The relationship between the rate of spread of HIV in the population and the virus evolutionary rate

Alexandre Coelho Silva
alexandre.coelho.s@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2019

Abstract

HIV-1 infection is nowadays a chronic disease when before was a death sentence. HIV-1 mutates quite easily and fast so that one of the barriers that still persist in finding a cure or a vaccine is the escape and evasion mechanisms of this retrovirus. Many factors influence the outcome of an epidemic and in order to assess any relationship between rate of spread of the disease and the virus evolutionary rate an agent-based model was used. Within-host evolution was integrated in the model in the form of genetic distance to further assess the relationship with across-host rate of spread. Different scenarios, contemplating only acute, only chronic or both transmissions were simulated and the influence of super-spreaders in the epidemics outcome was evaluated. The model was validated with data from the Latvia epidemic (1987-2010). In fact results suggest that there is an inverse relationship between rate of spread of the disease and evolutionary rate. Moreover, the speed of spread is enhanced by super-spreaders and epidemics are sustained by most infections occurring in acute phase, however acute phase transmissions alone are unlikely to sustain an epidemic. Higher divergence rates are associated with longer periods of time that virus replicates within-host before transmission.
Keywords: HIV-1 evolution, Epidemiology, Infection Dynamics, Modeling, Agent-Based, Viral divergence

1. Introduction

A current challenge in public health is the HIV pandemic, which accounts for more than 75 million infections so far [1]. In 2018, 1.7 million people became infected with HIV-1 and since the start of the epidemic, 30 million people died from AIDS[2]. Despite the decrease in AIDS related mortality and new infections, the disease is still a major cause of concern, with 37.9 million people living with HIV in 2018. [2] A very important aspect is that an infected person may not develop symptoms in the short term, being able nonetheless of spreading to other susceptible if any encounter takes place[3].

HIV-1 is a diploid retrovirus and targets CD4+ lymphocytes[4]. Once within a cell the viral genome is reverse transcribed to DNA by the virus' own transcriptase, which lacks proofreading activity, having mutation rates ranging from 10^{-6} to 10^{-4} mutations per base per replication cycle[5, 6]. This feature permits a rapid evolution of the virus[7]. Genetic diversity is driven by host pressures, mainly through the immune system[8, 6]. After replication, the new virions go on infecting new CD4+ T cells in a repetitive cycle[5]. In some cellular infections, the virus may remain silent, or in a latent phase[6, 5]. HIV-1 genetic information evolution is described in terms of diversity (genetic variation at a given time) and divergence (genetic distance to a reference point - most times, to a founder virion)[9].

According to the time since infection and the markers present in the blood it is possible to enumerate three disease stages. The **acute phase** is described by a high infectiousness probability due to a peak in plasma viremia, around 21 to 28 days post-infection, resultant from fast HIV-1 replication[10] - for this reason, a person in the acute phase is highly contagious (Figure 1)[11, 12]. By this time there is an observed destruction of helper T cells reservoirs with the establishment of viral latency - integration of the viral DNA in the T cell genome[11, 13]. After this phase, the infection enters the **chronic phase** when there is a stable level of the plasma viremia known as viral set-point. This measure can be used to predict disease progression - a low viral set-point is related to a lower CD4+ T cell depletion, meaning that the chronic phase will last longer, delaying the progression to AIDS[6]. In the course of the said stage the HIV-1 population experiences pressures from the immune system which drive and propel genetic evolution[10]. The chronic phase

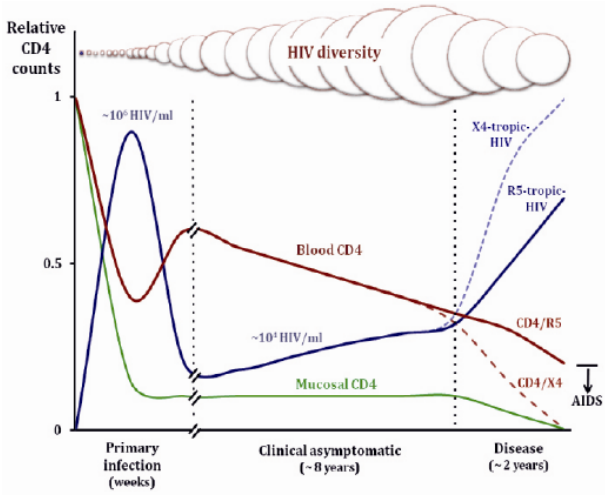


Figure 1: Relative amount of CD4+ T cells over the 3 stages of HIV-1 infection[6].

lasts on average 10 years, but is highly variable among individuals[14, 9]. The progression to **AIDS** happens when further depletion of the CD4+ T cell counts makes it reach values under 200/ml. In this phase, the immune system no longer performs its functions as it is completely devastated[10, 6].

2. HIV modeling

Mathematical modeling and simulation have become a widely used tool in epidemiology since the appearance of computers which allow to run complex and elaborate models in order to derive and extract knowledge from epidemiological data[15, 16].

It is possible to describe the infection dynamics either with compartment models (CM) or agent-based models (ABM). Compartment models divide the population in susceptible and infected, in the simplest case - SI model. Transitions between different compartments occur according to a certain law or mathematical description - transition rate. The main drawback of compartment models is that it is not possible to describe a heterogeneous population with it or explore the details of each member of the population. These types of models may be deterministic, being each compartment described by systems of differential equations, or incorporate stochastic aspects[17].

On the other hand, agent-based models reveal as a better fit solution to describe a heterogeneous population. This type of models simulates a collection of autonomous decision-making entities, the said agents which interactions are also considered by the model. The ABMs are tailored by object-oriented, discrete-event, rule-based and stochastic formalism. Each agent behaves according to a set of rules and stores its properties, for instance, in a set of variables[18, 19]. Thus, ABM captures a lower level of the system where it is possible to follow a single agent temporally and spatially in the whole simulation[16].

