# Parallel evolution of HIV-1 in a long-term experiment

Frederic Bertels[1,2,*], Christine Leemann[3,4], Karin J. Metzner[3,4,*], Roland Regoes[1,*]

[1]Department of Environmental Systems Sciences, ETH Zurich, Zurich

[2]Max-Planck-Institute for Evolutionary Biology, Department of Microbial Population Biology

[3]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich

[4]Insitute of Medical Virology, University of Zurich, Zurich

* To whom correspondence should be addressed: bertels@evolbio.mpg.de, metzner.karin@usz.ch, and roland.regoes@env.ethz.ch

## Abstract

One of the most intriguing puzzles in biology is the degree to which evolution is repeatable. The repeatability of evolution, or parallel evolution, has been studied in a variety of model systems, but has rarely been investigated with clinically relevant viruses. To investigate parallel evolution of HIV-1, we passaged two replicate HIV-1 populations for almost one year in each of two human T-cell lines. For each of the four evolution lines, we determined the genetic composition of the viral population at nine time points by deep sequencing the entire genome. Mutations that were carried by the majority of the viral population accumulated continuously over one year in each evolution line. Many majority mutations appeared in more than one evolution line, i.e. our experiments showed an extreme degree of parallel evolution. In one of the evolution lines, 62% of the majority mutations also occur in another line. The parallelism impairs our ability to reconstruct the evolutionary history by phylogenetic methods. We show that one can infer the correct phylogenetic topology by including minority mutations in our analysis. We also find that mutation diversity at the beginning of the experiment is predictive of the frequency of majority mutations at the end of the experiment.

## Introduction

In long-term evolution experiments, organisms are passaged for many generations in controlled environments in order to understand basic evolutionary principles (Elena and Lenski 2003). One of the most basic questions addressed with experimental evolution is to what extent the evolutionary path of an organism is predetermined, and hence can be predicted and repeated. Stephen J. Gould famously claimed, "[…] any replay of the tape of life would lead evolution down a pathway radically different from the one actually taken […]" thus rejecting the idea of evolution as a repeatable and predictable process (Gould 1989). According to Gould, any deviation from a common evolutionary path will lead to different evolutionary end points. This view is consistent with the neutral theory of evolution (Kimura 1983), according to which genetic changes accumulate mostly randomly without conferring fitness advantages or disadvantages.

Gould's hypothesis has been experimentally tested numerous times by letting populations of organisms evolve repeatedly in the same environment. In such experiments, the parallel evolution of similar traits and, in rare cases, also identical mutations, have been observed in bacteria and viruses (Bull et al. 1997; Wichman et al. 1999; Wichman et al. 2005; Barrick et al. 2009; Engel et al. 2011; Lobkovsky and Koonin 2012; de Visser and Krug 2014; Foll et al. 2014; Bailey et al. 2015; Lind et al. 2015; Lind et al. 2018; Gallie et al. 2019).

Particularly relevant experiments that study parallel evolution in viruses were performed by James Bull and Holly Wichman. In an early experiment phage ϕX174 was grown on two different bacterial hosts, *Escherichia coli* and *Salmonella typhimurium* (Bull et al. 1997). After 11 days of growth in two and three replicates respectively the authors observed 42% of all mutations in more than one replicate line. Due to the extensive parallelism, phylogenetic inference methods, applied to the final sequence data, failed to recover the known evolutionary history of the phage populations. Ever since, multiple follow up experiments have been performed with other phages exploring convergent and parallel evolution (Wichman et al. 1999; Wichman et al. 2000; Wichman et al. 2005; Miller et al. 2016).

There are also examples of experimental evolution studies with clinically relevant viruses. Experimental evolution has been applied to answer a wide range of question about the epidemiology and biology of these viruses. For example, studies have investigated: the effect of different drug concentrations on viral growth and evolution (Gatanaga et al. 2002; Maisnier-Patin and Andersson 2004); the effect of fitness recovery after bottlenecking (Lorenzo-Redondo et al. 2011); the effect of different environmental and genetic parameters on adaptation (Das et al. 1999; van Opijnen and Berkhout 2005; van Opijnen et al. 2007; Vasilakis et al. 2009; Bordería et al. 2015); and the shape of fitness landscapes (Acevedo et al. 2014; Lorenzo-Redondo et al. 2014). Hence, experimental evolution has been a valuable tool to understand the epidemiology and evolution of viruses.

Apart from the insights gained from a detailed analysis of a well-defined experiment, experimental studies of clinically relevant viruses can also take advantage of the vast amount of knowledge gained from clinical studies. Particularly, Human Immunodeficiency Virus (HIV) is one of the most relevant and deadly infections in the human population. In 2012 about 35.3 million people were infected with HIV (Joint United Nations Programme on HIV AIDS 2013). Due to its prevalence and a mortality rate of close to 100% if left untreated, HIV is extremely well studied. In PubMed alone, there are 185,490 articles with HIV in the title (2nd of May 2019).
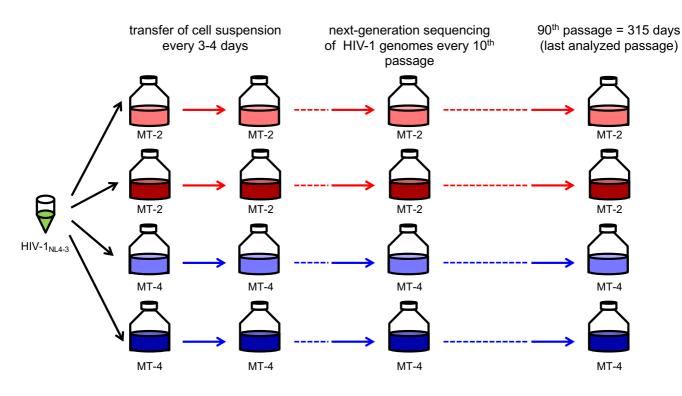
Part of the reason for the evolutionary success of HIV-1 in infecting the human population is its high mutation rate, which contributes to its ability to evade the human immune response and persist for years within the host (Coffin 1995; Koenig et al. 1995; Borrow et al. 1997; Goulder et al. 1997; Wei et al. 2003; Trkola et al. 2005; Liao et al. 2013). Over this time span HIV-1 evolves and acquires a large number of mutations (Shankarappa et al. 1999; Lemey et al. 2006; Keele et al. 2008; Poon et al. 2010). The distribution and type of these mutations is not entirely random but converges between different lineages that evolve in similar hosts or are treated with similar drugs

(Crandall et al. 1999; Duwe et al. 2001; Gatanaga et al. 2002; Lataillade et al. 2007). This convergence even allows the inference of the genetic setup of the human host from the viral sequence (Moore et al. 2002; Bhattacharya et al. 2007).

To study the extent of parallel evolution in a constant environment, we track the evolution of four independent HIV-1 populations for 315 days, which should correspond to about 180 generations (Srivastava et al. 1991; Ho et al. 1995; Wei et al. 1995; Mohammadi et al. 2013; Holmes et al. 2015). We identify all mutations that increase in frequency above that of the wildtype nucleotide, and compare mutations acquired in the four independent lineages. We observe a high degree of parallel evolution that causes phylogenetic reconstruction methods to infer an incorrect evolutionary history. This observation highlights the need to carefully assess the validity of phylogenetic methods that rely on the assumption of neutral evolution when analyzing within-host evolution of HIV-1, in particular phyloanatomic analyses (Salemi and Rife 2015; Lorenzo-Redondo et al. 2016; Bons and Regoes 2018). We also provide insights into how viral population sequencing data can be used to alleviate problems of phylogenetic reconstruction and analyze sequence diversity to predict evolutionary outcomes.

**Results**

We investigated the long-term evolution of HIV-1 in the laboratory by passaging HIV-1 NL4-3 for 315 days in the absence of antiviral drugs in two different human T-cell cultures (MT-2 and MT-4), with two replicates on each T-cell culture (**Fig. 1**). The viral populations were transferred to fresh culture two times every week (90 times in total), which should correspond to about 180 viral generations (Srivastava et al. 1991; Ho et al. 1995; Wei et al. 1995; Mohammadi et al. 2013; Holmes et al. 2015). Every $10^{th}$ transfer, we determined the population composition by Illumina sequencing for each of the four evolution lines.

**Fig. 1. Experimental design.** HIV-1 NL4-3 was serially transferred in MT-2 and MT-4 T-cells in two replicates each. Odd transfers occurred after 3 days, even transfers after 4 days. At each transfer we infected new T-cells with ~0.05% of the cell suspension from the previous culture (**Fig. S1**).

