Regular Article

Influence of Mutation and Recombination on HIV-1 in vitro Fitness Recovery

Miguel Arenas[1,2,*], Ramon Lorenzo-Redondo[3,#,*], and Cecilio Lopez-Galindez[3]

[1]Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP). Porto, Portugal.

[2]Centre for Molecular Biology "Severo Ochoa", Consejo Superior de Investigaciones Científicas (CSIC). Madrid, Spain.

[3]Centro Nacional de Microbiología (CNM), Instituto de Salud Carlos III. Majadahonda, Madrid, Spain.

[#]Present address: Division of Infectious Diseases, The Feinberg School of Medicine, Northwestern University. Chicago, Illinois, USA.

[*]MA and RLR contributed equally to this work.

Abstract

The understanding of the evolutionary processes underlying HIV-1 fitness recovery is fundamental for HIV-1 pathogenesis, antiretroviral treatment and vaccine design. It is known that HIV-1 can present very high mutation and recombination rates, however the specific contribution of these evolutionary forces in the "*in vitro*" viral fitness recovery has not been simultaneously quantified. To this aim, we analyzed substitution, recombination and molecular adaptation rates in a variety of HIV-1 biological clones derived from a viral isolate after severe population bottlenecks and a number of large population cell culture passages. These clones presented an overall but uneven fitness gain, mean of 3-fold, respect to the initial passage values. We found a significant relationship between the fitness increase and the appearance and fixation of mutations. In addition, these fixed mutations presented molecular signatures of positive selection through the accumulation of non-synonymous substitutions. Interestingly, viral recombination correlated with fitness recovery in most of studied viral quasispecies. The genetic diversity generated by these evolutionary processes was positively correlated with the viral fitness. We conclude that HIV-1 fitness recovery can be derived from the genetic heterogeneity generated through both mutation and recombination, and under diversifying molecular adaptation. The findings also suggest nonrandom evolutionary pathways for *in vitro* fitness recovery.

Keywords: HIV-1 molecular evolution; mutation; viral recombination; molecular adaptation; genetic heterogeneity; fitness recovery

1. Introduction

In HIV-1 natural infections, viral evolution is affected by different factors, among them is the viral population size (Coffin, 1995). Changes in the viral population size include severe population bottlenecks that may arise during a variety of processes like transmission (e.g., Keele et al., 2008), antiretroviral drug therapy (e.g., Arenas, 2015; Ibanez et al., 2000), invasion by a new virus (e.g., da Silva, 2012; Golani et al., 2007) or under the host-virus arm races (e.g., Daugherty and Malik, 2012). As a consequence, the study on the influence of population fluctuations and molecular mechanisms on viral fitness alterations (either increases or decreases) is fundamental for the understanding of viral evolution.

As expected after severe population bottlenecks, an active mutational process is required for fitness recovery (e.g., Lorenzo-Redondo et al., 2011; Poon et al., 2007). This process occurs thanks to the high HIV-1 mutation rate because of the lack of proof reading activity of the reverse transcriptase (Mansky and Temin, 1995). Additionally, the contribution of recombination to HIV-1 *in vitro* fitness recovery has not yet been completely identified. Handany and Beker (2003) showed a complex evolutionary advantage of fitness-associated recombination that could generate new and advantageous viral strains but also could break down existing good ones. Other authors showed the influence of recombination on fitness landscapes depending on the particular landscape topology, population size, mutation-selection balance and recombination rate (see Hadany and Beker, 2003; Moradigaravand and Engelstadter, 2012). Several works studied the role of recombination in the generation of antiviral drug resistant variants (Gheorghiu-Svirschevski et al., 2007; Kellam and Larder, 1994, 1995; Richman et al., 1991; Rouzine and Coffin, 2005), the presence of recombination in complex *in vivo* systems (Batorsky et al., 2010) or in theoretical studies that mimic

68 recombination in HIV evolution (Carvajal-Rodriguez et al., 2007; Rouzine and Coffin,
69 2010). Nevertheless, in order to clearly study influences of recombination on the fitness
70 of viral populations, *in vitro* studies are useful because they avoid *in vivo* processes that
71 could confound such influences (i.e., compartmentalization, transmission between
72 individuals or antiviral therapies).
73 HIV-1 presents very high recombination rates because of the pseudodiploid nature of
74 the virus (Jetzt et al., 2000a), which favors the generation of large viral diversity (Perez-
75 Losada et al., 2015; Smyth et al., 2012). Despite the wide diversity of recombinant
76 forms in HIV populations, it was found that only a minority of recombination events are
77 significant to viral evolution (Archer et al., 2008; Iglesias-Sanchez and Lopez-Galindez,
78 2002). In addition, most of recombinant forms present low fitness, especially if the viral
79 population evolves under strong selective pressures (Bretscher et al., 2004; Nijhuis et
80 al., 1998). In contrast, depending on parameters such as population size, recombination
81 and mutation rates, other authors found that recombination can accelerate adaptation
82 through both the fitness effects of individual mutations and epistatic fitness interactions
83 (Carobene et al., 2009; Moradigaravand et al., 2014).
84 Fitness recovery was observed in diverse RNA viruses (Elena et al., 1996; Novella et
85 al., 1995), reviewed in (Domingo et al., 2012). In order to study the effects of
86 population changes in HIV-1 evolution, several *in vitro* experiments were carried out in
87 our laboratory (Borderia et al., 2010; Lorenzo-Redondo et al., 2011; Yuste et al., 1999).
88 Initially, HIV-1 biological clones from a natural isolate were subjected to serial plaque-
89 to-plaque passages (Ebert, 1998) producing drastic fitness losses, known as the Muller
90 ratchet effect (Yuste et al., 1999), also named as mutation accumulation experiments
91 (Eyre-Walker and Keightley, 2007). This effect has been described in a number of RNA
92 viruses (Chao, 1990; Escarmis et al., 2009; Novella, 2004). Previous studies showed a

