

# Effects of epidemiological structure on the transient evolution of HIV virulence

Sang Woo Park and Ben Bolker

April 8, 2016

## 1 Introduction

How pathogen virulence evolves is a fundamental question in evolutionary biology, of both theoretical and (potentially) practical importance. The trade-off theory ([Ebert, 1999](#)) — which postulates that parasite virulence can be explained as the long-term evolutionary outcome of a saturating relationship between parasite clearance rate and transmission rate — has been criticized ([Ebert and Bull, 2003](#); [Alizon and Michalakis, 2015](#)), but has also been successfully applied in several important cases ([Dwyer et al., 1990](#)). One particularly interesting application of these ideas is the work by Fraser et al. showing that HIV appears to satisfy the prerequisites of the tradeoff theory: in a study of discordant couples (i.e. long-term sexual partnerships with one infected and one uninfected partner), HIV virulence as measured by the rate of progression to AIDS was both heritable and covaried with the set-point viral load (i.e., the characteristic virus load measured in blood during the intermediate stage of infection), which in turn predicted the probability of within-couple transmission ([Fraser et al., 2007, 2014](#)). Subsequent studies ([Shirreff et al., 2011](#); [Herbeck et al., 2014](#)) used these data to parameterize mechanistic models of HIV virulence evolution, suggesting that HIV invading a novel population would initially evolve increased virulence, peaking after approximately [XXX] years and then declining slightly to a long-stable virulence level.

The work of Shirreff et al., and particularly the predicted transient peak in HIV virulence midway through the epidemic, highlights the importance of interactions between epidemiological and evolutionary factors ([Day and Proulx, 2004](#); [Alizon, 2009](#)). However, despite the attention to mechanistic detail at the individual or physiological level, the epidemiological structures used in these models are relatively simple. As we discuss in detail below, the existing models of HIV eco-evolutionary dynamics either use implicit models that incorporate the average effects of within-couple sexual contact — without representing the explicit dynamics of pair formation and dissolution or accounting for extra-partnership contact — or use an agent-based formulation with parameters that effectively lead to random mixing among infected and uninfected individuals.

others? de Roode and Altizer?

read/check fraser+ 2014!; check vs. [Herbeck et al. \(2012\)](#)

don't forget to check other recent papers: esp [Payne et al. \(2014\)](#)

Here we explore the effects of incorporating *explicit* epidemiological structure in eco-evolutionary models.

We add complexity to the epidemiological model following the general approach of [Champredon et al. \(2013\)](#); individuals join and leave partnerships at a specified rate, and can have sexual contact both within and outside of established partnerships. At the same time, our analysis somewhat simplifies the models of Shirreff et al., for computational tractability; we check that our qualitative results are not sensitive to these simplifications. In order to explore how virulence evolution depends on epidemiological structure, we consider a series of models with increasing levels of complexity. In order to avoid dependence of the results on a particular set of parameters — as we explain below, finding matching sets of parameters across models with widely differing epidemiological structures is challenging — we evaluate our models across a wide range of parameters, again following Champredon et al. in using a Latin hypercube design. For each model run, we compute a set of metrics (peak virulence, timing of virulence peak, equilibrium virulence) that summarize the evolutionary trajectory of a simulated HIV epidemic.

## 2 Methods

### 2.1 Model formulation

As our primary goal is to explore how different epidemiological structures (i.e. partnership dynamics and contact structures) affect our conclusions about the evolution of virulence, our models use a simplified description of within-host dynamics and heritability derived from Shirreff et al.’s multi-strain evolutionary model. Like Shirreff et al., we use a simple susceptible-infected-susceptible demographic formulation; rather than modeling birth and death (or more specifically, recruitment into the sexually active population and death), we assume that whenever an individual dies from infection, another enters the susceptible compartment.

#### 2.1.1 Infection dynamics

Like Shirreff et al., we focus on the evolution of mean  $\log_{10}$  set-point viral load, SPVL (which we denote as  $\alpha$ ), rather than virulence (i.e. rate of progression to AIDS) itself. In contrast to Shirreff et al., we use a single-stage disease model instead of accounting explicitly for progression through the three main stages of HIV infection (primary, asymptomatic, and disease), and we use a simple exponentially distributed infectious period instead of a more realistic Weibull-distributed infectious period. We account for varying transmission rates and durations of each disease stage by summing the durations of three stages (again based on Shirreff et al.’s model) and taking the duration-weighted average of transmission rates of three stages. Thus the within-couple transmis-

was going to call this “within-host” but it includes transmission

sion rate,  $\beta$ , for our models is given by:

$$\beta(\alpha) = \frac{D_P\beta_P + D_A(\alpha)\beta_A(\alpha) + D_D\beta_D}{D_P + D_A(\alpha) + D_D} \quad (1)$$

where the duration of infection ( $D_P$  and  $D_D$ ) and rate of transmission ( $\beta_P$  and  $\beta_D$ ) of the **Primary** and **Disease** stages of infection are independent of the host's SPVL. Following Shirreff et al., the duration of infection ( $D_A$ ) and rate of transmission ( $\beta_A$ ) for the **Asymptomatic** stage are Hill functions of the SPVL:

$$\begin{aligned} D_A(\alpha) &= \frac{D_{\max}D_{50}^{D_k}}{V_{\alpha}^{D_k} + D_{50}^{D_k}}, \\ \beta_A(\alpha) &= \frac{\beta_{\max}V_{\alpha}^{\beta_k}}{V_{\alpha}^{\beta_k} + \beta_{50}^{\beta_k}}, \end{aligned} \quad (2)$$

where  $V_{\alpha} = 10^{\alpha}$ . The **uncoupled** and **extra-couple** transmission rates are scaled by multiplying the **within-couple** transmission rate  $\beta$  by the contact ratios  $c_u/c_w$  and  $c_e/c_w$ .