The main advantage of ABM over other modeling paradigms is the heterogeneity it conveys. Besides this, using ABMs it is possible to retrieve patterns at the macro level by analysing the interactions at the agents level[19]. One aspect where ABMs lack the most is in terms of computational requirements: having to cycle through millions of agents, and their properties, it may escalate beyond the computational power and ability to handle complexity. Therefore running multiple simulations in order to obtain representative results may not be feasible[19, 18, 16].

3. Methodology

The model used is an adaptation of Graw *et al* (2012)[14]. In the present work an ABM was implemented to investigate the relationship between the rate of spread of an epidemic and the virus evolutionary rate. It was empirically observed that faster epidemics result in slower viral evolution and slower epidemics on faster viral evolution. We hypothesize that this is related to the proportion of infections happening in acute and chronic phases.

The model was implemented in R language[20], and the IDE used was R Studio. Data analysis was also performed using R language, version 3.6.1. [20].

Each simulation starts with 3 infected agents and these index agents are not allowed to be super-spreaders. The simulation evolves in a time-step of one week and ends when the epidemic reaches 10000 infected agents or the simulation duration is attained.

When a susceptible agent is infected, a new infected agent is created and it is instantly calculated the time of death of the new infected agent. The infection event has an associated probability given by Poisson random deviates, with mean value of the product of the infecting agent infectivity and the number of susceptible individuals, in the infecting agent social group. The infected agents incorporate two different infectivities, one to characterize the acute phase infectivity and another for the chronic phase. Consequently, super-spreaders also have specific infectivities during acute and chronic phases. An infected agent is therefore described by an unique id, time of death and parameters of phase-specific infectivities.

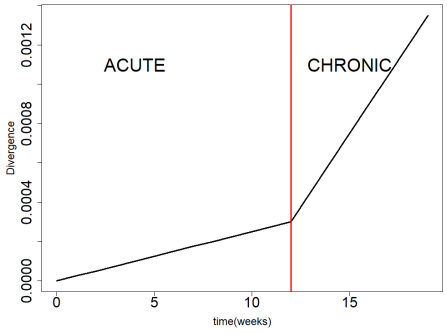
To measure the divergence Equation 1 below was used. This mathematical description is based on Lee et al [9], which inferred the relationship depicted in Equation 1 from Shankarappa and colleagues work (1999)[21]. The integration of intra-host virus evolution in the model represents the

Table 1: The scenario-specific parameters which minimize sum of squares error while originating considerable epidemics - base setting parameters.

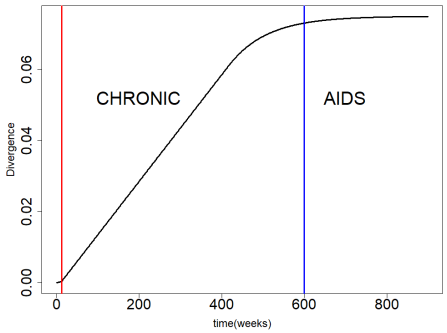
scenario	acute infectivity	chronic infectivity	ss acute infectivity	ss chronic infectivity
C	0	2e-5	0	0
A	2e-4	0	0	0
AC	8e-5	8e-6	0	0
CS	0	8e-6	0	4e-4
AS	9e-5	0	4.5e-3	0
ACS	6e-5	6e-6	3e-3	3e-4

novelty of the work here proposed in which the use of an ABM allows to track each individual of the model portraying epidemiological and disease spread dynamics.

Virus divergence, for each individual, was calculated through the divergence of the infecting virus plus the within-host accumulated divergence over the time of infection, which is given by Equation 1. That is, at each transmission event, a virus with the current divergence within the infecting agent (div_0) is transmitted to the newly infected individual. In the recipient, viral divergence accumulation again follows Equation 1, but adding to div_0. Thus, the div_0 of each agent is the sum of the virus divergence at infection time of the agent who infected him, plus the virus divergence at infection time of the agent who infected the agent who infected him, and so on, until one of the three index patients is reached.



(a)



(b)

Figure 2: The divergence curve profile, of arbitrary units, ($d(t)$ - Equation 1) for each infection stage that the virus will experience within-host 2(a) 0-15 weeks and 2(b) 0-900 weeks, total simulation time. Based on the work of Lee et al[9]

$$d(t) = \begin{cases} \mu_1 t & t < T_1 \\ \mu_1 T_1 + \mu_2 (t - T_1) & T_1 < t < T_2 \\ d_s + (d_{max} - d_s(1 - exp\left(\frac{-\mu_2 (t - T_2)}{d_{max} - d_s}\right))) & T_2 < t < T_3 \end{cases} \quad (1)$$

where $d_{max} = 0.075$, $d_s = \mu_1 T_1 + \mu_2 (T_2 - T_1)$, $\mu_1 = 0.0001/month$ and $\mu_2 = 0.0006/month$, $month = 4 \quad weeks$. The acute phase duration was of 12 weeks, $T_1 = 12$ and the chronic phase lasted until 2920/7 weeks post infection, $T_2 = 2920/7$.