93 recovery of fitness after a total of 20 large population passages of debilitated HIV-1
94 viruses (see Borderia et al., 2010; Lorenzo-Redondo et al., 2011). Viral populations
95 from other viruses tend to reach a fitness plateau after a given number of passages
96 (Escarmis et al., 1999), however it is unknown when such a plateau can be reached in
97 HIV-1 *in vitro* large population passages and thus the evolution of fitness with
98 additional passages should be analyzed.
99 In this study, we extended our previous works by including 10 additional large
100 population passages (reaching a total of 30 large population passages) and analyzed the
101 specific roles of mutation, recombination and molecular adaptation in HIV-1 *in vitro*
102 fitness recovery. We found that the new passages showed a non-uniform fitness
103 variation and each viral population evolved following a particular evolutionary
104 trajectory, suggesting a diverse and complex process of HIV-1 fitness recovery. Indeed,
105 we observed an increase in genetic diversity with the number of passages for all viral
106 populations with a prevalence of non-synonymous substitutions suggesting the presence
107 of diversifying (positive) selection at the molecular level where viruses evolve toward
108 new and more adapted variants. Interestingly, genetic recombination was detected in
109 several clones and it also increased with the passages until reaching a high fitness peak.
110 In general, both substitution and recombination rates positively correlated with HIV-1
111 *in vitro* fitness recovery, not only at the consensus sequence but also at the quasispecies
112 level.
113

114 ## 2. Materials and Methods

115 *2.1. Cells, viral populations, biological cloning, large population passages, viral*
116 *quasispecies and fitness*

117 The preparation of cells, viral populations, biological cloning and large population
118 passages of virus are described in (Lorenzo-Redondo et al., 2011; Yuste et al., 1999)
119 and in the Supplementary Material. A total of 10 clones D1, D2, E1, G1, G2, H1, I1, I5,
120 K1 and K2 derived from plaque-to-plaque transfers (severe population bottleneck) were
121 subjected to 30 large population passages by infection in $2.5x10^6$ MT-4 cells (Figure S1,
122 Supplementary Material). For each passage, fitness (replicative capacity) was calculated
123 in competition cultures against a common reference virus following (Borderia et al.,
124 2010; Lorenzo-Redondo et al., 2011; Yuste et al., 1999) and described in the
125 Supplementary Material. Global sequences and viral quasispecies were analyzed in the
126 genomic regions – *gag, vpu, V1-V2* and *V3-V4* in *env* – in at least 20 independent clones
127 for each viral population and passage.
128
129 *2.2. Estimation of genetic diversity and molecular adaptation*
130 Genetic diversity was estimated with the *p*-distance (proportion of nucleotide sites at
131 which the sequences being evaluated are different) calculated for every viral population
132 and passage with the maximum composite likelihood method (Tamura et al., 2004)
133 implemented in the *MEGA6* software (Tamura et al., 2013). *p*-distance was used as a
134 heterogeneity measure to allow the comparison of the different genomic regions (*vpu*,
135 *gag*, *env*). Mean values of *p*-distances from the genomic regions were used for the
136 analysis per passage.
137 Due to the quasispecies population structure, competition is always happening among
138 the different mutants existing and being generated in the QS during the cell cultures.
139 Without any external selection (*in vitro* system) the QS itself may generate evolutionary
140 pressure and selection. To explore this aim we analyzed synonymous (*dS*) and non-
141 synonymous (*dN*) mutations within quasispecies in each region with the modified Nei-

142 Gojobori method (Nei and Gojobori, 1986) implemented in *MEGA* (additional details
143 are shown in the supplementary material). The observed *dN* and *dS* for each viral
144 population were compared at different passages and tested for statistical differences.
145 Currently the best evidence of signatures of molecular adaptation in HIV evolution is
146 provided by *dN* and *dS* (e.g., Edwards et al., 2006; Perez-Losada et al., 2011; Perez-
147 Losada et al., 2009). Indeed, it is known that recombination can increase the number of
148 false positively selected sites through phylogenetic tree discordance but also it does not
149 influence molecular adaptation estimates at the global (sequence) level (Anisimova et
150 al., 2003; Arenas and Posada, 2010a, 2014). Here we provide estimates, by a method
151 that is not based on phylogenetic tree inference, at the global sequence level.
152
153 *2.3. Estimation of recombination rates*
154 Population recombination rates $\rho=4NrL$, where N is the effective population size, r is
155 the recombination rate per site and L is the sequence length in nucleotides, were
156 estimated with the *OmegaMap* program (Wilson and McVean, 2006). Following other
157 analysis of HIV-1 data (Arenas and Posada, 2010b; Lopes et al., 2014), we used a
158 uniform $U(0,100)$ prior distribution for ρ [also note that this prior covers estimated
159 values from diverse HIV-1 data (Carvajal-Rodriguez et al., 2006) and the estimated
160 values fell well within the prior (Results)] and all the other priors were set to the
161 software default, as recommended by the authors. A total of 5 independent runs were
162 performed for each genomic region and passage, and each run was based on 1,000,000
163 iterations following the author's recommendation. A burn-in of 10% of the total number
164 of iterations was applied also following the author's recommendation and we found that
165 after removing these first iterations, the effect of the starting values on the estimation is
166 eliminated (i.e. Figure S6, Supplementary Material), which can improve the estimation.

All independent runs reached convergence (assessed as a ρ distance between all the different runs lower than 1). Following this procedure we computed ρ for each passage and genomic region.

Additionally, clonal sequences of each passage and genomic region were used to infer quasispecies recombination networks [ancestral recombination graphs (Arenas, 2013; Griffiths and Marjoram, 1997)] in the different viral lineages with the program *SplitsTree4* (Huson, 1998). The number of reticulate nodes was considered as a measure of recombination events (Arenas, 2013; Arenas et al., 2010; Arenas et al., 2008; Huson and Bryant, 2006; Morrison, 2005).

2.4. Statistical analyses

Statistical analyses were performed with *Graphpad Prism* (http://www.graphpad.com/scientific-software/prism/) and *PasW Statistics17* (http://www.spss.com.hk/statistics/) software. For comparisons between passages, One-Way ANOVA with repeated measures and Bonferroni correction were used. Spearman rank correlation analyses were performed to study correlations of fitness with heterogeneity and recombination at passages levels. Note that this correlation analysis is informative for the representation of the relationship between the parameters from the ordered sets under a monotonic function.