### 2.1.2 Mutation

Like Shirreff et al. we incorporate a between-host mutation process in the SPVL, but simplify Shirreff et al.'s evolutionary model slightly by using a one-to-one genotype-phenotype mapping. The mutational process in our model is directly taken from Shirreff et al.. Over the course of infection, mutation occurs within the host. However, it is assumed that SPVL of an infected individual is determined by the SPVL at the time of infection for simplicity (and is not further affected by within-host mutation). Instead, the mutational effect takes place when an infected individual transmits the virus to a susceptible individual. First, the distribution of  $\log_{10}$  SPVL is discretized into a vector:

$$\alpha_i = (\alpha_{\max} - \alpha_{\min}) \frac{(i-1)}{n-1} + \alpha_{\min} \quad i = 1, 2, 3, \dots, n. \quad (3)$$

Then, we construct a mutational matrix,  $M$  — which is multiplied with the transmission term — so that  $M_{ij}$  is the probability that a newly infected individual will have  $\log_{10}$  SPVL of  $j$  given that the infector has  $\log_{10}$  of  $i$ . Finally, the probabilities are normalized so that each row sums to 1:

$$M_{ij} = \frac{\Phi(\alpha_j + d/2) - \Phi(\alpha_j - d/2)}{\Phi(\alpha_{\max} + d/2) - \Phi(\alpha_{\min} - d/2)}, \quad (4)$$

where  $\Phi(j)$  is the Gaussian cumulative distribution function with mean  $\alpha_i$  and variance of  $\sigma_M^2$ , and  $d = (\alpha_{\max} - \alpha_{\min})/n$ . Transmission rate and disease induced mortality rates are discretized into a vector as well:

$$\begin{aligned} \beta_i &= \beta(\alpha_i) \\ \lambda_i &= \frac{1}{D_P + D_A(\alpha_i) + D_D} \end{aligned} \quad (5)$$

### 2.1.3 Contact structure and partnership dynamics

We developed six multi-strain evolutionary models, designed to cover a gamut between Champredon et al.'s relatively realistic and Shirreff et al.'s relatively simplistic epidemiological structures, each of which is based on different assumptions regarding contact structure and partnership dynamics. Specifically, we focus on the effects of the assumptions of (1) instantaneous vs. non-instantaneous partnership formation and (2) zero vs. positive extra-partnership sexual contact and transmission on the evolution of mean  $\log_{10}$  SPVL.

Our first four models consider explicit partnership dynamics and are based on Champredon et al.'s model. Model 1 and 2 assume non-instantaneous partnership formation (i.e. individuals spend some time uncoupled, outside of partnerships) and consist of five states that are classified by infection status and partnership status.  $S$  is the number of single (uncoupled) susceptible individuals, and  $I$  is the number of single infected individuals.  $SS$  is the number of susceptible-susceptible couples,  $SI$  is the number of serodiscordant (susceptible-infected) couples, and  $II$  is the number of concordant positive (infected-infected) couples. Model 1 includes extra-partnership contact (with both uncoupled individuals and individuals in other partnerships) whereas model 2 only considers within-couple transmission. Models 3 and 4 assume instantaneous partnership formation and thus consist of only the three partnered states:  $SS$ ,  $SI$ , and  $II$ . Parallel to model 1 and 2, model 3 includes extra-partnership contact (now only with individuals in other partnerships, since uncoupled individuals don't exist in this model) and model 4 only considers within-couple transmission.

In contrast, models 5 and 6 are not explicitly structured. Model 5 is an implicit serial monogamy model based on the epidemiological model used by Shirreff et al.. It is actually a random mixing model that consist of only two states,  $S$  and  $I$ , and does not consider explicit partnership dynamics. However, to simulate the effect of instantaneous partnership formation, it uses an adjusted transmission rate that is derived from approximated basic reproduction number of a serial monogamy model [Hollingsworth et al. \(2008\)](#). Finally, model 6 is a simple random-mixing model.

is this the right ref? where does Shirreff's model originate?

## 2.2 Latin hypercube sampling

Despite considerable effort by many researchers (e.g. [Hollingsworth et al., 2008](#); [Champredon et al., 2013](#)), the parameters determining the rate and structure of sexual partnership change and contact are still very uncertain; this led Champredon et al. to adopt a Latin hypercube sampling (LHS) strategy ([Blower et al., 1991](#)) that evaluates model outcomes over a range of parameter values. In order to make sure that our comparisons among models apply across the entire space of reasonable parameter values, and in order to evaluate the differential sensitivity of different model structures to parameter values, we follow a similar protocol and perform LHS over a parameter set including both the transmission and duration parameters ( $\beta_P$ ,  $D_P$ ,  $\beta_D$ ,  $D_D$ ) and contact/partnership

I think we need a section here to describe partnership dynamics/contact structure, i.e. define the  $\rho$ ,  $c_u$ ,  $c_e$ ,  $c_w$  parameters - will be similar to Champredon et al.

parameters ( $\rho$ ,  $c$ ,  $c_u/c_w$ , and  $c_e/c_w$ ). We do not allow for uncertainties in parameters that are directly related to evolutionary process

i.e., ... ?

Latin hypercube sampling is done as in Champredon et al.. For each parameter,  $z$ , its range is divided into  $N$  equal intervals on a log scale, where  $N$  is the number of total runs:

give the number we actually used here

$$z_i = \exp(\log(z_{min}) + [\log(z_{max}) - \log(z_{min})] \frac{i-1}{N-1}) \quad i = 1, \dots, N. \quad (6)$$

For simplicity, we assume that all parameters follow logarithmic distribution. Following the vectorization of a parameter range, a matrix is constructed so that each column contains a vector of a parameter series which it represents ( $z_1, \dots, z_N$ ). Then, each column is replaced with a random permutation series of itself. Now, each row contains a different parameter set that is used for each simulation run.

Table 1 gives the ranges of the model parameters used for LHS. Parameters ranges regarding contact and partnership dynamics ( $\rho$ ,  $c$ , and  $c_e/c_w$ ) are taken from Champredon et al., whereas those regarding infection ( $\beta_P$ ,  $D_P$ ,  $\beta_D$ , and  $D_D$ ) are taken from Hollingsworth et al. (2008). Rest of the parameters are taken from Shirreff et al..

As we introduce a new parameter,  $c_u/c_w$ , to model uncoupled and extra-couple transmission dynamics more easily within multi-strain models (refer to appendix), it is important to pick a reasonable range for it. Champredon et al. assume that the effective within-couple contact rate and effective uncoupled contact rate have the same range of 0.05 - 0.25. Given Champredon et al.'s parameter range, it is trivial that the possible maximum and minimum values of  $c_u/c_w$  are 5 and 1/5. Therefore, we use 1/5-5 as the range for the parameter  $c_u/c_w$ . Although this adds more uncertainty to the parameter  $c_u$  — Champredon et al.'s range gives 5 fold difference whereas ours give 25 fold difference — as there is not much known about the uncoupled transmission rate, it is reasonable to use a wider range.

refer to Table 1 and describe where the new parameters come from; e.g. decisions we made about new scaling parameters

## 2.3 Simulation runs

One of the most difficult parts of model comparison is finding parameter sets that are commensurate against many different model structures. For the most part, our models are too complex to easily derive analytical correspondences among them. Given a numerical criterion, such as  $r$  (initial exponential growth rate) or  $\mathcal{R}_0$  (intrinsic reproductive number), we can adjust one or more parameters by brute force to ensure that all of the models match according to that criterion. While  $\mathcal{R}_0$  is often considered the most fundamental property of an epidemic, and might thus seem to be a natural goal, here we focus on matching the initial growth rate  $r$  for several reasons. First, our primary interest is in the transient evolutionary dynamics of virulence, which are more strongly affected by  $r$  than  $\mathcal{R}_0$  (CITE?). Second,  $r$  is in general more directly observable

in real epidemics;  $r$  can be estimated by simply fitting an exponential curve to the initial incidence or prevalence curves (CITE Ma et al.), while  $\mathcal{R}_0$  typically requires either (1) knowledge of *all* epidemic parameters or (2) relatively sophisticated back-calculation based on  $r$  and knowledge of the serial interval or generation interval of the disease (CITE Wallinga, Teunis, Lipsitch, etc.) Thus, we scale a parameter so that every run has an equal exponential rate of growth rate.