$$div_i(t) = div_0 + d(t - t_{inf}) \quad (2)$$

In order to obtain some realistic dynamics for the model, it was an objective to reproduce the observational data of Latvia epidemics (1987-2010) studied by Graw et al[14]. To simplify, in all our simulation scenarios, we chose the infectivity of the acute phase to be 10 times larger than the infectivity of the chronic phase, which is in agreement with higher viral loads in the acute phase. Some scenarios include the possibility of super-spreaders (ss), which are infectious people with larger infectivities than normal. In the initial scenarios, super-spreader infectivities are 50-fold higher than normal infectivities (both in the acute and chronic phases). Finally, we considered three extreme cases: i) simulations with no infections occurring in the acute phase; ii) simulations with no infections occurring in the chronic phase; and iii) simulations with infections both in the acute and chronic phases. Each case, with and without super-spreaders. These scenarios are labeled in Table 1 as: i) C; ii) A; iii) AC, without the super-spreaders and as CS, AS, ACS, respectively, with the super-spreaders. All these simulations kept the susceptible population (8500) constant. Several preliminary simulations were ran and the respective data analysis performed, which showed that no single set of infectivity parameters was able to give rise to consistent amount of total infected agents in all scenarios. To illustrate this, when considering a given set of infectivities, it was not possible to find a set that at the same time would give rise to epidemics in scenarios A and AS and to epidemics that did not grow too fast (reaching 10000 infected people in less than 4 weeks, for instance) in scenarios AC and ACS. Since no single set of parameters would generate credible epidemics for the whole set of scenarios, the approach was changed to finding a set of parameters that would lead to plausible epidemics, as similar as possible to the Latvia data, for each scenario - See Table 1.

Besides studying the scenarios originated by

Table 2: Settings for the different simulations scenarios to study super-spreaders and disease-phase stage influence on the epidemic behavior.

L1 - 85000 susceptible

Scenario	Super Spreaders	acute	chronic	ss acute	ss chronic
C	0	0	7.50E-07	0	1.25E-04
A	0	7.50E-06	0	1.25E-03	0
AC	0	7.50E-06	7.50E-07	1.25E-03	1.25E-04
CS	0.03	0	7.50E-07	0	1.25E-04
AS	0.03	7.50E-06	0	1.25E-03	0
ACS	0.03	7.50E-06	7.50E-07	1.25E-03	1.25E-04

L2 - 85000 susceptible

Scenario	Super Spreaders	acute	chronic	ss acute	ss chronic
C	0	0	7.50E-07	0	3.75E-06
A	0	7.50E-06	0	3.75E-05	0
AC	0	7.50E-06	7.50E-07	3.75E-05	3.75E-06
CS	0.03	0	7.50E-07	0	3.75E-06
AS	0.03	7.50E-06	0	3.75E-05	0
ACS	0.03	7.50E-06	7.50E-07	3.75E-05	3.75E-06

L3 - 85000 susceptible

Scenario	Super Spreaders	acute	chronic	ss acute	ss chronic
C	0	0	2.50E-06	0	1.25E-05
A	0	2.50E-05	0	1.25E-04	0
AC	0	2.50E-05	2.50E-06	1.25E-04	1.25E-05
CS	0.03	0	2.50E-06	0	1.25E-05
AS	0.03	2.50E-05	0	1.25E-04	0
ACS	0.03	2.50E-05	2.50E-06	1.25E-04	1.25E-05

S1 - 8500 susceptible

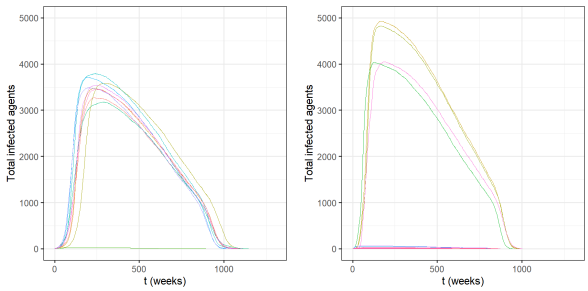
Scenario	Super Spreaders	acute	chronic	ss acute	ss chronic
C	0	0	2.50E-05	0	1.25E-04
A	0	2.50E-04	0	1.25E-03	0
AC	0	2.50E-04	2.50E-05	1.25E-03	1.25E-04
CS	0.03	0	2.50E-05	0	1.25E-04
AS	0.03	2.50E-04	0	1.25E-03	0
ACS	0.03	2.50E-04	2.50E-05	1.25E-03	1.25E-04

base setting fitting the simulations to the Latvia data, we were interested in varying other parameters in order to explore the effects of different susceptible population size, varying super-spreaders infectivity and examining transmission profiles with more extreme infectivities (looking for any patterns linked to this variable). The extra simulated scenarios are presented in Table 2. The parameters were adjusted after performing preliminary simulations taking into account that the simulations should give rise to epidemics that do not die out. For each setting, we simulated the six scenarios C, A, AC, CS, AS, ACS.

4. Results

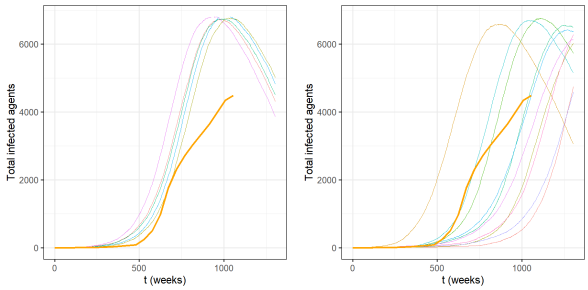
In general, we only analyzed simulations that overcame a certain threshold of infected individuals, typically more than 100. The two main variables of interest were the rate of epidemic growth and the rate of divergence of the virus. These were calculated, for each simulation, in sliding windows of 500 infected individuals each. The rate of epidemic growth is expressed in new infected agents/week and the rate of divergence in the same units as the divergence from Equation 1 per week. In addition, we also analyzed the proportion of transmissions happening at each disease stage.

