5-2022

Mathematical Modeling Suggests Cooperation of Plant-Infecting Viruses

Joshua Miller
University of Tennessee, Knoxville, jmill233@vols.utk.edu

Vitaly V. Ganusov
University of Tennessee, Knoxville

Tessa Burch-Smith

Mathematical Modeling Suggests Cooperation of Plant-Infecting Viruses

Joshua Miller[1*], Vitaly V. Ganusov[1,3], Tessa M. Burch-Smith[2]

[1]Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA
[2]The Donald Danforth Plant Science Center, St. Louis, MO 63132, USA
[3]Department of Microbiology, University of Tennessee, Knoxville, TN 37996, USA
*Corresponding author: jmill233@vols.utk.edu

This paper is submitted to satisfy the thesis requirement of the Chancellor's Honors Program.

Department of Mathematics
University of Tennessee, Knoxville
United States
3/28/2022

Abstract

Viruses are major pathogens of agricultural crops. Viral infections often start after the virus enters the outer layer of a tissue, and many successful viruses, after local replication in the infected tissue, are able to spread systemically. Quantitative details of virus dynamics in plants, however, are poorly understood, in part, because of the lack of experimental methods which allow the accurate measurement of the degree of infection in individual plant tissues. Recently, Tromas et al. PLoS Genetics (2014) followed the kinetics of infection of individual cells in leaves of *Nicotiana tabacum* plants using Tobacco etch virus (TEV), labeled with either Venus or blue fluorescent protein (BFP) to produce to different strains. Assuming that viral spread occurs from lower to upper leaves, the authors fitted a simple mathematical model to the frequency of cellular infection by the two viral variants found using flow cytometry. While the original model could accurately describe the kinetics of viral spread locally and systemically, we found that many alternative versions of the model, for example, if viral spread starts at upper leaves and progresses to lower leaves or when virus dissemination is stopped due to an immune response, fit the data with reasonable quality, and yet with different parameter estimates. These results strongly suggest that experimental measurements of the virus infection in individual leaves may not be sufficient to identify the pathways of viral dissemination between different leaves and reasons for viral control. We propose experiments that may allow discrimination between the alternatives. By analyzing the kinetics of coinfection of individual cells by Venus and BFP strains of TEV we found a strong deviation from the random infection model, suggesting cooperation between the two strains when infecting plant cells. Importantly, we showed that many mathematical models on the kinetics of coinfection of cells with two strains could not adequately describe the data, and the best fit model needed to assume i) different susceptibility of uninfected cells to infection by two viruses locally in the leaf vs. systemically from other leaves, and ii) decrease in the infection rate depending on the fraction of uninfected cells which could be due to a systemic immune response. Our results thus demonstrate the difficulty in reaching definite conclusions from extensive and yet limited experimental data and provide evidence of potential cooperation between different viral variants infecting individual cells in plants.

Introduction

With a burgeoning human population expected to reach between 7 and 13 billion by 2100, humans' lifeblood, food and water, will be evermore difficult to protect and sustain over time [1–3]. Under these circumstances, dependence on agriculture will only increase [4]. Food crops, however, are vulnerable to numerous biotic stresses, including but not limited to animal pests, fungi, bacteria, and viruses. Viral infections especially can devastate food-crops, with documentation of such infections being identified as early as eighth-century Japan; however it was not until the nineteenth century that it was known and accepted that microscopic agents like viruses could cause diseases in plants [5, 6].

Mathematical models have been widely used to understand virus-plant interactions. For example, early studies investigated how the virus concentration in the inoculum influences the number of lesions formed by the virus on plant leaves [7–10]. More recent studies investigated virus dynamics in individual plant cells or in the whole plant [11–14]. Most studies, however, focused on understanding epidemiological spread of viral disease in plant populations with the aim to control viral spread and to limit damage to agricultural crops [15–25].

Mechanisms of viral spread within individual plants remain incompletely understood. Usually infection of a single cell or a small group of cells occurs via mechanical means or by an animal or insect vector. After replication in the inoculation site, virions move to neighboring cells through plasmodesmata – pores between individual cells in the leaf [26]. The replication-movement process is repeated until the virus enters the vasculature. It has been experimentally demonstrated that viral distribution via the vasculature follows the path of sugar distribution, i.e., from source to sink tissues, with strong sinks like roots receiving a larger portion of the viral cargo [27]. Once arriving at sink tissue, the virus exits the vasculature via the plasmodesmata and enters neighboring cells. From there viruses use plasmodesmata once again to invade the ground tissue [27, 28]. In some cases, however, viruses can be introduced directly into the vasculature resulting in rapid infection of sink tissues.

Different methods have been used to measure the degree of infection of a given leaf in the plant including ELISA for viral proteins and PCR for viral genomes [12, 29]. However, these methods are semi-quantitative and typically do not allow measurement of the degree of infection of individual cells in the leaf. Recently, a new method to measure the frequency of infection of cells in plant leaves through the use of flow cytometry was developed [14]. In their experiments, Tromas *et al.* [14] infected lower (3rd) leaves of 4 week old *Nicotiana tabacum* (henceforth referred to as "tobacco") plants with two strains of Tobacco etch virus (TEV), TEV-Venus and TEV-BFP, carrying different fluorescent proteins. At different times after the infection cells (protoplasts) were isolated from individual leaves, and the fraction of protoplasts infected with either or both viral variants was quantified using flow cytometry [14]. Flow cytometry allowed the measurement of virus infection in thousands of individual cells, thus providing unique quantitative information about kinetics of TEV infection in tobacco plants.

Tromas *et al.* [14] performed several important analyses including calculation of basic reproductive number and multiplicity of infection of cells (MOI) by different viruses. In addition, the authors developed a detailed mathematical model of how the virus spreads over time from the 3rd leaf to other leaves and fitted the model to experimental data. Importantly, the model was able to accurately describe virus dissemination and predicted that viral spread kinetics was similar within the leaves. One major difference between infection levels in individual leaves was due to different import rates of the virus from the lower to upper leaves [14].

1

Here we built upon this pioneering work and further analyzed experimental data of Tromas *et al.* [14] with use of mathematical models. The main objectives of our study were to understand the details of dissemination of TEV in tobacco plants and to determine if coinfection of individual plant cells with two TEV variants occur independently. Specifically, because exact pathways of TEV dissemination in tobacco plants have not been unequivocally identified and may depend on the age of plants and details of virus inoculation, we investigated whether mathematical models of TEV dissemination, alternative to the Tromas *et al.* [14] model, may be also consistent with the data. Surprisingly, we found that indeed many different routes of TEV dissemination (e.g., when the initial infection first spreads in top (7th) leaf and then disseminates to lower leaves) are quite consistent with experimental data, even though some such models fitted the data with slightly reduced quality (as evaluated by AIC, [30]). By analyzing kinetics of coinfection of individual cells by two TEV variants we found that coinfection does not proceed randomly; rather, cells are more likely to be coinfected with two viruses than infected with either of the variants suggesting cooperativity in infection (or that plant cells vary in susceptibility to infection). Our results suggest that understanding pathways of virus dissemination in plants will be difficult using only data on virus infection in individual leaves and may likely require specific experiments that determine the systemic distribution of virions in host tissues over the time course of infection.

Materials & Methods

Data

Specific details of how infection of plants had been performed are given in the previous publication [14]. In short, 4-week-old *Nicotiana tabacum L. cv. Xanthi* plants, a widely used model plant host [31], were inoculated into the 3rd leaf with an equal mixture of TEV-BFP and TEV-Venus. These two viral strains express blue and yellow fluorescent proteins, respectively. Preliminary work demonstrated that expression of these proteins does not impair growth kinetics of the viral variants [14]. To measure the kinetics of viral dissemination 3rd, 5th, 6th, and 7th true leaves of individual plants were removed at days 3, 5, 7 and 10 post inoculation; five plants per time point were analyzed. Leaf 4 was skipped because it did not show any infection under the experimental conditions. From these leaves, plant cells with their cell walls removed (protoplasts) were isolated and the number of protoplasts expressing none, one, or both of the two fluorescent proteins was measured with flow cytometry. The data have been formatted and are available as a supplement to this paper.

Mathematical models

Original virus dissemination model of Tromas *et al.* [14]

To predict kinetics of infection of the inoculated leaf and dissemination of infection to other leaves in the plant Tromas *et al.* [14] developed a novel mathematical model. The model tracks the fraction of infected cells in a kth leaf, I_k, over time with S_k being the fraction of susceptible cells. In the model a cell infected with either of two viral variants or both viral variants is considered to be infected. The model assumes that infection starts at leaf 3 and then proceeds in the leaf $k = 3$ at a rate β and disseminates to upper leaves (leaves 5, 6, 7) at a rate proportional to the total infection rate of the leaves below a given leaf k at a rate χ_k (**Figure 1**A). When the virus reaches other leaves, infection

also proceeds locally at a rate β. Local virus dissemination at a kth leaf stops when the fraction of infected cells reaches a critical level ψ_k:

$$\frac{dI_3}{dt} = \beta I_3 S_3, \tag{1}$$

$$\frac{dI_k}{dt} = \beta I_k S_k + \chi_k S_k \sum_{j=3, j \neq 4}^{k-1} I_j, \qquad k = 5, 6, 7, \tag{2}$$

$$S_k = \begin{cases} 1 - \frac{I_k}{\psi_k}, & I_k < \psi_k, \\ 0, & I_k \geq \psi_k, \end{cases} \qquad k = 3, 5, 6, 7, \tag{3}$$

where β is the rate of infection of uninfected cells in the leaf during to local viral spread in the leaf, χ_k is the rate of virus infection of other (upper) leaves, and ψ_k is the level at which infection of new cells in leaf k stops (see eqns. (S.1)–(S.4) for full set of equations of the model). Initial conditions for the model are $I_k(0) = I_0$ if $k = 3$ and $I_k(0) = 0$, otherwise, and $S_k(0) = 1$. In total, this model has 9 parameters to be estimated from the data.

Alternative virus dissemination models for the total leaf infection

While the original mathematical model of Tromas *et al.* [14] seems logical we sought to investigate whether alternative mathematical models of virus dynamics within individual leaves and virus dissemination to other leaves in the plant may be consistent with experimental data. In most of these alternative models we use the same nomenclature for the model parameters (I_0, β, χ_k, and ψ_k) as in Tromas *et al.* [14].

- **Alternative model 1**. In this model the dynamics of infection of the leaf 3 is given in eqn. (1), and instead of summing the infection from all the leaves below, we suppose that only the leaf immediately below the one in question can infect it. The dynamics of uninfected leaves is given by eqn. (3). Dynamics of infection in other leaves is described by the following equations (see **Figure 1**B):

$$\frac{dI_k}{dt} = \beta I_k S_k + \chi_k S_k I_{k-1}, \qquad k = 5, 6, 7. \tag{4}$$

Initial conditions for the model are $I_k(0) = I_0$ if $k = 3$ and $I_k(0) = 0$, otherwise, and $S_k(0) = 1$.

- **Alternative model 2**. Leaf 3 infects only leaf 5 which then infects leaves 6 and 7. Leaf 6 also contributes to the infection of leaf 7. Infection for leaf 3 is given by eqn. (1) and dynamics of uninfected leaves is given by eqn. (3). Dynamics of infection in other leaves is described by the following equations:

$$\frac{dI_5}{dt} = \beta I_5 S_5 + \chi_5 S_5 I_3, \tag{5}$$

$$\frac{dI_k}{dt} = \beta I_k S_k + \chi_k S_k I_5, \qquad k = 6, 7. \tag{6}$$

Initial conditions for the model are $I_k(0) = I_0$ if $k = 3$ and $I_k(0) = 0$, otherwise, and $S_k(0) = 1$.
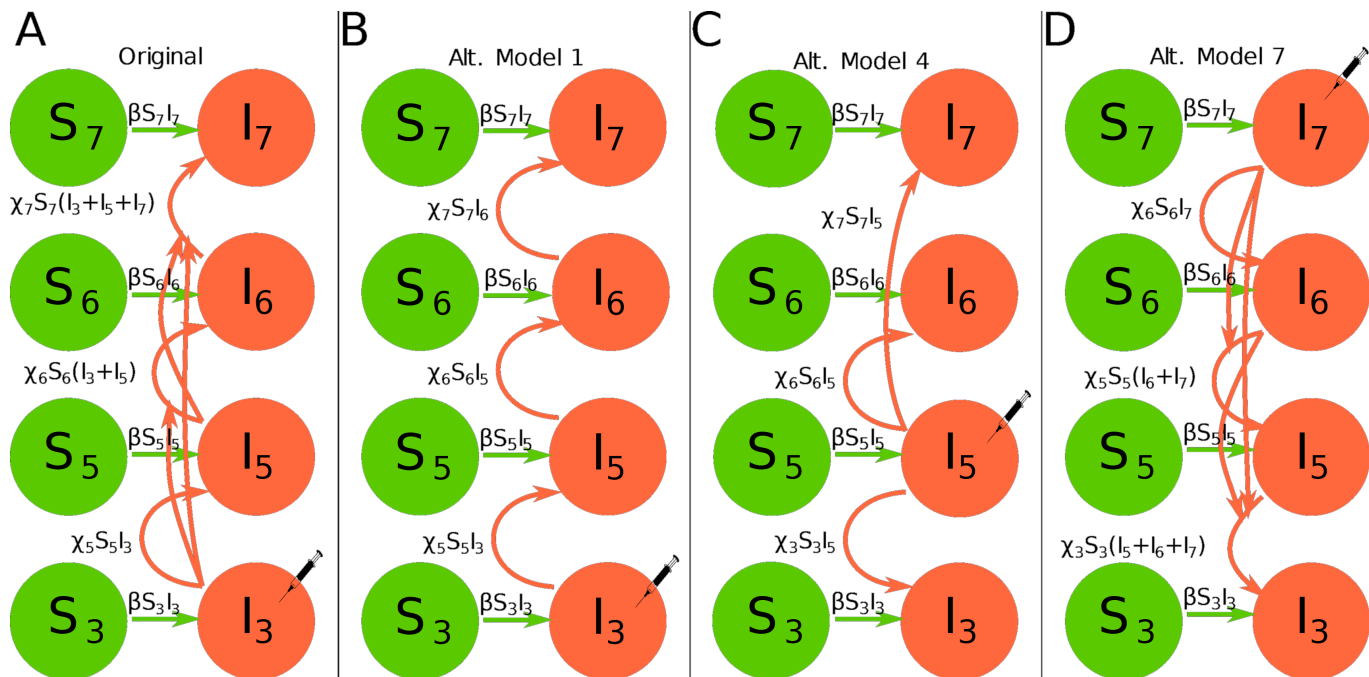
3

