within parentheses. Bacteriophage ΦX174 has a circular genome but is shown linear for convenience. Details about the location of each mutation and its fitness effect can be found in supplementary table S1, Supplementary Material online.

to GenBank accession number GQ153931). All of these five mutations were retested using appropriate controls to ensure that the observed effect was not caused by additional changes appearing elsewhere in the genome (see Methods). In ΦX174, mutations had an average fitness effect of $+0.0051$, with variance 0.0008, and only one mutation of 26 proved to be significantly deleterious (A955G in sequence GQ153915; $s = -0.076 \pm 0.049$).

Synonymous mutations appear thus to be more frequently deleterious in Qβ than in ΦX174, although some of the assayed mutants might actually be nonneutral but with an effect too small to be detected in our assays. Direct comparison of average s values, which does not present this problem, confirms that the fitness effects of synonymous mutations are significantly greater in Qβ than in ΦX174 (Mann–Whitney test: $P = 0.006$). We cannot discard that, in some cases, compensatory mutations might have become fixed prior to fitness assays, leading to an underestimation of s, but this should occur more frequently in Qβ because of its higher mutation rate and therefore does not affect the validity of our conclusion.

To assess the generality of our results, we considered four previous site-directed mutagenesis studies carried out with the vesicular stomatitis virus (VSV) (Sanjuán et al. 2004),

tobacco etch virus (TEV) (Carrasco et al. 2007), and the bacteriophages Qβ, ΦX174, and f1 (Domingo-Calap et al. 2009; Peris et al. 2010), that is, ssRNA and ssDNA viruses belonging to different families and infecting widely different hosts (bacteria, plants, and animals). Including the present study, the data set consists of 322 mutations, 122 synonymous, and 200 nonsynonymous (table 1). Among the three ssRNA viruses examined, the fitness effects of synonymous substitutions were roughly five times smaller on average than those of nonsynonymous substitutions ($\bar{s} = -0.090$ vs. $\bar{s} = -0.492$; $P < 0.001$), and the fraction of lethal mutants was eight times smaller (5% vs. 41%; Fisher's exact test: $P < 0.001$). In DNA viruses, the fitness effects of synonymous substitutions were roughly 70 times smaller on average than those of nonsynonymous substitutions ($\bar{s} = -0.005$ vs. $\bar{s} = -0.360$; $P < 0.001$), and none of the 67 synonymous substitutions assayed was lethal, as opposed to a 23% lethality among nonsynonymous substitutions ($P < 0.001$). These data confirm that synonymous substitutions are significantly more deleterious in ssRNA viruses than in ssDNA viruses ($P < 0.001$).

We conclude that, although selection is mainly exerted at the protein level both in ssRNA and ssDNA viruses, RNA-level selection plays also a significant role in ssRNA viruses.

Table 1. Fitness Effects of Random Synonymous (S) and Nonsynonymous (NS) Substitutions in Several Viruses[a].

Group	Virus	Number		Mean[b]		Variance[b]		Lethal Fraction (%)	
		S	NS	S	NS	S	NS	S	NS
ssRNA	VSV	8	35	−0.143	−0.532	0.122	0.194	13	43
	TEV[c]	11	52	−0.173	−0.509	0.120	0.217	9	44
	Qβ[d]	36	33	−0.052	−0.424	0.029	0.184	3	33
	Total	55	120	−0.090	−0.492	0.060	0.200	5	41
ssDNA	ΦX174[d]	38	28	−0.007	−0.352	0.003	0.170	0	21
	f1	29	52	−0.003	−0.364	0.002	0.163	0	24
	Total	67	80	−0.005	−0.360	0.002	0.164	0	23
Total	—	122	200	−0.043	−0.439	0.030	0.189	2	34

[a] Defined as the growth rate ratio minus one (see Methods).
[b] Lethal mutations included.
[c] Fitness values from the original study were transformed to the units used here. See a previous article for details on this unit conversion (Sanjuán 2010).
[d] Data from the present study and a previous work (Domingo-Calap et al. 2009) have been pooled. There were no significant differences between the mean effects of synonymous substitutions obtained in the two studies ($P > 0.5$ for both Qβ and ΦX174).