Infection dynamics with only acute phase transmissions from the simulations performed with the parameters explicit in Table 1, without and with super-spreaders, scenarios A and AS, are shown in Figure 3



(a) scenario A (b) scenario AS

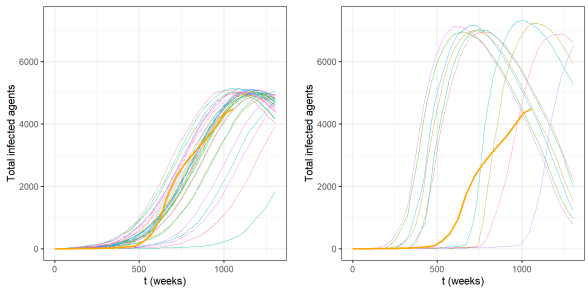
Figure 3: Infection dynamics of the scenarios A and AS with only acute phase infectivity with and without super-spreaders.



(a) scenario C (b) scenario CS

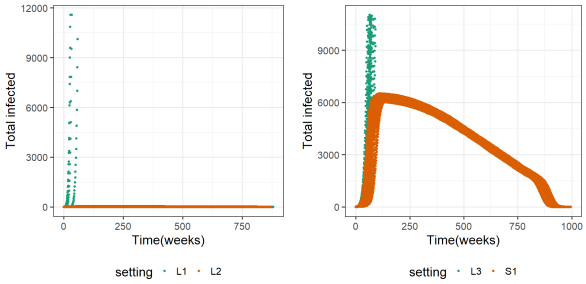
Figure 4: Infection dynamics of the base setting, scenarios C and CS with only chronic phase infectivity with and without super-spreaders. The orange thick line is the Latvia epidemic, for comparison purposes.

Interesting enough is that epidemics with only acute phase transmission have only the profile shown, with a fast growing phase and then a phase where the epidemic dies out. No other shape of epidemic was seen in the hundreds of simulations run. Therefore, in this scenario it was not possible to replicate the Latvia epidemic curve shape. This observation suggests that predominantly acute phase transmissions give rise to epidemics growing relatively fast. It is possible to see



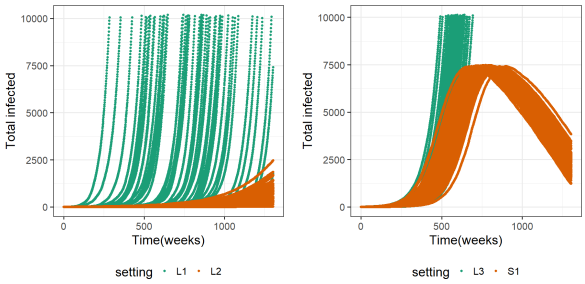
(a) scenario AC (b) scenario ACS

Figure 5: Infection dynamics of the scenarios AC and ACS with acute and chronic phases infectivities, without and with super-spreaders - base setting. The orange thick line is the Latvia epidemic, for comparison purposes.



(a) settings L1 and L2 (b) settings L3 and S1

Figure 6: Infection dynamics of the shown settings, in scenarios with only acute phase infectivity, scenarios AS, with super-spreaders



(a) settings L1 and L2 (b) settings L3 and S1

Figure 7: Infection dynamics of the shown settings, in scenarios with only chronic phase infectivity, scenarios CS, with super-spreaders

that scenario AS (Figure 3(b)) originates larger epidemics, with more infected agents, which is understandable, given the presence of super-spreaders.

The results from the remaining model settings are shown in Figure 6. Regarding L1 and L2 settings, where both settings had the same acute infectivity for non-super spreaders infected agents, only L1 produced epidemics with meaningful numbers of infected agents. This indicates that super-spreaders infectivities influence the outcome of the epidemic. In particular, super-spreaders' infectivities similar to those of L1 are able to sustain an epidemic.

In relation to L3 and S1 settings, Figure 7(b), we see that both settings have the same initial behaviour until the S1 setting starts to slow down.

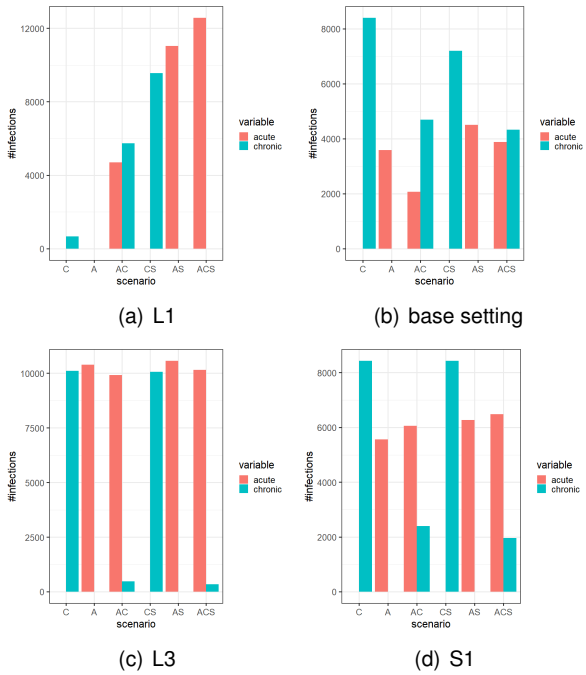
Comparing L3 with S1, these settings differ in susceptible population size and in infectivities (super-spreaders infectivities kept constant), we clearly see that a higher population size associated with lower infectivities better originates epidemics than a lower population size with higher infectivities - L3 leads to bigger epidemics than S1. We compare these two settings, L3 and S1, because they differ in their population size and infectivities, in complementary ways - Table 2 - the susceptible population size of L3 is 10 times higher than S1's but the L3 infectivities are 10 times lower than S1's. We conclude that the trade-off between number of susceptibles and infectivity is responsible for the same initial behavior in the two settings. The overall conclusions apply to the remaining scenarios of these two settings (data not shown).

Infection dynamics with only chronic phase transmissions of the base setting, scenarios C and CS (without and with super-spreaders, respectively), are shown in Figure 4. These results show that in the absence of ss there is a lower stochasticity regarding the initiation of an epidemic, as revealed by the smaller interval that includes the whole set of epidemics of scenario C and a much broader time window in scenario CS. This is likely related with the higher infectivity of non-super spreaders in the C vs. CS scenario. Moreover, both scenarios reach about the same amount of total infected agents, probably also for that same reason.