In order to allow for all models to have equal initial exponential growth rate,  $r$ , we need to pick a parameter,  $s$ , so that  $\lim_{s \rightarrow 0} r(s) = 0$  and  $\lim_{s \rightarrow \infty} r(s) = \infty$ . As adjusting either partnership change rate (i.e. partnership formation and dissolution rate) or transmission rate does not fulfill this requirement for certain models, we decided to scale partnership change rate and dissolution rate by the same factor of  $\gamma$ :  $\beta_{\text{adj}} = \gamma\beta_{\text{base}}$ ,  $c_{\text{adj}} = \gamma c_{\text{base}}$ ,  $\rho_{\text{adj}} = \gamma\rho_{\text{base}}$ . Since transmission rate is adjusted by the scale of  $\gamma$ , uncoupled and extra-couple transmission rates are adjusted as well. For model 3, 4, and 5, all of which assume instantaneous partnership, only the transmission rate and partnership dissolution rate are adjusted.

We run each model for 1000 times with different parameter sets calculated from Latin hypercube sampling [CITE Blower et al], with fixed starting conditions of mean  $\log_{10}$  SPVL of 3 and epidemic size of  $10^{-4}$  (refer to appendix). An initial run of After each run, initial exponential growth rate is calculated. Then, parameters are scaled so that the initial exponential growth rate is scaled to 0.04, which is approximately equal to that of Shirreff et al's model.

For each model we derive the following summary statistics: peak virulence, peak time, equilibrium virulence, and relative peak virulence. The transient phase of an epidemic is often characterized by high virulence, and we define peak virulence as the maximum virulence during this phase. It is simply calculated by taking the maximum value from the virulence trajectory, and peak time is the time at which the maximum value is reached. Once the epidemic enters the endemic phase, evolution of virulence stabilizes and reaches equilibrium. Equilibrium virulence is calculated by taking the mean virulence at 4000 years. Although most simulations reach equilibrium much earlier, we set our time horizon at a much later date as some simulation runs have slow rate of evolution depending on the parameter set and model assumptions.

We focus on these statistics for the following two reasons. First of all, knowing the possible ranges for the peak virulence allows us to estimate the worst-case scenario for the HIV and other sexually transmitted disease epidemics. When an epidemic begins, it is often the case that the pathogen has already evolved towards high virulence by the time it is observed [CITE]. Understanding how virulent a pathogen can evolve before an epidemic begins can be helpful for controlling the disease. Furthermore, knowing the initial virulence, peak virulence, timing of the peak virulence, and equilibrium virulence provide sufficient detail to identify the shape of the virulence trajectory. During an epidemic outbreak, it is difficult to observe virulence evolution. Specifically,

mention that baseline parameters are geometric means of ranges

Do we need to talk about the baseline parameters? We never use it... We only use it for the first set of graphs but we need to mention that we are using Shirreff's partnership rate just for comparison...So we actually never use baseline parameters...

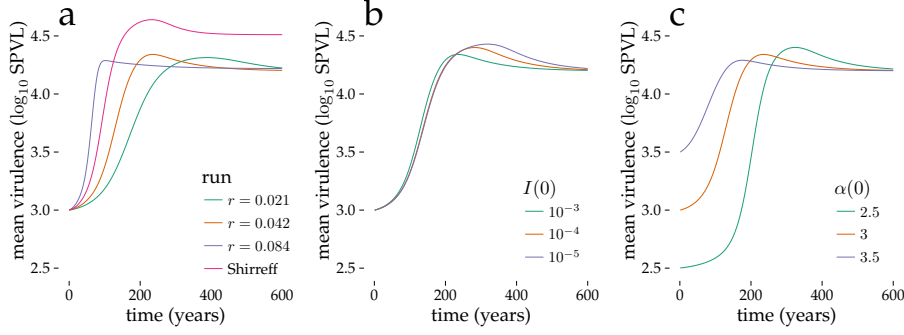


Figure 1: Baseline dynamics. Time series of mean population virulence ( $\log_{10}$  SPVL). (a) Shirreff model, effects of varying  $r$ . (b) Effects of varying initial infectious density  $I(0)$ . (c) Effects of varying initial mean virulence  $\alpha(0)$

in the case of HIV and other sexually transmitted diseases, slow evolutionary time-scale makes observing changes in the mean virulence even more challenging. Knowing the ranges of these statistics can help real-time virulence evolution prediction during an epidemic less troublesome.

### 3 Results

Our simplifications of Shirreff et al.'s model reproduce its qualitative behaviour reasonably well; as  $r$  decreases from 0.084 to 0.42 (the latter value matching the initial rate of increase in prevalence in Shirreff et al.'s full model) the initial trajectory of increasing virulence brackets the rate from the original model (Figure 1a). However, our model produces lower peak virulence ( $\approx?$  vs.  $\approx?$ ) and equilibrium virulence  $\approx?$  vs.  $\approx?$ ) than Shirreff's, even for matching initial incidence (i.e.,  $r = 0.042$ ).

Changing the initial infectious density ( $I(0)$ ), while it produces the expected changes in the initial epidemic trajectory (Supplementary material), has little effect on the virulence trajectory, making the virulence peaks slightly later and larger as  $I(0)$  decreases and allows a longer epidemic phase before the transition to endemic dynamics (Figure 1b). Decreasing the initial virulence similarly but more strongly leads to progressively later, larger peaks in virulence (Figure 1c).