3. Results

3.1. Fitness recovery in HIV-1 viral clones

A previous study with HIV-1 biological clones, debilitated due to plaque-to-plaque passages (Yuste et al., 1999), showed a slow fitness increase after 21 passages. In this study we found a general increase of fitness after 10 additional serial large population

192 passages (Table 1). While the mean fitness in passage 1 was 0.52, it raised to 0.75 at
193 passage 11, 1.03 at passage 21 and 1.57 at passage 31 (Table 1). This final increase is
194 around 3-fold over the initial values.
195 Interestingly, we observed a great variability in the recovery patterns among the viral
196 populations (Table 1). Individual fitness increases ranged from 7.5-fold in clone G1, or
197 5.6× in H1.5, to a non-significant increase or stability in clones E1.5 and K2 (Table 1).
198 Most of the viral populations showed significant fitness increases at passage 31 (Table
199 1). Although two viral populations showed fitness losses in the 10 final passages (D1.5
200 and G1.5, Table 1) that could be associated with a temporal excessive specialization
201 after 20 passages. The specialization was referred to viral replication, which is the main
202 property under which viruses are evolving in tissue cultures. Viruses D1.5 and G1.5
203 showed very high titers at passage 20, the highest of all lineages, showing that there
204 were replicating very efficiently. This process has resulted in a posterior
205 homogenization of these viral populations where the signal of the advantageous variants
206 could be removed. Thus we hypothesize that this homogenization of the viral population
207 could be the cause of a fitness decrease in the following 10 passages. The different
208 alterations detected in viral populations also suggest distinct evolutionary patterns of *in*
209 *vitro* HIV-1 recovery of fitness.
210
211 *3.2. Influence of mutations on fitness recovery*
212 The consensus viral genomic sequences obtained from the large population passages
213 were systematically analyzed for the detection of genetic changes behind fitness
214 recovery. In the first 10 recovery passages, only 24 mutations were observed, including
215 some viral populations without fixed mutations. In passage 31, after an exponential
216 increase in the fixed mutations at passage 21, a moderate increase of the total number of

mutations took place with a total of 100 mutations in the global sequences (Figure 1A). Overall, 80% of the substitutions appearing during the serial passages were non-synonymous (Figure 1A) suggesting the predominance of positive selection in the consensus sequences.

The fixation of mutations (substitutions) during the passages was also analyzed. Among the changes detected, more than 60% of the mutations appearing in the 10 initial passages persisted at passages 21 and 31 (Figure 1B). Interestingly, only 26% of the mutations fixed at passage 21, persisted at passage 31 (Figure 1B), which suggests an increase of selective constraints (although the number of non-synonymous mutations decreased from passage 21 to 31 (Figure 1A), this number was still much higher than the corresponding number of synonymous mutations). On the other hand, the other observed mutations always reverted to the original nucleotide present in the virus before the recovery passages. Altogether, these findings suggest an important role of the mutations fixed during the initial passages, which according to their maintenance, constitute the genetic basis for the subsequent fitness recovery of the final viruses.

3.3. Analysis of molecular adaptation in viral quasispecies during fitness recovery

The investigation of the evolutionary mechanisms operating during HIV-1 fitness recovery was also carried out at the quasispecies level. Quasispecies represents the swarm of different variants present in any RNA viral population according to the model proposed by Eigen (1971) and applied to viruses by Domingo et al. (1985). The mutant spectrum of the different viral populations was here analyzed in different genomic regions (see Methods and Supplementary Material). Taking into account all passages and lineages, quasispecies heterogeneity in these regions was estimated by the mean of

241 the genetic distance expressed in number of substitutions per site with a maximum
242 likelihood method (see Methods) and represented per passage in Figure 2.
243 For the investigation of the evolutionary forces operating in viral quasispecies, we
244 analyzed the type of mutations (synonymous and non-synonymous) within the mutant
245 spectra along the large population passages. In this analysis, synonymous mutations
246 were constant during the passages with an average of 4 mutations per region (data not
247 shown). We found variation of *dN-dS* among genomic regions and viral populations but,
248 in general, the last passages presented the highest signatures of positive selection in
249 most of viral populations (Figure S2, Supplementary Material). In addition, when
250 correlation analyses were performed between *dN* or *dS* substitution rates and fitness
251 values, this correlation was only significant for *dN* (*p*-value = 0.0004) (Figure 3). This
252 preponderance of non-synonymous mutations within quasispecies indicated that not
253 only the consensus sequences, but also the quasispecies composition evolved under
254 positive selection.
255
256 *3.4. Influence of recombination on fitness recovery*
257 Taking into account the limitations of the analytical method (see Methods and
258 Supplementary Material), the estimated recombination rates in most of viral populations
259 (8 out of 12; D1.5, D2, G1, G1.5, G2, H1.5, I5 and K1) showed that recombination
260 might have also contributed to the fitness recovery (Figure 4). Concerning genomic
261 regions, the highest recombination rates were detected in *gag* and *env*, whilst *vpu*
262 showed recombination rates close to 0 (Figure S3, Supplementary Material).
263 Interestingly, the highest recombination rates were observed in the different viral
264 populations at passages immediately previous to the passages with the highest fitness
265 values (Figure 4). Similar results were obtained from the number of reticulate nodes of

the inferred recombination networks (not shown). Nevertheless, the viral populations D1 and I1 increased fitness without presenting large recombination rates and the viral population K2 showed the opposite pattern (Figure 4). Globally throughout the large population passages (from 1 to 31), the estimated recombination rates and the number of reticulations positively correlated with fitness by Spearman correlation (nonparametric monotonic function) with *p*-values 0.0002 and 0.008, respectively (see also Figure S4, Supplementary Material).

3.5. Influence of genetic diversity in viral quasispecies on fitness recovery

We found variable genetic diversity among genomic regions and viral populations but, in general, the last passages (p21-p31) presented the highest contribution to the final genetic heterogeneity (Figure S5, Supplementary Material). The quasispecies heterogeneity of the viral populations throughout the large population passages was also analyzed with the Spearman, Pearson and mixed-effects tests of correlation. All of them generated a significant positive correlation between heterogeneity and fitness with an increase from 0.325 at passage 1 to 0.4391 substitutions per site in passage 31 (p-value < 0.05) (Figure 2). The obtained p-values were 0.03, 0.006 and 0.07 from the Pearson correlation, Spearman correlation and best mixed-effects methods, respectively.