Figure 1: Examples of several alternative mathematical models of virus spread in plants analyzed in this paper. In experiments of Tromas *et al.* [14], two different viruses ("Venus" and "BFP") were rubbed into leaf 3 and the fraction of infected leaf cells (protoplasts) was followed by flow cytometry over time (see Materials and Methods for more detail). In schematics, S_k and I_k denote uninfected and infected cells in the kth leaf, respectively, and the syringe indicates the primary place where infection started in the model. Arrows denote the process of leaf infection (at a rate β) and transmission of infection between leaves (at a rate χ). In the original Tromas *et al.* [14] model (A, eqns. (1)–(3)), infection starts at leaf 3 and is then transported to other leaves at a rate proportional to the total fraction of infected cells in leaves below. In alternative model 1, infection starts with leaf 3 but upper leaves are only infected by the leaves just below them (B, eqn. (4)). In alternative model 4, infection starts in leaf 5 and then proceeds to leaves above or below leaf 5 similar to the alternative model 1 (C, eqn. (11)). Finally, in the alternative model 7 infection starts at the upper leaf 7 and proceeds to lower leaves in a manner similar to the original Tromas *et al.* [14] model (D, eqn. (15)). Other alternative models are described in the Materials and Methods.

- **Alternative model 3**. Infection for leaf 3 is given by eqn. (1) and dynamics of uninfected leaves is given by eqn. (3). Leaf 3 is the only leave that contributes to infections of higher leaves. Dynamics of infection in other leaves is described by the following equations:

$$\frac{dI_k}{dt} = \beta I_k S_k + \chi_k S_k I_3, \qquad k = 5, 6, 7. \tag{7}$$

Initial conditions for the model are $I_k(0) = I_0$ if $k = 3$ and $I_k(0) = 0$, otherwise, and $S_k(0) = 1$.

- **Alternative model 4**. The initial infection occurs on leaf 5 which contributes to infections of leaves 3, 6, and 7. All imported virions for these leaves come exclusively from leaf 5. Dynamics of uninfected leaves is given by eqn. (3). Dynamics of infection in other leaves is described by the following equations (see **Figure 1**C):

$$\frac{dI_5}{dt} = \beta S_5 I_5, \tag{8}$$

$$\frac{dI_k}{dt} = \beta I_k S_k + \chi_k S_k I_5, \qquad k = 3, 6, 7. \tag{9}$$

Initial conditions for the model are $I_k(0) = I_0$ if $k = 5$ and $I_k(0) = 0$, otherwise, and $S_k(0) = 1$.

- **Alternative model 5**. The initial infection occurs on leaf 6 which contributes exclusively to the infections of leaves 3, 5, and 7. Dynamics of uninfected leaves is given by eqn. (3) and dynamics of infection in other leaves is described by the following equations:

$$\frac{dI_6}{dt} = \beta S_6 I_6, \tag{10}$$

$$\frac{dI_k}{dt} = \beta I_k S_k + \chi_k S_k I_6, \qquad k = 3, 5, 7. \tag{11}$$

Initial conditions for the model are $I_k(0) = I_0$ if $k = 6$ and $I_k(0) = 0$, otherwise, and $S_k(0) = 1$.

- **Alternative model 6**. The initial infection occurs on leaf 7 which contributes exclusively to the infections of leaves 3, 5, and 6. Dynamics of uninfected leaves is given by eqn. (3) and dynamics of infection in other leaves is described by the following equations:

$$\frac{dI_7}{dt} = \beta S_7 I_7, \tag{12}$$

$$\frac{dI_k}{dt} = \beta I_k S_k + \chi_k S_k I_7, \qquad k = 3, 5, 6. \tag{13}$$

Initial conditions for the model are $I_k(0) = I_0$ if $k = 7$ and $I_k(0) = 0$, otherwise, and $S_k(0) = 1$.

- **Alternative model 7**. The initial infection occurs on leaf 7 and virus accrues downward; it is essentially the model by Tromas *et al.* [14] being inverted. Dynamics of uninfected leaves is given by eqn. (3) and dynamics of infection in other leaves is described by the following equations (see **Figure 1**D):

$$\frac{dI_7}{dt} = \beta S_7 I_7, \tag{14}$$

$$\frac{dI_k}{dt} = \beta I_k S_k + \chi_k S_k \sum_{j=7}^{k} I_j, \qquad k = 3, 5, 6. \tag{15}$$

Initial conditions for the model are $I_k(0) = I_0$ if $k = 7$ and $I_k(0) = 0$, otherwise, and $S_k(0) = 1$.

- **Alternative model 8**. The model assumes that infection starts in all leaves and proceeds independently (aka "logistic" model for individual leaves). Dynamics of infection in all leaves is described by the following equations:

$$\frac{dI_k}{dt} = \beta_k I_k \left(1 - \frac{I_k}{\psi_k}\right), \qquad k = 3, 5, 6, 7, \tag{16}$$

  Initial conditions are $I_k(0) = I_{0_k}$, $k = 3, 5, 6, 7$. This model has 12 parameters to be estimated from the data.

- **Alternative model 9**. In all previous models virus dissemination within a given leaf stops when the fraction of infected cells reaches ψ_k (e.g., eqn. (3)). This stop of infection is also observed in the data. However, specific mechanisms of why the infection stops while not all cells in the leaf are infected were not fully investigated. Therefore, in our alternative model we assume that the dynamics of virus infection in a given leaf are not infection level-dependent but instead time-dependent. We define T_k to be the time that the kth leaf accumulates the "immune response" to stop the spread of the virus inside it, and n_k represents how quickly this immune response kicks in [32]. The dynamics of the infection is given by the same equations as in the Tromas *et al.* [14] model (eqns. (1)–(2)), and the dynamics of uninfected cells available for infection due to generation of the immune response in the kth leaf is given by

$$S_k = \frac{1}{1 + \left(\frac{t}{T_k}\right)^{n_k}}, \tag{17}$$

  where the initial conditions for the model are $I_k(0) = I_0$ if $k = 3$ and $I_k(0) = 0$, otherwise. This model has 4 extra parameters as compared to other alternative models but the model can be reduced in size by assuming that some of the parameters (e.g., T_k or n_k) to be leaf number-independent (see Main text for results). In such cases, the model has 10 parameters to be estimated from the data.

Virus dissemination models for the infection/coinfection with two viral variants

In the experiments, the plants were infected with an equal mixture of two viral variants, TEV-Venus and TEV-BFP [14]. However, the original model of Tromas *et al.* [14] and our previous alternative models did not discriminate between infection of the cells with two variants. The following alternative models now make this distinction. In these models we denote V_k and B_k as the fraction

of Venus-infected and BFP-infected cells, respectively, and the fraction of coinfected cells is denoted as M_k. Because our analysis illustrated that the specific pathway of TEV dissemination in 4-week-old tobacco plants cannot be fully resolved using infection data alone, we assume the dissemination pathway of Tromas *et al.* [14]. Then the dynamics of infection of plant leaves with the two viral variants we use the following equations:

$$\frac{dV_k}{dt} = \beta_V S_k V_k + \chi_k S_k \sum_{i=3}^{k-1} V_i, \qquad k = 3, 5, 6, 7, \tag{18}$$

$$\frac{dB_k}{dt} = \beta_B S_k B_k + \chi_k S_k \sum_{i=3}^{k-1} B_i, \qquad k = 3, 5, 6, 7, \tag{19}$$

$$S_k = \begin{cases} 1 - \frac{I_k(t)}{\psi_k}, & I_k(t) < \psi_k, \\ 0, & I_k(t) \geq \psi_k. \end{cases} \qquad k = 3, 5, 6, 7, \tag{20}$$

where β_B and β_V are the within-leaf infection rates for BFP and Venus viruses, respectively, and $I_k(t) = V_k + B_k + M_k$. Note that we assume that virus dissemination to upper leaves is strain-independent. The initial conditions for all the following models are $V_k(0) = V_0$, $B_k = B_0$, and $M_k(0) = M_0$ if $k = 3$ and 0 otherwise. To describe the kinetics of viral coinfection we consider several alternative mathematical models.

- **1-alpha coinfection model**. In this model, we describe the coinfection growing as dependent on the within-leaf spread dynamics of both viruses. Here, and in other models $V_k B_k$ is proportional to the rate at which coinfections are expected to arise by chance. We sum these these rates assuming that cells are first infected by one variant and then coinfected with another, and use a scaling factor α to indicate synergy ($\alpha > 1$) or inhibition ($\alpha < 1$) of the coinfection process as compared to random, mass action-like infection process:

$$\frac{dM_k}{dt} = \alpha(\beta_B V_k B_k + \beta_V V_k B_k), \qquad k = 3, 5, 6, 7. \tag{21}$$

This model has 12 parameters.

- **2-alpha coinfection model**. We assume that the rate of coinfection may proceed differently by the two viral strains denoted by α_1 and α_2 which is a simple extension of the 1-alpha coinfection model (eqn. (21)):

$$\frac{dM_k}{dt} = \alpha_1 \beta_B V_k B_k + \alpha_2 \beta_V V_k B_k, \qquad k = 3, 5, 6, 7. \tag{22}$$

This model has 13 parameters.

- **Probabilistic model**. Because V_k and B_k measure the fraction of cells infected by the particular virus in the kth leaf, then for fraction of coinfected cells, M_k, we can think of the probability of a cell being infected by both strains as being determined by $B_k V_k$. We can then use parameter α to measure how much more or less often coinfection is happening as compared

with random chance: $\alpha = 1$ means coinfection is behaving like a random process; $\alpha < 1$ means coinfection is occurring less often than it would by random chance, and $\alpha > 1$ means coinfection is occurring with greater frequency than random chance [33]. Multiplying by α the product $B_k V_k$ and differentiating it with respect to t gives:

$$\frac{d}{dt}(\alpha B_k V_k) = \alpha \left(B_k \frac{dV_k}{dt} + V_k \frac{dB_k}{dt} \right), \tag{23}$$

which then with the use of eqns. (18)–(19) results in the following model for the dynamics of coinfected cells:

$$\frac{dM_k}{dt} = \alpha \left[S_k V_k B_k (\beta_B + \beta_V) + \chi_k S_k \left(B_k \sum_{i=3}^{k-1} V_i + V_k \sum_{i=3}^{k-1} B_i \right) \right], \quad k = 3, 5, 6, 7. \tag{24}$$

This model has 12 parameters. Note that in contrast with previous models (e.g., eqn. (22)), in this model coinfection within the leaf depends on the fractions of uninfected (S_k) and virus-infected cells in the leaf (V_k and B_k).

- **2-alpha probabilistic model**. As in the original Tromas *et al.* [14] model, the equation for coinfection in the probabilistic model is composed of two parts (eqn. (24)): the first term with parameters β_B and β_V represents the within-leaf spread, and the second term with the parameter χ_k represents the leaf-to-leaf spread. It seemed reasonable that coinfection may be driven more by one form of spread or the other, so we used α_1 and α_2 to measure their respective contributions:

$$\frac{dM_k}{dt} = \alpha_1 S_k V_k B_k (\beta_B + \beta_V) + \alpha_2 \chi_k S_k \left(B_k \sum_{i=3}^{k-1} V_i + V_k \sum_{i=3}^{k-1} B_i \right), \quad k = 3, 5, 6, 7. \tag{25}$$

This model has 13 parameters.

