(Rise and fall of synonymous mutations)

Fabio Zanini and Richard A. Neher

November 28, 2012

Abstract

Intrapatient HIV evolution is goverened by selection on the protein level in the arms race with the immune system (killer T-cells and antibodies). Synonymous mutations do not have an immunity-related phenotype and are often assumed to be neutral. In this paper, we show that synonymous changes in epitope-rich regions are often deleterious but still reach frequencies of order one. We analyze time series of viral sequences from the V1-C5 part of *env* within individual hosts and observe that synonymous derived alleles rarely fix in the viral population. Simulations suggest that such synonymous mutations have a (Malthisuan) selection coefficient of the order of -0.001, and that they are brought up to high frequency by linkage to neighbouring beneficial nonsynonymous alleles (genetic draft). As far as the biological causes are concerned, we detect a negative correlation between fixation of an allele and its involvement in evolutionarily conserved RNA stem-loop structures. This phenonenon is not observed in other parts of the HIV genome, in which selective sweeps are less dense and the genetic architecture less constrained.

1 Introduction

HIV evolves rapidly within a single host during the course of the infection. The driving forces shaping this process are the high mutation rate and the strong selection imposed by the host immune system via a wealth of mechanisms, notably killer T cells (CTLs) and neutralizing antibodies (Pantaleo and Fauci, 1996).

In a nutshell, when the host develops a CTL or antibody respose against a particular viral epitope, rare HIV variants carrying mutated versions of the epitope, called *escape mutants*, acquire a fitness advantage and spread rapidly in the viral population, within a few months (see Fig. 1, solid lines). During chronic infection, the (Malthusian) effect size of this beneficial mutations is of the order of 0.01 (Neher and Leitner, 2010). The viral *env* gene shows the fastest rates of

adaptation, because is both rich of CTL epitopes and targeted by antibodies; its sequence diverges at rates of the order of 1% per year (Shankarappa *et al.*, 1999).

Many nucleotide polymorphisms are escape mutations, and in particular are nonsynonymous, i.e. they appear in protein coding regions and change the amino acid sequence. Nonetheless, nucleotide changes unrelated to immune escape are seen, in *env* and elsewhere, and some of them become abundant alike, often rapidly. In particular, it is not uncommon for synonymous mutations to reach frequencies of order one within months from their first appearance (see Fig. 1, dashed lines). The biological function of these mutations in the economy of HIV is not well understood. By definition, the immunological phenotype, which is decided at the protein level, is unaffected, but other biological and ecological aspects of the viral lifestyle might be involved. In practice, a couple of RNA-level phenotypes are known. For example, within *env* a certain RNA sequence, called *rev* response element (RRE), is used by HIV to enhance nuclear export of some of its transcripts (Fernandes *et al.*, 2012). Another case is the interaction between viral reverse transcriptase, viral ssRNA, and the host $\text{tRNA}^{\text{Lys3}}$: the latter is required for priming viral replication and bound by a specifical pseudoknotted RNA structure in the viral 5' untranslated region (Barat *et al.*, 1991; Paillart *et al.*, 2002).

Crucially for evolutionary studies, the minor phenotypes caused by synonymous mutations might have an effect on viral fitness. For instance, recent studies have shown that genetically engineered HIV strains with skewed codon usage bias (CUB) patterns towards more or less abundant tRNAs replicate better or worse, respectively (Ngumbela *et al.*, 2008; Li *et al.*, 2012). In this study, we try to characterize the fitness effects of synonymous polymorphisms that, at some point during the infection, become abundant in the viral population.

One simple way to assess the neutrality of synonymous mutations is to look at their level of conservation. Deleterious mutations at functional sites are expected to be absent or rare across the viral population; vice versa, mutant alleles that reach high frequencies are expected to be neutral. Confirmatorily, population genetics shows that the equilibrium frequency of a deleterious allele with fitness $-s$ is $\mu/|s|$, where μ is the mutation rate per site per generation; neutral alleles have no equilibrium frequency and can slowly fix via genetic drift (Ewens, 2004). This approach, albeit intuitive, works only under the assumption of independent sites. If the focal synonymous mutant is linked to another, nonneutral allele, its frequency is the result of the combined fitness effects of both sites. Since recombination in HIV is known to be rather rare (Neher and Leitner, 2010; Batorsky *et al.*, 2011), the genetic context of the synonymous change at hand must be taken into account.