4. Discussion

Fitness recovery can occur in nature in a variety of evolutionary scenarios where population size contractions and expansions affect genetic diversity (Arenas et al., 2012; Ibanez et al., 2000). For example, in HIV-1 low genetic diversification can be observed under antiretroviral drug therapy (Ercoli et al., 1997) or as a consequence of immune system-virus arms races (Hamoudi et al., 2013). After such population contractions,

291 genetic diversity and fitness can be recovered (Escarmis et al., 1999; Hadany and Beker,
292 2003; Moradigaravand and Engelstadter, 2012). In this study we presented the
293 underlying evolutionary mechanisms of *in vitro* HIV-1 fitness recovery. Overall the
294 results indicate that, in most of studied viral populations, the replication of HIV-1
295 during 30 large population passages results in a general viral fitness increase, especially
296 during the last passages.
297 The fitness increase was achieved in different, but specific, evolutionary pathways
298 depending on the viral clone (although no parallel cultures were carried out) and
299 associated with the accumulation of genetic changes, the fixation of non-synonymous
300 mutations and, often, the emergence of recombination events. Then, as expected, such
301 evolutionary processes increased the genetic heterogeneity among viral quasispecies.
302 The accumulation of non-synonymous substitutions indicated episodes of positive
303 selection during the fitness recovery that suggests an active evolution toward different
304 and more adapted viruses. Altogether, our results indicate that the generation of HIV-1
305 genetic heterogeneity through mutation, often contributed by the shuffling of genetic
306 material through recombination, is fundamental for HIV-1 *in vitro* fitness recovery.
307
308 *4.1. Diversifying selection during fitness recovery*
309 We detected evidence of positive selection operating at the consensus sequence level
310 (Figure 1A), but more interestingly, we also observed that the quasispecies composition
311 was shaped by positive selection (Figure 3). The signatures of molecular adaptation
312 varied among viral populations and genomic regions (Figure S2) although we found an
313 overall increase of positive selection at last passages (Figure S2, total). This suggests
314 that the different viral populations and regions present different evolutionary trajectories
315 to finally increase diversity of the encoded proteins. Most of non-synonymous

316 mutations did not become dominant in the population, but they seem to provide an
317 important improvement of the quasispecies because of the accumulation of higher
318 fitness variants based on a better mutant background (see also Borderia et al., 2010;
319 Yusim et al., 2001). The strong preponderance of non-synonymous changes in the
320 global populations and in the viral quasispecies also indicates a non-random
321 evolutionary process for fitness increase.
322
323 *4.2. Recombination can promote fitness recovery*
324 The estimation of recombination rates can be influenced by the amount of genetic
325 diversity (Posada and Crandall, 2001). Under low genetic diversity a recombination
326 event could be not detected but also its resulting recombinant sequence will be very
327 similar to the parental sequences and probably, with similar properties. Therefore, in
328 this study, the observed recombination could lead to fitness variation whilst
329 recombination that was not detected (because of the limitations of the analytical
330 method), probably did not. Recombination is a very important viral evolutionary force
331 but its incidence during HIV-1 *in vitro* fitness recovery has not been yet quantified. We
332 observed that recombination was more frequent in *gag* and *env* than in *vpu* genes
333 (Figure S3). Recombination events can be present along the entire genome (Archer et
334 al., 2008; Jetzt et al., 2000b). However, as noted by Archer et al. (2008), only a minority
335 of recombinant events (in particular, short regions within *gag*, *pol* and *env*) are of
336 significance to the evolution of the virus because of the action of natural selection
337 (Archer et al., 2008). We believe that our *in vitro* study also suggest this finding where
338 only the favored recombination forms, through the operation of natural selection along
339 the population passages, are finally established.

In all, recombination rate correlated with fitness recovery under a Spearman correlation (see Results), but we also noted a non-linear relationship based on several particularities. Unexpectedly, many viral populations presented high recombination rates at passage 21 whereas the higher fitness values were detected in passage 31 (Figure 4). A possible explanation for the fact that recombination does not increase fitness immediately, is probably related to the requirement of posterior mutations for the optimization of the new recombinant forms to gain fitness. Additionally, a new recombinant form needs to reach a certain frequency in the viral population to have a detectable effect on fitness (Galli et al., 2010; Iglesias-Sanchez and Lopez-Galindez, 2002). Also, viral populations D1 and I1 recovered fitness without the need of recombination events (Figure S2). We believe that in these situations, the mutation process could be sufficient to generate improved viral forms and thus, most of recombination events do not provide better viral forms and/or are deleterious, as described by some authors (see Background and Archer et al., 2008; Iglesias-Sanchez and Lopez-Galindez, 2002). Indeed, viral population K2 presented large recombination at passage 21 with no fitness increase. In this case, we believe that the generated recombinant forms could also be removed by selection and therefore do not contribute to the fitness recovery (see Archer et al., 2008; Bretscher et al., 2004; Nijhuis et al., 1998). Altogether, our results suggest that, as expected, only advantageous recombination events during HIV-1 *in vitro* fitness recovery can lead to fitness increase.

4.3. Genetic diversity drives fitness recovery

We found that genetic diversity varies among viral populations and genomic regions, which also suggests different evolutionary trajectories among viral populations. Overall we found a general trend to increase the amount of genetic diversity during the last large

365 population passages (p21-p31, Figure S5). Therefore, our results on the role of genetic
366 heterogeneity in fitness recovery point to the importance of the generation of diversity
367 to improve fitness. Note that Novella *et* al. (1995) showed an exponential fitness
368 increase due to large populations passages in vesicular stomatitis virus (VSV) mutants.
369 The strong correlation between increase of genetic heterogeneity and fitness recovery is
370 in concordance with previous works of our group showing phenotypic effects of
371 quasispecies composition for HIV-1 during a few populations passages (Borderia et al.,
372 2010; Lorenzo-Redondo et al., 2011) and with other studies based on diverse RNA
373 viruses (reviewed in Domingo et al., 2012). In addition, studies performed with HIV-1
374 *in vivo* also showed positive correlation of genetic diversity with viral fitness and
375 disease progression (Ragonnet-Cronin et al., 2012; Shankarappa et al., 1999).
376

377 ## 5. Conclusions

378 We conclude that HIV-1 fitness recovery is a dynamic and complex process driven by
379 quasispecies heterogeneity that can be generated by mutation and recombination and
380 under positive selection for the better-adapted variants to the new environment. Thus
381 the process follows a "Darwinian" dynamic with the generation of variability and
382 selective pressure towards the fittest variants. This knowledge can be of great help for
383 the understanding of the course of HIV-1 natural infection and the evolutionary
384 mechanisms by which, for example, viral strains with resistance mutations arise after
385 severe population bottlenecks derived from antiviral treatments (Condra et al., 1995).
386