The HIV-1 population grows exponentially between transfers. At the start of each transfer, our cultures contain approximately 400,000 uninfected T-cells. With a multiplicity of infection (MOI) of 0.0001-0.004 (**Fig. S1**), between 40 and 1600 cells are infected during the first generation after each inoculation. The number of initially infected cells (i.e., 40 – 1600) constitute the population bottlenecks in our experiment. Since we transfer about 0.05% of the previous culture into the new culture and we do not observe any obvious population growth or death across our experiment we can assume that each HIV population amplifies 2000-fold between transfers (**Fig. S1**). There are two parameters that determine the level of population amplification in our experiment: the generation time and the number of infective viral offspring ($R_0$). $R_0$ of about 10 (Ribeiro et al. 2010) and the generation time of about two days (Srivastava et al. 1991; Ho et al. 1995; Wei et al. 1995; Mohammadi et al. 2013; Holmes et al. 2015) that are cited in the literature cannot explain the virus amplification in our experiment. These parameters would only allow for a 100-fold amplification of the virus population and hence would lead to the HIV populations dying out in our experiment. To increase the population size 2000-fold we need to either assume $R_0$ to be 44 while the HIV population completes two generations or assume $R_0$ to be 12 while the HIV population completes three generations. The truth may lie somewhere in the middle. Either way we expect the MOI to reach a value close to one at the end of each transfer, which would allow for recombination to happen in our experiment.

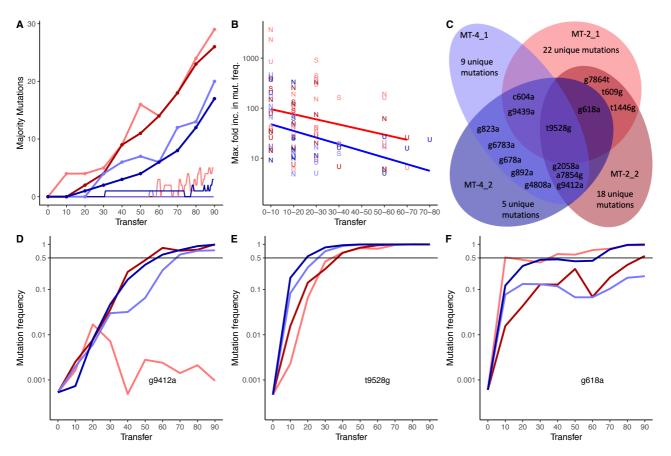**Majority mutations accumulate linearly over time**

We observed a total of 92 mutations across the four evolution lines (69 unique mutations) that became more frequent in the population than the wildtype nucleotide (**Fig. 2A**). We call these mutations majority mutations. Although we expected the accumulation of majority mutations to decelerate when a population approaches an adaptive peak, we did not observe any deceleration during the 90 transfers we considered in this study.

To understand why mutations continue to accumulate at a high rate even towards the end of the experiment we determined the time point of the maximum fold increase of the frequencies for each majority mutation (**Fig. 2B**). These maximum fold increases should correspond to "fitness" because a large increase in the frequency of a novel mutation could be indicative of greater reproductive success of the viruses that carry this mutation. However, one has to keep in mind that the reproductive effect of a novel mutation is not the only factor affecting the frequency increase of that mutation. Random fluctuations (especially for low frequency mutants), clonal interference, recombination, and linkage can also substantially affect mutation frequency differences. Nevertheless, our analysis shows that the "fitness gains" of majority mutations significantly decrease over time at similar rates. However, the fitness gains for majority mutations in MT-2 evolution lines initially are much larger than the fitness gains of majority mutations in MT-4 evolution lines (fitted regression lines in **Fig. 2B**). The difference in the fitness gains suggests that the viruses growing on MT-4 evolution lines might be closer to an adaptive peak than the viruses growing on MT-2 evolution lines.

For all four evolution lines we observe that fitness gains of novel majority mutations decrease over time. This could be due to diminishing fitness effects, or more clonal interference as the sequence diversity increases in the population. Diminishing fitness effects are expected for populations that are adapting to a new environment (Lenski et al. 1991; Wiser et al. 2013). Hence, the fitness effects we observe in our experiment of new mutations are decreasing as expected. Nevertheless, the number of new majority mutations observed at each time step does not decrease (**Fig. 2A**). This pattern is similar to the linear accumulation of mutations observed in the long term evolution experiment in *E. coli*, which the authors explained through an accumulation of a large number of small effect mutations (Barrick et al. 2009). However, for our experiment there is an alternative explanation for the observed pattern. Due to the smaller population sizes in our experiment neutral majority mutations are expected to accumulate towards the end of the experiment. These neutral majority mutations could maintain a constant accumulation rate of majority mutations.

The appearance of neutral majority mutations at the end of the experiment is supported by simulations. We performed simulations of neutrally evolving viruses using experimental data on population sizes at the start of every 10 transfers (**Fig. 2A**, **Fig. S1**). We simulate the evolution of individual HIV-1 genomes for 180 generations with a periodic bottleneck every two generations. Between bottlenecks the simulated HIV-1 populations replicate exponentially ($R_0$=44). Our results suggest that neutral majority mutations appear at the end of the experiment because it takes time for neutral mutations to accumulate in a population before they can become majority mutations (thin lines in **Fig. 2A**).

We also analyzed the effect of APOBEC3G on the accumulation of majority mutations. APOBEC3G is only expressed in MT-2 T-cells and not in MT-4 T-cells (Borman et al. 1995). Interestingly, the proportion of G to A majority mutations at the end of the experiment is higher in HIV-1 growing in MT-4 T-cells (24 G to A mutations or 67%) than in HIV-1 growing in MT-2 T-cells (24 G to A mutations or 44%). Among these mutations five were found in the APOBEC3G motif (GG to GA) in MT-2 and eight were found in MT-4 HIV-1 populations. Hence, it is quite clear that APOBEC3G expression does not drive the accumulation of majority mutations through an increased mutation rate. However, it is not clear whether there is an indirect effect of APOBEC3G expression (e.g. stronger selective pressure) that drives the faster accumulation of majority mutations in MT-2 virus populations.
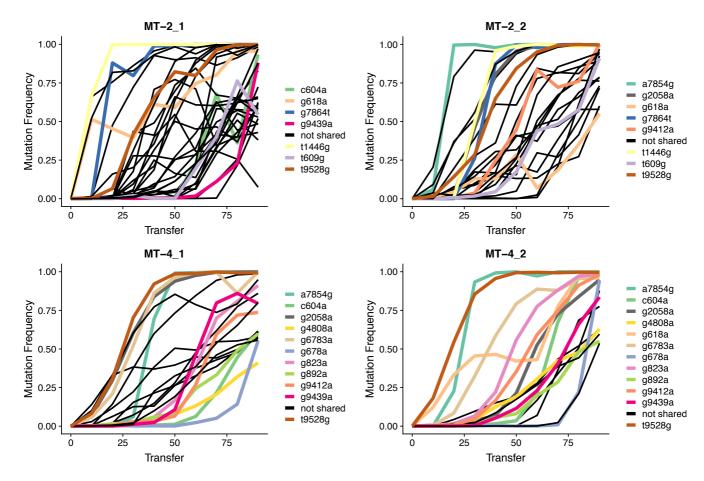


**Fig. 2. Accumulation, maximum mutation frequency slopes (fitness), parallelism, and dynamics of majority mutations. (A)** Accumulation of majority mutations over time (thick lines), accumulation of neutral majority mutations in simulations over 180 generations (thin lines). **(B)** For each majority mutation observed in the evolution experiment we show the maximum fold mutation frequency increase, plotted against the time frame at which the frequency increase occurred. The maximum fold increase of the mutation frequencies is a surrogate of the fitness advantage of a mutation. Synonymous mutations are displayed as "S", non-synonymous mutations are displayed as "N", and mutations in untranslated regions are shown as "U". Linear regressions on the fitness advantages in MT-2 and MT-4 T-cells are displayed in red and blue, respectively. The decline in fitness advantages for evolution lines grown in MT-2 (red) is marginally significant from 0 (p-Value=0.05) and significantly different

for MT-4 (blue) evolution lines (p-Value = 0.008). There is no majority mutation, for which the maximum mutation frequency increase occurred after transfer 80. **(C)** Venn diagram of the presence of majority mutations in the four evolution lines at transfer 90. **(D)** to **(F)** show mutation frequencies of three selected majority mutations for the four evolution lines. The black solid line indicates a mutation frequency of 50%. Mutation frequencies are unlikely to be informative once they fall below the Illumina sequencing error rate of ~0.1%. The evolution lines are colored as in **(C)**. **(D)** The mutation g9412a becomes a majority mutation in three populations in parallel except for MT-2_1. **(E)** Mutation frequencies of mutation, t9528g, increase almost simultaneously for all four evolution lines. **(F)** In all four evolution lines the mutation g618a increases in frequency, in two lines it becomes a majority mutation at the beginning of the experiment and in one line towards the end of the experiment. For the frequency dynamics of other majority mutations please see **Fig. S2**.

### Large extent of parallel evolution across evolutionary lines

Consistent with the large observed fitness effects of the majority mutations, we see many of them arise in parallel in more than one evolution line (**Fig. 2C**, **2D-F** and colored lines in **Fig. 3**). To test whether this observation is expected for neutrally evolving viruses with even mutation rates across the genome, we performed simulations (see Materials and Methods). For each evolution line we distributed the observed majority mutations at transfer 90 across the genome (i.e. an observed A to G majority mutations will remain an A to G substitution albeit at another position in the genome). We then determined how many of the mutations occurred in more than one evolution line. We repeated the simulation 100 times and found that in only 28 of 100 simulations there was a single mutation that occurred in more than one evolution line. In the remaining 72 simulations we found no mutations in more than one evolution line. Hence, we expect that almost all parallel mutations we observe in our experiment are adaptive or due to mutational hotspots (Cuevas et al. 2015; Sackman et al. 2017; Lind et al. 2019).