- **Logistic model for coinfection growth**. The details of how plant cells become coinfected by two different viruses during the local spread are not fully understood. Because typically plant viruses spread to adjacent cells via plasmodesmata, a coinfected cell may be a source of both viral strains when infecting neighboring cells. In this alternative model we therefore assume that the frequency of coinfected cells increases randomly due to viral dissemination systemically from other leaves and logistically due to local, within-leaf spread:

$$\frac{dM_k}{dt} = \alpha \left[S_k M_k (\beta_B + \beta_V) + \chi_k S_k \left(B_k \sum_{i=3}^{k-1} V_i + V_k \sum_{i=3}^{k-1} B_i \right) \right], \quad k = 3, 5, 6, 7. \tag{26}$$
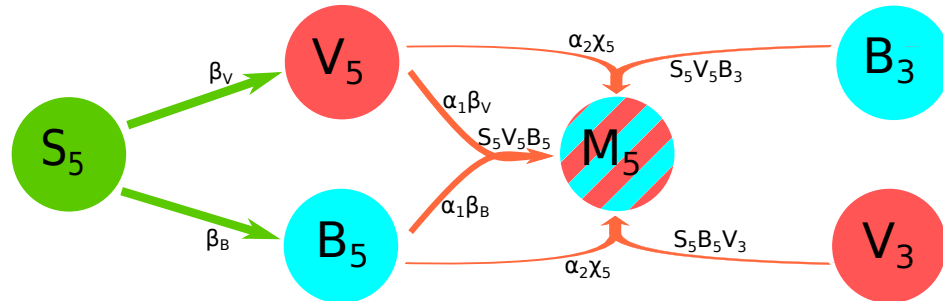
This model has 12 parameters.
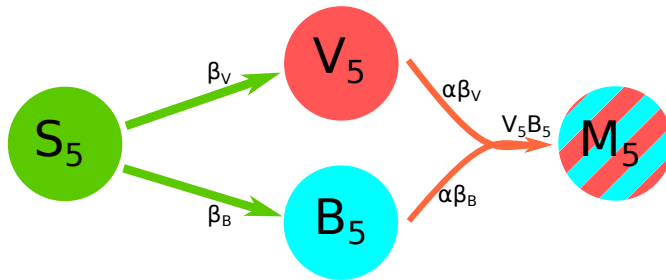
Figure 2: Examples of schematics of two alternative mathematical models for coinfection with two viruses. These diagrams show the infection pathways of infection by TEV-Venus (V_k) or TEV-BFP (B_k) in the leaf 5 of the plants, and how these strains combine to form coinfected cells (M_k) in the 1-alpha coinfection model (A, eqn. (21)) or 2-alpha probabilistic model (B, eqn. (25)). Major model parameters such as β and χ_k have the same meaning as in the previous models (e.g., **Figure 1**). We only show what happens in leaf 5 because in the 2-alpha probabilistic model (B, eqn. (22)), the connections between higher leaves become very complicated and difficult to illustrate in a figure such as this one. Like **Figure 1**, arrows represent the transmission of virions. In the 1-alpha coinfection Model (A, eqn. (21)), coinfection comes from the combination of the Venus and BFP viruses within leaf 5 only. In the 2-alpha probabilistic model (B, eqn. (22)), coinfection comes also from the combination of Venus and BFP virions in leaf five, but is also fed by the combination of virions imported from leaf 3 and combining with their opposite, e.g. Venus from leaf 3 combining with BFP from leaf 5. Two separate alpha terms are used to distinguish dynamics between within-leaf growth and infection from virions imported from lower leaves.

- **2-alpha logistic model for coinfection growth**. Similarly to the 2-alpha probabilistic model, the rate of coinfection may be different between local and systemic viral spread (eqn. (25)). Therefore, we use α_1 and α_2 to differentiate between coinfection occurring as within-leaf and leaf-to-leaf/systemic spread, respectively:

$$\frac{dM_k}{dt} = \alpha_1 S_k M_k (\beta_B + \beta_V) + \alpha_2 \chi_k S_k \left(B_k \sum_{i=3}^{k-1} V_i + V_k \sum_{i=3}^{k-1} B_i \right), \quad k = 3, 5, 6, 7. \quad (27)$$

This model has 13 parameters.

Statistical treatment

To fit models to data we used two alternative approaches. Tromas *et al.* [14] proposed to use the following binomial distribution-based likelihood to fit the models to data

$$L(I_{k,p,t}|A_{k,p,t}, V_{k,p,t}) = \prod_{k,t,p} I_{k,p,t}^{V_{k,p,t}} (1 - I_{k,p,t})^{A_{k,p,t}-V_{k,p,t}}, \tag{28}$$

where L is the likelihood of the model given the data, $I_{k,t,p}$ is the model prediction for the frequency of infection (by either or both viral variants) of the particular leaf k and time point t of a plant p, $V_{k,p,t}$ is the number of infected cells observed in a sample, $A_{k,p,t}$ is the total number of cells observed in the sample (k is the leaf number, $p = 1 \ldots 5$ is the plant replicate number, and t is the day on which the observation was made). The model parameters are estimated by minimizing the negative log likelihood nll

$$nll = -\sum_{k,t,p} \left(V_{k,p,t} \log(I_{k,p,t}) + (A_{k,p,t} - V_{k,t,p}) \log(1 - I_{k,t,p}) \right). \tag{29}$$

In the "coinfection" models we track the dynamics of cells infected with individual viral strains as well as coinfected cells. In these models $I_{V_{k,p,t}}$, $I_{B_{k,p,t}}$, $I_{M_{k,p,t}}$ represent the model predictions for the frequency of Venus- or BFP-infected, or coinfected cells, respectively. Therefore, to fit the coinfection models to data we extended the binomial distribution-based likelihood in the following way. We let $V_{V_{k,t,p}}$, $V_{B_{k,t,p}}$, $V_{M_{k,t,p}}$ be the number of cells infected by Venus, BFP, or both, respectively, as was measured experimentally. Note that $V_{k,t,p} = V_{V_{k,t,p}} + V_{B_{k,t,p}} + V_{M_{k,t,p}}$. Then we let

$$nll_V = -\sum_{k,t,p} \left(V_{V_{k,p,t}} \log(I_{V_{k,p,t}}) + (A_{k,p,t} - V_{V_{k,t,p}}) \log(1 - I_{V_{k,t,p}}) \right), \tag{30}$$

$$nll_B = -\sum_{k,t,p} \left(V_{B_{k,p,t}} \log(I_{B_{k,p,t}}) + (A_{k,p,t} - V_{B_{k,t,p}}) \log(1 - I_{B_{k,t,p}}) \right), \tag{31}$$

$$nll_M = -\sum_{k,t,p} \left(V_{M_{k,p,t}} \log(I_{M_{k,p,t}}) + (A_{k,p,t} - V_{M_{k,t,p}}) \log(1 - I_{M_{k,t,p}}) \right), \tag{32}$$

and nll is simply

$$nll = nll_V + nll_B + nll_M, \tag{33}$$

where the best fit parameters are found by minimizing the nll.

Binomial distribution-based likelihood takes into account the number of cells (protoplasts) extracted from each leaf. The total number of extracted cells varied dramatically between leaves (by up to 8 fold). It was therefore possible that different numbers of cells in the data may skew the likelihood-based estimates towards measurements with more cells. We therefore aimed to investigate whether other methods, e.g., assuming normally distributed data, i.e., normal distribution-based likelihood or least squares, can be used to fit the models to data. We tried several different ways of how least squares could be used to fit the models to data.

One approach is to use the frequency of infected cells $I_{k,t,p}$ as predicted by the mathematical model with the data $V_{k,t,p}/A_{k,t,p}$. For the models that only consider uninfected and infected cells

(i.e., cells infected with either viral variant or coinfected with both variants), the sum of squared residuals (SSR) was then calculated as follows:

$$SSR = \sum_{k,t,p} \left(\frac{V_{k,t,p}}{A_{k,t,p}} - I_{k,t,p} \right)^2. \tag{34}$$

In our analyses we found that such a method does not typically result in normally distributed residuals (see Results section for details). Given large variability in the frequency of infected cells over time we applied log-transformation to the data and the model predictions and calculated the SSR using the following formula:

$$SSR_{Log} = \sum_{k,t,p} \left(\log \left(\frac{V_{k,t,p}}{A_{k,t,p}} \right) - \log \left(I_{k,t,p} \right) \right)^2, \tag{35}$$

where the notations are the same as in eqn. (34). Log-transformation of the data, however, is problematic because in 2 cases of leaf 5 infection, the measured frequency of infected cells was 0. One approach was to remove such data points from the analysis but data removal can generate biases in the model fits, and therefore, we opted for a more appropriate approach whereby we replaced zeros in the data and the model predictions with the limit of detection (LOD). LOD in the data for infected cells was defined as the lowest value of the frequency of infected cells found in the data (for infected cells LOD $= 5.12 \times 10^{-4}$).

Similarly to eqn. (33) we used the following definition for SSR to fit the coinfection models to the data on the frequency of cellular infection with Venus ($V_{V_{k,t,p}}$), BFP ($V_{B_{k,t,p}}$) or both viruses (Mixed, $V_{M_{k,t,p}}$)

$$SSR = \sum_{k,t,p} \left(\frac{V_{V_{k,t,p}}}{A_{k,t,p}} - I_{V_{k,t,p}} \right)^2 + \sum_{k,t,p} \left(\frac{V_{B_{k,t,p}}}{A_{k,t,p}} - I_{B_{k,t,p}} \right)^2 + \sum_{k,t,p} \left(\frac{V_{M_{k,t,p}}}{A_{k,t,p}} - I_{M_{k,t,p}} \right)^2, \tag{36}$$

and the following is the log-transformed variant (eqn. (37)):

$$\begin{aligned} SSR_{Log} &= \sum_{k,t,p} \left(\log \left(\frac{V_{V_{k,t,p}}}{A_{k,t,p}} \right) - \log \left(I_{V_{k,t,p}} \right) \right)^2 + \sum_{k,t,p} \left(\log \left(\frac{V_{B_{k,t,p}}}{A_{k,t,p}} \right) - \log \left(I_{B_{k,t,p}} \right) \right)^2 + \\ &\quad \sum_{k,t,p} \left(\log \left(\frac{V_{M_{k,t,p}}}{A_{k,t,p}} \right) - \log \left(I_{M_{k,t,p}} \right) \right)^2, \end{aligned} \tag{37}$$

where data in which the frequency of infected cells was zero, we replaced these zero values with the LOD for frequency of cells infected with different viral variants as $LOD_{Venus} = 8.43 \times 10^{-5}$, $LOD_{BFP} = 3.26 \times 10^{-4}$, and $LOD_{Mixed} = 3.2 \times 10^{-5}$.

For binomial distribution-based likelihood, confidence intervals for best fit parameters were estimated by bootstrapping the data with replacement (sampling a given plant) 1000 times [34]. For

11

least squares, we used routine `minimize` from the python library `lmfit` that provided 95% confidence intervals for the estimated parameters.

To compare alternative mathematical models we used Akaike Information Criterion, AIC, that are calculated differently for binomial distribution- and normal distribution-based (least squares) likelihoods [30]:

$$AIC_{Lik} = 2N_{par} - 2\ln(L) = 2nll + 2N_{par},$$ (38)

$$AIC_{LS} = N\log\left(\frac{SSR}{N}\right) + 2N_{par},$$ (39)

$$AIC_{LS_{log}} = N\log\left(\frac{SSR_{Log}}{N}\right) + 2N_{par},$$ (40)

where N is the number of data points in the sample (in this case $N = 80$), and N_{par} is the number of model parameters estimated by fitting the model to the data. Note that AIC differences of 0-4 are typically considered to be small while a difference of 10 indicates inferiority of the model in describing the data [30].

If plant cells are infected randomly by two different strains of the virus we expect that the frequency of coinfections with two viruses should be proportional to the product of the frequency of infections with single viral strains. To estimate the deviation from the random coinfection we used Odds Ratio of infection (OR) proposed previously to estimate deviation from random coinfection for HIV [33]:

$$OR = \frac{(A_{k,t,p} - V_{k,t,p}) \times M_{k,t,p}}{V_{V_{k,t,p}} \times V_{B_{k,t,p}}},$$ (41)

where $A_{k,t,p} - V_{k,t,p}$ is the number of uninfected cells and $V_{k,t,p} = V_{V_{k,t,p}} + V_{B_{k,t,p}} + V_{M_{k,t,p}}$ is the total number of infected cells in the data for the kth leaf, time point t, and plant p.

Programming details

All major analyses were done in Python (ver. 3.7.2) and some analyses were repeated in R (ver. 3.9.1). Python libraries used were `matplotlib` (ver. 3.3.2), `Pandas` (ver. 1.1.3), `NumPy` (ver. 1.19.0), `lmfit` (ver. 1.0.1), and `SciPy` (ver. 1.5.2). To solve the ODE-based models we used the `odeint` routine from `scipy.integrate` package. To fit models to data we used a differential evolution algorithm when the goodness of fit metric was nll, and when minimizing least squares residuals we used the Levenberg-Marquardt algorithm with a trust region. Both methods are part of Python's `lmfit` library. To ensure reproducibility of our results as a part of this publication we share the data and the code to fit the original virus dissemination model to data using either binomial distribution-based likelihood or least squares, and the code to illustrate the impact of various parameters on the virus dynamics according to the 2-alpha probabilistic model (eqn. (25)).