From L1 and L2 settings analysis, Figure 7(a), and in particular regarding L1 setting, we only make the observation that: the behaviour seems to be the same across all curves, however what is different is the time that each epidemics need to take off. This is related to stochastic effects and to the fact that the chronic phase comprehends a much larger duration than the acute phase, therefore the behaviour is spread over a much broader time window. Another aspect to note is that L1 scenario shows faster exponential growing epidemics. Looking at Figure 7(b), we see that L3 and S1 settings initially reproduce the same behavior and then S1 wanes due to the trade-off between number of susceptibles and infectivity, as explained in the infection dynamics with only acute phase transmissions section of this paper.

Infection dynamics with acute and chronic transmissions, scenarios AC and ACS, are shown in Figure 5, and infectivities in Table 1. By analysing the simulations of scenario AC one easily sees that better fit the Latvia data, than those of scenario ACS. As previously mentioned, the scenario with ss, scenario ACS, leads to higher numbers of infected agents, however the peak appears around the simulation half-time for the majority of simulations, which is also identifiable in previously analysed epidemics, without infecting the whole population. This occurs because as the epidemic spreads out, potential new infections are less likely to occur since the susceptible population undergoes a reduction. Overall, the different settings studied give rise to a wide variety of epidemic profiles, which allow us to study in the following sections several properties of the epidemics under a range of profiles.

The **proportion of infections in acute vs chronic phases** are presented in Figure 8.



(a) L1 (b) base setting

(c) L3 (d) S1

Figure 8: Proportion of transmissions occurring in acute vs chronic phases.

Figure 9 **portrays the relationship between the virus evolutionary rate (divergence over time) and rate of spread of the disease** (increase in number of infected over time). One can beforehand distinguish 2 groups of scenarios in the multiple analyses made to the different windows: a upper group (scenarios C, AC and CS), in the upper left of the figures, and a lower group (scenarios A, AS and ACS) spreading over the center of the same figures. For this reason, we zoom into each of these groups in detail respectively in the middle and right columns of Figure 9.

For scenarios A, AS and ACS, right column, we observe an inverse relationship between divergence over time and speed of the epidemic across the studied windows (Figure 9). These scenarios allow, respectively, transmissions in only acute phase w/o ss, only acute phase w/ ss, and acute and chronic phases with ss. Recalling Figure 8(b), we see that scenario AS is the one with the highest number of acute infections, followed very close by scenario ACS and finally scenario A. For this reason, it is understandable that scenario AS is the scenario with lowest evolutionary rate and highest speed of spread since most transmissions happen in the acute phase - Figure 2(a) shows us a non-linear within-host evolutionary rate, therefore it is logical that infections resulting of mainly acute transmissions give rise to low evolutionary rates. That being said, since scenario ACS is the only one with chronic transmissions in the lower group (see that chronic phase transmissions are associated with higher evolutionary rates within-host - Figure 2(b)) it is expected that this scenario appears in Figure 9 associated with higher evolutionary rate and lower speed of spread, comparing with scenario AS, since the difference is that scenario AS only has acute infections, while scenario ACS has both acute and chronic infections - each type of infection occurs in the same proportion in both scenarios.

Regarding the upper group, composed of scenarios C, AC and CS - scenarios contemplating mainly chronic transmissions - we see in Figure 9, middle column, that scenario CS consistently presents a positive relationship between evolutionary rate and speed of spread. Scenario CS only allows chronic transmissions, that is, the virus will always evolve through the chronic portion of Figure 2(b), before transmission. This translates into a direct relationship between the evolutionary rate and speed of spread - the divergence will increase directly with the number of infected agents. In the same upper group, concerning scenarios C and AC we see that initially in the lowest presented windows, 1500-2000 (Figure 9(b)) these scenarios verify an inverse relationship between the virus evolutionary rate and the speed of spread of the infection. However, as the windows progress, in particular, in 2500-3000 (Figure 9(e)) we see a change in the behavior in which a direct relationship is verified instead (scenarios C and AC). Looking at Figure 8 we see that scenario C originates a high number of infected agents, whose infections were caused by agents in chronic phases, and scenario AC is characterized by a number of acute infections which is half of the chronic infections. As a result, no clear (positive or negative) relationship is verified in scenario AC, since both transmissions are allowed. Thus, one can only suppose that the inverse relationship between the aforementioned quantities verified in Figure 9(b) was originated because of mainly acute infections occurring early on in the AC epidemics and the in a later stage. Later on, when the epidemic accounted for more than 2500 infected agents the main type of transmissions taking place, in scenario AC, was chronic and there the mentioned direct relationship was verified - Figures 9(e) and 9(h). Surprisingly, scenario C, with the exception of the window of 2500-3000, shows an inverse relationship between the evolutionary rate and the speed of spread of the disease. Perhaps it should be expected, after the discussion above, that such a scenario, which only allows chronic transmissions, should originate epidemics with the opposite relation. However, the observed results may indicate that another cause of the inverse relation between divergence and speed of spread is slow growth due to transmis-