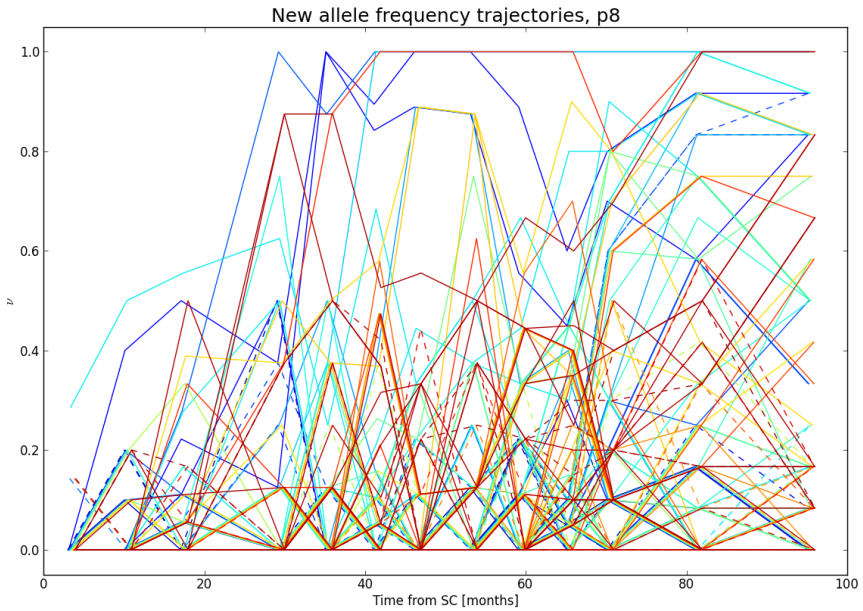

Fig. 1: Allele frequency trajectories of typical patient, C3-V5, nonsynonymous (solid) and synonymous mutations (dashed lines). Most synonymous mutations are not fixed. Colors are set according to the position of the site along the C3-V5 region (red to blue). Data from Ref. Shankarappa *et al.* (1999).

2 Results

We start from time series of viral nucleotide sequences from single patients, which span several years of chronic infection (Shankarappa *et al.*, 1999; Bunnik *et al.*, 2008; Liu *et al.*, 2006). Plotting the allele frequencies against time for all polymorphic sites, it is evident that, although nonsynonymous changes are widespread, synonymous ones are also present in various occasions at high frequency (see Fig. 1).

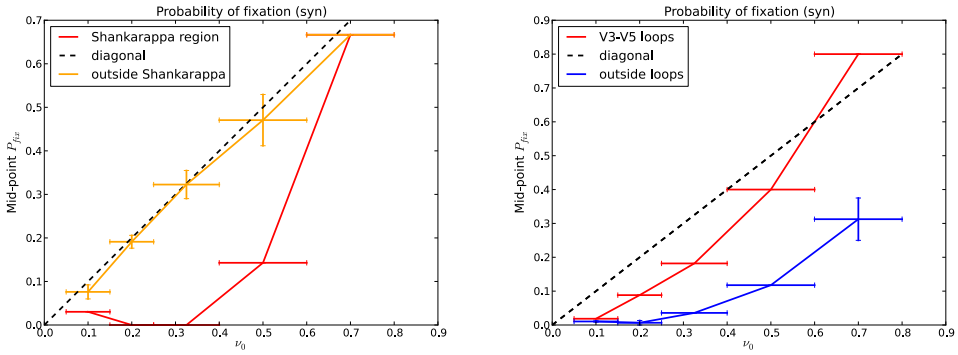

Fig. 2: Fixation probability of derived synonymous alleles is strongly suppressed in C3-V5 versus other parts of the *env* gene (left panel). Especially hard is fixation of new alleles in conserved regions flanking the V loops (right panel). The black dashed line is the prediction from neutral theory, for comparison purposes. Data from Refs. Shankarappa *et al.* (1999); Bunnik *et al.* (2008).

3 Discussion

4 Methods

Acknowledgements

References

Barat, C., Grice, S. F. J. L., and Daelix, J.-L. (1991). Interaction of HIV-1 reverse transcriptase with a synthetic form of its replication primer, tRNALys,3. *Nucleic Acids Research*, **19**(4), 751–757.

Batorsky, R., Kearney, M. F., Palmer, S. E., Maldarelli, F., Rouzine, I. M., and Coffin, J. M. (2011). Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(14), 5661–6.

Bunnik, E., Pisas, L., Van Nuenen, A., and Schuitemaker, H. (2008). Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype b human immunodeficiency virus type 1 infection. *Journal of virology*, **82**(16), 7932.

Ewens, W. J. (2004). *Mathematical Population Genetics: I. Theoretical Introduction*. Springer.