**Fig. 3. Frequency dynamics of majority mutations show decoupling of mutations.** We observe decoupling when the mutation frequency dynamics of two mutations changes from being identical to becoming different. For example, in MT-2_1 there is a yellow and a black mutation that show the exact same frequency increase from transfer 0 to 10. At transfer 20 the mutation frequencies are different. Hence at transfer 20 the two mutations decoupled probably through recombination. Colored thick lines in each of the four graphs indicate mutations that are acquired in more than one evolution line in parallel. The same mutations are drawn in the same colors. Black thin lines are majority mutations that are only found in a single line.

In the most extreme case, out of 17 majority mutations observed in one of the MT-4 evolution lines (MT-4_2) at transfer 90, 12 were shared with the other evolution lines. Of the remaining five private majority mutations from MT-4_2, two were also present at frequencies above 10% in other evolution lines (g909a at 15% in MT-4_1 and g614a at 43% in MT-2_1, **Table S1**). In MT-4_1, 11 of 20 majority mutations also reached majority status in at least one other line. Of the remaining nine mutations two occurred at frequencies higher than 10%.

In viral populations grown on MT-2 T-cells, parallelism was not as high. Only 7 of 29 majority mutations in MT-2_1 occur in another evolution line, and only 8 of 26 majority mutations for MT-2_2. Additionally, five private majority mutations in MT-2_1 and three such mutations in MT-2_2 are present at frequencies above 10% in other
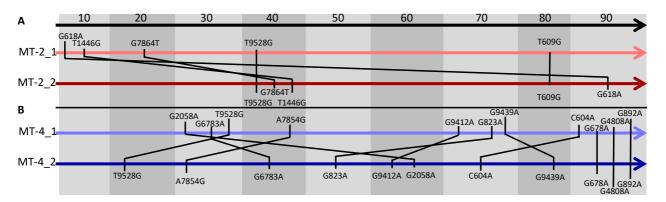
evolutionary lines. Notably, we observed one majority mutation (t9528g) that occurred across all four replicates and six mutations that occurred in three replicates in parallel (**Fig. 2C**).

Interestingly, there was one synonymous mutation that also evolved in parallel (g4808a) of a total of 16 synonymous mutations. Out of 41 non-synonymous mutations we observed seven majority mutations in more than one line. Hence the ratio of parallel mutations to private mutations is about three times as high in non-synonymous mutations (17%) compared to synonymous mutations (6.2%). This difference, however, is not significant (Fisher's exact test: p-Value=0.42). The degree of parallelism is higher for majority mutations in untranslated regions of the genome: Of 26 mutations 7 were shared (27%, but again not significantly different to both, parallel synonymous (p=0.21) and non-synonymous majority mutations (p=0.53), Fisher's exact test).

Apart from highlighting parallel mutations, **Fig. 3** also shows that there are a few instances where recombination has decoupled the increase in frequency of a few mutations from each other. For example, the first mutation to attain a frequency larger than 50% in MT-2_1 (t1446g, yellow) rises simultaneously with a private mutation (not parallel, g3078a, black). Both mutations reach a frequency of 66% at transfer 10. At transfer 20, the two mutations decouple in their increase: while t1446g has risen to 99%, g3078a has only increased to 75%. Similarly, g618a (orange) in MT-2_1 first increases in parallel with the private mutation t8700c (black) to reach 51% of the population at transfer 10. Subsequently the mutations decouple. But in this case, interestingly, the private mutation keeps increasing in frequency to 82% at transfer 20 while g618a decreases to 45%. There are other examples in **Fig. 3** showing similar patterns. These patterns strongly suggest that recombination occurs frequently in our evolution experiment.

**Majority mutations appear in a different order, suggesting low levels of sign epistasis between parallel majority mutations**

Although many majority mutations occur in parallel across evolution lines, their chronological order is different (**Fig. 4**). For example, one of the mutations, that exceed the frequency of the wildtype nucleotide at transfer 10 in MT-2_1 (g618a), becomes a majority mutation only at transfer 90 in MT-2_2. Similar permutations of the chronological order are observed for the MT-4 evolution lines. A different order of mutations suggests that they are beneficial in different genetic backgrounds. Although we cannot exclude that the extent of the benefit is modulated by the genetic background, the benefit is apparently not turned into a deleterious effect. This means that there is no sign epistasis between the parallel majority mutations in our experiment.
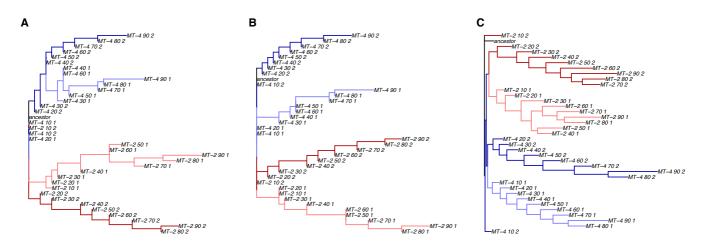
**Fig. 4. Order of majority mutations shared between evolution lines from transfer 10 to 90. (A)** All majority mutations shared by MT-2_1 and MT-2_2. Black lines connect identical mutations. **(B)** Majority mutations shared by MT-4_1 and MT-4_2.

## Implications of parallelism for the reconstruction of the evolutionary history

The high level of parallelism we find in our experiment poses a challenge for the reconstruction of the evolutionary history of our viral populations. When inferring the evolutionary history for the consensus sequences at every tenth transfer, the history cannot be reconstructed correctly for MT-4 evolution lines, where we observe the highest level of parallel evolution. Instead, the sequences cluster by the environment they evolved in (**Fig. 5A**). The mixing of MT-4_1 and MT-4_2 evolution lines in this tree would be wrongly interpreted as a signature of viral migration in a phylogeographic analysis.

The correct evolutionary history (**Fig. 5B**) can be inferred from simulated sequence data, in which the observed majority mutations are randomly redistributed across the HIV-1 genome (see Materials and Methods "*Construction of neutral phylogenies*"). When statistically comparing the correct phylogeny with the one reconstructed from the sequences, the hypothesis that our viral populations evolved neutrally can be rejected with high confidence (Likelihood ratio = $3.5 \times 10^{-69}$, see Materials and Methods "*Construction of neutral phylogenies*").

After failing to reconstruct the correct phylogeny for our evolution lines from majority mutations, we decided to include minority mutations in our analysis. Unfortunately, commonly used phylogenetic methods cannot take the frequency of minority mutations into account when reconstructing phylogenies. Hence, we developed our own method. This method calculates the genetic distance between two populations by summing up the absolute differences in frequency of all wild type nucleotides and all three possible point mutations. We calculated the distances between all sample pairs and with this distance matrix inferred a phylogeny with Neighbor Joining (Saitou and Nei 1987). The inferred tree recapitulates the evolutionary history better than the majority mutation tree (**Fig. 5C**). Both MT-4 lines are now clustered separately on the tree. This means that the two evolution lines genetically diverge early at the level of minority mutations but not at the level of majority mutations.

**Fig. 5. Phylogenetic trees. (A)** Phylogeny inferred from the observed consensus sequences. **(B)** A phylogenetic tree, of which the topology reflects the correct evolutionary history. The tree was generated by randomly (for each evolution line independently) distributing the number of observed majority mutations across the HIV-1 genome. The phylogenies in **A** and **B** were inferred with PhyML (Guindon et al. 2010). **(C)** Phylogeny is based on the differences between all mutation frequencies that were higher than 1% in our data set. For all pairwise differences we constructed a distance matrix. We used the distance matrix to infer the phylogeny *via* Neighbor joining as implemented in the R package ape (Paradis et al. 2004). The trees were colored with phytools (Revell 2012). The colors correspond to the four evolution lines and are consistent between A, B, and C.