The contribution of this component to the total fitness effects can be estimated as the ratio between the s values obtained for synonymous and nonsynonymous substitutions. This ratio is 18% (0.090/0.492) for ssRNA viruses and only 1.4% (0.005/0.360) for ssDNA viruses. In contrast, selection at the protein level, calculated as the net fitness effect of nonsynonymous substitutions (i.e., after subtracting the average effect of synonymous substitutions from that of nonsynonymous substitutions in table 1), is remarkably constant for the five viruses studied, ranging from $\bar{s} = -0.336$ to $\bar{s} = -0.389$. As a result, and as previously shown (Domingo-Calap et al. 2009), the overall mutational robustness of ssRNA and ssDNA viruses is similar, albeit slightly greater in the former.

The differences between the fitness effects of synonymous substitutions in ssRNA and ssDNA viruses seem to be not related with codon usage bias because, although the bias is significant for the above five viruses (χ^2 test: $P < 0.0001$), it is more marked in ssDNA viruses than in ssRNA viruses, as determined from the effective numbers of codons (ΦX174: 47.6; f1: 51.2; VSV: 56.3; TEV: 56.3; and Qβ: 55.2). The fact that the functional relevance of genomic secondary structures has been more extensively demonstrated in ssRNA viruses than in ssDNA offers a more plausible explanation for our results. Since the selective pressures associated with RNA structure are also exerted in nonsynonymous substitutions, they should leave an imprint on the evolution of RNA virus proteins, a prediction that has been recently validated using the empirically determined secondary structure of the HIV-1 genome (Sanjuán and Bordería 2011).

Methods

Bacteriophage Qβ was derived from an infectious clone kindly provided by Dr René C. Olsthoorn (Leiden University). Bacteriophage ΦX174 and the *Escherichia coli* C strain IJ1862 were obtained from Prof. James J. Bull (University of Texas). Site-directed mutagenesis and fitness assays were carried out as described previously (Domingo-Calap et al. 2009; Peris et al. 2010; Sanjuán 2010). Briefly, the whole viral genome was amplified by polymerase chain reaction using primers carrying the desired mutations, and the product was used to transfect competent cells. Individual plaques were picked and sequenced to verify the presence of the mutation and that no additional changes were present in a 500-base region flanking the target site. Three plaques from each mutant were pooled before fitness assays to minimize the effects of potential mutations appearing elsewhere in the genome, and fitness assays were performed in three independent blocks. Fitness was calculated as $W = r_i/r_0$, where r_i is the growth rate of mutant i and is r_0 the growth rate of the wild-type virus determined in the same block, and the fitness effect of each mutation was then obtained as $s = W - 1$, such that s is negative for deleterious mutations and positive for beneficial mutations. Mutants showing a growth rate significantly different from that of the wild type ($P < 0.05$) were classified as

candidate nonneutrals. For each candidate, we designed a primer identical to the one used for mutagenesis except that it carried no mutation (control primer) and performed the mutagenesis protocol in parallel for the mutant and its corresponding control. Then, we reassayed the growth rate of each mutant and control in the same experimental block (three replicates of each) to confirm that nonneutrality was due only to the engineered mutation. The reported s values were obtained by pooling the initial and confirmatory assays. Codon usage bias calculations were performed from files containing a concatenate of all protein-coding sequences (excluding regions with overlapping reading frames) using ENCprime (http://www.eeb.ucla.edu/Faculty/Novembre/software/software.html). The effective number of codons (N_c) has been defined previously (Novembre 2002). $N_c = 61$ (the set of all sense codons) indicates no codon bias, whereas lower values indicate that some codons are preferred.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org).

Acknowledgments

References

Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux IJ. 1997. Exceptional convergent evolution in a virus. *Genetics* 147:1497–1507.

Carrasco P, de la Iglesia F, Elena SF. 2007. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in tobacco etch virus. *J Virol.* 81:12979–12984.

Chamary JV, Hurst LD. 2005. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol.* 6:R75.