387 ## Acknowledgments

388 We want to thank Daniel Wilson for helpful comments. We also want to thanks two
389 anonymous reviewers for their insightful comments. Work in CNM is supported by

grants SAF SAF 2010-17226 from MICIN Spain and grant FIS (PI 13/02269) from the Fondo de Investigaciones Sanitarias (ISCIII) and in part by the RETIC de Investigación en SIDA (Red de grupos 173) of the Fondo de Investigaciones Sanitarias (FIS). This work has been partially funded by the RD12/0017/0036 Project as part of the Plan Nacional R + D + I and cofinanced by ISCIII, Subdirección General de Evaluación and Fondo Europeo de Desarrollo Regional (FEDER). MA was supported by the Spanish Government through the "Juan de la Cierva" fellowship JCI-2011-10452 and by the Portuguese Government through the FCT Starting Grant IF/00955/2014.

References

Anisimova, M., Nielsen, R., Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164, 1229-1236.

Archer, J., Pinney, J.W., Fan, J., Simon-Loriere, E., Arts, E.J., Negroni, M., Robertson, D.L., 2008. Identifying the important HIV-1 recombination breakpoints. PLoS Comput Biol 4, e1000178.

Arenas, M., 2013. The Importance and Application of the Ancestral Recombination Graph. Front Genet 4, 206.

Arenas, M., 2015. Genetic Consequences of Antiviral Therapy on HIV-1. Computational and Mathematical Methods in Medicine 2015, 9.

Arenas, M., Patricio, M., Posada, D., Valiente, G., 2010. Characterization of phylogenetic networks with NetTest. BMC Bioinformatics 11, 268.

Arenas, M., Posada, D., 2010a. Coalescent simulation of intracodon recombination. Genetics 184, 429-437.

Arenas, M., Posada, D., 2010b. The effect of recombination on the reconstruction of ancestral sequences. Genetics 184, 1133-1139.

Arenas, M., Posada, D., 2014. The influence of recombination on the estimation of selection from coding sequence alignments. In: Fares, M.A. (Ed.), Natural Selection: Methods and Applications. CRC Press/Taylor & Francis, Boca Raton, pp. 112-125.

Arenas, M., Ray, N., Currat, M., Excoffier, L., 2012. Consequences of Range Contractions and Range Shifts on Molecular Diversity. Mol Biol Evol 29, 207-218.

Arenas, M., Valiente, G., Posada, D., 2008. Characterization of reticulate networks based on the coalescent with recombination. Mol Biol Evol 25, 2517-2520.

Batorsky, R., Kearney, M.F., Palmer, S.E., Maldarelli, F., Rouzine, I.M., Coffin, J.M., 2010. Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. Proc Natl Acad Sci U S A 108, 5661-5666.

Borderia, A.V., Lorenzo-Redondo, R., Pernas, M., Casado, C., Alvaro, T., Domingo, E., Lopez-Galindez, C., 2010. Initial fitness recovery of HIV-1 is associated with quasispecies heterogeneity and can occur without modifications in the consensus sequence. PLoS One 5, e10319.

Bretscher, M.T., Althaus, C.L., Muller, V., Bonhoeffer, S., 2004. Recombination in HIV and the evolution of drug resistance: for better or for worse? Bioessays 26, 180-188.

Carobene, M.G., Rodriguez Rodrigues, C., De Candia, C.A., Turk, G., Salomon, H., 2009. In vitro dynamics of HIV-1 BF intersubtype recombinants genomic regions involved in the regulation of gene expression. Virol J 6, 107.

Carvajal-Rodriguez, A., Crandall, K.A., Posada, D., 2006. Recombination estimation under complex evolutionary models with the coalescent composite-likelihood method. Mol Biol Evol 23, 817-827.

Carvajal-Rodriguez, A., Crandall, K.A., Posada, D., 2007. Recombination favors the evolution of drug resistance in HIV-1 during antiretroviral therapy. Infect Genet Evol 7, 476-483.

Chao, L., 1990. Fitness of RNA virus decreased by Muller's ratchet. Nature 348, 454-455.

Coffin, J.M., 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science 267, 483-489.

Condra, J.H., Schleif, W.A., Blahy, O.M., Gabryelski, L.J., Graham, D.J., Quintero, J.C., Rhodes, A., Robbins, H.L., Roth, E., Shivaprakash, M., et al., 1995. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. Nature 374, 569-571.

da Silva, J., 2012. The dynamics of HIV-1 adaptation in early infection. Genetics 190, 1087-1099.

Daugherty, M.D., Malik, H.S., 2012. Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet 46, 677-700.

Domingo, E., Martinez-Salas, E., Sobrino, F., de la Torre, J.C., Portela, A., Ortin, J., Lopez-Galindez, C., Perez-Brena, P., Villanueva, N., Najera, R., et al., 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance--a review. Gene 40, 1-8.

Domingo, E., Sheldon, J., Perales, C., 2012. Viral quasispecies evolution. Microbiol Mol Biol Rev 76, 159-216.

Ebert, D., 1998. Experimental evolution of parasites. Science 282, 1432-1435.

Edwards, C.T., Holmes, E.C., Pybus, O.G., Wilson, D.J., Viscidi, R.P., Abrams, E.J., Phillips, R.E., Drummond, A.J., 2006. Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection. Genetics 174, 1441-1453.

Eigen, M., 1971. Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58, 465-523.

Elena, S.F., Gonzalez-Candelas, F., Novella, I.S., Duarte, E.A., Clarke, D.K., Domingo, E., Holland, J.J., Moya, A., 1996. Evolution of fitness in experimental populations of vesicular stomatitis virus. Genetics 142, 673-679.

Ercoli, L., Sarmati, L., Nicastri, E., Giannini, G., Galluzzo, C., Vella, S., Andreoni, M., 1997. HIV phenotype switching during antiretroviral therapy: emergence of saquinavir-resistant strains with less cytopathogenicity. Aids 11, 1211-1217.

Escarmis, C., Davila, M., Domingo, E., 1999. Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet. J Mol Biol 285, 495-505.