Results

The experimental dataset of the kinetics of TEV spread

In their original study, Tromas *et al.* [14] manually introduced two different strains of TEV to the third leaf of the 4 week old tobacco plants and counted the number of infected and uninfected cells in different leaves ($k = 3, 5, 6, 7$) of the infected plants over time ($t = 3, 5, 7, 10$ days). Given that plant cells are immotile and are surrounded by cellulosic cell walls, viruses can infect other cells in the leaf via two ways: 1) by passing through pores in the cells' membranes and cell walls (called plasmodesmata) creating portals between adjacent cells, or 2) by entering the vasculature and migrating with phloem to other (sink) leaves of the plant [35]. Over time, the viral infection disseminates unequally between the leaves (**Figure 3** and **Supplemental Figure S1**). In particular, only about 10% of all cells in the originally inoculated leaf 3 become infected by 10 days of infection (**Figure 3**A), while on average 30% of cells become infected in leaves 6 and 7 (**Figure 3**C-D). Interestingly, leaf 5 becomes minimally infected (**Figure 3**B), and infection did not spread to leaf 4 [14]. There was great variability between infection of leaves in individual plants; for example, in leaf 7 by day 10 less than 10% of cells were infected in one plant while over 40% were infected in another plant (**Figure 3**D).
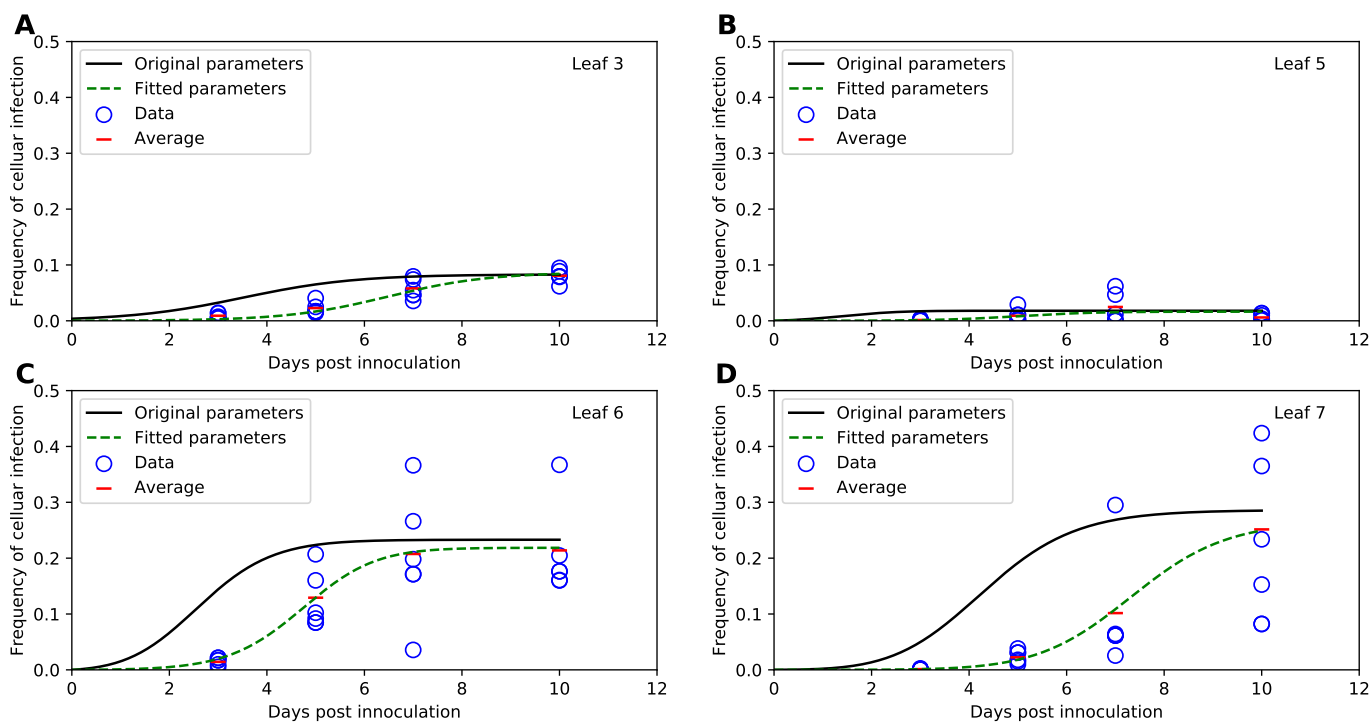
Figure 3: Previously published parameter estimates in Tromas *et al.* [14] do not provide a reasonable description of the data. We simulated basic mathematical model for viral spread in plants developed by Tromas *et al.* [14] (given in eqns. (1)–(3) and **Figure 1**A) using parameter values provided in the original publication (solid lines), or fitted the model to the data using binomial distribution-based likelihood method (eqn. (29), dashed lines). Data for the fraction of infected cells are shown by markers for leaf 3 (A), leaf 5 (B), leaf 6 (C), and leaf 7 (D) with red horizontal lines denoting average fraction of infected cells per time point. Parameters for the model fits are shown in **Table 1**.

Model with Tromas *et al.* [14] parameter values does not match the data

To estimate basic parameters determining kinetics of TEV spread in tobacco plants, Tromas *et al.* [14] developed a mathematical model assuming that virus infection proceeds locally in each leaf and spreads from lower to upper leaves (**Figure 1**A). Via several model iterations, the model in which within-leaf virus spread was leaf number-independent but the virus transport to upper leaves from the lower leaves was leaf number-dependent, fitted the data with best quality [14].

	Original parameters (nll)		New parameters (nll)		New parameters ($log\ LS$)	
Parameter	Estimate	95% CIs	Estimate	95% CIs	Estimate	95% CIs
I_0	0.00372	(0.001,0.017)	0.00022	(0.00011,0.00108)	0.0005	(0.0001,0.0008)
β, 1/day	0.871	(0.257,1.66)	0.950	(0.549,1.183)	0.902	(0.730,1.618)
χ_5, 1/day	0.724	(0.033,0.813)	0.167	(0.042,17.291)	0.063	(0.023,0.162)
χ_6, 1/day	1.38	(0.580,2.340)	1.046	(0.228,1.487)	0.691	(0.387,1.053)
χ_7, 1/day	0.107	(0.050,0.263)	0.029	(0.009,0.160)	0.029	(0.015,0.051)
ψ_3	0.083	(0.053,0.147)	0.080	(0.074,0.134)	0.074	(0.051,0.096)
ψ_5	0.018	(0.002,0.050)	0.016	(0.005,0.024)	0.006	(0.004,0.010)
ψ_6	0.233	(0.155,0.345)	0.224	(0.203,0.287)	0.204	(0.181,0.234)
ψ_7	0.286	(0.234,0.346)	0.269	(0.130,0.418)	0.224	(0.092,0.557)

Table 1: New parameter estimates for the basic mathematical model of virus spread in plants. We list the parameter estimates and 95% confidence intervals (CIs) of the basic mathematical model of viral spread in plants (eqns. (1)–(3)) as provided by Tromas *et al.* [14] ("Original parameters (nll)") or by fitting the model to data in this work using using binomial distribution-based likelihood ("New parameters (nll)", eqn. (29)) or using least squares with a logarithmic transform ("New parameters ($log\ LS$)", eqn. (35)). Fits of the mathematical model for two sets of model parameters are given in **Figures 3 and 4**. Confidence intervals for best fit parameters were generated using bootstrap by resampling the data (for likelihood-based fits) or were provided by the routine `minimize` in from python library `lmfit` for least square-based fits (see Materials and methods for more detail).

To verify these results we simulated virus spread dynamics using Tromas *et al.* [14] published model equations (eqns. (1)–(3)) and parameter values (**Table 1**) and compared model predictions with the data (provided by Tromas *et al.* [14]). Surprisingly, the model predictions did not match the average infection levels observed in the data (solid lines in **Figure 3**). While we did not fully know the exact reasons for this discrepancy, we found that if we were to shift the infection trajectories predicted by the model by 3 days, the model predictions matched the data relatively well (**Supplemental Figure S2**). We therefore hypothesize that when numerically solving the model, Tromas *et al.* [14] may have initiated the solver starting at day 3 post infection given that it is the first time point at which experimental measurements were taken. (It is typical to obtain model predictions for times as given in the data, and solvers in R or python typically take the first time point as the time at which initial conditions are provided and not at the time 0 as is often assumed in models.)

To check that the virus dissemination model of Tromas *et al.* [14] is consistent with experimental data we fitted the model to the data using binomial distribution-based likelihood (see Materials and Methods for more detail). Importantly, the model fitted the data visually with good quality (dashed lines in **Figure 3**) indicating consistency of the model with the data. Interestingly, while some model parameters, such as ψ_k, varied little between the original and corrected values, others such as I_0 or χ_k differed substantially (**Table 1**). While confidence intervals for newly estimated parameters of

the Tromas *et al.* [14] model are a bit large, we found that there is large difference in AIC_{Lik} for this model when used with previously published Tromas *et al.* [14] parameters and our new estimates ($\Delta_{Lik} > 100$, results not shown). Thus, our analysis provided updated and correct estimates of parameters characterizing kinetics of TEV spread in tobacco plants in the Tromas *et al.* [14] model.

Fitting the models using binomial distribution-based likelihood or normal distribution-based likelihood (least squares) delivers similar parameter estimates

In their study, Tromas *et al.* [14] proposed the use of binomial distribution-based likelihood to fit the models to data. In this approach, the probability of a plant cell being infected was treated as a Bernoulli trial in which A total cells are sampled, and the number of infected cells V is determined. While it seemed reasonable it was not fully justified why such a likelihood is a good choice. There may be several potential issues with it. First, because there was a large variability in the total number of cells recovered from different leaves (from minimal 4314 to maximal 32168 protoplasts/leaf), the data are unbalanced. Sources of such variability, however, are not entirely clear and may be due to variation of leaf sizes but also may be related to difficulty of isolating protoplasts from leaves [36]. Parameter estimates may be biased if the fit favors better description of the data with the larger number of isolated cells. Second, while the large number of cells isolated may indicate certainty in estimation of the frequency of infected cells in a sample, there is a great variability in frequency of infected cells in the same leaf number between individual plants (e.g., **Figure 3**D), and binomial distribution-based likelihood may not adequately take such variability into account. Third and finally, given that a relatively large number of cells was measured in each leaf ($> 10^3$), the distribution of the fraction of infected cells per central limit theorem may approach normal distribution, and therefore, one could use a normal distribution-based likelihood (least squares) for fitting models to data.

Therefore, we fitted the Tromas *et al.* [14] model (eqns. (1)–(3)) to the data using several different versions of least squares (see eqns. (34)–(35) and Materials and Methods for more detail). Surprisingly, independent of the method used, the model predictions of the binomial distribution-based fits or least squares fits were nearly identical (e.g., **Figure 4**) and with a minimal, statistically non-significant difference in the parameter estimates for both fits (**Table 1**). Therefore, this result suggests that it may be reasonable to use least squares (or more generally, normal distribution-based likelihood) to fit virus dissemination models to these data. We did, however, find that not all least squares-based methods were appropriate. In particular, least squares with the frequency of infected cells resulted in skewed, non-normally distributed residuals (Shapiro-Wilk test, $W = .785$, $p = 1.887 \times 10^{-9}$). Some of the traditional approaches, for example the $\arcsin(\sqrt{x})$ transformation for the frequency of infected cells did not normalize the residuals ($W = 0.803$, $p = 5.745 \times 10^{-9}$), however, log-transformation in which zero values were replaced with the limit of detection (LOD, see Materials and Methods for more detail) nearly did ($W = 0.963$, $p = 0.021$). Therefore, this analysis suggests that log-transformation of the data and model predictions is a viable alternative to the binomial distribution-based likelihood method of Tromas *et al.* [14] that may better account for variability in the frequency of infected cells between individual plants.
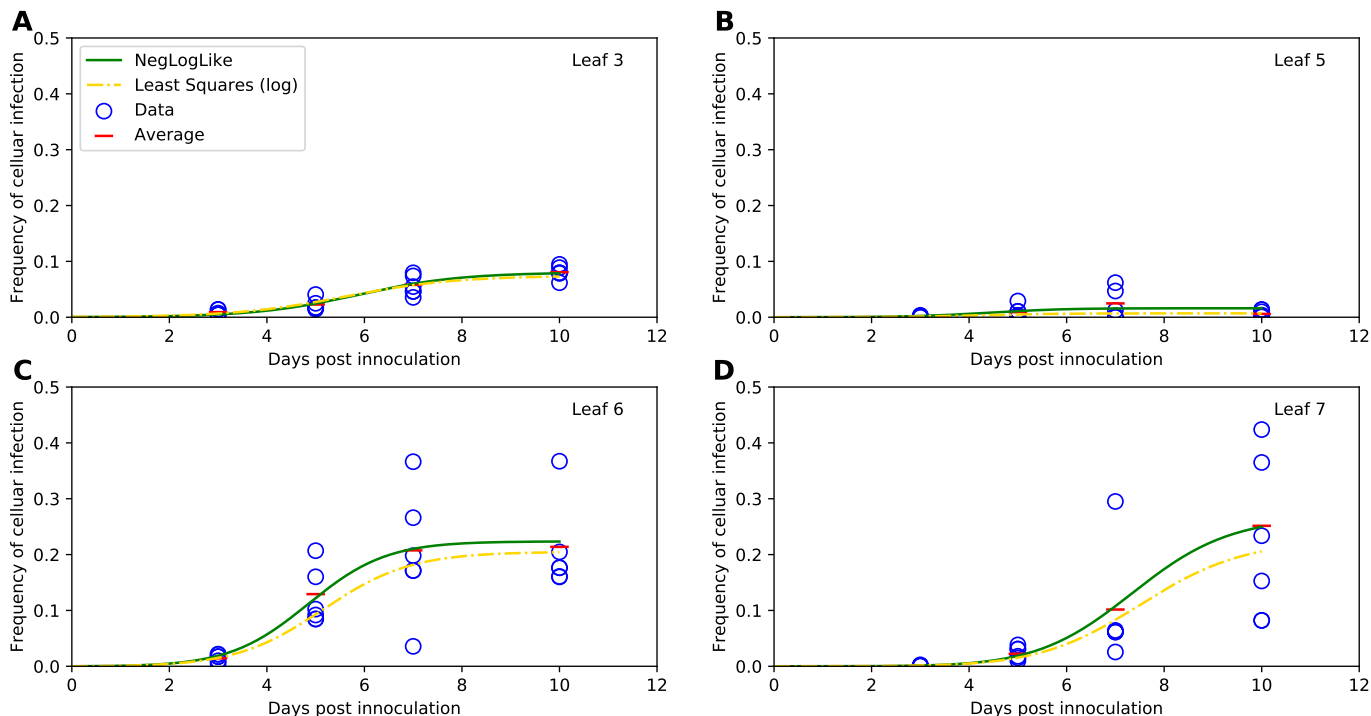
15

Figure 4: Best fits of the basic mathematical model found using either binomial distribution-based likelihood or least squares are nearly identical. We used either binomial distribution-based likelihood method (eqn. (29), [14]) or least squares for log-transformation of the data and model predictions (eqn. (35)) to fit the basic mathematical model (eqns. (1)–(3)) to the virus spread data for leaf 3 (A), leaf 5 (B), leaf 6 (C) and leaf 7 (D). Data on proportion of virus-infected cells are shown by markers and lines are the predictions of best fit models. Parameters for the model fit using likelihood and least squares with the log transform are given in **Table 1** ("New parameters (nll)" and "New parameters ($Log\ LS$)" columns, respectively). In fitting the models using least squares for log-transformed data, the limit of detection (LOD) was LOD $= 5.12 \times 10^{-4}$.