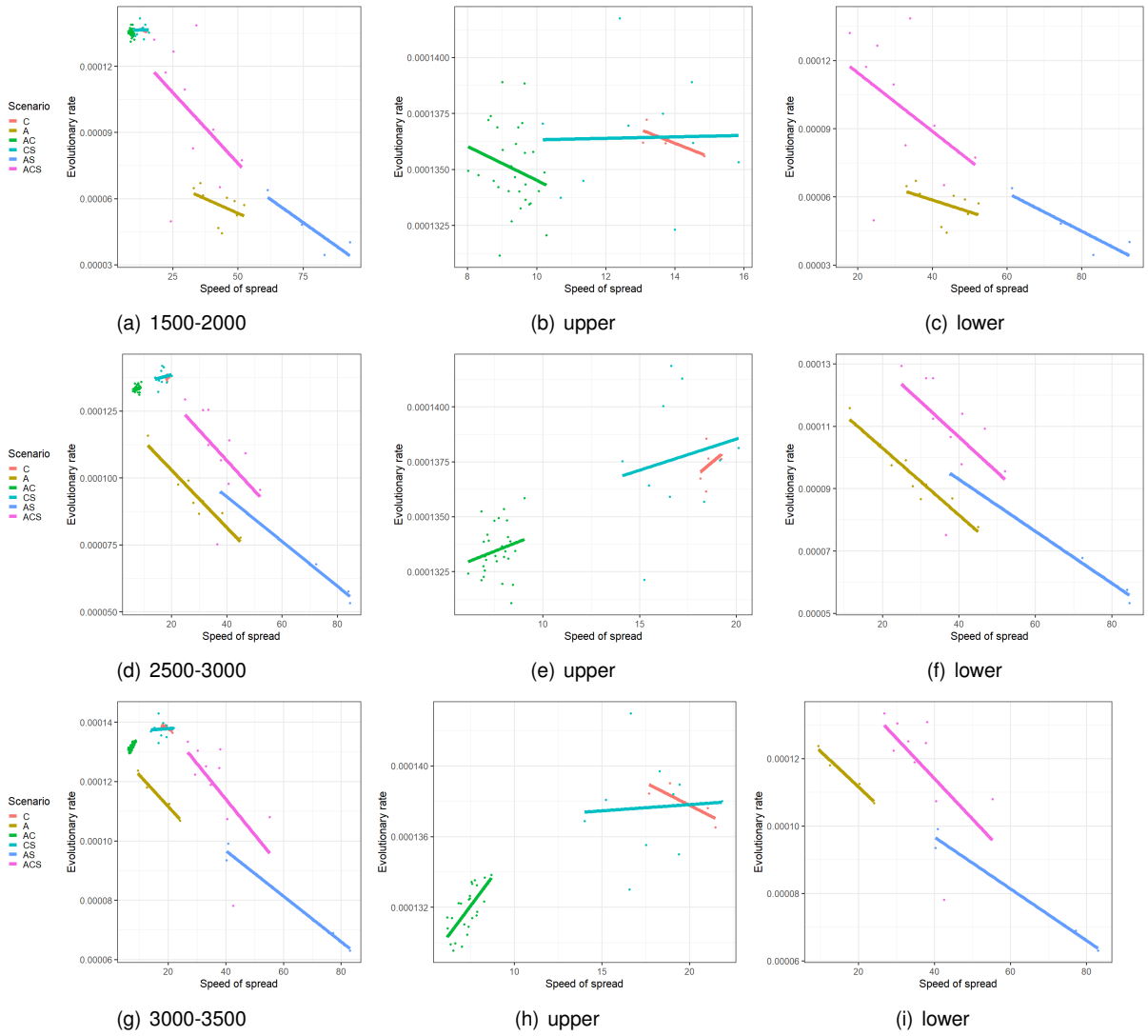
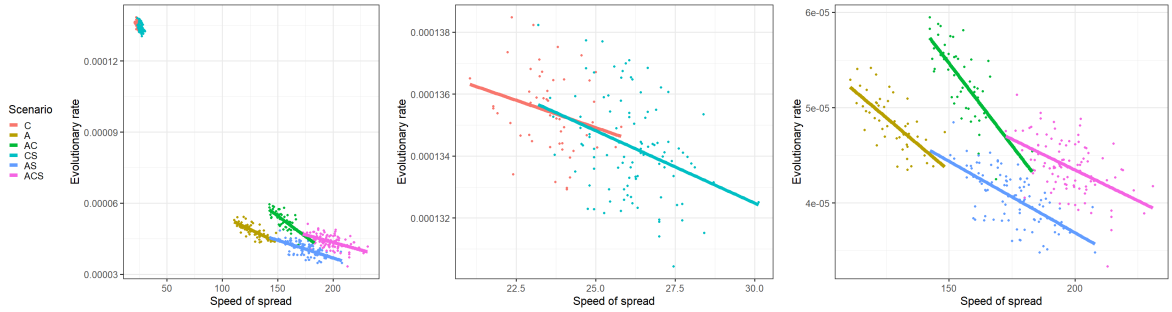


Figure 9: Basic setting divergence over time vs number of infected over time in the depicted windows of size of 500 infected in which i)left: accounts for all scenarios; ii)middle: scenarios C, AC and CS; iii)right: scenarios A, AS and ACS.

sions in the chronic phase, which are accompanied by more time for viral divergence within host. To further study this relation between divergence and epidemic growth, we plot in Figure 10 the results for simulations of S1 setting. Here we only show one window because the same conclusions are drawn for other windows, because scenarios do not change the behavior regarding evolutionary rate and speed of spread of the disease. In the window we can distinguish two groups in the graphic. There is the one composed by scenarios C and CS, upper group, and then there is the one composed by the remaining, scenarios A, AC, AS, ACS, lower group. Clearly, the first group is characterized by a higher divergence rate and low epidemic growth. Analyzing the second group requires more attention. In this group, there is a pseudo-continuity, as the regressions were only separated by a tiny discontinuity, between the pairs of scenarios A-AS and AC-ACS. Not surprisingly, each element of these pairs of scenarios differs only from the other concerning the presence (or absence) of super-spreaders. That is, scenario AS has ss, hence appears as a continuity of scenario A - ss here allow a higher number of infected agents, in a sense - and scenario ACS appears as a continuity of scenario AC - see in detail in Figure 10(c). In addition, we can observe that the regression line of scenarios A-AS is in a lower position than the one of scenarios AC-ACS. This is synonym of the lower divergence speed of scenarios A-AS when comparing with scenarios AC-ACS. In other words, scenarios with only acute phase transmissions, scenarios A and AS (Illustrated by Figure 8(d)) are linked to lower divergence rate, while scenarios AC-ACS, despite originating higher numbers of infected, also contemplate about 1/4 of total infections occurring in the chronic phase,