Escarmis, C., Perales, C., Domingo, E., 2009. Biological effect of Muller's Ratchet: distant capsid site can affect picornavirus protein processing. J Virol 83, 6748-6756.

Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. Nat Rev Genet 8, 610-618.

Galli, A., Kearney, M., Nikolaitchik, O.A., Yu, S., Chin, M.P., Maldarelli, F., Coffin, J.M., Pathak, V.K., Hu, W.S., 2010. Patterns of Human Immunodeficiency Virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. J Virol 84, 7651-7661.

Gheorghiu-Svirschevski, S., Rouzine, I.M., Coffin, J.M., 2007. Increasing sequence correlation limits the efficiency of recombination in a multisite evolution model. Mol Biol Evol 24, 574-586.

Golani, D., Azzurro, E., Corsini-Foka, M., Falautano, M., Andaloro, F., Bernardi, G., 2007. Genetic bottlenecks and successful biological invasions: the case of a recent Lessepsian migrant. Biol Lett 3, 541-545.

Griffiths, R.C., Marjoram, P., 1997. An ancestral recombination graph. In: Donelly, P., Tavaré, S. (Eds.), Progress in population genetics and human evolution. Springer-Verlag, Berlin, pp. 257-270.

Hadany, L., Beker, T., 2003. On the evolutionary advantage of fitness-associated recombination. Genetics 165, 2167-2179.

Hamoudi, M., Simon-Loriere, E., Gasser, R., Negroni, M., 2013. Genetic diversity of the highly variable V1 region interferes with Human Immunodeficiency Virus type 1 envelope functionality. Retrovirology 10, 114.

Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14, 68-73.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23, 254-267.

Ibanez, A., Clotet, B., Martinez, M.A., 2000. Human immunodeficiency virus type 1 population bottleneck during indinavir therapy causes a genetic drift in the env quasispecies. J Gen Virol 81, 85-95.

Iglesias-Sanchez, M.J., Lopez-Galindez, C., 2002. Analysis, quantification, and evolutionary consequences of HIV-1 in vitro recombination. Virology 304, 392-402.

Jetzt, A., Yu, H., Klarmann, G.J., Ron, Y., Preston, B.D., Dougherty, J.P., 2000a. High rate of recombination throughout the human immunodeficiency virus type 1 genome. J Virol 74, 1234-1240.

Jetzt, A.E., Yu, H., Klarmann, G.J., Ron, Y., Preston, B.D., Dougherty, J.P., 2000b. High rate of recombination throughout the human immunodeficiency virus type 1 genome. J Virol 74, 1234-1240.

Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J.L., Gao, F., Anderson, J.A., Ping, L.H., Swanstrom, R., Tomaras, G.D., Blattner, W.A., Goepfert, P.A., Kilby, J.M., Saag, M.S., Delwart, E.L., Busch, M.P., Cohen, M.S., Montefiori, D.C., Haynes, B.F., Gaschen, B., Athreya, G.S., Lee, H.Y., Wood, N., Seoighe, C., Perelson, A.S., Bhattacharya, T., Korber, B.T., Hahn, B.H., Shaw, G.M., 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105, 7552-7557.

Kellam, P., Larder, B.A., 1994. Recombinant virus assay: a rapid, phenotypic assay for assessment of drug susceptibility of human immunodeficiency virus type 1 isolates. Antimicrob Agents Chemother 38, 23-30.

Kellam, P., Larder, B.A., 1995. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. J Virol 69, 669-674.

Lopes, J.S., Arenas, M., Posada, D., Beaumont, M.A., 2014. Coestimation of Recombination, Substitution and Molecular Adaptation rates by approximate Bayesian computation. Heredity 112, 255-264.
Lorenzo-Redondo, R., Borderia, A.V., Lopez-Galindez, C., 2011. Dynamics of in vitro fitness recovery of HIV-1. J Virol 85, 1861-1870.
Mansky, L.M., Temin, H.M., 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol 69, 5087-5094.
Moradigaravand, D., Engelstadter, J., 2012. The effect of bacterial recombination on adaptation on fitness landscapes with limited peak accessibility. PLoS Comput Biol 8, e1002735.
Moradigaravand, D., Kouyos, R., Hinkley, T., Haddad, M., Petropoulos, C.J., Engelstadter, J., Bonhoeffer, S., 2014. Recombination accelerates adaptation on a large-scale empirical fitness landscape in HIV-1. PLoS Genet 10, e1004439.
Morrison, D.A., 2005. Networks in phylogenetic analysis: new tools for population biology. Int J Parasitol 35, 567-582.
Nei, M., Gojobori, T., 1986. Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3, 418-426.
Nijhuis, M., Boucher, C.A., Schipper, P., Leitner, T., Schuurman, R., Albert, J., 1998. Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. Proc Natl Acad Sci U S A 95, 14441-14446.
Novella, I.S., 2004. Negative effect of genetic bottlenecks on the adaptability of vesicular stomatitis virus. J Mol Biol 336, 61-67.
Novella, I.S., Duarte, E.A., Elena, S.F., Moya, A., Domingo, E., Holland, J.J., 1995. Exponential increases of RNA virus fitness during large population transmissions. Proc Natl Acad Sci U S A 92, 5841-5844.
Perez-Losada, M., Arenas, M., Galan, J.C., Palero, F., Gonzalez-Candelas, F., 2015. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. Infect Genet Evol 30C, 296-307.
Perez-Losada, M., Jobes, D.V., Sinangil, F., Crandall, K.A., Arenas, M., Posada, D., Berman, P.W., 2011. Phylodynamics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand. PLoS One 6, e16902.
Perez-Losada, M., Posada, D., Arenas, M., Jobes, D.V., Sinangil, F., Berman, P.W., Crandall, K.A., 2009. Ethnic differences in the adaptation rate of HIV gp120 from a vaccine trial. Retrovirology 6, 67.
Poon, A.F., Kosakovsky Pond, S.L., Richman, D.D., Frost, S.D., 2007. Mapping protease inhibitor resistance to human immunodeficiency virus type 1 sequence polymorphisms within patients. J Virol 81, 13598-13607.
Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci U S A 98, 13757-13762.
Ragonnet-Cronin, M., Aris-Brosou, S., Joanisse, I., Merks, H., Vallee, D., Caminiti, K., Rekart, M., Krajden, M., Cook, D., Kim, J., Malloch, L., Sandstrom, P., Brooks, J., 2012. Genetic diversity as a marker for timing infection in HIV-infected patients: evaluation of a 6-month window and comparison with BED. J Infect Dis 206, 756-764.
Richman, D.D., Shih, C.K., Lowy, I., Rose, J., Prodanovich, P., Goff, S., Griffin, J., 1991. Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture. Proc Natl Acad U S A 88, 11241-11245.