Across the entire range of parameters covered by the LHS analysis, all of the classes of models we considered produce qualitatively similar virulence trajectories (Figure 2). Although the speed of virulence evolution varies, leading to wide variation in the peak virulence (means ranging from approximately 3.75 to 4.5  $\log_{10}$  SPVL), virulence peaks in all models between 200 and 300 years.

pairs plot: instswitch +epc is the most different from others uncoupled mixing rate 1/5 to 5: rate of uncoupled contact relative to within-pair contact. If

note that in Fig 1 we're using Shirreff's parameters

add Shir. params to table 1: Beta1=2.76, Dur1 = 0.25, Beta3 = 0.76, Dur3 = 0.75, c = 1.25, alpha={2,7,0.1}

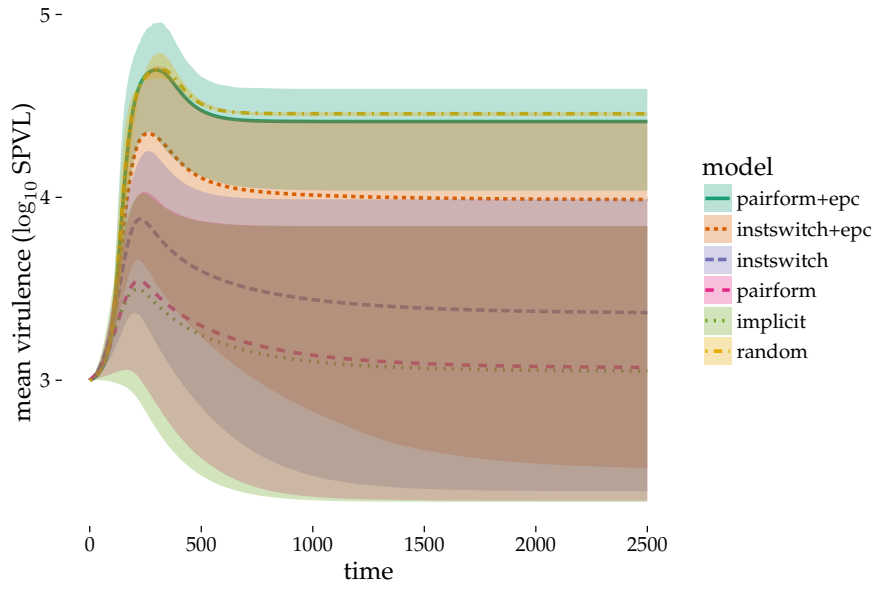


Figure 2: Envelopes of virulence trajectories under all models

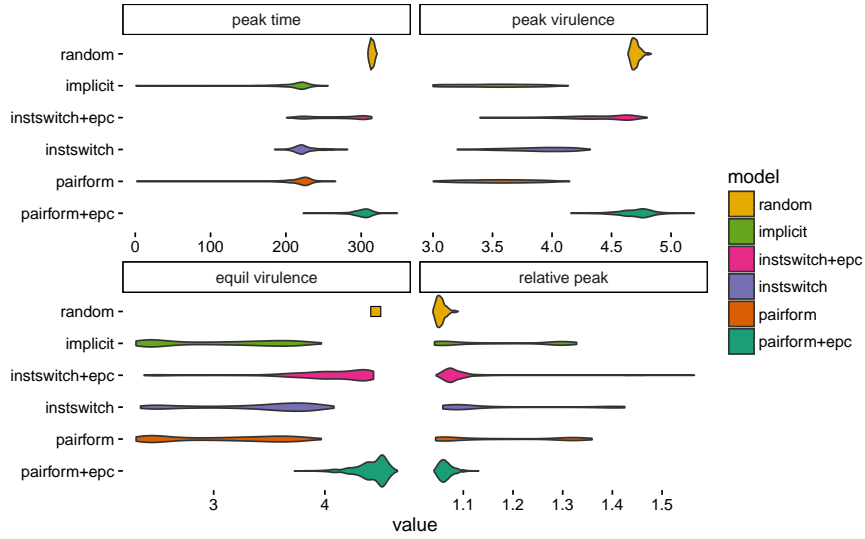


Figure 3: Univariate distributions of summary statistics. Note: point for eqvir/random is a placeholder, will have to be explained



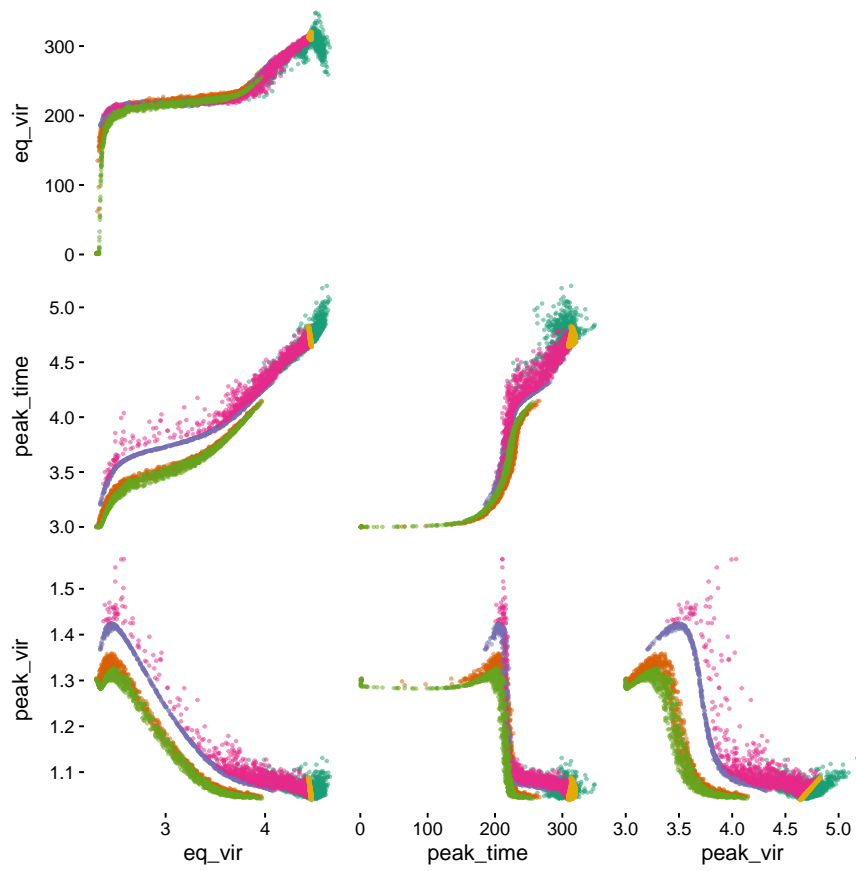


Figure 4: pairs plot

it were much lower instswitch+epc would look more like instswitch \*without\* epc (uncoupled mixing acts like random mixing)

instswitch is intermediate between (pairform + implicit)

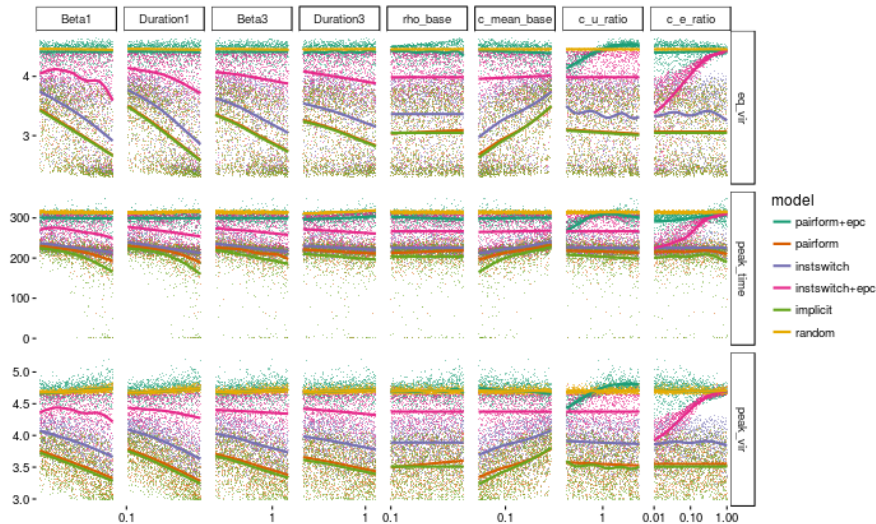


Figure to-do: maybe go back to bw theme?

