The Fitness Effects of Synonymous Mutations in DNA and RNA Viruses

José M. Cuevas, Pilar Domingo-Calap, and Rafael Sanjuán

1Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, València, Spain
2Departament de Genètica, Universitat de València, València, Spain
3Unidad 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.

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; Rodriguez-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 lineages, 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 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
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 Qβ174, mutations had an average fitness effect of \( \pm 0.0051 \), 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 (\( s = -0.090 \) vs. \( s = -0.492; P < 0.001 \)), 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 (\( s = -0.005 \) vs. \( s = -0.360; P < 0.001 \)), 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.

<table>
<thead>
<tr>
<th>Group</th>
<th>Virus</th>
<th>Number</th>
<th>Mean(^b)</th>
<th>Variance(^b)</th>
<th>Lethal Fraction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td>NS</td>
</tr>
<tr>
<td>ssRNA</td>
<td>VSV</td>
<td>8</td>
<td>35</td>
<td>-0.143</td>
<td>-0.532</td>
</tr>
<tr>
<td></td>
<td>TEV(^d)</td>
<td>11</td>
<td>52</td>
<td>-0.173</td>
<td>-0.509</td>
</tr>
<tr>
<td>ssRNA</td>
<td>Qβ(^d)</td>
<td>36</td>
<td>33</td>
<td>-0.052</td>
<td>-0.424</td>
</tr>
<tr>
<td>ssRNA</td>
<td>Total</td>
<td>55</td>
<td>120</td>
<td>-0.090</td>
<td>-0.492</td>
</tr>
<tr>
<td>ssRNA</td>
<td>ΦX174(^d)</td>
<td>38</td>
<td>28</td>
<td>-0.007</td>
<td>-0.352</td>
</tr>
<tr>
<td>ssRNA</td>
<td>f1</td>
<td>29</td>
<td>52</td>
<td>-0.003</td>
<td>-0.364</td>
</tr>
<tr>
<td>ssRNA</td>
<td>Total</td>
<td>67</td>
<td>80</td>
<td>-0.006</td>
<td>-0.360</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Total</td>
<td>122</td>
<td>200</td>
<td>-0.043</td>
<td>-0.439</td>
</tr>
</tbody>
</table>

\(^a\) 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 $s = -0.336$ to $s = -0.389$. 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, although the bias is significant for the above five viruses ($\chi^2$ test: $P < 0.0001$), it is more marked in ssDNA viruses than in ssRNA viruses, as determined from the effective numbers of codons (DX174: 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 Borderia 2011).

**Methods**

Bacteriophage Qβ was derived from an infectious clone kindly provided by Dr René C. Osthoorn (Leiden University). Bacteriophage FX174 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**

We thank Eddie Holmes for helpful comments. This work was financially supported by grant BFU2008-03978/BMC and the Ramón y Cajal research program from the Spanish Ministerio de Ciencia e Innovación (MICINN) to R.S., the Juan de la Cierva research program from MICINN to J.M.C., and a predoctoral fellowship from the Generalitat Valenciana to P.D.-C.

**References**