577 Rouzine, I.M., Coffin, J.M., 2005. Evolution of human immunodeficiency virus under
578 selection and weak recombination. Genetics 170, 7-18.
579 Rouzine, I.M., Coffin, J.M., 2010. Multi-site adaptation in the presence of infrequent
580 recombination. Theor Popul Biol 77, 189-204.
581 Shankarappa, R., Margolick, J.B., Gange, S.J., Rodrigo, A.G., Upchurch, D.,
582 Farzadegan, H., Gupta, P., Rinaldo, C.R., Learn, G.H., He, X., Huang, X.-L.,
583 Mullins, J.I., 1999. Consistent viral evolutionary changes associated with the
584 progression of human immunodeficiency virus type 1 infection. J Virol 73, 10489-
585 10502.
586 Smyth, R.P., Davenport, M.P., Mak, J., 2012. The origin of genetic diversity in HIV-1.
587 Virus Res 169, 415-429.
588 Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by
589 using the neighbor-joining method. Proc Natl Acad Sci U S A 101, 11030-11035.
590 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6:
591 Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30, 2725-
592 2729.
593 Wilson, D.J., McVean, G., 2006. Estimating diversifying selection and functional
594 constraint in the presence of recombination. Genetics 172, 1411-1425.
595 Yusim, K., Peeters, M., Pybus, O.G., Bhattacharya, T., Delaporte, E., Mulanga, C.,
596 Muldoon, M., Theiler, J., Korber, B., 2001. Using human immunodeficiency virus
597 type 1 sequences to infer historical features of the acquired immune deficiency
598 syndrome epidemic and human immunodeficiency virus evolution. Philos Trans R
599 Soc Lond B Biol Sci 356, 855-866.
600 Yuste, E., Sanchez-Palomino, S., Casado, C., Domingo, E., Lopez-Galindez, C., 1999.
601 Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck
602 events. J Virol 73, 2745-2751.

603

604

Figure captions

Figure 1. Analysis of mutations appearing during the recovery passages. A) Classification by type of mutations including the total number, synonymous (Syn), non-synonymous (Non-syn), non-coding (NC) and new mutations arising in the viral populations at passages (p) 1, 11, 21 and 31. B) Percentage of mutations emerged at passages (p) 11, 21 and 31 that are maintained in subsequent passages.

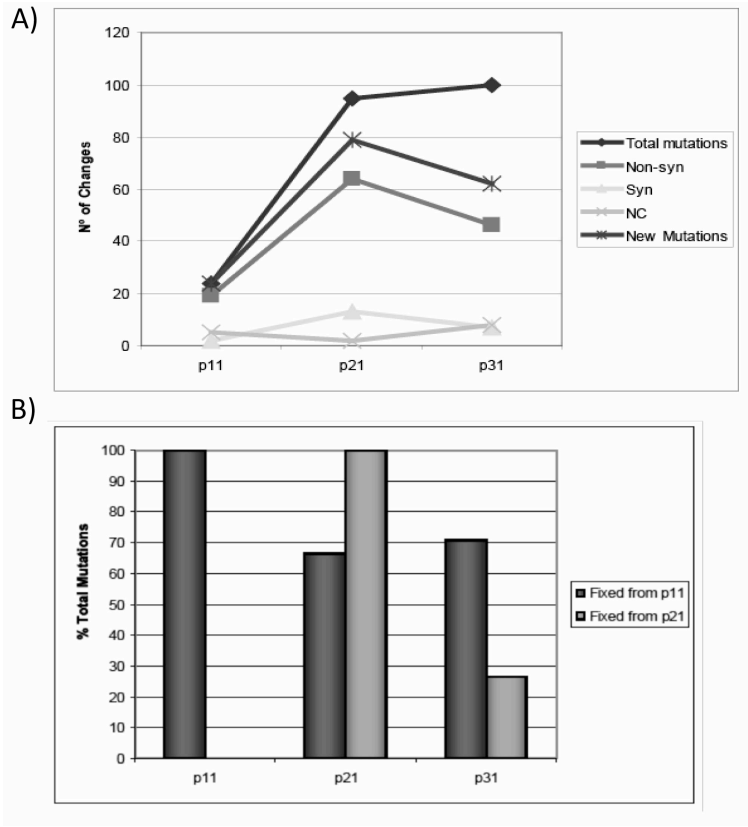


Figure 2. Correlation between quasispecies heterogeneity and fitness throughout the passages. Quasispecies heterogeneity (genetic distance) was analysed using mean values of each viral population in the genomic regions examined. Heterogeneity was measured as the mean genetic distance (number of substitutions per site). Significant correlation of heterogeneity and fitness was found (see Results) and a trend line is depicted to better visualize the studied association. Big grey circles group the average fitness values and average heterogeneity (radius of the circles) at passages 1 (circles), 11 (squares), 21 (up triangles) and 31 (down triangles).

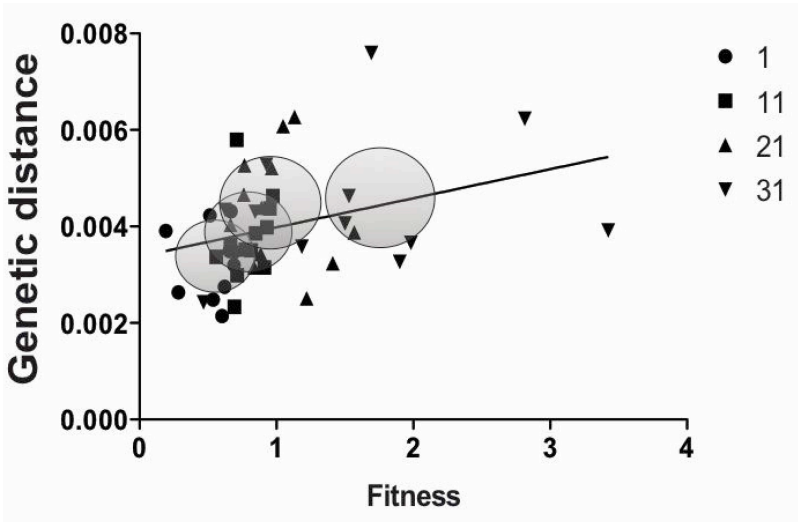


Figure 3. Genetic diversity and molecular adaptation as a function of fitness. Correlation of non-synonymous and synonymous mutations with fitness in the quasispecies data. The linear correlation is derived from a non parametric Spearman test with *p*-value = 0.0004 and the trend lines are depicted to better visualize the association.