- fig 1. tweak size? remove redundant y axes/labels?
- fig 2. 6-panel figure? log scale? boundary lines instead of ribbons for edges of envelopes? move labels away from axes?
- fig 3. change order (Random after pairform)
- fig 4. add legend, fix axis labels; maybe transparent
- fig 5. sensitivity plot (unscaled plot); take out plots that are not related ... implicit model doesn't use rho, cu or ce; random doesn't use rho, cmean?, cu, ce; instswitch doesn't have rho or cu or ce; models without epc don't have ce or cu; instswitch + epc doesn't have rho and cu (no uncoupled compartment). Maybe squash panels together, higher resolution

## 4 Discussion

## 5 To do

- clean up code and results
- incorporate pictures
- finish writing!
- hard parts of the results

- explaining similarities among the models: why does implicit look most like pair formation model? (see R0comparison HTML)
- explaining sensitivity plot

See [Bolker JRSI Google scholar cites](#) ([alternative link](#))

## A Model details

Since we use multi-strain models, in which the distribution of  $\log_{10}$  SPVL has been discretized into a vector, we use a matrix notation to describe our models. Five states described in the method section is replaced with the following notations:  $S, I_i, SS, SI_i, II_{ij}$ . Subscript indicates a strain that an individual is infected with. For example,  $I_i$  is number of infected individuals with  $\log_{10}$  SPVL of  $\alpha_i$ , and  $II_{ij}$  is the number of concordant couples, in which two partners have  $\log_{10}$  SPVL of  $\alpha_i$  and  $\alpha_j$ .  $II_{ij}$  is equivalent to  $II_{ji}$ . Lastly, we define Kronecker delta as follows:

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases} \quad (7)$$

In our models, we introduce a non-standard use of Kronecker delta as exponent, such as  $2^{\delta_{ij}}$ . In a single-strain model, both individuals in a concordant partnership,  $II$ , enter single-infected compartment,  $I$ , when a partnership dissolves:  $I' = 2cII'$ . However, this is not the case in a multi-strain model. If infected individuals in an  $II$  partnership have different strains, they will enter different compartment. On the other hand, if both partners in an  $II$  partnership are infected with same strains, both will enter the same compartment. Therefore, we use Kronecker delta as exponent to distinguish such difference. Further details are explain below.

explain non-standard use of Kronecker delta as exponent!

### Model 1 and 2 - Partnership dynamic

Single individuals acquire partners at rate  $\rho$ :  $S' = -\rho S$  and  $I'_i = -\rho I_i$ . We follow Champredon et al's results and assume that single individuals are distributed into couple states through binomial distribution:

$$\begin{aligned} SS' &= \frac{\rho S \cdot S}{2(S + \sum_k I_k)} \\ SI'_i &= \frac{\rho S \cdot I_i}{S + \sum_k I_k} \\ II'_{ij} &= \left(\frac{1}{2}\right)^{\delta_{ij}} \cdot \frac{\rho I_i \cdot I_j}{S + \sum_k I_k} \end{aligned} \quad (8)$$

We introduce Kronecker delta above to differentiate the partnership formation rate for  $II_{ij}$  when  $i = j$  from that of  $I \neq j$ . When  $i = j$  — like the partnership

formation rate of  $SS$  — partnership formation rate becomes  $II'_{ii} = \frac{\rho I_i \cdot I_i}{2(S + \sum_k I_k)}$  due to binomial distribution. On the other hand, when  $i \neq j$ , partnership formation rate becomes  $II'_{ij} = \frac{\rho I_i \cdot I_j}{S + \sum_k I_k}$ .

Partnerships dissolve at rate  $c$ :  $SS' = -cSS$ ,  $SI'_i = -cSI_i$ , and  $II'_{ij} = -cII_{ij}$ . Unlike single strain model, where both individuals leaving the  $II$  partnership would enter  $I$ , we have to account for strains which the individuals in concordant partnership are infected with (i.e. both partners in  $II_{ii}$  enter  $I_i$  whereas only one partner in  $II_{ij}$  enter  $I_i$ ).

$$\begin{aligned} S' &= 2cSS + \sum_k cSI_k \\ I'_i &= cSI_i + \sum_k 2^{\delta_{ik}} cII_{ik} \end{aligned} \quad (9)$$

Combining partnership formation and dissolution process yields the following equation:

$$\begin{aligned} S' &= -\rho S + 2cSS + \sum_k cSI_k \\ I'_i &= -\rho I_i + cSI_i + \sum_k 2^{\delta_{ik}} cII_{ik} \\ SS' &= \frac{\rho S \cdot S}{2(S + \sum_k I_k)} - cSS \\ SI'_i &= \frac{\rho S \cdot I_i}{S + \sum_k I_k} - cSI_i \\ II'_{ij} &= \left(\frac{1}{2}\right)^{\delta_{ij}} \cdot \frac{\rho I_i \cdot I_j}{S + \sum_k I_k} - cII_{ij} \end{aligned} \quad (10)$$

## Model 1 and 2 - Infection

Within-couple transmission occurs in both models. An infected partner in  $SI$  partnership transmits virus to a susceptible partner, and partnership state becomes  $II$ :  $SI'_i = -\beta_i SI_i$ . Since we assume that mutation occurs,  $II_{ij}$ , where  $i \neq j$ , can be formed from both  $SI_i$  and  $SI_j$  partnership:  $II'_{ij} = M_{ij}\beta_i SI_i + M_{ji}\beta_j SI_j$ . On the other hand,  $II_{ii}$  can only be formed from an  $SI_i$  partnership:  $II'_{ii} = M_{ii}\beta_i SI_i$ . Using the Kronecker delta notation, we obtain following set of equations that describe within-couple transmission dynamics:

$$\begin{aligned} SI'_i &= -\beta_i SI_i \\ II_{ij} &= \left(\frac{1}{2}\right)^{\delta_{ij}} \cdot (M_{ij}\beta_i SI_i + M_{ji}\beta_j SI_j) \end{aligned} \quad (11)$$