Cuevas JM, Elena SF, Moya A. 2002. Molecular basis of adaptive convergence in experimental populations of RNA viruses. *Genetics* 162:533–542.

de la Torre JC, Giachetti C, Semler BL, Holland JJ. 1992. High frequency of single-base transitions and extreme frequency of precise multiple-base reversion mutations in poliovirus. *Proc Natl Acad Sci U S A.* 89:2531–2535.

Domingo-Calap P, Cuevas JM, Sanjuán R. 2009. The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet.* 5:e1000742.

Groeneveld H, Thimon K, van Duin J. 1995. Tanslational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding? *RNA* 1:79–88.

Kaempfer R. 2003. RNA sensors: novel regulators of gene expression. *EMBO Rep.* 4:1043–1047.

Klovins J, Berzins V, van Duin J. 1998. A long-range interaction in Qbeta RNA that bridges the thousand nucleotides between the M-site and the 3' end is required for replication. *RNA* 4:948–957.

Klovins J, Tsareva NA, de Smit MH, Berzins V, van Duin J. 1997. Rapid evolution of translational control mechanisms in RNA genomes. *J Mol Biol.* 265:372–384.

Klovins J, van Duin J, Olsthoorn RC. 1997. Rescue of the RNA phage genome from RNase III cleavage. *Nucleic Acids Res.* 25: 4201–4208.

Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324:255–258.

Le SY, Chen JH, Braun MJ, Gonda MA, Maizel JV. 1988. Stability of RNA stem-loop structure and distribution of non-random structure in the human immunodeficiency virus (HIV-I). *Nucleic Acids Res.* 16:5153–5168.

Le SY, Chen JH, Chatterjee D, Maizel JV. 1989. Sequence divergence and open regions of RNA secondary structures in the envelope regions of the 17 human immunodeficiency virus isolates. *Nucleic Acids Res.* 17:3275–3288.

Novella IS, Zárate S, Metzgar D, Ebendick-Corp. 2004. Positive selection of synonymous mutations in vesicular stomatitis virus. *J Mol Biol.* 342:1415–1421.

Novembre JA. 2002. Accounting for background nucleotide composition when measuring codon usage bias. *Mol Biol Evol.* 19:1390–1394.

Olsthoorn RC, van Duin J. 1996. Evolutionary reconstruction of a hairpin deleted from the genome of an RNA virus. *Proc Natl Acad Sci U S A.* 93:12256–12261.

Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuán R. 2010. Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. *Genetics* 185:603–609.

Rodríguez-Alvarado G, Roossinck MJ. 1997. Structural analysis of a necrogenic strain of cucumber mosaic cucumovirus satellite RNA in planta. *Virology* 236:155–166.

Sanjuán R. 2010. Mutational fitness effects in RNA and ssDNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos Trans R Soc Lond B Biol Sci.* 365:1975–1982.

Sanjuán R, Bordería AV. 2011. Interplay between RNA structure and protein evolution in HIV-1. *Mol Biol Evol.* 28:1333–1338.

Sanjuán R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A.* 101:8396–8401.

Shields DC, Sharp PM, Higgins DG, Wright F. 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol.* 5:704–716.

Simmonds P, Smith DB. 1999. Structural constraints on RNA virus evolution. *J Virol.* 73:5787–5794.

Tucker BJ, Breaker RR. 2005. Riboswitches as versatile gene control elements. *Curr Opin Struct Biol.* 15:342–348.

Tuplin A, Wood J, Evans DJ, Patel AH, Simmonds P. 2002. Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* 8:824–841.

Voinnet O. 2009. Origin, biogenesis, and activity of plant micro-RNAs. *Cell* 136:669–687.

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460:711–716.

Xu Z, Choi J, Yen TS, Lu W, Strohecker A, Govindarajan S, Chien D, Selby MJ, Ou J. 2001. Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. *EMBO J.* 20:3840–3848.

Yoshida K, Nakamura M, Ohno T. 1997. Mutations of the HIV type 1 V3 loop under selection pressure with neutralizing monoclonal antibody NM-01. *AIDS Res Hum Retroviruses.* 13:1283–1290.