An alternative model with variable within-leaf replication kinetics is consistent with observed viral spread kinetics

In their analysis Tromas *et al.* [14] investigated which parameters of the virus dissemination model may vary with the leaf number. By comparing alternative models they found that ψ and χ must be leaf-dependent to explain the data accurately. However, in that analysis they did not investigate if differences in virus dissemination may be due to variable within-leaf replication kinetics, determined by the parameter β, and not due to virus dissemination rate between leaves χ. Interestingly, we found that the alternative model (based on eqns. (1)–(3)) in which β_k and ψ_k vary with the leaf number k (i.e., virus dynamics in a given leaf is determined mainly by the local spread in the leaf) while systemic dissemination of the virus to upper leaves is constant ($\chi_k = \chi$) fitted the data with similar quality (as judged by SSR or AIC) as the original model. This alternative model has an extra parameter because of four β for four leaves studied while in the original model χ was defined for three leaves only. Yet, this result already suggested that the data on variable virus accumulation in different leaves can be explained equally well by differences in how much virus is delivered to upper leaves (χ_k) or by differences in how the virus replicates and spreads in individual leaves (β_k).

16

Parameter	Original	Alt. Model 1	Alt. Model 2	Alt. Model 3	Alt. Model 4	Alt. Model 5	Alt. Model 6	Alt. Model 7
Fitted with least squares method (Log transformation & $0 \equiv$ LOD)								
I_0	0.0005	0.0008	0.0006	0.0001	0.00006	0.0008	0.0005	0.0005
β, 1/day	0.902	0.744	0.871	1.299	1.050	1.028	1.116	1.159
χ_3, 1/day	N/A	N/A	N/A	N/A	1.672	0.133	1.541	0.286
χ_5, 1/day	0.063	0.059	0.059	0.322	N/A	0.025	0.277	0.050
χ_6, 1/day	0.691	8.201	8.025	3.987	3.663	N/A	3.079	3.263
χ_7, 1/day	0.029	0.073	0.749	0.477	0.332	0.026	N/A	N/A
ψ_3	0.074	0.083	0.075	0.072	0.060	0.059	0.055	0.051
ψ_5	0.006	0.006	0.006	0.005	0.006	0.006	0.005	0.005
ψ_6	0.204	0.204	0.199	0.194	0.201	0.215	0.189	0.184
ψ_7	0.224	0.228	0.238	13.000	0.215	0.211	0.218	0.209
SSR_{Log}	52.713	51.991	52.255	55.612	54.468	55.148	55.486	56.309
$AIC_{SSR_{Log}}$	-15	-16	-16	-11	-13	-12	-11	-10
Δ_{AIC}	1	0	0	5	3	4	5	6
W	0.963	0.959	0.960	0.973	0.969	0.973	0.972	0.973
p	0.021	0.012	0.013	0.099	0.049	0.091	0.080	0.084

Table 2: Several alternative models provide similar fits of the virus spread data with different parameter sets. We fit seven alternative models for virus spread kinetics (given in eqns. (4)–(15)) to the data on viral spread in plants using least squares with a logarithmic transform (see eqn. (35)). Along with parameter values for every model we provide the total error (SSR_{Log}), $AIC_{LS_{Log}}$, and Δ_{AIC} (difference in AIC between the model with the lowest AIC and all other models). We also show the results of the Shapiro-Wilk normality test (W and p value) applied to the residuals of the fitted models.

Alternative models with differing patterns of viral dissemination are largely consistent with observed viral spread kinetics

We next questioned whether a specific pattern of virus dissemination from the inoculated leaf 3 to the upper leaves can be determined from these experimental data. While there is a general understanding of how viruses in plants disseminate after a local infection (e.g., [28]) details of the dissemination may vary by the plant species, age, conditions in which the plant was grown, the virus species, inoculation method, and many other details. For example, the time when individual leaves become sources or sinks for sugar transport – which will influence virus dissemination pathways – depends on many environmental and developmental factors [27]. Because many of these details are unknown for a specific experimental set-up, we investigated if the information provided by the experimental data on the fraction of infected cells in individual leaves over time is sufficient to establish a pattern for systemic viral dissemination.

Therefore, we developed a series of alternative mathematical models in which the pattern of virus dissemination differed in multiple ways from the original dissemination model of Tromas *et al.* [14] (**Figure 1**B-D and eqns. (4)–(17)) and fitted these models to data. For example, alternative model 1 assumed that virus dissemination to upper leaves occurs only from the leaf below it, i.e., from leaf 3 to leaf 5, and then from leaf 5 to leaf 6 and so on (**Figure 1**B). Alternative model 7 assumed that even though virus inoculation occurred at leaf 3, via access to vasculature, the virus immediately disseminated to leaf 7, and then spread to lower leaves (**Figure 1**D and see Materials and Methods for details for other models). Some of these alternative models should not be necessarily considered as inappropriate because, for example, at day 3 after infection, leaf 6 on average had already nearly twice the frequency of infected cells as leaf 3 (0.014 vs. 0.009).

Finding the best fit model depended strongly on the statistical method used for fitting models to data. For example, using binomial distribution-based likelihood method suggested that best fit is provided by the alternative model 2 with the Tromas *et al.* [14] model fitting the data significantly worse (**Supplemental Table S1**, $\Delta_{AIC} = 290$). We hypothesize that this result arose because of the high sensitivity of such a likelihood function to the experimental measurements, especially at the low frequency of infected cells. In contrast, fitting the models to data using least squares (eqn. (34)) provided fits of all models with identical quality (results not shown). This result was driven by the need of the models to more accurately fit the data with high frequency of infected cells in leaves 6 and 7 at later time points, at the expense of poorer fits of other data. These fits, however, were not adequate due to non-normally distributed residuals as was observed when fitting Tromas *et al.* [14] model to data (see above). Finally, fitting the models to log-transformed data (and replacing the zero values with the LOD) provided a more graded classification of alternative models (**Table 2**). In particular, three models (original and alternative models 1&2) assuming that virus dissemination starts from leaf 3 provided better fits (based on AIC) than the models assuming that spread starts from upper leaves (e.g., alternative model 7). Interestingly, the quality of the model fits deteriorated as the models assumed virus dissemination did not originate from leaf 3 — i.e., the models in which dissemination started at leaf 5 or 6 fitted the data with better quality than the model in which dissemination started at leaf 7 (**Table 2** and **Supplemental Figure S3**). This result suggests that the data on virus dissemination does contain the signal indicating the virus most likely starts spreading from leaf 3 upwards; however, the strength of such a statement from our mathematical modeling-based analysis is relatively weak. Thus, these experimental data do not provide strong evidence for a specific route of TEV dissemination in tobacco plants. There is some good news, however. Some parameters appear to be robustly estimated in all the models such as β and ψ_k; that

18

is perhaps unsurprising given that these parameters determine within-leaf viral spread.

Alternative models incorporating independent replication or immune responses are also consistent with observed viral spread kinetics

We tested two additional alternative models for how well they may describe the data. Alternative model 8 assumed that upon virus inoculation, virus disseminates to all leaves and then replicates in individual leaves independently of other leaves, as described by the logistic equation (eqn. (16)). Alternative model 9 assumed that reduction in the fraction of susceptible cells in a leaf is not determined by the fraction of infected cells but by the time since infection (eqn. (17)). The rationale for this modification is that it is possible that infection induces generation of a local or systemic immune response after a delay T_k which renders uninfected cells resistant to infection [32]. Both of these alternative models fitted the data well based on SSR or AIC metrics (**Supplemental Table S2**). Interestingly, the time-dependent cell susceptibility model suggested that differences in how quickly cells become resistant is leaf-dependent (**Supplemental Table S2**); however, this could be due to differences in the timing of initiation of immune responses and/or virus "arrival" in a given leaf (determined by T_k) or the speed at which uninfected cells in the leaf are rendered resistant (determined by n_k). Taken together, our results strongly suggest that multiple pathways of TEV dissemination and growth in individual leaves in the tobacco plants are consistent with the data and additional experiments and/or data need to be involved to eliminate unreasonable models [37].

Odds ratio test implies a higher than random rate of coinfection

Our modeling-based analysis so far and that of Tromas *et al.* [14] treated cells in our data as infected or uninfected. However, in their experiments Tromas *et al.* [14] measured the fraction of cells infected with either or both of two viral strains of TEV, Venus or BFP (see Materials and Methods for more detail). Virus coinfection may impact many facets of viral dynamics and growth. A paramount consequence of two or more virions infecting the same cell simultaneously is that it may result in production of recombinant variants, which has been well documented for human immunodeficiency virus (HIV) [38, 39]. In particular, in acute HIV infection, variants representing recombinants of infecting/founding strains, arose rapidly within a few months; interestingly, a simple mathematical model predicted that accumulation of the variants can be simply due to random coinfection of the susceptible cells by two viral variants [40]. Dang *et al.* [33] investigated whether infection of CD4 T cells in culture occurs randomly by two different HIV variants, HIV-eGFP and HIV-IHSA. The authors proposed an odds ratio (OR) metric to estimate deviation of the rate of cell coinfection with two viruses as compared to single infections (eqn. (41)). Interestingly, in all their experiments with 2 HIV strains and different types of target T cells $OR > 1$ (typically, $OR = 2 - 8$), suggesting that coinfections were observed more often than single infections [33]. The authors explained this result by variability in CD4 T cell susceptibility to infection with susceptible cells being more easily infected with the two variants. A similar result was found later in another study [41]. Given our rich dataset on the dynamics of coinfection of plant cells with two variants of TEV we calculated the OR (eqn. (41)) for every leaf and every time point in our data.

Interestingly, we found very high values for OR for most of the data, all exceeding one, with many values being in the range 10-100 (**Figure 5**). Note that in some cases, mostly for leaf 5, we could not calculate OR due to absence of coinfected cells (**Figure 5**B). OR of 10 to 100 is much higher than that found previously for HIV [33]. There may be several reasons for that. First, it is
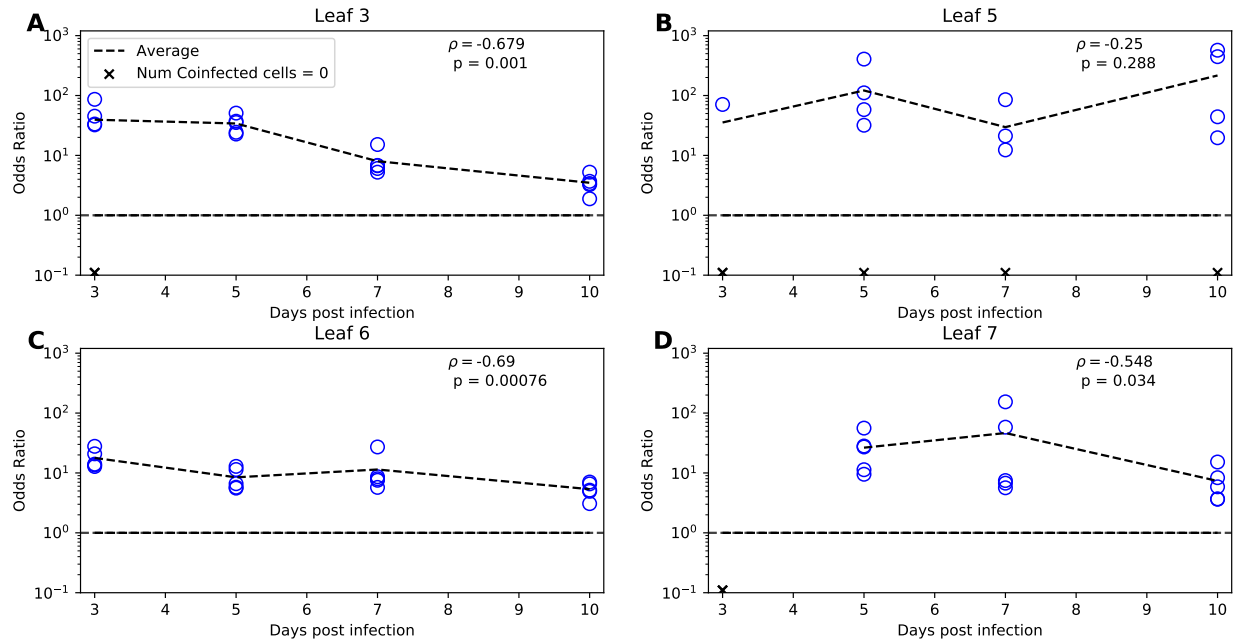
19

Figure 5: There is a high degree of coinfection of individual leaf cells by two different viruses. For each leaf we calculated the odds ratio (the relative probability of a cell being coinfected by two different viruses as compared to infection rate of cells by individual viruses, eqn. (41)) using a previously published method [33]. Note that when infection proceeds randomly, the expected odds ratio is 1. Resulting odds ratio for individual plants are shown for leaf 3 (A), leaf 5 (B), leaf 6 (C), and leaf 7 (D); lines connect the average values per time point. Missing values (when odds ratio could not be calculated) are denoted as crosses. Spearman-Rank correlation ρ of the change in odds ratio with time per leaf and p-values from the test ($\rho = 0$) are shown on individual panels (when calculating ρ missing values were excluded).

possible that there is a high degree of variability in susceptibility of different plant cells to infection, and cells that are highly susceptible get infected with both variants easily. We also found that there is a significant decline in OR with time of infection for all but leaf 5; this decline is consistent with the hypothesis that initially highly susceptible cells are infected resulting in high OR which declines as more resistant cells are infected (**Figure 5**). Alternatively, the mode of virus transmission within the leaf may have played a major role. Indeed, in plants viruses are transmitted from the infected cell to adjacent cells via plasmodesmata, and if a cell is coinfected with two variants, it is possible that all new infections occur by both variants simultaneously [35]. Finally, if infection of cells occurs sequentially, infection with one variant may suppress any potential antiviral activity in the cell, allowing that cell to be coinfected with another variant [42–44]. To further test these hypotheses we used mathematical modeling.