**dN/dS method identifies parallel majority mutations**

To identify positively selected substitutions, one commonly applies dN/dS methods. These methods assume synonymous substitutions as selectively neutral (Kimura 1977). This assumption allows the estimation of selection by comparing the number of expected synonymous substitutions with the expected number of non-synonymous substitutions (dN/dS) (Yang and Bielawski 2000). An excess of non-synonymous substitutions indicates positive selection, an excess of synonymous substitutions purifying selection. More sophisticated methods, such as CodeML (Yang 2007) also take the phylogenetic relationship between sequences into account to identify positively selected codons.
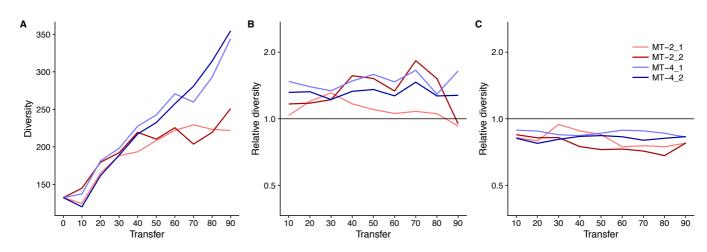
We used CodeML to infer positively selected codons. In our case we can provide CodeML with the consensus sequences of the four evolution lines at all time points and the three different trees from **Figure 5**. Based on our sequence analyses, we expect that CodeML, a package from the PAML suite of programs, will identify the seven majority mutations that occurred in parallel across evolution lines as positively selected, if we provide CodeML with the tree that reflects the true evolutionary history (**Figure 5B**). As expected CodeML identifies all seven parallel non-synonymous majority mutations as likely positively selected (**Figure 5B, Table S2**). However, it also infers two more mutations that appear, disappear and reappear as majority mutations due to changes in mutation frequency. These fluctuations in mutation frequency are incorrectly interpreted as independent substitution events because CodeML is designed to analyze fully speciated organisms and not evolving populations. These two

mutations are also inferred as positively selected when providing the other two phylogenies (**Figure 5A** and **5C**). CodeML performs worse when supplying the inferred phylogeny and the minority mutation phylogeny. Hence, our analysis shows that CodeML also identifies parallelly evolving non-synonymous sites as positively selected.

**The distribution of minority mutations as an indicator for selection strength**
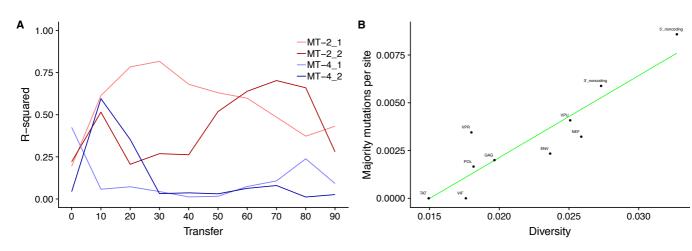
In our experiment we do not have to solely rely on majority mutations to gauge selection strength, we can also include all mutation frequencies in the population by using common diversity indices. We find that over time the nucleotide diversity (measured as the Shannon entropy summed over all nucleotide sites) in all four HIV-1 populations increases (**Fig. 6A**). This mirrors the pattern that has previously been observed for HIV-1 *in vivo* (Shankarappa et al. 1999). We observe that diversity increases faster in MT-4 T-cells than in MT-2 T-cells late in the experiment.

We also measure nucleotide diversity for each gene separately (**Fig. 6B** and **C** and **Fig. S3** and **S4**). Interestingly, genes show similar levels of diversity in the different evolution lines. Across the experiment diversity is greatest in *nef* and the 5' and 3' non-coding regions (**Fig. 6B, Fig. S4**). *Nef* is particularly diverse in HIV-1 populations growing on MT-4 T-cells, whereas the 5' non-coding region is the most diverse region in MT-2 virus populations. The genes with the lowest diversity during the experiment are *gag* and *tat* (**Fig. 6C**).



**Fig. 6. Nucleotide diversity observed across time for all four evolutionary lines. (A)** We applied the Shannon Entropy as diversity measure, which we calculated for each nucleotide site with coverage of more than 1000 sequence reads and then summed it up across the HIV-1 genome. **(B)** Relative diversity measures for the 3' non-coding region, the most diverse region in the HIV-1 genome in our experiment**.** To calculate relative diversity, we divide the average diversity across one gene by the average diversity across the entire genome. Hence, values above one indicate higher than average diversity and below one a lower than average diversity. **(C)** Like **(B)** just for the least diverse gene *gag* in our experiment.

We hypothesize that genes with high nucleotide diversity should accumulate large numbers of parallel majority mutations and genes with low nucleotide diversity should accumulate low numbers of parallel majority mutations. When we correlate the nucleotide diversity at transfer 90 with the number of parallel majority mutations at transfer 90, we find relatively weak correlations ranging from $R^2$ values of 0.03 for MT-4_2 to 0.43 in MT-2_1, with only MT-2_1 showing a significant p-Value of 0.04 (**Fig. 7A**). However, when we correlate nucleotide diversity of all the earlier time points with the number of majority mutations that accumulate at transfer 90, we find much higher correlations (**Fig. 7A**). For MT-2_1 we observe the best correlation at transfer 30 (p-Value = 0.0003, $R^2$=0.79, **Fig. 7B**). Interestingly, for each evolution line the time point is different at which the correlation between nucleotide diversity and number of parallel majority mutations at transfer 90 is highest. The $R^2$ is highest at transfer 70, 0 and 10 for MT-2_2 ($R^2$=0.70, p-Value=0.002), MT-4_1 ($R^2$=0.42, p-Value=0.04) and MT-4_2 ($R^2$=0.60, p-Value=0.009), respectively. Hence, nucleotide diversity at earlier time points is predictive of the number of parallel majority mutations observed at the end of our experiment, at a gene by gene basis.



**Fig. 7. Mutation diversity is predictive of observed parallel mutations at transfer 90. (A)** Correlation between mutation diversity from all measured time points and parallel majority mutations observed at transfer 90 for all genes and noncoding regions in the HIV-1 genome of all four lines. $R^2$ values are a measure of how close the data are to the fitted regression line. High $R^2$ values indicate good correlations, low $R^2$ values indicate bad correlations. **(B)** The linear model fitted for the diversity observed for all genes and noncoding regions at transfer 30 and the number of parallel majority mutations in the same regions for MT-2_1 at transfer 90.

## Discussion

In this study, we investigate the long-term evolutionary dynamics of HIV-1 NL4-3 in four independent evolutionary lines. We find that the four viral populations rapidly and continuously diverge from the ancestor over the course of almost an entire year. The mutation dynamics suggest that at least the first mutations that accumulate in our experiment confer large fitness benefits to the virus. Hence, our four evolution lines respond to large selective pressures despite the absence of antiviral drugs or the immune system. The selective pressures the viral populations

respond to could be due to fluctuating selection (alternating transfers last three or four days), adaptation to different T-cells, or adaptation to growing in flask cell cultures.

Detailed analyses of the mutation dynamics suggest that mutational fitness effects are decreasing over the course of our evolution experiment, similar to diminishing fitness effects found in the Lenski experiment (Wiser et al. 2013). We find that the fitness effects decrease at similar rates for HIV-1 populations growing on MT-2 and MT-4 T-cells. However, viruses growing on MT-2 cells show greater fitness effects early on in the experiment, indicating that HIV-1 NL4-3 is closer to a local optimum in the MT-4 fitness landscape than in the MT-2 fitness landscape. In line with this hypothesis, mutations sweep through the MT-4 viral populations at a lower rate. Being closer to a local optimum in a fitness landscape also means that there are fewer beneficial mutations available (Silander et al. 2007). If there are fewer beneficial mutations available then it is more likely that, in two independent populations, the same mutations are selected to become majority mutations. This is exactly what we observe: a higher proportion of sites that evolve in parallel between the two MT-4 viral populations. Also relevant in this context are the less stringent bottlenecks (larger $N_e$) that MT-4 viral populations undergo (**Fig. S1**). This increases the balance between drift and selection in favor of selection in the MT-4 evolution lines and thereby increases the proportion of parallel mutations.

Since our analysis suggests that parallel mutations are likely positively selected we mined the existing literature to determine the biological relevance of these parallel mutations. The most common parallel non-synonymous mutation is a7864g (also known as D547G in ENV, or D36G in gp41). This mutation is a reversion to the database consensus of glycine at position 547 in the ENV protein. Aspartate at position 36 has been shown to confer resistance to enfuvirtide, a fusion inhibitor (Marconi et al. 2008). There is one more non-synonymous mutation (g2058a or M423I in GAG, M46I in p7) that occurs in three evolution lines in parallel. This mutation is also extremely common in other evolution experiments that were performed in the presence of protease inhibitors (Koh et al. 2009). Unfortunately, these *in vitro* evolution experiments rarely comprise a drug free control. Hence it cannot be determined if the mutation has been selected by the drug, or if it constitutes an adaptation to the host cell type. Our experiment suggests that it is a general adaptation to the host cell type or laboratory environment. We found another GAG mutation (t1446g or H219Q in GAG or H87Q in p24) that only occurs in evolution lines growing on MT-2 cells. This mutation has also been reported previously and found to increase replicative capacity in cyclophilin-A rich cells (such as MT-2) (Gatanaga et al. 2006). MT-4 cells are not cyclophilin-A rich and hence it is not surprising that we did not observe this mutation in our MT-4 lines. There are four more non-synonymous mutations that we found in more than one evolution line (E12K and V35I in GAG, S190N and Q550H in ENV). Except for one (S190N in ENV), all of them were reported in previous evolution experiments: one emerged in the absence of drug pressure (V35I in GAG (Lorenzo-Redondo et al. 2011)), the other two were found to evolve in the presence of antiviral drugs (Nameki et al. 2005; Aoki et al. 2009). The emergence of adaptive mutations in our long-term, drug-free HIV evolution experiment emphasizes that drug free control lines in *in vitro* experiments are

required to ascertain that a mutation confers drug resistance rather than an adaptive advantage at growing in the host cell line.