Champredon et al define the proportion of infectious extra-couple and uncoupled contact through the following term:

$$P = \frac{c_u I + c_e(SI + 2II)}{c_u(S + I) + 2c_e(SS + SI + II)}. \quad (12)$$

Effective uncoupled,  $c_u$ , and extra couple,  $c_e$ , contact rate can be divided into two terms: uncoupled/extra-couple contact rate  $\times$  rate of transmission per contact. Therefore, transmission rate per contact term in  $c_u$  and  $c_e$  is canceled out in the equation above. Using this property, we modify the equation above as follows:

$$P = \frac{r_u I + r_e (SI + 2II)}{r_u (S + I) + 2r_e (SS + SI + II)}, \quad (13)$$

where  $r_u = c_u/c_w$  and  $r_e = c_e/c_w$  are the relative uncoupled/extra-couple contact rates. This simplification is useful in a multi-strain model since we cannot multiply a vector with a single value (e.g.  $c_u S$  in denominator) if we use Champredon et al's equation as it is. Extending the above equation to multi-strain model so that  $P_i$  represents the proportion of the extra-couple and uncoupled contact of an infected individual with strain  $i$ , we obtain the following equation:

$$P_i = \frac{r_u I_i + r_e (SI_i + \sum_k (II_{ik} + \delta_{ik} II_{ik}))}{r_u (S + \sum_k I_k) + r_e (2SS + \sum_k 2SI_k + \sum_l \sum_k (1 + \delta_{lk}) II_{lk})}, \quad (14)$$

Using the equation above, we can model uncoupled and extra-couple mixing. For convenience, uncoupled and extra-couple transmission rates,  $c_u$  and  $c_e$ , will be replaced with  $U_i = r_u \beta_i$  and  $E_i = r_e \beta_i$  from now on.

Single susceptible individuals become infected and enter single infected compartment at the total rate of  $\sum_k P_k U_k S$ . Through mutation, newly infected individuals are distributed into each single infected compartments with different strains:  $I'_i = \sum_k M_{ki} P_k U_k S$ . Either partners in  $SS$  partnership can be infected and the partnership state can become  $SI$  partnership at the total rate of  $\sum_i 2P_i E_i SS$ . Formation of  $SI_i$  partnership is similar to the process through which single susceptible individuals are distributed into single infected compartments:  $SI'_i = \sum_k 2M_{ki} P_k E_k SS$ . Lastly, susceptible partner of an  $SI$  partnership can be infected due to uncoupled/extra-couple contacts and partnership can move to an  $II$  partnership. Like previous cases,  $SI_i$  partnership moves out of the compartment at a rate of  $\sum_k P_k E_k SI_i$ . Mutation process is similar to that of within-couple transmission. The only difference is that the  $\log_1 0$  SPVL of a newly infected partner is not determined by its original partner but from an extra couple partner (i.e. the term  $P_i$ ):  $II'_{ij} = (1 - \frac{\delta_{ij}}{2})(\sum_k (M_{kj} P_k E_k SI_i + M_{ki} P_k E_k SI_j))$ . Combining these equations we get the following set of equations that describe all the transmission dynamics

## Model 1 and 2 - Disease induced mortality

Disease induced death rate,  $\lambda$ , is given by taking the reciprocal of the total duration of the infection:  $\lambda_i = 1/(D_A + D_P(\alpha_i) + D_D)$ . Since we assume SIS formulation, where infected individuals that die from infection enter single susceptible compartment, we obtain the following equation for the single infected individuals:

$$\begin{aligned}
S' &= \sum_k \lambda_k I_k \\
I_i' &= -\lambda_i I_i
\end{aligned} \tag{15}$$

If an infected individual in a partnership dies, partnership dissolves. Thus, an  $SI_i$  partnership dissolves at a rate  $-\lambda_i$ , and the susceptible partner enters single susceptible compartment at rate  $\lambda_i SI_i$  (infected partner that dies enter single susceptible compartment at an equal rate as well due to SIS formation):

$$\begin{aligned}
S' &= \sum_k 2\lambda_k SI_k \\
SI_i' &= -\lambda_i SI_i
\end{aligned} \tag{16}$$

Similarly,  $II_{ij}$  partnership dissolves at a rate  $-(\lambda_i + \lambda_j)$ , but two cases, when  $i \neq j$  and  $i = j$ , must be considered separately. When an  $II_{ij}$  partnership dissolves due to disease induced mortality, where  $i \neq j$ , death of partner with strain  $i$  causes its partner to enter  $I_j$  compartment at rate  $\lambda_j II_{ij}$ , and vice versa. When an  $II_{ii}$  partnership dissolves, death of either partner causes the other partner to enter  $I_i$  compartment at rate  $\lambda_i II_{ii}$ , which sums up to  $2\lambda_i II_{ii}$ . Combining these dynamics yield the following equation:

$$\begin{aligned}
S' &= \sum_l \sum_k 2^{\delta_{lk}} \lambda_k II_{lk} \\
I_i' &= \sum_k 2^{\delta_{ik}} \lambda_k II_{ik} \\
II_{ij}' &= -(\lambda_i + \lambda_j) II_{ij}
\end{aligned} \tag{17}$$

Finally, combining all these equations give us the full model, which is model 1. We can simply take out the uncoupled and extra couple transmission term to obtain equation 2:

$$\begin{aligned}
S' &= -\rho S + 2cSS + \sum_k cSI_k - \sum_k P_k U_k S + \sum_k \lambda_k I_k \\
&\quad + \sum_k 2\lambda_k SI_k + \sum_l \sum_k 2^{\delta_{lk}} \lambda_k II_{lk} \\
I'_i &= -\rho I_i + cSI_i + \sum_k 2^{\delta_{ik}} cII_{ik} + \sum_k M_{ki} P_k U_k S - \lambda_i I_i \\
&\quad + \sum_k 2^{\delta_{ik}} \lambda_k II_{ik} \\
SS' &= \frac{\rho S \cdot S}{2(S + \sum_k I_k)} - cSS - \sum_i 2P_i E_i SS \\
SI'_i &= \frac{\rho S \cdot I_i}{S + \sum_k I_k} - cSI_i - \beta_i SI_i + \sum_k 2M_{ki} P_k E_k SS - \sum_k P_k E_k SI_i \\
&\quad - \lambda_i SI_i \\
II'_{ij} &= \left(\frac{1}{2}\right)^{\delta_{ij}} \cdot \frac{\rho I_i \cdot I_j}{(S + \sum_k I_k)} - cII_{ij} + \left(\frac{1}{2}\right)^{\delta_{ij}} \cdot (M_{ij}\beta_i SI_i + M_{ji}\beta_j SI_j) \\
&\quad + \left(\frac{1}{2}\right)^{\delta_{ij}} \cdot \left(\sum_k (M_{kj} P_k E_k SI_i + M_{ki} P_k E_k SI_j)\right) - (\lambda_i + \lambda_j) II_{ij}
\end{aligned} \tag{18}$$