A probability-based coinfection model performs best compared to other coinfection models

Given that many alternative mathematical models are consistent with the pathway of systemic virus dissemination (**Table 2**) to investigate potential mechanisms of TEV coinfection dynamics in different leaves we decided to fix the details of virus dissemination between leaves to those provided in the previous study [14], i.e., we let the virus infection to be initiated in the leaf 3 and dissemination to upper leaves to depend on the infection frequency of leaves below (**Figure 1**A and eqns. (1)–(3)). To describe how coinfected cells are generated we developed six alternative "coinfection" models (see

20

⁵¹⁴ **Figure 2** for 2 examples, eqns. (21)–(27), and Materials and Methods for more detail). In the first,
⁵¹⁵ 1-alpha coinfection model, dynamics of coinfected cells are driven only by the within-leaf frequency
⁵¹⁶ of cells infected with either of two variants with the parameter α determining deviations of the
⁵¹⁷ coinfection from random (**Figure 2**A and eqn. (21)). A simple extension of this model was to allow
⁵¹⁸ for different efficacies of coinfection depending of which virus infected the susceptible cell first (eqn.
⁵¹⁹ (22)). Two other models assumed that coinfection may happen via two different pathways: local,
⁵²⁰ within-leaf infection dependent on the frequency of single-infected cells and uninfected cells and via
⁵²¹ between leaf virus dissemination, with either identical (α) or different (α_1 and α_2) weights for this
⁵²² coinfection processes (**Figure 2**B and eqns. (24)–(25)). Finally, the third set of two models assume
⁵²³ that coinfection due to within-leaf dynamics occurs due to coinfected cells transmitting both viral
⁵²⁴ variants to susceptible cells, and due to between-leaf dynamics occurs similarly as in the previous
⁵²⁵ model. We similarly assume that these two processes may proceed with different deviations from a
⁵²⁶ random process which is captured by parameters α_1 and α_2 (eqns. (26)–(27)).

Parameters	1-a Prob. eqn. (24)	2-a Prob. eqn. (25)	1-a Coin. eqn. (21)	2-a Coin. eqn. (22)	1-a Log. eqn. (26)	2-a Log. eqn. (27)
nll	512112	511728	519132	525576	515892	517656
AIC_{Like}	1024246	1023480	1038286	1051176	1031806	1035336
Δ_{AIC}	766	0	14806	27696	8326	11856
SSR_{Log}	339.459	269.521	335.169	514.719	421.662	477.806
$AIC_{SSR_{Log}}$	109	56	106	211	161	193
Δ_{AIC}	53	0	50	155	105	137

Table 3: The 2-alpha probabilistic model fits the coinfection data with best quality. We fitted a series
of mathematical models (see Materials and methods and **Figure 2**) that make different assumptions on
how coinfection of individual cells with two different viruses occur to the data on viral spread. The models
were fitted using the binomial distribution-based likelihood method (eqn. (33)) or the least squares method
with a log transform of the data (eqn. (37)). AICs were calculated using eqn. (38) and eqn. (40), for the
likelihood and least squares methods respectively. Values for nll, AIC_{Lik}, and $AIC_{LS_{Log}}$ were rounded to
the nearest whole number. Δ_{AIC} for both methods are calculated by taking the AIC score from the model
and method in question and subtracting it from the lowest AIC in its corresponding row. In fitting models
using least squares to log-transformed data we used the following values for the limit of detection of the
frequency of infected cells: $\mathrm{LOD}_{Venus} = 8.43 \times 10^{-5}$, $\mathrm{LOD}_{BFP} = 3.26 \times 10^{-4}$, and $\mathrm{LOD}_{Mixed} = 3.2 \times 10^{-5}$.

⁵²⁷ We fitted these models to experimental data using two alternative approaches, log-transformed
⁵²⁸ least squares (with LOD replacements of zero values) and binomial distribution-based likelihood,
⁵²⁹ both extended to account for singly and co-infected cells in each leaf (see eqn. (37) and eqn. (33) in
⁵³⁰ Materials and methods for more detail). The 2-alpha Probabilistic model (eqn. (22)) was the best
⁵³¹ performing model when fitted by either method (**Table 3**). Importantly, with both methods the
⁵³² basic models assuming that coinfections occur randomly, due to within-leaf coinfection of cells poorly
⁵³³ described the data (**Table 3** and **Supplemental Figure S4**).

⁵³⁴ We also fitted the models using the least squares method for raw, untransformed frequencies of
⁵³⁵ infected cells, but these fits poorly described the dynamics of coinfected cells (see "AllMaterialsAnd-
⁵³⁶ Methods.xlsx" in https://github.com/Plant-Virus-Spread/Models-And-Tools/tree/7553fd98261d1c4b4e75bd3f0
⁵³⁷ We reasoned that this is because there are typically fewer coinfected cells than single-infected cells,
⁵³⁸ and this least squares method favored fitting the dynamics of single-infected cells with better quality
⁵³⁹ (due to their higher abundance). In this specific case, a statistical model based on untransformed
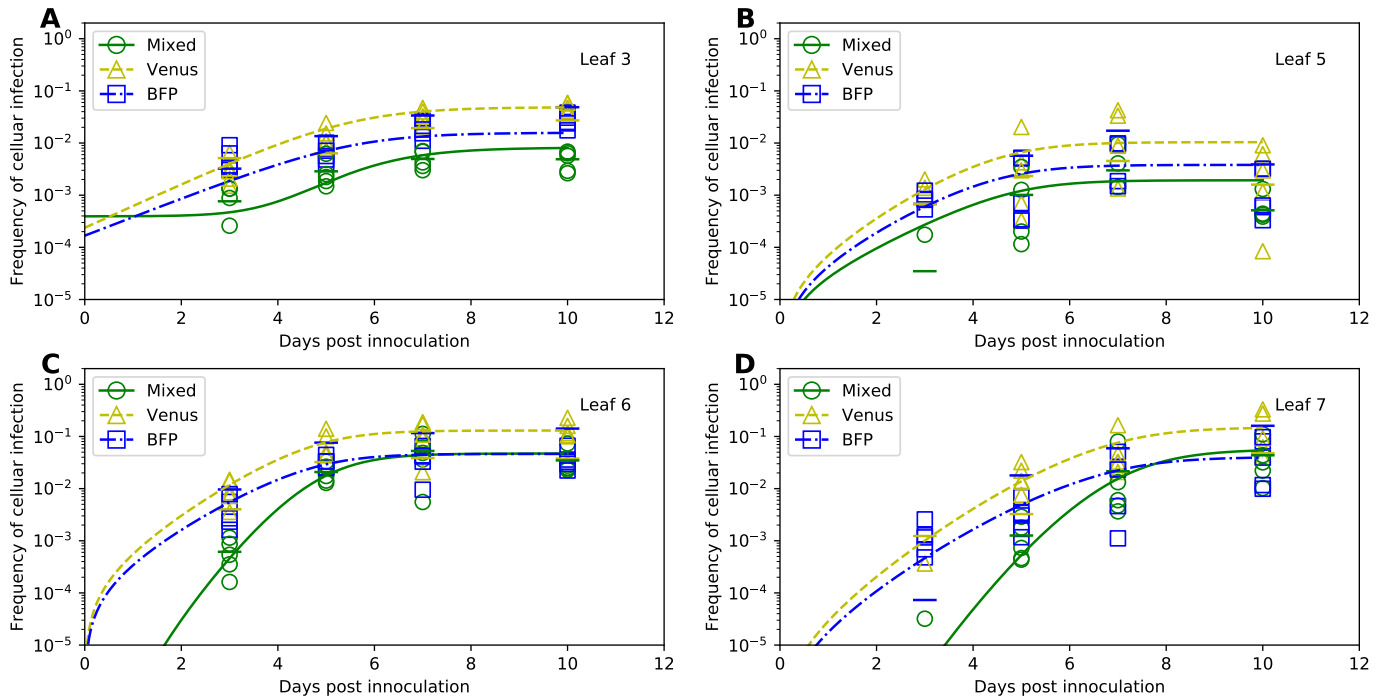
Figure 6: The 2-alpha Probabilistic Model fits the coinfection data with best quality. The 2-alpha Probabilistic Model (eqns. (18)–(19) and eqn. (25)) assumes that coinfection of individual cells by two different strains depends on the level of uninfected cells in the leaf (S_k) and that local coinfections in the leaf occur at different kinetics that coinfection between leaves (**Figure 2**B). We fitted this model to the data on infection of cells by either individual viruses or coinfection of the same cell by different viruses. The model was fitted using the binomial distribution-based likelihood method (eqn. (29)). Markers show frequency of cells infected with Venus or BFP viruses or coinfected with both viruses ("Mixed"), and lines are predictions of the mathematical model. The short horizontal bars show the average infection rate for a given virus variant for a particular day and infected cell type. The parameters providing the best fit and their 95% confidence intervals (estimated using by boostrapping the data) are as follows: $V_0 = 0.0002$ $(2 \cdot 10^{-5}, 0.001)$, $B_0 = 0.0002$ $(2 \cdot 10^{-8}, 0.001)$, $M_0 = 0.0008$ $(0.0, 0.001)$, $\beta_V = 0.975$ $(0.606, 10)$/day, $\beta_B = 0.835$ $(0.426, 10)$/day, $\chi_5 = 0.116$ $(0.004, 9.522)$/day, $\chi_6 = 0.858$ $(0.0001, 10)$/day, $\chi_7 = 0.031$ $(0.0001, 10)$/day, $\psi_3 = 0.073$ $(0.040, 0.118)$, $\psi_5 = 0.016$ $(0.003, 0.260)$, $\psi_6 = 0.223$ $(0.124, 0.260)$, $\psi_7 = 0.247$ $(0.067, 0.400)$, $\alpha_1 = 10.120$ $(3.686, 16.275)$, $\alpha_2 = 0.814$ $(0, 20)$.

541 With both of the appropriate methods we found that the 2-alpha probabilistic model fits the data
542 with best quality, and the next best, 1 alpha probabilistic model performed significantly worse (per
543 *AIC* scores, **Table 3**). Indeed, the best fit model could very accurately describe the dynamics of
544 single- and co-infected cells and predicted a more rapid increase in the coinfected cells for leaf 6 and
545 7 than that for single-infected cells (**Figure 6**). Unfortunately, we found relatively wide confidence
546 intervals for estimates of many of these parameters except ψ_k suggesting that the amount of data
547 available was relatively low, and increasing the number of time points and/or plant repeats may have
548 allowed for more precise estimates. We should note, however, that mean estimates for within-leaf
549 infection rates β_V and β_B and between-leaf spread rates χ_k were very similar to those found when
550 fitting Tromas *et al.* [14] model to the data on infected cell dynamics (**Table 1**) lending some support
551 that our model parameters are not unrealistic. Excitingly, we found that for within-leaf virus spread,
552 coinfection rate was much higher than cell infection by single viruses ($\alpha_1 = 10.1$) supporting our
553 analysis using odds ratio (**Figure 5**). The between-leaf coinfection rate was not different from the

random model ($\alpha \approx 1$) suggesting that most coinfection events were driven by within-leaf dynamics and not due to transfer of viruses systemically. This, perhaps, makes sense because locally it is easier for one cell to be coinfected by 2 viruses while when viruses enter the leaf at random locations due to systemic dissemination, coinfection is expected to be rare.

Both probabilistic models assume that the dynamics of coinfection within the leaf depends on the product of frequency of cells infected with either of two viral variants and the frequency of uninfected cells in the leaf (eqn. (25)). We found that removing S_k term in these models resulted in significantly poorer fit of the data (results not shown). Intuitively, the frequency of uninfected cells drives the dynamics of infection and when S_k approaches 0, infection of the leaf mostly stops, thus over-predicting the data. However, when such a term is absent in the equation for coinfected cells M_k, co-infection would proceed even when single infections stop. With this mechanistic/mathematical insight it was difficult to come up with a biological explanation for why coinfections are dependent on the frequency of uninfected cells. One possibility that infections stop not because the number of uninfected cells declines to zero, but because of leaf-specific immune response makes uninfected cells in the leaf resistant to infection – similar to the alternative model 9 for the dynamics of infected/uninfected cells that we considered earlier (eqn. (17)).

We found it interesting that the model in which the frequency of coinfected cells due to within-leaf dynamics grows logistically (eqns. (26)–(27)) could not well describe the data (**Table 3**). The model underestimated the frequency of coinfected cells at early time points (results not shown). This result argues that the high odds ratio for the coinfection of cells observed in our data is not likely to arise exclusively due to adjacent cells being coinfected with the two TEV variants at once. This model prediction can be tested experimentally, for example, by using microscopy and examining spatial distribution of foci of cells infected with individual viral variants or with both variants [45].