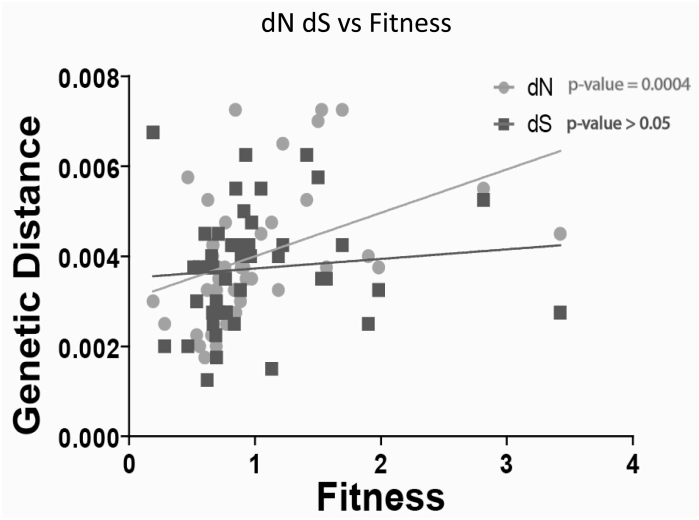
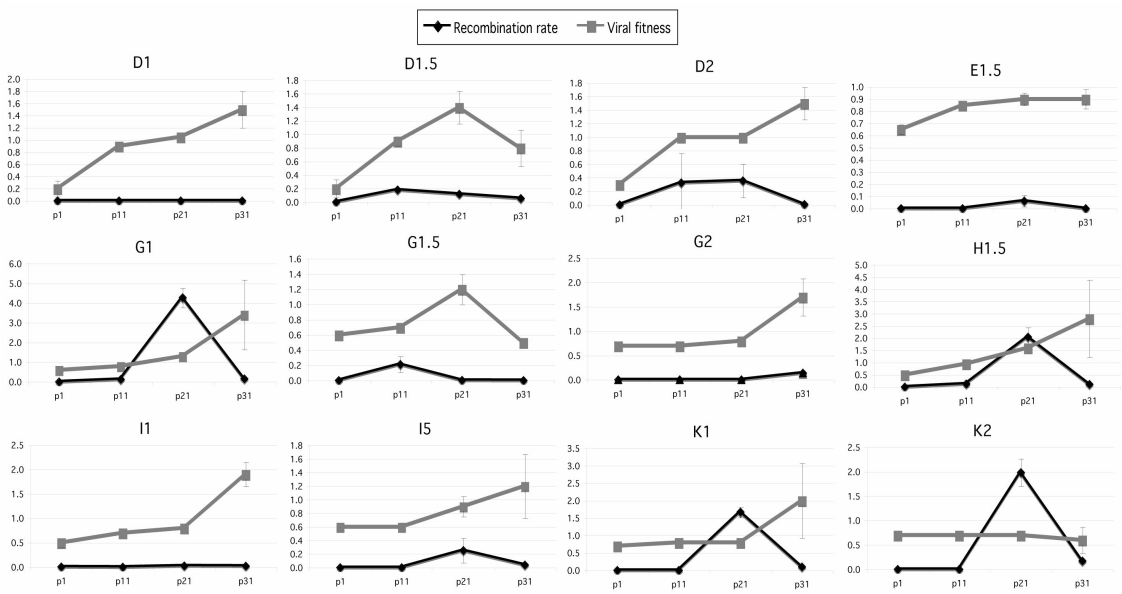


Figure 4. Recombination rate and fitness recovery per viral population and passage. Population recombination rates (rate per codon per 2Ne generations, where Ne is the effective population size) and viral fitness at 1, 11, 21 and 31 large population recovery passages (p). The point estimates are the mean of the estimates from the 5 different runs (see Methods) and error bars indicate credible intervals considering the different runs.



641 # Table captions

642 **Table 1. Fitness variation during the recovery passages.** Fitness values are
643 shown "±" standard deviations. "Total increase" (last column) refers to the fold
644 increase between the initial and the final passages and indicates a fitness increase in
645 most of viral populations. Notice that the passage 31 presents viral populations with
646 significant higher fitness (bold), significant lower fitness (italic) or not significant
647 variation in relation to passage 21. In addition, fitness at passage 31 presents a
648 much higher dispersion than in previous passages suggesting that viral populations
649 evolve under different evolutionary trajectories. Biological fitness values were
650 obtained by competition assays (see Supplementary Material).

651

Viral population	Passages				
	1	11	21	31	Total Increase (×)
D1	0.2±0.13	0.9±0.06	1.05±0.05	**1.5±0.30**	7.5×
D1.5	0.2±0.13	0.9±0.07	1.4±0.24	*0.8±0.27*	4.0×
D2	0.3±0.03	1.0±0.02	1.0±0.01	**1.5±0.24**	5.0×
E1.5	0.65±0.04	0.85±0.03	0.9±0.05	0.9±0.08	1.4×
G1	0.6±0.05	0.8 ±0.03	1.3±0.18	**3.4±1.77**	5.6×
G1.5	0.6±0.05	0.7 ±0.02	1.2±0.20	*0.5±0.04*	0.8×
G2	0.7±0.03	0.7 ±0.01	0.8±0.08	**1.7±0.38**	2.6×
H1.5	0.5±0.04	0.95±0.01	1.6±0.15	2.8±1.58	5.6×
I1	0.5±0.09	0.7±0.02	0.8±0.01	**1.9±0.25**	3.8×
I5	0.6±0.03	0.6±0.03	0.9±0.15	1.2±0.47	2.0×
K1	0.7±0.00	0.8±0.00	0.8±0.02	**2.0±1.07**	2.9×
K2	0.7±0.04	0.7 ±0.04	0.7±0.05	0.6±0.27	0.9×

652