(a) 2500-3000 (b) upper (c) lower

Figure 10: S1 setting divergence over time vs number of infected over time in the depicted windows of size of 500 infected in which i)left: accounts for all scenarios; ii)middle: scenarios C, AC and CS; iii)right: scenarios A, AS and ACS.

therefore the higher divergence rate of scenarios AC-ACS in comparison with scenarios A-AS. Scenarios C-CS dominate the others in terms of divergence rate and interestingly show little dispersion, comparing with the other scenarios - that is, scenarios with only chronic transmissions (C and CS) have almost no variation in the growth rate of number of infected individuals or in the divergence rate (note the narrow range of the x and y axes). As expected and is shown in Figure 10 scenarios with super-spreaders have higher variation concerning total number of infected individuals than the homologous scenarios. Finally, we see in Figure 10 that scenarios with only chronic transmissions allowed, scenarios C and CS, follow an inverse relationship between evolutionary rate and speed of spread of the epidemic. To sum up, our findings suggest that super-spreaders increase the number of infected people in an epidemic and the speed of spread of the disease. Also, the divergence rate of a given epidemic is, typically, inversely related to the growth rate of the epidemic. If most transmissions take place in the acute phase, we find a slow virus evolution and a fast spread of the disease. On the other hand, when most transmissions occur in the chronic phase, we expect a slow spread of the disease and fast virus evolution. Thus in both situations, we end up with the inverse relationship observed. Only when there is a mixture of both types of transmissions, during acute and chronic infection, this relation is masked.

5. Discussion

HIV-1 sequences from injecting drug users (IDU) epidemics differ from sexual epidemics (HET) in terms of sequence variability. IDU sequences are quite homogeneous over time, contrarily to sexual epidemics[10].

In the present work, fast epidemics IDU can be represented by the scenarios with mostly (or only) acute phase transmissions and the slower epidemics HET can be represented by scenarios with mostly (or only) chronic phase transmissions. This is a rough approximation but it is intended to express the fast behaviour of IDU epidemics by restricting transmissions to the acute phase only, and the analogous to the HET epidemics. Scenarios with both transmissions are considered hybrid, with transmissions allowed during the whole period of the disease. HET transmissions - slow spread - occur predominantly in the chronic stage whereas in an IDU epidemics the infected individuals are mainly in the acute stage of infection[10].

Maljkovic-Berry et al (2007), using a standard compartment model with a Susceptible Infected (SI model), concluded that IDU epidemic fast spread in the former Sovietic Union had a lower evolutionary rate (8.4 times) than the slow spread heterosexual African epidemics[10]. In the base setting (Figure 9) and in the S1 setting (Figure 10) we reached the conclusion that scenarios with only chronic phase, C and CS (HET), are associated with higher evolutionary rates and slower epidemic growth, and scenarios with only acute phase, A and AS (IDU), are associated with slower evolutionary rates but higher epidemic speed of growth - these are corroborating results. When HIV-1 transmissions occur before the pressure of the host immune system, the evolutionary rate of the virus is expected to be lower, comparing IDU and HET epidemics. Hence, an IDU epidemic involves virus very similar to each other[10]. Furthermore, the slower spreading of HET based epidemics is understandable, given the necessity of time consuming task of establishment of new social contacts to transmit the virus. During that time the virus experiences the pressure of the host immune system, acquiring genetic variability - intra-host evolution[10]. Also from Maljkovic[10] model analysis, it was identified a relationship between the susceptible population size and the evolutionary rate in early infections stage. In detail, it was suggested that since exponential spread has a direct relationship to population size, it would be expected that a large susceptible population had a higher rate of spread in

the beginning of the infection. On the other hand, it has simply a lower evolutionary rate during that period[10]. In our work, we identified too a relationship between susceptible population size and acute transmissions - comparison of L3 and S1 settings approached this observation regarding infectivities and susceptible population size. In particular, if acute infectivities are high enough to produce infections in the short time period that is the acute phase, then in the beginning of the infection it is naturally expected that the evolutionary rate will be slower.

Cohen et al (2011)[13] reported that early infections by infected people in early stage greatly shape the epidemics' outcome. It was thought that early infections are responsible for a great proportion of HIV-1 transmissions, therefore targeting people in this phase is crucial. A study, by Kimberly et al[22] found that 38% of transmissions occurring in the established epidemic in Malawi were from donors that were in an early infection stage. As we concluded too, acute infections greatly impact the outcomes of an epidemic and should be a major concern targeting and preventing infections of this kind.

There is an difference regarding HIV evolution within and across hosts. Within host evolution is driven by immune system response and it is fueled by the selection of the immune pressures, resulting in new strains or the extinction of existing ones. A major concern is to investigate how within host evolution influences evolution at the population level, in particular, how immune resistance and immune escape mutations spread at this level[23]. It is not fully understood the mechanisms by which within host evolution influences evolution at the population level. Ideally, given the existence of real data, phylogenetics analyses on sequences of real and large transmission chains would be the best scenario to study those relationships. In any case, epidemiological modeling is always necessary to interpret those type of analyses and validate putative conclusions. The present work is a step in this direction of understanding the connections between within host and population level evolution, using epidemiological modeling and multiple parameter settings.