### Model 3 and 4 - Partnership dynamic

Since model 3 and 4 assume instantaneous partnership formation, there are only three states:  $SS$ ,  $SI_i$ , and  $II_{ij}$ . Partnership dissolution is equal to that of model 1 and 2:  $SS' = -cSS$ ,  $SI'_i = -cSI_i$ , and  $II'_{ij} = -II_{ij}$ . Once the individuals leave partnership, they enter temporary compartments and are distributed into a partnership through binomial distribution:

$$\begin{aligned}
X &= 2cSS + \sum_k cSI_k \\
Y_i &= cSI_i + \sum_k 2^{\delta_{ik}} cII_{ik} \\
SS' &= -cSS + \frac{X^2}{2(X + \sum_k Y_k)} \\
SI'_i &= -cSI_i + \frac{XY_i}{X + \sum_k Y_k} \\
II'_{ij} &= -cII_{ij} + \left(\frac{1}{2}\right)^{\delta_{ij}} \frac{Y_i Y_j}{X + \sum_k Y_k}.
\end{aligned} \tag{19}$$

### Model 3 and 4 - Infection

Model 3 and 4 share the within-couple transmission term with model 1 and 2. Since there is no single state, only extra couple transmission exists:

$$P_i = \frac{r_e(SI_i + \sum_k (II_{ik} + \delta_{ik} II_{ik}))}{r_e(2SS + \sum_k 2SI_k + \sum_l \sum_k (II_{lk} + \delta_{kl} II_{lk}))}. \quad (20)$$

Movement from  $SS$  state to  $SI$  state and  $SI$  to  $SS$  is modeled through the same equation that is used in model 1 and 2.

### Model 3 and 4 - Disease induced mortality

Disease induced mortality is modeled similar to model 1 and 2. However, as single state does not exist in model 3 and 4, individuals that has left their partnerships due to death of their partners enter temporary compartments and form partners instantly:

$$\begin{aligned} X &= \sum_k 2\lambda_k SI_k + \sum_l \sum_k 2^{\delta_{lk}} \lambda_k II_{lk} \\ Y_i &= \sum_k 2^{\delta_{ik}} \lambda_k II_{ik} \\ SS &= \frac{X^2}{2(X + \sum_k Y_k)} \\ SI_i &= -\lambda_i SI_i + \frac{XY_i}{X + \sum_k Y_k} \\ II'_{ij} &= -(\lambda_i + \lambda_j) II_{ij} + \left(\frac{1}{2}\right)^{\delta_{ij}} \cdot \frac{Y_i Y_j}{X + \sum_k Y_k} \end{aligned} \quad (21)$$

Combining all these dynamics, we have equation 3. If we remove extra-couple transmission, we have equation 4.

$$\begin{aligned} X &= 2cSS + \sum_k cSI_k + \sum_k 2\lambda_k SI_k + \sum_l \sum_k 2^{\delta_{lk}} \lambda_k II_{lk} \\ Y_i &= cSI_i + \sum_k 2^{\delta_{ik}} cII_{ik} + \sum_k 2^{\delta_{ik}} \lambda_k II_{ik} \\ SS &= -cSS + \frac{X^2}{2(X + \sum_k Y_k)} - \sum_i 2P_i E_i SS \\ SI_i &= -cSI_i + \frac{XY_i}{X + \sum_k Y_k} - \beta_i SI_i + \sum_k 2M_{ki} P_k E_k SS \\ &\quad - \sum_k P_k E_k SI_i - \lambda_i SI_i \\ II_{ij} &= cII_{ij} + \left(\frac{1}{2}\right)^{\delta_{ij}} \frac{Y_i Y_j}{X + \sum_k Y_k} + \left(\frac{1}{2}\right)^{\delta_{ij}} \cdot (M_{ij} \beta_i SI_i + M_{ji} \beta_j SI_j) \\ &\quad + \left(\frac{1}{2}\right)^{\delta_{ij}} \cdot \left(\sum_k (M_{kj} P_k E_k SI_i + M_{ki} P_k E_k SI_j)\right) - (\lambda_i + \lambda_j) II_{ij} \end{aligned} \quad (22)$$



## Model 5

Model 5 is an implicit instantaneous partnership formation model, which uses adjusted transmission rate,  $\beta'$ , that is derived from Hollingsworth et al's approximated basic reproduction number:

$$\beta'_i = \frac{c\beta_i}{c + \beta_i + \lambda_i}. \quad (23)$$

Thus, we get the following model:

$$\begin{aligned} S' &= \sum_k \lambda_k I_k - \sum_k \beta'_k S I_k \\ I_i &= \sum_k M_{ki} \beta'_k S I_k - \lambda_i I_i \end{aligned} \quad (24)$$

## Model 6

Model 6 is a random mixing model. It is modeled in a same way as model 6 without the adjusted transmission rate:

$$\begin{aligned} S' &= \sum_k \lambda_k I_k - \sum_k \beta_k S I_k \\ I_i &= \sum_k M_{ki} \beta_k S I_k - \lambda_i I_i \end{aligned} \quad (25)$$

## Initial distribution of infected individuals

We follow Champredon et al's result to calculate the initial distribution of infected individuals. For model 1 and 2, we have disease equilibrium state of  $S^* = \frac{c}{c+\rho}$  and  $SS^* = \frac{1-S^*}{2}$ . We let  $\epsilon = 10^{-4}$ , which is the total number of infected individuals in the beginning of simulation and  $D$  be the vector such that  $D_i$  represent the proportion of individuals with  $\log_{10}$  SPVL of  $i$ .  $Y_i$  is taken from normal distribution with mean 3 and is normalized so that  $\sum_i D_i = 1$ . Then, we have the following initial distribution of each states:

$$\begin{aligned} S(0) &= (1 - \epsilon)S^* \\ SS(0) &= (1 - \epsilon)^2 SS^* \\ SI_i(0) &= 2\epsilon(1 - \epsilon)SS^* D_i \\ I_i(0) &= \epsilon S^* D_i \\ II_{ij}(0) &= \left(\frac{1}{2}\right)^{\delta_{ij}} 2\epsilon^2 SS^* D_i D_j. \end{aligned} \quad (26)$$

Since model 3 and 4 do not have single state,  $SS^* = 1$  at the disease free equilibrium and the initial distribution becomes as follows:

$$\begin{aligned}
SS(0) &= (1 - \epsilon)^2 SS^* \\
SI_i(0) &= 2\epsilon(1 - \epsilon)SS^* D_i \\
II_{ij}(0) &= \left(\frac{1}{2}\right)^{\delta_{ij}} 2\epsilon^2 SS^* D_i D_j.
\end{aligned} \tag{27}$$

Lastly, as model 5 is an implicit model, which does not consider different stages of partnership, we have the following initial distribution.

$$\begin{aligned}
S(0) &= 1 - \epsilon \\
I_i(0) &= \epsilon D_i.
\end{aligned} \tag{28}$$

Model 6 also shares the same distribution of initial infected individuals as model 5.