Most of our parallel mutations (7 out of 15) occurred in the untranslated part of the HIV-1 genome, which only makes up 5.5% of the HIV-1 genome. Unfortunately, there is little information on the function of these mutations. However, there is evidence that the host cell environment significantly affects replication rates for different LTR (long terminal repeat) sequences (van Opijnen et al. 2004). Interestingly, five of the seven parallel majority mutations occur in three of the four evolution lines and only two are host cell specific. Hence, our data suggests that HIV-1 NL4-3 adapts to *in vitro* growth conditions rather than to the specific host cell environment. Alternatively, one could argue that the virus adapts most to environmental features that are common between MT-2 and MT-4 cells.

The large extent of parallel evolution has also a strong effect on our ability to infer the evolutionary history from the sequence data at the end of the experiment. If parallel evolution also plays a significant role *in vivo*, then we expect HIV-1 phylogenies to also reflect host similarities, particularly when the viruses have evolved for long periods of time within these hosts. Indeed, parallel evolutionary changes have been observed *in vivo* in genetically identical twins infected with the same viral strain as well as in HIV-1 populations during early infection and to a smaller extent also in other studies (Draenert et al. 2006; Wood et al. 2009; Bertels et al. 2018).

Nevertheless, reconstructed transmission trees have been found to be consistent with a small number of epidemiologically confirmed transmission chains (Leitner et al. 1996). Although the reliability of such reconstructions is still being investigated (Leitner et al. 1996; Romero-Severson et al. 2014), it is conceivable that transmission histories can be reconstructed accurately if each host exerts a specific selection pressure, thus leaving a signature in the viral genome unique to the infected host (Carlson et al. 2016). Such an effect is also apparent in our experiment: The extent of parallel evolution between populations growing on different T-cell lines is relatively low (**Fig. 2C**). But one has to keep in mind that even a small number of parallel mutations could cause distortions in the branch lengths and the topology of a phylogenetic tree (Bertels et al. 2014).

The levels of parallel evolution we found may be particularly relevant for phyloanatomical approaches (Salemi and Rife 2015; Bons and Regoes 2018) that have received a lot of attention recently (Lorenzo-Redondo et al. 2016). Phyloanatomy entails the application of phylogeographic methods to viral genetic data sampled from different anatomical compartments within infected individuals. The aim is to reconstruct replication rates within each anatomical compartment, and the rates of migration between these compartments. The basis for estimating the migration rates is the presence of the same mutation in more than one compartment, which, according to a neutral model, is most likely due to a virus migrating from one compartment to another. With high rates of parallel evolution, however, such an inference cannot be drawn. Indeed, the phylogenetic tree constructed with the majority sequences obtained from our four independently evolving viral populations wrongly suggests migration events

between the populations. The implications of our experiments for phyloanatomical analyses are more immediate than for the reconstruction of epidemiological transmission history of the virus because selection by the immune system in different hosts might lead to divergent viral sequences, making the inference of the transmission history from viral sequences possible. Within a single host in contrast, immune selection is likely to be more homogeneous across different compartments, and will not generate divergent lineages. It may generate even more parallel evolution (Vanderford et al. 2011), which will further confound a phyloanatomical analysis.

We tried to alleviate this issue by building a tree from a distance matrix calculated by subtracting mutation frequencies between different sequencing samples. The resulting phylogeny separates the two MT-4 lines better than the tree that was only built from majority mutations. Nevertheless, there are still numerous inaccuracies of the tree topology, which will lead to issues in downstream analyses as shown with our CodeML analysis. However, since most studies nowadays take advantage of deep sequencing this method could supplement classical phylogenetic analyses when the evolutionary history of sampled populations is not known.

The extent of parallel evolution in our experiment is likely to be determined by the number of high fitness effect mutations available during adaptation (Orr 2005; Bons et al. 2018). Interestingly, the majority mutations that evolve in more than one line do not necessarily appear in the same chronological order in different evolutionary lines. This suggests that these mutations not only have a high fitness effect but also that the effect is independent of the presence of other majority mutations that have been acquired previously. Technically speaking, there appears to be only low levels of sign epistasis between these mutations. If the number of mutations with high fitness effect were large then picking the same mutation twice is unlikely. Similarly, if their effect strongly depends on the genetic background, populations would diverge on different evolutionary trajectories. Hence, we can explain high parallelism and different mutation orders with a relatively simple argument: for each population, mutations are independently drawn from the same small pool of large effect mutations.

Apart from low levels of sign epistasis and large fitness effects, parallel evolution could also be explained by differences in mutation rates across the genome, which could for example happen when APOBEC3G affects viral evolution (Cuevas et al. 2015; Bertels et al. 2018). High mutation rates at certain positions in the genome could lead to increased levels of parallel evolution at these locations in different evolution lines (Bauer and Gokhale 2015). Similar observations have been made in other model organisms (Sackman et al. 2017; Lind et al. 2018). However, it is very challenging to disentangle the contributions of mutation rate differences and selection to the levels of parallel evolution in our experiment.

We also show that the level of nucleotide diversity in each gene early in the experiment is predictive of the number of majority mutations that accumulate in these genes by the end of the experiment (**Fig. 7**). Interestingly the point at which the correlation between diversity and number of majority mutations was highest was different between the replicate lines. The cause of these differences is unclear, but one contributing factor could be the changing

bottleneck sizes in our experiment. We postulate at least two causes for the correlation between diversity and the accumulation of majority mutations. Either fitness effects or mutation rates (or both) may be unevenly distributed across the genome. In future studies we hope to improve our understanding of this phenomenon.

There are only few long-term evolution experiments with viruses (Holland et al. 1979; Kearney et al. 1999; Wichman et al. 2005). An interesting experiment with viruses was conducted with phage $\phi$X174 growing on *E. coli* (Wichman et al. 2005). The phage population grew in a chemostat for six months and approximately 13,000 generations. At the end of the experiment 137 substitutions were identified, which is about 4- to 8-fold higher than in our experiment. In our experiment we observe between 17 (MT-4_2) and 29 (MT-2_1) majority mutations after about one year, but considering that HIV-1 undergoes only 180 generations this difference is not surprising. Despite the 70-fold difference in the number of generations we only observe an about 7-fold lower number of majority mutations in HIV-1. There are at least two factors affecting this difference, first the five-fold higher mutation rate of HIV-1 ($\sim 2 \times 10^{-5}$ (Mansky 1996)) compared to $\phi$X174 ($\sim 1 \times 10^{-6}$ (Cuevas et al. 2009)), and the second probably more important factor is the much smaller population size in HIV-1 (bottleneck sizes between 40 and 1600) compared to $\phi$X174 ($\sim 10^9$). Smaller population sizes lead to significantly shorter fixation times. Neutral mutations sweep through the population in about $2N_e$ generations (Kimura 1983), which is a tremendously long period of time in the $\phi$X174 experiment ($\sim 76,000$ years) but likely to occur in our HIV-1 experiment ($> 160$ days).

In conclusion our experiment suggests that the high level of parallel evolution we observe is the result of a limited number of large effect mutations with low levels of sign epistasis between them. The high level of parallel evolution together with the observation that genomic regions of high nucleotide diversity early in the experiment accumulate more majority mutations late in the experiment indicates that HIV-1 evolution may be predictable.

**Materials and Methods:**

*Passaging of HIV-1*

The human T-cell leukemia cell lines MT-2 and MT-4 (Harada et al. 1985) were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Douglas Richman. Cells were maintained in RPMI 1640 medium containing 10 % fetal calf serum, 100 U/ml penicillin, and 100 $\mu$g/ml streptomycin. The HIV-1 full-length plasmid pNL4–3 was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Malcolm Martin (Adachi et al. 1986). The virus stock HIV-1 NL4-3 was generated and characterized as previously described (Di Giallonardo et al. 2014). The cultures were grown in separate cell culture flasks. At day 0: $4 \times 10^5$ cells per replicate and per cell line were infected with HIV-1 NL4-3 at an MOI (multiplicity of infection measured on peripheral blood mononuclear cells) of 0.01 resulting in four independent T-cell cultures (**Fig. 1**). Virus passaging was performed twice a week (odd transfers after three days, even transfers after four days) as follows: Infected cell cultures were resuspended. $4 \times 10^5$ uninfected cells were inoculated with 2 $\mu$l cell suspension. Every 10th transfer, cell-free supernatant was stored at -

80°C. In the first 25 transfers, higher volumes (30-3 µl) were transferred based on the extent of cytopathic effects microscopically observed in the cell cultures.

*Ancestral sequence*
The HIV-1 NL4-3 ancestor has been sequenced and assembled previously (Di Giallonardo et al. 2014). Differences to the HIV-1 NL4-3 reference strain (accession AF324493) are listed in Supplementary Table 3 of the same paper (Di Giallonardo et al. 2014).