Fernandes, J., Jayaraman, B., and Frankel, A. (2012). The HIV-1 rev response element: An RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. *RNA Biology*, **9**(1), 4–9.

Li, M., Kao, E., Gao, X., Sandig, H., Limmer, K., Pavon-Eternod, M., Jones, T. E., Landry, S., Pan, T., Weitzman, M. D., and David, M. (2012). Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature*.

Liu, Y., McNevin, J., Cao, J., Zhao, H., Genowati, I., Wong, K., McLaughlin, S., McSweyn, M., Diem, K., Stevens, C., *et al.* (2006). Selection on the human immunodeficiency virus type 1 proteome following primary infection. *Journal of virology*, **80**(19), 9519.

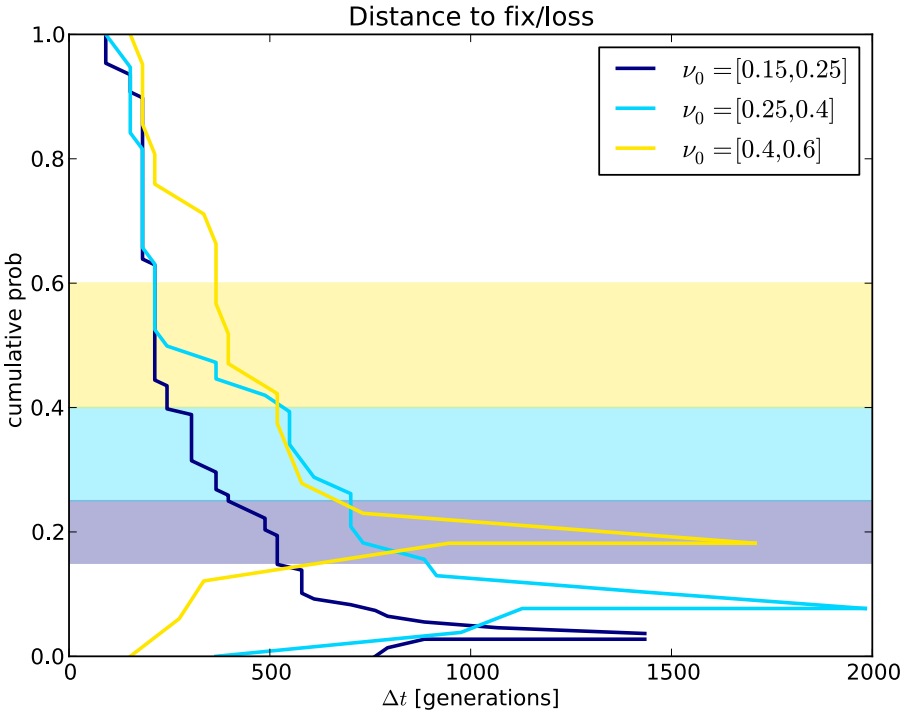

Fig. 3: Fixation or extinction times for synonymous alleles starting from intermediate frequencies. The colored bands are the final fixation probabilities expected from neutral theory; the observed alleles are fixed less frequently than expected. The timescale of fixation/extinction is approximately 500 days, corresponding to a selective effect of ~ -0.001.

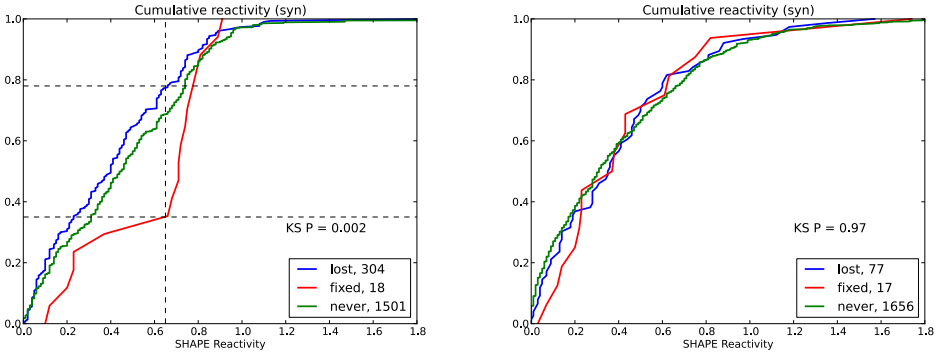

Fig. 4: Watts et al. have measured the reactivity of HIV nucleotides to *in vitro* chemical attack and shown that some nucleotides are more likely to be involved in RNA secondary folds. C1-C5 regions, in particular, show conserved stem-loop structures (Watts *et al.*, 2009). We show that among all derived alleles in those regions reaching frequencies of order one, there is a negative correlation between fixation and involvement in a base pairing in a RNA stem (left panel). The rest of the genome does not show any correlation (right panel). There might be too few silent polymorphisms in the first place, or the signal might be masked by a lot of non-functional RNA structures. Data from Refs. Shankarappa *et al.* (1999); Bunnik *et al.* (2008); Liu *et al.* (2006).