Dynamics of coinfected cells compared to singly-infected cells

While our analysis provided solid evidence that coinfection of plant cells by two TEV variants does not proceed randomly we sought to investigate how coinfection rate varies with the frequency of single-infected cells. Previous mathematical modeling-based work on HIV infection of target cells suggested that the frequency of doubly-infected cells should scale as square of the frequency of single-infected cells [46]. As far as we are aware such prediction has not been tested for plant-infecting viruses. For every leaf we therefore plotted the relationship between the frequency of coinfected cells versus the frequency of cells infected with Venus (**Figure 7**A-D) or BFP (**Figure 7**E-H) strains of TEV and compared these data with predictions of the two alternative probabilistic models. We also fitted a line to log-log transformed frequencies and estimated the slope n of the relationship (**Figure 7**). Several interesting results emerged.

First, we found that the relationship between frequency of coinfected cells and cells infected with a single virus is either sub-linear or linear for lower leaves (leaves 3 and 5, respectively, **Figure 7**A-B and E-F). This is not fully consistent with the results found using odds ratio (**Figure 5**A-B) suggesting that different ways of data analysis may result in different conclusions. However, for upper leaves we found strong deviation from the linear relationship whereby coinfection frequency increased more rapidly than linearly with increasing frequency of single-infected cells ($n > 1$, **Figure 7**C-D and G-H). This is consistent with what we found using odds ratio (**Figure 5**C-D). Predictions of our best fit 2-alpha probability model were mostly consistent with the data except for the leaves 6-7 and cells, singly infected in BFP variant (**Figure 7**G-H). Finally, we noticed that at later time points ($\sim$ 7-10 days post infections), all of the curves in **Figure 7** approximate lines. To understand
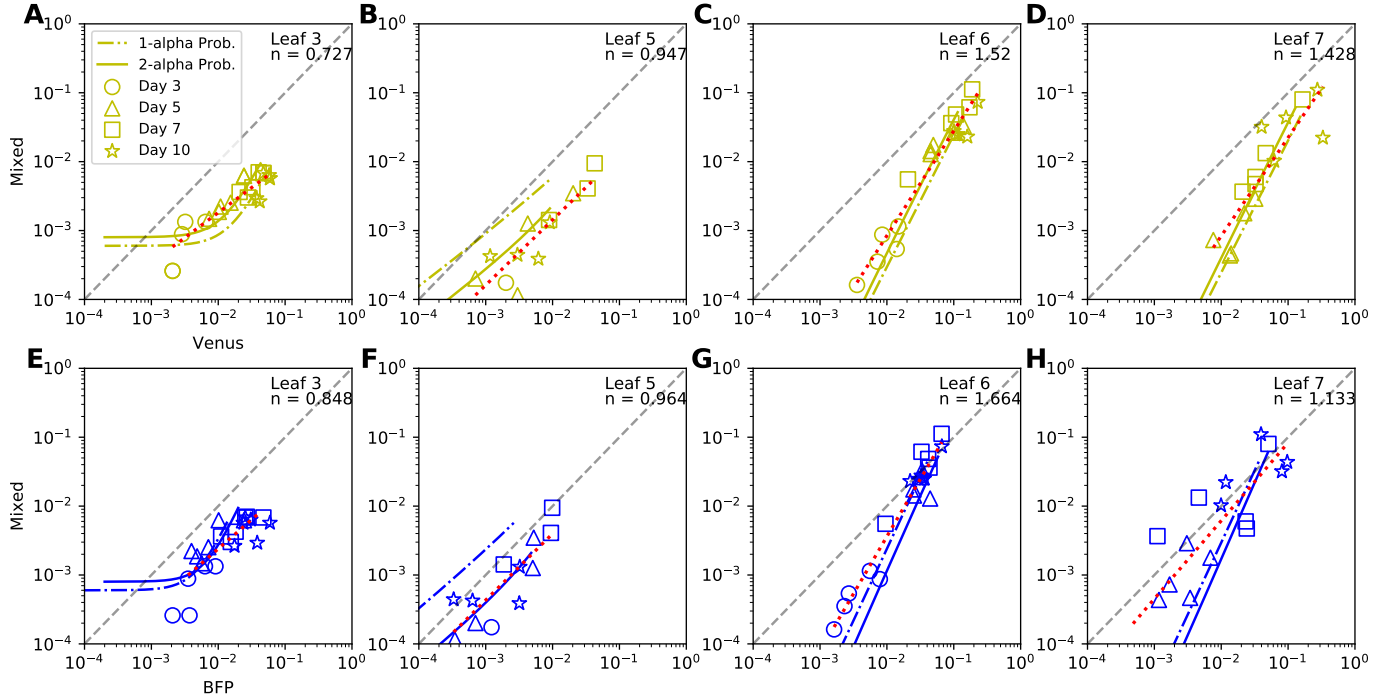
Figure 7: The two-alpha probabilistic model accurately describes the relationship between frequency of coinfected cells and those infected with a single virus for most data. We plot the relationship between frequency of cells coinfected with Venus and BFP strains of TEV and the frequency of cells infected with a single strain (A-D for Venus and E-H for BFP) for various leaves of the plant (different panels) and different days since infection (shown by markers). Some of the data are not shown on a log-log plot due to zeros of the number of infected or coinfected cells. Solid lines are the predictions of 2-alpha probabilistic model (eqn. (25)) and dash-dotted lines are the predictions of 1-alpha probabilistic model (eqn. (24)). The red dotted lines are power functions fitted to the data. The exponents of these functions are shown in the top right corners below the leaf number. Asymptotic relationship between the frequency of coinfected and single infected cells appears to become a straight line for all three models, a feature which we examine in eqns. (S.41)–(S.48). We also provide a python-based script in which these data and model predictions can be explored further (**Supplemental Figure S5**).

why this occurs, and what the slopes of these lines are, we performed additional analyses (shown in Supplemental Information).

Discussion

In this paper we performed extensive analyses of the recently published data on the kinetics of infection of tobacco plants with two variants of TEV [14]. We found that the pathway of virus dissemination in the plant could not be robustly determined directly from the data on the change in frequency of infected cells in different leaves over time — several alternative models that assumed slightly different pathways of dissemination fitted the data with very similar quality. The model assuming that viral dissemination starts from the upper leaf 7, however, fitted the data poorer than the models assuming that dissemination starts with lower (3rd) leaf suggesting that these data do contain some information on the direction of virus dissemination.

The best performing model in our analysis was dependent on the method of how the models were fitted to data; fitting the models using binomial distribution-based likelihood (eqn. (29)) suggested that alternative model 2 (eqn. (6)) was the best (**Supplemental Table S1**). On the other hand, when total infection models were fitted using the least squares method based on log-transformed data and model predictions (eqn. (35)), Tromas *et al.* [14] model and alternative models 1&2 (eqn. (4) and eqn. (6)) fitted the data with the best quality (**Table 2**). The way experimental measurement errors influence the data remains poorly understood, and therefore, which statistical model – log-transformed least squares or binomial distribution-based likelihood – are more appropriate in fitting the models to such data remains undefined. The way forward is to understand better sources of errors in experiments measuring the fraction of infected protoplasts by flow cytometry.

It is generally unknown why not all cells in the leaves were infected 10 days post infection; for example, leaf 3 had less than 10% of its cells infected by the end of experiment (**Figure 3**). One possibility is that not enough time has passed for all cells to be infected. Tracking virus infection at longer than 5.5 week periods may be complicated because at this time plant physiology changes dramatically due to development of flowers. Tromas *et al.* [14] assumed that infection stops after the fraction of infected cells in a leaf exceeds some critical value ψ_k, but how the physiological aspects of the plant, or the virus infecting it, determine the value of ψ_k remain a mystery. We showed that an alternative model in which infection of a given leaf slows down due to a time-dependent factor and not directly due to increase in the fraction of infected cells, can describe the data with similar quality (**Supplemental Table S2**). Such time-dependent factors may be an immune response such as RNAi generation and dissemination via plasmodesmata that may render cells in the leaf resistant to infection. Another factor could be changes to plasmodesmata themselves, like the accumulation of callose at the pores, that prevent the local cell-to-cell movement of the virus [47]. Physiological changes in the leaves in a growing plant may also contribute to the increased resistance of some plant cells to infection.

Our main findings, however, are about coinfection of cells with two different variants of TEV. Interestingly, by using odd ratio metrics [33] we found significantly higher frequency of coinfections of leaf cells by two viruses, in some cases with $OR = 100$ or more that is much higher than that observed in other systems [33] and is in contrast with another study finding suppression of coinfections [45]. Importantly, we developed a series of novel mathematical models that track the coinfection dynamics; the best fit model also predicted higher rates of coinfections of plant cells with two viruses for the within-leaf virus spread but not for virus dissemination to other leaves (**Figure 6**).

25

Additional analysis showed that at least for the upper leaves (leaf 6&7) the frequency of coinfected cells increases more rapidly than linear with frequency of single-infected cells (**Figure 7**), and we show analytically that this is not expected in the random infection model. It has been proposed that deviation of coinfection frequency from random is likely to result from heterogeneity in target cell susceptibility to infection [33]. However, given the mechanics of virus spread in plants via plasmodesmata, ability of multiple viruses to enter the same cell, and thus increase chances of coinfection, remains a possibility (although the model assuming this mechanism did not fit the data with best quality, (**Table 3**))[48, 49]. Given that virus coinfection of leaf cells in other systems can be high and that virus coinfections may result in higher virus production by infected cells [50, 51], impact of coinfections on virus evolution has received considerable attention [52–54].

As far as we are aware, Tromas *et al.* [14] performed the first comprehensive analyses of virus dissemination in plants, and so far, no similar works (experiments and modeling) on virus dissemination within and between multiple tissues have not been performed in animals. However, several studies have investigated how, for example, hepatitis C virus (HCV) spreads locally in the liver [55–57]. There is also evidence for local spread of influence A virus in humans and animals (reviewed in [58]), and mathematical models that take into account physiology of the lung tissue to study virus spread have been proposed [59]. Our observation of potential cooperativity between viruses infecting individual cells extends the results found with animal viruses such as HIV or vaccinia virus [33, 46, 60, 61]. Our analyses thus illustrate that additional insights can be generated by experiments in which infection accumulation (and loss) are tracked over time systematically in the whole organism; using barcoded viruses may be particularly useful in this regard [62].

Our study has several limitations. In our analysis we ignored the complexity of the growing, 4 week old tobacco plants, and changes that occur with leaves in the growing plant. Plants do not have pumping systems like animals, and therefore systemic movement of viruses must follow the already established pathways provided by the phloem, typically, from source to sink tissues such as leaves. However, it is not always obvious based on visual appearance when a given leaf changes from being a sink to being a source (or vice versa). Viruses can manipulate source-sink relationships in their hosts; e.g., some viruses can convert source tissues into sinks [27, 63]. While we had information on the fraction of infected cells in different leaves, spatial aspects of the infection process were lost during protoplast extraction. Better understanding of virus dissemination kinetics is likely to benefit when such spatial details are also recorded, along with the high throughput flow cytometry-based measurements.

While we provided evidence that coinfection occurs at higher frequency than predicted by the random infection hypothesis, we were unable to provide a solid explanation for this effect. Variability in susceptibility of cells to viral infection, local, cell-to-cell virus transmission via plasmodesmata, or cooperation between viral variants may be contributing.

We showed that inference of the best fit model depends on the method used to fit models to data. Given limited understanding of the sources of errors in these data, the most appropriate statistical models that take into account measurement errors will need to be developed. In particular, by fitting the models to data using binomial distribution-based likelihood we found large differences in quality of how alternative models fitted the data (based on AIC values). We hypothesize that binomial distribution-based likelihood amplifies small differences in the infection frequency of individual leaves at early time points, leading to significant favoring of one model over the other. However, this method does not truly account for experimental noise in extraction efficiency of protoplasts from the leaves and false positives when detecting fluorescence signals from individual cells by flow cytometry. Therefore, we believe that the finding that there is one best fit model among the alternative models

when models are fitted using binomial distribution-based likelihood is insufficient to choose a specific model. Additional experiments that better address experimental errors in measuring the fraction of infected cells in different leaves will be needed to derive a better statistical model to fit our dynamical models to such data.

Similarly to Tromas *et al.* [14] we ignored the fact that infection occurs in a plant, and pooled all infection-per-leaf data together without tracking infection per plant. It is clear, however, that some plants may have more infection in all leaves than others (e.g., **Supplemental Figure S1**) and fitting the models to such "paired" data may provide additional insights into details of viral spread locally and systemically. Finally, we showed that the pathway of virus dissemination in plants cannot be easily determined from experiments that measured virus accumulation in different leaves over time, although this result was dependent on the way the models were fitted to data.

Biases introduced by extraction of protoplasts for use with flow cytometry, remain unclear. For example, infected cells may preferentially die during the extraction process which would reduce the fraction of infected cells measured. In immunology, one potential way to understand such biases has been by comparing the flow cytometry-based measurements with microscopy-based measurements [64, 65].