## References

- Alizon, S. (2009, May). The Price equation framework to study disease within-host evolution. *Journal of Evolutionary Biology* 22(5), 1123–1132.
- Alizon, S. and Y. Michalakis (2015, January). Adaptive virulence evolution: the good old fitness-based approach. *Trends in Ecology & Evolution* 30(5), 248–254.
- Blower, S. M., D. Hartel, H. Dowlatabadi, R. M. Anderson, and R. M. May (1991, February). Drugs, sex and HIV: A mathematical model for New York City. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 331(1260), 171–187.
- Champredon, D., S. Bellan, and J. Dushoff (2013, 12). HIV sexual transmission is predominantly driven by single individuals rather than discordant couples: A model-based approach. *PLoS ONE* 8(12), e82906.
- Day, T. and S. R. Proulx (2004, April). A general theory for the evolutionary dynamics of virulence. *The American Naturalist* 163(4), E40–E63.
- Dwyer, G., S. Levin, and L. Buttel (1990). A simulation model of the population dynamics and evolution of myxomatosis. *Ecol Monog* 60, 423–447.
- Ebert, D. (1999). The evolution and expression of parasite virulence. In S. C. Stearns (Ed.), *Evolution in Health & Disease*, Chapter 14, pp. 161–172. New York: Oxford University Press, Oxford, UK.
- Ebert, D. and J. J. Bull (2003). Challenging the trade-off model for the evolution of virulence: is virulence management feasible? *Trends Microbiol* 11(1), 15–20.
- Fraser, C., T. D. Hollingsworth, R. Chapman, F. de Wolf, and W. P. Hanage (2007). Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *PNAS* 104, 17441–17446.

- Fraser, C., K. Lythgoe, G. E. Leventhal, G. Shirreff, T. D. Hollingsworth, S. Alizon, and S. Bonhoeffer (2014, March). Virulence and Pathogenesis of HIV-1 Infection: An Evolutionary Perspective. *Science* 343(6177), 1243727.
- Herbeck, J. T., J. E. Mittler, G. S. Gottlieb, and J. I. Mullins (2014, June). An HIV epidemic model based on viral load dynamics: Value in assessing empirical trends in HIV virulence and community viral load. *PLoS Comput Biol* 10(6), e1003673.
- Herbeck, J. T., V. Mller, B. S. Maust, B. Ledergerber, C. Torti, S. Di Giambenedetto, L. Gras, H. F. Gnthard, L. P. Jacobson, J. I. Mullins, and G. S. Gottlieb (2012, January). Is the virulence of HIV changing? A meta-analysis of trends in prognostic markers of HIV disease progression and transmission. *AIDS (London, England)* 26(2), 193–205.
- Hollingsworth, T., R. Anderson, and C. Fraser (2008, September). HIV-1 Transmission, by Stage of Infection. *Journal of Infectious Diseases* 198(5), 687–693.
- Payne, R., M. Muenchhoff, J. Mann, H. E. Roberts, P. Matthews, E. Adland, A. Hempenstall, K.-H. Huang, M. Brockman, Z. Brumme, M. Sinclair, T. Miura, J. Frater, M. Essex, R. Shapiro, B. D. Walker, T. Ndungu, A. R. McLean, J. M. Carlson, and P. J. R. Goulder (2014, December). Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. *Proceedings of the National Academy of Sciences* 111(50), E5393–E5400.
- Shirreff, G., L. Pellis, O. Laeyendecker, and C. Fraser (2011, October). Transmission selects for HIV-1 strains of intermediate virulence: A modelling approach. *PLoS Computational Biology* 7(10), e1002185. WOS:000297262700019.

Table 1: Parameter ranges/values. Note that  $c$  and  $\rho$  values are doubled from Champredon et al. because we keep track of individuals, while they keep track of couples

Notation	Description	Range/Value	Source
$\rho$	Partnership formation rate	1/10-2/5	<a href="#">Champredon et al. (2013)</a>
$c$	Partnership dissolution rate	1/15-1/5	<a href="#">Champredon et al. (2013)</a>
$c_u/c_w$	Relative contact rate uncoupled	1/5-5	Assumption
$c_e/c_w$	Relative contact rate extra-couple	0.01-1	<a href="#">Champredon et al. (2013)</a>
$\beta_P$	Rate of transmission during primary infection	1.31-5.09	<a href="#">Hollingsworth et al. (2008)</a>
$\beta_D$	Rate of transmission during high transmission disease stage	0.413-1.28	<a href="#">Hollingsworth et al. (2008)</a>
$D_P$	Duration of primary infection	1.23/12-6/12	<a href="#">Hollingsworth et al. (2008)</a>
$D_D$	Duration of high transmission disease stage	4.81/12-14/12	<a href="#">Hollingsworth et al. (2008)</a>
$\beta_{\max}$	Maximum rate of transmission during asymptomatic stage	0.317	<a href="#">Shirreff et al. (2011)</a>
$\beta_{50}$	SPVL at which infectiousness is half maximum	13938	<a href="#">Shirreff et al. (2011)</a>
$\beta_k$	Hill coefficient: steepness of increase in infectiousness as a function of SPVL	1.02	<a href="#">Shirreff et al. (2011)</a>
$D_{\max}$	Duration of primary infection	25.4	<a href="#">Shirreff et al. (2011)</a>
$D_{50}$	SPVL at which duration of asymptomatic infection is half maximum	3058	<a href="#">Shirreff et al. (2011)</a>
$D_k$	Hill coefficient: steepness of decrease in duration as a function of SPVL	0.41	<a href="#">Shirreff et al. (2011)</a>
$\sigma_M$	Mutation standard deviation of $\log_{10}$ SPVL	0.12	<a href="#">Shirreff et al. (2011)</a>
$\alpha_{\min}$	Minimum $\log_{10}$ SPVL	2	<a href="#">Shirreff et al. (2011)</a>
$\alpha_{\max}$	Maximum $\log_{10}$ SPVL	2	<a href="#">Shirreff et al. (2011)</a>
$n$	Number of strains	21	Assumption