6. Conclusions

The present study focused on the relationship between rate of spread of HIV-1 in a population with the rate of divergence accumulation by the virus within-host. It had been previously empirically observed an inverse relationship between these two quantities and we demonstrated it using an ABM validated by empirical data of the Latvia epidemic (1987-2010). The hypothesis was that this relationship had to do with the number of infections occurring in acute and chronic phases. In order to assess the hypothesis several settings and scenarios were formulated and simulated. In the end, we showed that, in fact, the data suggests an inverse relationship between the speed of spread of HIV-1 and the virus evolutionary rate, which had not been found before in such a detailed model. These findings are in agreement with the results in Maljkovic-Berry et al[10].

Concluding, the authors expect that this novel approach hereby presented provide a better comprehension of influence of super-spreaders in the epidemic's outcomes, as well as some enlightenment regarding evolutionary rate across the population.

References

[1] Nuno R. Faria, Andrew Rambaut, Mare A. Suchard, Guy Baele, Trevor Bedford, Melissa J. Ward, Andrew J. Tatem, João D. Sousa, Nimalan Arinaminpathy, Jacques Pépin, David Posada, Martine Peeters, Oliver G. Pybus, and Philippe Lemey. The early spread and epidemic ignition of HIV-1 in human populations. *Science*, 346(6205):56–61, 2014.

[2] UNAIDS. Global HIV & AIDS statistics -2019 fact sheet, 2019.

[3] Shula Marks. An Epidemic Waiting to Happen ? The Spread of HIV / AIDS in South Africa in Social and Historical Shula Marks. 0184(770885180), 2002.

[4] Samuel Baron, editor. *Medical Microbiology*. University of Texas Medical Branch at Gavelston, Texas, 4th edition, 1996.

[5] Viviana Simon, David D Ho, and Quarraisha Abdool Karim. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. *Lancet (London, England)*, 368(9534):489–504, aug 2006.

[6] G. Bocharov, V. Chereshnev, I. Gainova, S. Bazhan, B. Bachmetyev, J. Argilaguet, J. Martinez, and A. Meyerhans. Human immunodeficiency virus infection: From biological observations to mechanistic mathematical modelling. *Mathematical Modelling of Natural Phenomena*, 7(5):78–104, 2012.

[7] Donald S. Burke. Recombination in HIV: An Important Viral Evolutionary Strategy. *Emerging Infectious Diseases*, 3(3):253–259, 1997.

[8] Enzo Z Poirier and Marco Vignuzzi. Virus population dynamics during infection. *Current Opinion in Virology*, 23:82–87, 2017.

[9] Ha Youn Lee, Alan S. Perelson, Su Chan Park, and Thomas Leitner. Dynamic correlation between intrahost HIV-1 quasispecies evolution and disease progression. *PLoS Computational Biology*, 4(12):20–22, 2008.

[10] I. M. Berry, R. Ribeiro, M. Kothari, G. Athreya, M. Daniels, H. Y. Lee, W. Bruno, and T. Leitner. Unequal Evolutionary Rates in the Human Immunodeficiency Virus Type 1 (HIV-1) Pandemic: the Evolutionary Rate of HIV-1 Slows Down When the Epidemic Rate Increases. *Journal of Virology*, 81(19):10625–10635, 2007.

[11] Paola Paci, Federico Martini, Massimo Bernaschi, Gianpiero D'Offizi, and Filippo Castiglione. Timely HAART initiation may pave the way for a better viral control. *BMC Infectious Diseases*, 11(1):56, 2011.

[12] Centers for Disease Control and Prevention. about hiv/aids — hiv basics — hiv/aids. `http://aiweb.techfak.uni-bielefeld.de/content/bworld-robot-control-software/`, 2019. [Online; accessed 19-September-2019].

[13] Myron S Cohen, George M Shaw, Andrew J McMichael, and Barton F Haynes. Acute HIV-1 Infection. *The New England journal of medicine*, 364(20):1943–1954, 5 2011.

[14] Frederik Graw, Thomas Leitner, and Ruy M. Ribeiro. Agent-based and phylogenetic analyses reveal how HIV-1 moves between risk groups: Injecting drug users sustain the heterosexual epidemic in Latvia. *Epidemics*, 4(2):104–116, 2012.

[15] Herbert Hethcote. *Modeling HIV Transmission and AIDS in the United States*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992.

[16] David Barnes. *Introduction to modeling for biosciences*. Springer, London New York, 2010.

[17] Institute for Disease Modeling. Compartmental models and EMOD — HIV Model documentation. `http://idmod.org/docs/hiv/model-compartments.html`.

[18] Ricardo Pereira de Magalhães Cruz. *Travels into several remote models of HIV-1 Immune response. In four parts.* PhD thesis, FCUP, 2015.

[19] S M Niaz Arifin, G R Madey, and F H Collins. *Spatial Agent-Based Simulation Modeling in Public Health: Design, Implementation, and Applications for Malaria Epidemiology*. 2016.

[20] R Core Team. R: A Language and Environment for Statistical Computing, 2014.

[21] R Shankarappa, J B Margolick, S J Gange, A G Rodrigo, D Upchurch, H Farzadegan, P Gupta, C R Rinaldo, G H Learn, X He, X L Huang, and J I Mullins. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *Journal of virology*, 73(12):10489–502, 1999.

[22] Kimberly A Powers, Azra C Ghani, William C Miller, Irving F Hoffman, Audrey E Pettifor, Gift Kamanga, Francis E A Martinson, and Myron S Cohen. The Role of Acute and Early HIV Infection in the Spread of HIV-1 in Lilongwe, Malawi: Implications for "Test and Treat" and Other Transmission Prevention Strategies. *Lancet*, 378(9787):256–268, 2011.

[23] T Leitner, D Escanilla, C Franzén, M Uhlén, and J Albert. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 93(20):10864–10869, 10 1996.