*Sequencing of near full-length HIV-1 genomes*
Near full-length genomes of the virus stock HIV-1 NL4-3 (ancestor) and transfers 10, 20, 30, 40, 50, 60, 70, 80, and 90 were sequenced using the Illumina MiSeq next-generation sequencing platform as previously described (Di Giallonardo et al. 2014). Briefly, HIV-1 RNA was isolated from 150 µl virus stock HIV-1 NL4-3 or cell-free supernatant; four samples per transfer. Five overlapping amplicons were generated by RT-PCR covering almost the full genome of HIV-1 per sample. In total, 185 amplicons were obtained (4 samples/transfer x 9 transfers x 5 amplicons/transfer + 5 amplicons of the ancestral virus) after one round of PCR. The five amplicons per sample were pooled and libraries were prepared with the Nextera XT DNA Sample Preparation Kit (Illumina, San Diego) according to the manufacturer's description. Next-generation sequencing was performed using a MiSeq Benchtop Sequencer generating paired-end reads of 2×250 bp length (v2 kit). To minimize the risk of cross-contamination, samples from each replicate line were processed separately. Samples were pooled after barcoding and the following samples were sequenced on one chip: Chip 1) virus transfers 10-60; chip 2) virus transfers 70-90; and chip 3) the virus stock among other samples.

*Determining 50% tissue culture infectious dose (TCID50) and multiplicity of infections for every 10th transfer*
We determined the multiplicity of infection for every 10th transfer in our evolution experiment. For each of these nine time points we first approximated the number of infectious viruses per ml for each of the four evolution lines. The number of infectious viruses can be determined by plating cell-free supernatant of our evolution lines on MT-2 and MT-4 cell cultures respectively. To determine infectious viruses we plate increasingly high dilutions of the cell free supernatant on 1,000 cells each. For each of the nine time points and four evolution lines we perform four replicates. After seven days we determine the number of cell cultures that were successfully infected by checking for cytopathic effects under the microscope. This procedure is also called the TCID50 (Reed and MUENCH 1938). Using the TCID50 and the number of cells (400,000) and transfer volumes (2-10 µl) in our experiment we can calculate the average number of viruses that successfully infect a single T-cell in the first generation of every 10th transfer in our experiment.

*Mutation identification from Illumina sequencing data*
Illumina sequences were aligned to the ancestral sequence *via* Bowtie 2 (Langmead and Salzberg 2012). All aligned nucleotides with a quality score of greater than 35 (*i.e.,* error probability of <0.00032) were considered. The first 80

bp and last 30 bp of the HIV-1 genome were excluded from all analyses due to low and spurious coverage. The remaining genome was covered in all data sets by at least 1000 nucleotides. Mean coverage for any sequence sample was at least 2912 nucleotides (**Fig. S5)**.

Consensus sequences were determined by selecting the most common base at each nucleotide site.

*Construction of neutral phylogenies*

For each transfer we simulated sequence evolution under a neutral evolutionary model. We randomly distributed the same number of majority mutations that occurred from one transfer to the next for each evolution line across the HIV-1 genome. When introducing mutations we maintained the GC content of the ancestral HIV genome. Hence at the end of the simulation the sequences contained the exact same number of majority mutations as the consensus sequence sampled in our experiment. From these sequences we built the phylogenetic tree shown in **Fig. 5B**.

To determine the robustness of the neutral phylogenetic tree we repeated the sequence simulation 100 times. The log-likelihood of the correct trees (inferred from randomly distributed mutations) for the observed consensus sequence alignment is -13,375.7 with a standard deviation of 1.1. We call trees correct that resemble the true setup of the experiment, i.e. trees expected for neutrally and independently evolving sequences with even mutation rates across the genome. In contrast the most likely tree for our consensus sequences has a log-likelihood of -13,218.1, which is about $3.5 \times 10^{69}$ times greater than the correct trees. The most likely tree does not correspond to the real evolutionary history of the sequences because it clusters MT-4 strains together, although in the experiment they kept completely separate from each other. In **Fig. 5B** we only showed a single example for illustrative purposes.

*Phylogenetic analyses*

We inferred all phylogenetic trees with PhyML (Guindon et al. 2010) under a general time reversible nucleotide substitution model and gamma distributed rate variations.

*Phylogeny based on minority mutations*

To calculate a high quality phylogeny we first eliminated all mutations that occurred at frequencies of less than 1% in the population and included only sites with more than 1,000-fold sequence coverage. For the remaining mutations we calculated the absolute distances $D_{ij}$ between all populations at all time points:

$$D_{ij} = \sum_{o=1}^{l} \sum_{p \ in \ [ATCG]} |f_{iop} - f_{jop}|$$

$D_{ij}$ *The absolute distance between population i and population j.*

$l$    *Length of the HIV genome.*

$p$  *Nucleotide at position o.*

$f$  *Frequency of the nucleotide p at position o in population i or j.*

From the distance matrix $D_{ij}$ we calculated a Neighbor-joining tree in R with the ape package (Paradis et al. 2004).

*Likelihood ratio between the correct and the inferred phylogeny*

To assess the likelihood that the sequences of our experiment evolved under a neutral model of evolution, we determined the likelihood of the simulated correct tree for the observed consensus sequence and compared it to the likelihood of the tree inferred from the consensus sequence alignment.

*Determining maximum fitness*

For each mutation that became a majority mutation at transfer 90 we determined the maximum fitness by calculating the maximum increase in mutation frequency between two consecutive transfers. We also distinguished between viruses growing on MT-2 and MT-4 T-cells. We then fitted a linear model to the three data sets and got the following results. For the entire data set the decline slope ($\log_{10}$(fitness) declines by 0.006779 per transfer) was insignificantly different from zero (p-Value = 0.4). For MT-2 the decline slope was even smaller ($\log_{10}$(fitness) declines by -0.001345 per transfer) and also not significantly different from zero (p-Value= 0.9). For MT-4, however, the decline slope was larger ($\log_{10}$(fitness) declines by 0.02762 per transfer) and significantly different from zero (p-Value = 0.02762).

*dN/dS calculations*

dN/dS calculations were done using the Ka_Ks calculator (Zhang et al. 2006). We applied the calculator with standard settings (model averaging) to the open reading frames of the ancestor aligned to consensus sequences at transfer 90.

*Calculating diversity*

We calculated diversity by measuring the Shannon entropy for each position in the HIV-1 genome that has a coverage of more than 1000bp. Then the diversity at position $j$ at transfer $t$ is:

$$D_j(t) = \sum_{i\ in\ \{A,T,G,C\}} f_{ji}(t) \log\left(f_{ji}(t)\right)$$

$f_{ji}$ is the relative frequency of base $i$ at position $j$ at transfer $t$

The per site diversity $\overline{D}$ for gene $x$ of length $l$ is then:

$$\overline{D_x(t)} = \frac{\sum_{i=1}^{l} D_i(t)}{l}$$

and for the whole genome with length $gl$:

$$\overline{D_w(t)} = \frac{\sum_{i=1}^{gl} D_i(t)}{gl}$$

We defined the accumulated diversity $\overline{CD_x(t)}$ as:

$$\overline{CD_x(t)} = \sum_{i=0}^{t} \overline{D_x(\iota)}$$

The relative diversity for gene $x$, $Drel_x$, is then:

$$Drel_x(t) = \frac{\overline{D_x(t)}}{\overline{D_w(t)}}$$

*Calculating correlation between diversity and per site majority mutations*

We defined the per site majority mutation numbers as:

$$\overline{MM_x(t)} = \frac{MM_x(t)}{length(x)}$$

$MM_x(t)$ is the number of majority mutations found in gene $x$ at transfer $t$

For **Fig. 7A** we combined all $\overline{MM_x(90)}$ from the four different lines and correlated it with all $\overline{D_x(t)}$ (where $t$ runs from 0 to 90) individually, by inferring a linear model in R. We also correlated $\overline{MM_x(90)}$ with the accumulated diversity $\overline{CD_x(t)}$, which slightly improved the fit.

For **Fig. 7B** and **C** we inferred a linear model for all $\overline{D_x(t)}$ and $\overline{MM_x(90)}$ from each line separately as well as between $\overline{CD_x(t)}$ and $\overline{MM_x(90)}$.

*Simulation of neutral mutations (**Fig. 2A**)*

To predict how many neutral mutations would accumulate in our experiments, we simulated the evolution of individual HIV-1 genomes of length 9,130 for 180 generations. We started our simulation with 400 individuals. Each individual genome is assumed to produce 44 offspring genomes per generation ($R_0$=44). The viral population is allowed to replicate exponentially for two generations. Consistent with the experimental setup, we simulate a bottleneck every second viral generation, by randomly selecting a small number of individuals from the previous population. The size of the bottleneck, i.e. the number of viruses transferred at every passage, was empirically determined. To this end, we determined bottleneck sizes every 20 generations in **Figure S1**. Because we do not have more detailed information on the number of transferred viruses we maintain the number for 20 viral generations which corresponds to ten transfers. Every newly produced viral genome can acquire a mutation with a

probability of $2.16\times10^{-5}$ (Lee et al. 2009).