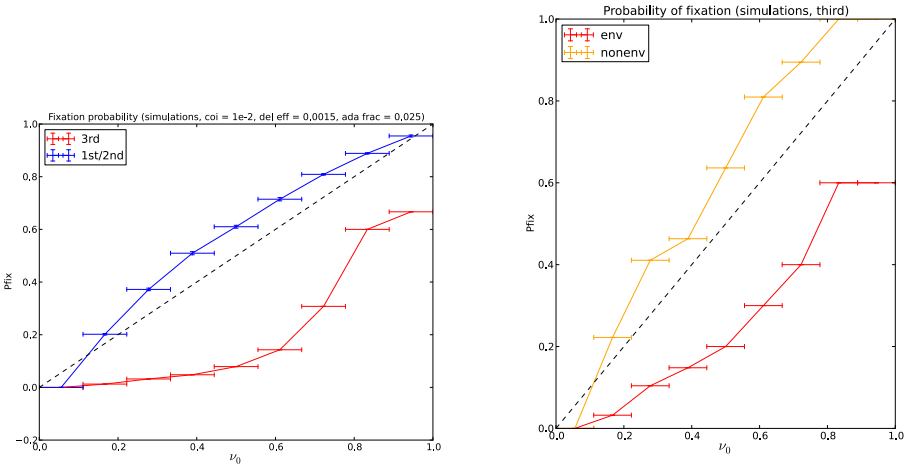

Fig. 5: Simulations show that the suppression of fixation probability can be generated by linkage to sweeping nonsynonymous alleles nearby. Two possible scenarios are competition between escape mutants (left panel) and time-dependent selection due to immune sytem recognition (right panel).

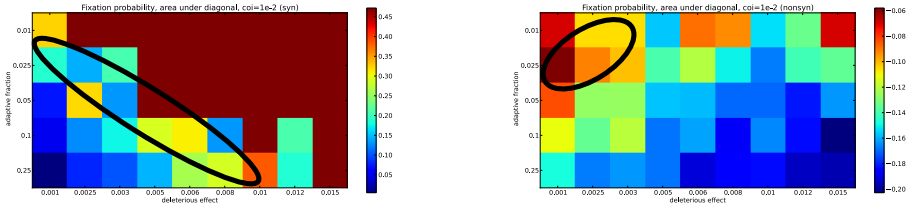

Fig. 6: Simulations on the escape competition scenario show that the density of selective sweeps and the size of the deleterious effects of synonymous mutations are the main driving forces of the phenomenon. A convex fixation probability is recovered, as seen in the data, along the diagonal (left panel): more dense sweeps can support more deleterious linked mutations. The density of sweeps is limited, however, by the nonsynonymous fixation probability, which is quite close to neutrality (right panel). Moreover, strong competition between escape mutants is required, so that several escape mutants are "found" by HIV within a few months of antibody production.

Neher, R. and Leitner, T. (2010). Recombination rate and selection strength in HIV intra-patient evolution. *PLoS Comput Biol*, **6**(1), e1000660.

Ngumbela, K. C., Ryan, K. P., Sivamurthy, R., Brockman, M. A., Gandhi, R. T., Bhardwaj, N., and Kavanagh, D. G. (2008). Quantitative effect of suboptimal codon usage on translational efficiency of mRNA encoding HIV-1 gag in intact t cells. *PLoS ONE*, **3**(6), e2356.

Paillart, J.-C., Skripkin, E., Ehresmann, B., Ehresmann, C., and Marquet, R. (2002). In vitro evidence for a long range pseudoknot in the 5′-untranslated and matrix coding regions of HIV-1 genomic RNA. *Journal of Biological Chemistry*, **277**(8), 5995–6004.

Pantaleo, G. and Fauci, A. S. (1996). Immunopathogenesis of hiv infection1. *Annual Review of Microbiology*, **50**(1), 825–854. PMID: 8905100.

Shankarappa, R., Margolick, J., Gange, S., Rodrigo, A., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C., Learn, G., He, X., *et al.* (1999). Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *Journal of Virology*, **73**(12), 10489.

Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Jr, J. W. B., Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**(7256), 711–716.