*Determining extent of parallel evolution in neutrally evolving HIV-1 evolution lines*

We determined the level of parallel evolution that we would observe for neutrally evolving sequences. For this purpose we randomly distribute the same number of mutations that we observed in the four evolution lines across the HIV-1 genome. When distributing substitutions we first randomly chose a position in the genome. The mutation introduced at this position was then determined from transition probabilities inferred from the mutations observed in full length *env* sequence data during early infection (Keele et al. 2008). Hence we distributed 29 and 26 mutations for the MT-2 evolution lines and 20 and 17 mutations for the MT-4 evolution lines randomly across 9,130 unique HIV-1 NL4-3 nucleotide sites. Once the mutations were distributed we determined the number of mutations that occurred in more than one evolution line in parallel. We then repeated the entire simulation 100 times.

*Running CodeML*

To identify positively selected sites in the translated part of the HIV-1 genome we supplied CodeML with an alignment of the translated regions of all evolution lines and sequenced time points in the experiment where majority mutations are substituted for the wild type allele. We ran CodeML three times each time supplying a different phylogeny, i.e. the inferred phylogeny (**Figure 5A**), the true phylogeny (**Figure 5B**) and the phylogeny inferred from the comparison of minority as well as majority mutations (**Figure 5C**). We used the same settings as the HIVNSsites Sweden example. Briefly, the parameter values that were set to a non-zero value are: the CodonFreq model is set to F3X4, kappa is 0.3, omega is 1.3, ncatG is 10, cleandata is 1 and Small_diff is .45e-6.

*Data availability*

The Illumina sequencing data was deposited in the NCBI short read archive under the accession number PRJNA506879.

**Supporting Information**

**Figure S1. Multiplicity of infection and corresponding bottleneck size for the beginning of every 10 transfers in the evolution experiment.**

**Figure S2. Mutation frequencies of all experimental lines for all identified majority mutations during the experiment.** Each plot shows mutation frequencies of the four experimental lines. The y-axis of each plot shows the position and type of mutation in the ancestral sequence (i.e. c6515a means we are looking at the relative frequencies of a mutation that changed a C at position 6515 in the ancestral sequence to an A) and the frequency of that mutation on a log scale. If mutations were not observed in the sequence data then a pseudo count of 1 was added. A solid black line indicates the 50% threshold. The x-axis shows the transfer number, i.e. time.

**Figure S3**. **Per nucleotide site diversity measures for each HIV-1 gene and noncoding regions.**

**Figure S4. Relative per nucleotide site diversity measures for each HIV-1 gene and noncoding regions.**

**Figure S5. Mean coverage per nucleotide for each sequence sample**. The sample names are on the bottom of the graph. The first part of the name denotes the cell line the viruses evolved in, the second part shows the transfer that the sample was sequenced at and the last part the number of the biological repeat.

**Table S1. Majority mutation frequencies observed at transfer 90.**

**Table S2. Probability of positively selected codons inferred by CodeML.**

**References**

Acevedo A, Brodsky L, Andino R. 2014. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature 505:686–690.

Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, Martin MA. 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J. Virol. 59:284–291.

Aoki M, Venzon DJ, Koh Y, Aoki-Ogata H, Miyakawa T, Yoshimura K, Maeda K, Mitsuya H. 2009. Non-cleavage site gag mutations in amprenavir-resistant human immunodeficiency virus type 1 (HIV-1) predispose HIV-1 to rapid acquisition of amprenavir resistance but delay development of resistance to other protease inhibitors. J. Virol. 83:3059–3068.

Bailey SF, Rodrigue N, Kassen R. 2015. The effect of selection environment on the probability of parallel evolution. Mol. Biol. Evol. 32:1436–1448.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461:1243–1247.

Bauer B, Gokhale CS. 2015. Repeatability of evolution on epistatic landscapes. Sci Rep 5:9607.

Bertels F, Metzner KJ, Regoes RR. 2018. Convergent evolution as an indicator for selection during acute HIV-1 infection. bioRxiv:168260.

Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. Mol. Biol. Evol. 31:1077–1088.

Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, Kadie C, Carlson J, Yusim K, McMahon B, Gaschen B, et al. 2007. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. Science 315:1583–1586.

Bons E, Bertels F, Regoes RR. 2018. Estimating the mutational fitness effects distribution during early HIV infection. Virus Evol 4:vey029.

Bons E, Regoes RR. 2018. Virus dynamics and phyloanatomy: Merging population dynamic and phylogenetic approaches. Immunol Rev 285:134–146.

Bordería AV, Isakov O, Moratorio G, Henningsson R, Agüera-González S, Organtini L, Gnädig NF, Blanc H, Alcover A, Hafenstein S, et al. 2015. Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype.Coyne CB, editor. PLoS Pathog. 11:e1004838.

Borman AM, Quillent C, Charneau P, Dauguet C, Clavel F. 1995. Human immunodeficiency virus type 1 Vif-mutant particles from restrictive cells: role of Vif in correct particle assembly and infectivity. J. Virol. 69:2058–2067.

Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, Nelson JA, Gairin JE, Hahn BH, Oldstone MB, et al. 1997. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nat. Med. 3:205–211.

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.

Carlson JM, Du VY, Pfeifer N, Bansal A, Tan VYF, Power K, Brumme CJ, Kreimer A, DeZiel CE, Fusi N, et al. 2016. Impact of pre-adapted HIV transmission. Nat. Med. 22:606–613.

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

Crandall KA, Kelsey CR, Imamichi H, Lane HC, Salzman NP. 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. Mol. Biol. Evol. 16:372–382.

Cuevas JM, Duffy S, Sanjuán R. 2009. Point mutation rate of bacteriophage PhiX174. Genetics 183:747–749.

Cuevas JM, Geller R, Garijo R, López-Aldeguer J, Sanjuán R. 2015. Extremely High Mutation Rate of HIV-1 In Vivo.Rowland-Jones SL, editor. PLoS Biol 13:e1002251.

Das AT, Land A, Braakman I, Klaver B, Berkhout B. 1999. HIV-1 evolves into a nonsyncytium-inducing virus upon prolonged culture in vitro. Virology 263:55–69.

de Visser JAGM, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. Nat. Rev. Genet. 15:480–490.

Di Giallonardo F, Töpfer A, Rey M, Prabhakaran S, Duport Y, Leemann C, Schmutz S, Campbell NK, Joos B, Lecca MR, et al. 2014. Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. Nucleic Acids Res 42:e115–e115.

Draenert R, Allen TM, Liu Y, Wrin T, Chappey C, Verrill CL, Sirera G, Eldridge RL, Lahaie MP, Ruiz L, et al. 2006. Constraints on HIV-1 evolution and immunodominance revealed in monozygotic adult twins infected with the same virus. J. Exp. Med. 203:529–539.

Duwe S, Brunn M, Altmann D, Hamouda O, Schmidt B, Walter H, Pauli G, Kücherer C. 2001. Frequency of genotypic and phenotypic drug-resistant HIV-1 among therapy-naive patients of the German Seroconverter Study. J. Acquir. Immune Defic. Syndr 26:266–273.

Elena SF, Lenski RE. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. 4:457–469.

Engel P, Salzburger W, Liesch M, Chang C-C, Maruyama S, Lanz C, Calteau A, Lajus A, Médigue C, Schuster SC, et al. 2011. Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen Bartonella. PLoS Genet. 7:e1001296.

Foll M, Poh Y-P, Renzette N, Ferrer-Admetlla A, Bank C, Shim H, Malaspinas A-S, Ewing G, Liu P, Wegmann D, et al. 2014. Influenza virus drug resistance: a time-sampled population genetics perspective.Akey JM, editor. PLoS Genet. 10:e1004185.

Gallie J, Bertels F, Remigi P, Ferguson GC, Nestmann S, Rainey PB. 2019. Repeated phenotypic evolution by different genetic routes in Pseudomonas fluorescens SBW25.Malik H, editor. Mol. Biol. Evol.

Gatanaga H, Das D, Suzuki Y, Yeh DD, Hussain KA, Ghosh AK, Mitsuya H. 2006. Altered HIV-1 Gag protein interactions with cyclophilin A (CypA) on the acquisition of H219Q and H219P substitutions in the CypA binding loop. J. Biol. Chem. 281:1241–1250.

Gatanaga H, Suzuki Y, Tsang H, Yoshimura K, Kavlick MF, Nagashima K, Gorelick RJ, Mardy S, Tang C, Summers MF, et al. 2002. Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. J. Biol. Chem. 277:5952–5961.

Gould SJ. 1989. Wonderful Life: The Burgess Shale and the Nature of History. W. W. Norton & Co.

Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA, Giangrande P, Luzzi G, Morgan B, Edwards A, et al. 1997. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat. Med. 3:212–217.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biol. 59:307–321.

Harada S, Koyanagi Y, Yamamoto N. 1985. Infection of HTLV-III/LAV in HTLV-I-carrying cells MT-2 and MT-4 and application in a plaque assay. Science 229:563–566.

Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373:123–126.

Holland JJ, Grabau EA, Jones CL, Semler BL. 1979. Evolution of multiple genome mutations during long-term persistent infection by vesicular stomatitis virus. Cell 16:495–504.

Holmes M, Zhang F, Bieniasz PD. 2015. Single-Cell and Single-Cycle Analysis of HIV-1 Replication.Krausslich H-G, editor. PLoS Pathog. 11:e1004961.

Joint United Nations Programme on HIV AIDS. 2013. Global Report: UNAIDS report on the global AIDS epidemic 2013. Available from:

http://www.unaids.org/sites/default/files/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf

Kearney CM, Thomson MJ, Roland KE. 1999. Genome evolution of tobacco mosaic virus populations during long-term passaging in a diverse range of hosts. Arch. Virol. 144:1513–1526.

Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, et al. 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.

Kimura M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275–276.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge University Press

Koenig S, Conley AJ, Brewah YA, Jones GM, Leath S, Boots LJ, Davey V, Pantaleo G, Demarest JF, Carter C. 1995. Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. Nat. Med. 1:330–336.

Koh Y, Das D, Leschenko S, Nakata H, Ogata-Aoki H, Amano M, Nakayama M, Ghosh AK, Mitsuya H. 2009. GRL-02031, a novel nonpeptidic protease inhibitor (PI) containing a stereochemically defined fused cyclopentanyltetrahydrofuran potent against multi-PI-resistant human immunodeficiency virus type 1 in vitro. Antimicrobial Agents and Chemotherapy 53:997–1006.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Meth 9:357–359.

Lataillade M, Chiarella J, Kozal MJ. 2007. Natural polymorphism of the HIV-1 integrase gene and mutations associated with integrase inhibitor resistance. Antivir. Ther. (Lond.) 12:563–570.

Lee HY, Giorgi EE, Keele BF, Gaschen B, Athreya GS, Salazar-Gonzalez JF, Pham KT, Goepfert PA, Kilby JM, Saag MS, et al. 2009. Modeling sequence evolution in acute HIV-1 infection. J. Theor. Biol. 261:341–360.

Leitner T, Escanilla D, Franzén C, Uhlén M, Albert J. 1996. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci U S A 93:10864–10869.

Lemey P, Rambaut A, Pybus OG. 2006. HIV evolutionary dynamics within and among hosts. AIDS Rev 8:125–140.

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. Am. Nat.

Liao H-X, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, Fire AZ, Roskin KM, Schramm CA, Zhang Z, et al. 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496:469–476.

Lind PA, Farr AD, Rainey PB. 2015. Experimental evolution reveals hidden diversity in evolutionary pathways. Elife 4.

Lind PA, Libby E, Herzog J, Rainey PB. 2018. Predicting mutational routes to new adaptive phenotypes. bioRxiv:335711.

Lind PA, Libby E, Herzog J, Rainey PB. 2019. Predicting mutational routes to new adaptive phenotypes. Elife 8:5.

Lobkovsky AE, Koonin EV. 2012. Replaying the tape of life: quantification of the predictability of evolution. Front Genet 3:246.

Lorenzo-Redondo R, Bordería AV, López-Galíndez C. 2011. Dynamics of in vitro fitness recovery of HIV-1. J. Virol. 85:1861–1870.

Lorenzo-Redondo R, Delgado S, Morán F, López-Galíndez C. 2014. Realistic Three Dimensional Fitness Landscapes Generated by Self Organizing Maps for the Analysis of Experimental HIV-1 Evolution.Vartanian J-P, editor. PLoS ONE 9:e88579.

Lorenzo-Redondo R, Fryer HR, Bedford T, Kim E-Y, Archer J, Kosakovsky Pond SL, Chung Y-S, Penugonda S, Chipman JG, Fletcher CV, et al. 2016. Persistent HIV-1 replication maintains the tissue reservoir during therapy. Nature 530:51–56.

Maisnier-Patin S, Andersson DI. 2004. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res. Microbiol. 155:360–369.

Mansky LM. 1996. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. AIDS Res. Hum. Retroviruses 12:307–314.

Marconi V, Bonhoeffer S, Paredes R, Lu J, Hoh R, Martin JN, Deeks SG, Kuritzkes DR. 2008. Viral dynamics and in vivo fitness of HIV-1 in the presence and absence of enfuvirtide. J. Acquir. Immune Defic. Syndr 48:572–576.

Miller CR, Nagel AC, Scott L, Settles M, Joyce P, Wichman HA. 2016. Love the one you're with: replicate viral adaptations converge on the same phenotypic change. PeerJ 4:e2227.

Mohammadi P, Desfarges S, Bartha I, Joos B, Zangger N, Muñoz M, Günthard HF, Beerenwinkel N, Telenti A, Ciuffi A. 2013. 24 hours in the life of HIV-1 in a T cell line.Emerman M, editor. PLoS Pathog. 9:e1003161.

Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science 296:1439–1443.

Nameki D, Kodama E, Ikeuchi M, Mabuchi N, Otaka A, Tamamura H, Ohno M, Fujii N, Matsuoka M. 2005. Mutations conferring resistance to human immunodeficiency virus type 1 fusion inhibitors are restricted by gp41 and Rev-responsive element functions. J. Virol. 79:764–770.

Orr HA. 2005. The probability of parallel evolution. Evolution 59:216–220.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290.

Poon AFY, Swenson LC, Dong WWY, Deng W, Kosakovsky Pond SL, Brumme ZL, Mullins JI, Richman DD, Harrigan PR, Frost SDW. 2010. Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1. Mol. Biol. Evol. 27:819–832.

Reed LJ, MUENCH H. 1938. A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS. Am J Epidemiol 27:493–497.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Ribeiro RM, Qin L, Chavez LL, Li D, Self SG, Perelson AS. 2010. Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. J. Virol. 84:6096–6102.

Romero-Severson E, Skar H, Bulla I, Albert J, Leitner T. 2014. Timing and order of transmission events is not directly reflected in a pathogen phylogeny. Mol. Biol. Evol. 31:2472–2482.

Sackman AM, McGee LW, Morrison AJ, Pierce J, Anisman J, Hamilton H, Sanderbeck S, Newman C, Rokyta DR. 2017. Mutation-Driven Parallel Evolution during Viral Adaptation. Mol. Biol. Evol. 34:3243–3253.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Salemi M, Rife B. 2015. Phylogenetics and Phyloanatomy of HIV/SIV Intra-Host Compartments and Reservoirs: The Key Role of the Central Nervous System. Curr. HIV Res.

Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, et al. 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J. Virol. 73:10489–10502.

Silander OK, Tenaillon O, Chao L. 2007. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects.Barton NH, editor. PLoS Biol 5:e94.

Srivastava KK, Fernandez-Larsson R, Zinkus DM, Robinson HL. 1991. Human immunodeficiency virus type 1 NL4-3 replication in four T-cell lines: rate and efficiency of entry, a major determinant of permissiveness. J. Virol. 65:3900–3902.

Trkola A, Kuster H, Rusert P, Joos B, Fischer M, Leemann C, Manrique A, Huber M, Rehr M, Oxenius A, et al. 2005. Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies. Nat. Med. 11:615–622.

van Opijnen T, Berkhout B. 2005. The host environment drives HIV-1 fitness. Rev. Med. Virol. 15:219–233.

van Opijnen T, de Ronde A, Boerlijst MC, Berkhout B. 2007. Adaptation of HIV-1 depends on the host-cell environment.Maldarelli F, editor. PLoS ONE 2:e271.

van Opijnen T, Jeeninga RE, Boerlijst MC, Pollakis GP, Zetterberg V, Salminen M, Berkhout B. 2004. Human immunodeficiency virus type 1 subtypes have a distinct long terminal repeat that determines the replication rate in a host-cell-specific manner. J. Virol. 78:3675–3683.

Vanderford TH, Bleckwehl C, Engram JC, Dunham RM, Klatt NR, Feinberg MB, Garber DA, Betts MR, Silvestri G. 2011. Viral CTL escape mutants are generated in lymph nodes and subsequently become fixed in plasma and rectal mucosa during acute SIV infection of macaques. PLoS Pathog. 7:e1002048.

Vasilakis N, Deardorff ER, Kenney JL, Rossi SL, Hanley KA, Weaver SC. 2009. Mosquitoes Put the Brake on Arbovirus Evolution: Experimental Evolution Reveals Slower Mutation Accumulation in Mosquito Than Vertebrate Cells.Andino R, editor. PLoS Pathog. 5:e1000467.

Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, et al. 2003. Antibody neutralization and escape by HIV-1. Nature 422:307–312.

Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, Lifson JD, Bonhoeffer S, Nowak MA, Hahn BH. 1995. Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373:117–122.

Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ. 1999. Different trajectories of parallel evolution during viral adaptation. Science 285:422–424.

Wichman HA, Millstein J, Bull JJ. 2005. Adaptive molecular evolution for 13,000 phage generations: a possible arms race. Genetics 170:19–31.

Wichman HA, Scott LA, Yarber CD, Bull JJ. 2000. Experimental evolution recapitulates natural evolution. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 355:1677–1684.

Wiser MJ, Ribeck N, Lenski RE. 2013. Long-term dynamics of adaptation in asexual populations. Science 342:1364–1367.

Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, Gaschen B, Daniels M, Ferrari G, Haynes BF, McMichael A, et al. 2009. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC.Douek DC, editor. PLoS Pathog. 5:e1000414.

Yang Z, Bielawski J. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. (Amst.) 15:496–503.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 4:259–263.