Conclusions and future directions

Our study opens avenues for future research. In particular, similar analyses may need to be performed for other plant viruses. TEV is a potyvirus, one of the largest classes of viruses in plants, [66], and together with the geminiviruses, they are responsible for the majority of disease in commercial agriculture. Understanding how these viruses disseminate in their hosts may bring practical benefits through improved interventions (e.g., [67]). To understand better details of local virus dissemination it will be necessary to combine measurements of spatial virus spread in individual leaves with flow cytometry-based measurements of the fraction of infected cells. Local virus spread can be measured by confocal microscopy and larger spread by light microscopy [26, 45, 68, 69]; previous studies have developed frameworks of how such local viral spread may be modeled [57]. Future studies should better understand why infection of a given leaf stops when not all cells are infected. Whether this is related to changes in leaf physiology (moving from sink to source) or immune responses in the leaf or systemically needs to be tested in experiments and modeled appropriately. Whether measurement of infection in leaves is sufficient to accurately predict virus dissemination kinetics is unclear. For example, roots are typical sink tissues in plants [70]. Thus, it is likely that virus accumulation in the roots precedes or coincides with systemic virus dissemination to upper above-ground structures. Future experiments and modeling studies may benefit to include the dynamics of virus-infected cells in the plant roots. Finally, more precise understanding of the pathway of virus dissemination will benefit from additional data in which infection is initiated in different leaves and experiments in which some leaves are removed after a specific time period (e.g., [28]). Such experiments are not without caveats because removing a leaf may induce systemic changes in the plant that in turn may influence virus dissemination kinetics. Therefore, such experiments would be helped by mathematical models that can make quantitative predictions on the impact of different leaf removal on the virus dissemination kinetics, and these models can be tested and some falsified [37]. Ultimately, a combination of well designed experiments to test specific hypotheses and quantitative mathematical models is likely to bring novel insights into how viruses disseminate in their plant hosts. Such knowledge will be critical for development of novel strategies for limiting

Data sources

The data for our analyses have been provided by S. Elena [14]. Formatted data are available with this publication as a supplement (csv file) and via Github (https://github.com/Plant-Virus-Spread/Models-And-Tools/tree/7553fd98261d1c4b4e75bd3f0cfac4fb49067174).

Codes

We performed most of our analyses in python and provide several key codes to ensure reproducibility of our results on github (https://github.com/Plant-Virus-Spread/Models-And-Tools/tree/7553fd98261d1c4b4e75bd3f0cfac4f Specifically, we provide the codes to 1) plot the predictions of Tromas *et al.* [14] model with their parameter estimates, 2) fitting [14] model to the data using either binomial distribution-based likelihood or least squares, 3) fitting the 2-alpha probabilistic model to the coinfection data using binomial distribution-based likelihood, and 4) sliders code allowing to explore the relationship between coinfected and singly-infected cells (in the data and predictions of 2 alpha probabilistic model).

Author's contributions

VVG had the initial idea of the study and obtained data from the published study. VVG and JM developed the mathematical models. JM performed all the data analyses, mathematical model development and fitting to data. VVG verified some of the results in R. TBS contributed biological insights into viral spread in plants and critical overview of the modeling results. JM wrote the first draft of the paper which was sequentially edited by VVG and JM. All authors contributed to writing the final version of the paper.

Acknowledgements

References

1. Lutz, W., Sanderson, W. & Scherbov, S. 2001 The end of world population growth. *Nature*, **412**, 543–545. doi:10.1038/35087589.

2. Hoornweg, D. & Pope, K. 2017 Population predictions for the world's largest cities in the 21st century. *Environment and Urbanization*, **29**(1), 195–216. doi:10.1177/0956247816663557.

3. United Nations 2019 Population facts.
   https://www.un.org/en/development/desa/population/publications/pdf/popfacts/PopFacts_-
   2019-6.pdf.

4. Walker, R. J. 2016 Population growth and its implications for global security. *The American journal of economics and sociology*, **4**, 980–1004.

5. Hull, R. 2013 *Plant Virology*. Elsevier Science & Technology, 5th edn.

6. Jones, R. A. C. 2021 Global plant virus disease pandemics and epidemics. *Plants*, **10**. doi: 10.3390/plants10020233.

7. Youden, W. J. 1935 Statistical aspect of the production of primary lesions by plant viruses. *Nature*, **135**, 1075–1075. doi:10.1038/1351075b0.

8. Bald, J. G. 1935 Statistical aspect of the production of primary lesions by plant viruses. *Nature*, **135**, 996–996. doi:10.1038/135996a0.

9. Furumoto, W. A. & Mickey, R. 1967 A mathematical model for the infectivity-dilution curve of tobacco mosaic virus: experimental tests. *Virology*, **32**(2), 224–233.

10. Furumoto, W. A. & Mickey, R. 1967 A mathematical model for the infectivity-dilution curve of tobacco mosaic virus: theoretical considerations. *Virology*, **32**, 216–223. doi:10.1016/0042-6822(67)90271-1.

11. Martínez, F., Sardanyés, J., Elena, S. F. & Daròs, J.-A. 2011 Dynamics of a plant rna virus intracellular accumulation: stamping machine vs. geometric replication. *Genetics*, **188**, 637–646. doi:10.1534/genetics.111.129114.

12. Lafforgue, G., Sardanyés, J. & Elena, S. F. 2011 Differences in accumulation and virulence determine the outcome of competition during tobacco etch virus coinfection. *PloS one*, **6**, e17 917. doi:10.1371/journal.pone.0017917.

13. Miljkovic, D., Depolli, M., Stare, T., Mozetič, I., Petek, M., Gruden, K. & Lavrač, N. 2014 Plant defence model revisions through iterative minimisation of constraint violations. *International journal of computational biology and drug design*, **7**, 61–79. doi:10.1504/IJCBDD.2014.058588.

14. Tromas, N., Zwart, M. P., Lafforgue, G. & Elena, S. F. 2014 Within-host spatiotemporal dynamics of plant virus infection at the cellular level. *PLoS genetics*, **10**, e1004 186. doi: 10.1371/journal.pgen.1004186.

15. Zhang, X. S., Holt, J. & Colvin, J. 2000 Mathematical models of host plant infection by helper-dependent virus complexes: why are helper viruses always avirulent? *Phytopathology*, **90**, 85–93. doi:10.1094/PHYTO.2000.90.1.85.

16. Amaku, M., Burattini, M. N., Coutinho, F. A. B. & Massad, E. 2010 Modeling the competition between viruses in a complex plant-pathogen system. *Phytopathology*, **100**, 1042–1047. doi: 10.1094/PHYTO-10-09-0289.

17. Moore, S. M., Manore, C. A., Bokil, V. A., Borer, E. T. & Hosseini, P. R. 2011 Spatiotemporal model of barley and cereal yellow dwarf virus transmission dynamics with seasonality and plant competition. *Bulletin of mathematical biology*, **73**, 2707–2730. doi:10.1007/s11538-011-9654-4.

18. Jeger, M. J., Chen, Z., Powell, G., Hodge, S. & van den Bosch, F. 2011 Interactions in a host plant-virus-vector-parasitoid system: modelling the consequences for virus transmission and disease dynamics. *Virus research*, **159**, 183–193. doi:10.1016/j.virusres.2011.04.027.

19. Neofytou, G., Kyrychko, Y. N. & Blyuss, K. B. 2016 Mathematical model of plant-virus interactions mediated by rna interference. *Journal of theoretical biology*, **403**, 129–142. doi:10.1016/j.jtbi.2016.05.018.

20. Hamelin, F. M., Allen, L. J. S., Prendeville, H. R., Hajimorad, M. R. & Jeger, M. J. 2016 The evolution of plant virus transmission pathways. *Journal of theoretical biology*, **396**, 75–89. doi:10.1016/j.jtbi.2016.02.017.

21. Jeger, M. J., Madden, L. V. & van den Bosch, F. 2018 Plant virus epidemiology: Applications and prospects for mathematical modeling and analysis to improve understanding and disease control. *Plant disease*, **102**, 837–854. doi:10.1094/PDIS-04-17-0612-FE.

22. Arias, J. H., Gómez-Gardeñes, J., Meloni, S. & Estrada, E. 2018 Epidemics on plants: Modeling long-range dispersal on spatially embedded networks. *Journal of theoretical biology*, **453**, 1–13. doi:10.1016/j.jtbi.2018.05.004.

23. Allen, L. J. S., Bokil, V. A., Cunniffe, N. J., Hamelin, F. M., Hilker, F. M. & Jeger, M. J. 2019 Modelling vector transmission and epidemiology of co-infecting plant viruses. *Viruses*, **11**. doi:10.3390/v11121153.

24. Kendig, A. E., Borer, E. T., Boak, E. N., Picard, T. C. & Seabloom, E. W. 2020 Host nutrition mediates interactions between plant viruses, altering transmission and predicted disease spread. *Ecology*, **101**, e03 155. doi:10.1002/ecy.3155.

25. Al Basir, F., Kyrychko, Y. N., Blyuss, K. B. & Ray, S. 2021 Effects of vector maturation time on the dynamics of cassava mosaic disease. *Bulletin of mathematical biology*, **83**, 87. doi:10.1007/s11538-021-00921-4.

26. Reagan, B. C. & Burch-Smith, T. M. 2020 Viruses reveal the secrets of plasmodesmal cell biology. *Molecular plant-microbe interactions*, **33**, 26–39. doi:10.1094/MPMI-07-19-0212-FI.

27. Lemoine, R., La Camera, S., Atanassova, R., Dédaldéchamp, F., Allario, T., Pourtau, N., Bonnemain, J.-L., Laloi, M., Coutos-Thévenot, P. *et al.* 2013 Source-to-sink transport of sugar and regulation by environmental factors. *Frontiers in plant science*, **4**, 272. doi:10.3389/fpls.2013.00272.

28. Roberts, A. G., Cruz, S. S., Roberts, I. M., Prior, D. A. M., Turgeon, R. & Oparka, K. J. 1997 Phloem unloading in sink leaves of nicotiana benthamiana: Comparison of a fluorescent solute with a fluorescent virus. *The Plant cell*, **9**, 1381–1396. doi:10.1105/tpc.9.8.1381.

29. Wang, Y. & Hajimorad, M. R. 2016 Gain of virulence by soybean mosaic virus on rsv4-genotype soybeans is associated with a relative fitness loss in a susceptible host. *Molecular plant pathology*, **17**, 1154–1159. doi:10.1111/mpp.12354.

30. Burnham, K. P. & Anderson, D. R. 2002 *Model selection and multimodel inference: a practical information-theoretic approach.* Springer-Verlag, New York.

31. Creager, A. N. 2002 *The life of a virus : tobacco mosaic virus as an experimental model, 1930-1965*. University of Chicago Press.

32. Gupta, N., Reddy, K., Bhattacharyya, D. & Chakraborty, S. 2021 Plant responses to geminivirus infection: guardians of the plant immunity. *Virology journal*, **18**, 143. doi:10.1186/s12985-021-01612-1.

33. Dang, Q., Chen, J., Unutmaz, D., Coffin, J. M., Pathak, V. K., Powell, D., KewalRamani, V. N., Maldarelli, F. & Hu, W.-S. 2004 Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways. *PNAS*, **101**, 632–637. doi:10.1073/pnas.0307636100.

34. Efron, B. & Tibshirani, R. 1993 *An introduction to the bootstrap*. New York: Chapman & Hall.

35. Wang, A. 2021 Cell-to-cell movement of plant viruses via plasmodesmata: a current perspective on potyviruses. *Current opinion in virology*, **48**, 10–16. doi:10.1016/j.coviro.2021.03.002.

36. Wu, F.-H., Shen, S.-C., Lee, L.-Y., Lee, S.-H., Chan, M.-T. & Lin, C.-S. 2009 Tape-Arabidopsis sandwich - a simpler Arabidopsis protoplast isolation method. *Plant methods*, **5**, 16. doi:10.1186/1746-4811-5-16.

37. Ganusov, V. V. 2016 Strong Inference in Mathematical Modeling: A Method for Rob ust Science in the Twenty-First Century. *Front Microbiol*, **7**, 1131.

38. Coffin, J. M. 1995 HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science*, **267**, 483–489. doi:10.1126/science.7824947.

39. Letvin, N. L. & Walker, B. D. 2003 Immunopathogenesis and immunotherapy in AIDS virus infections. *Nature medicine*, **9**, 861–866. doi:10.1038/nm0703-861.

40. Song, H., Giorgi, E. E., Ganusov, V. V., Cai, F., Athreya, G., Yoon, H., Carja, O., Hora, B., Hraber, P. *et al.* 2018 Tracking HIV-1 recombination to resolve its contribution to HIV-1 evolution in natural infection. *Nature communications*, **9**, 1928. doi:10.1038/s41467-018-04217-5.

41. Haqqani, A. A., Marek, S. L., Kumar, J., Davenport, M., Wang, H. & Tilton, J. C. 2015 Central memory CD4+ T cells are preferential targets of double infection by HIV-1. *Virology journal*, **12**, 184. doi:10.1186/s12985-015-0415-0.

42. Wang, M.-B. & Metzlaff, M. 2005 Rna silencing and antiviral defense in plants. *Current opinion in plant biology*, **8**, 216–222. doi:10.1016/j.pbi.2005.01.006.

43. Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A. C., Smith, T. H. & Vance, V. B. 1998 A viral suppressor of gene silencing in plants. *PNAS*, **95**, 13 079–13 084. doi:10.1073/pnas.95.22.13079.

44. Csorba, T., Kontra, L. & Burgyán, J. 2015 viral silencing suppressors: Tools forged to fine-tune host-pathogen coexistence. *Virology*, **479-480**, 85–103. doi:10.1016/j.virol.2015.02.028.

45. Gutiérrez, S., Pirolles, E., Yvon, M., Baecker, V., Michalakis, Y. & Blanc, S. 2015 The multiplicity of cellular infection changes depending on the route of cell infection in a plant virus. *Journal of virology*, **89**, 9665–9675. doi:10.1128/JVI.00537-15.

46. Dixit, N. & Perelson, A. 2005 HIV dynamics with multiple infections of target cells. *Proc Natl Acad Sci USA*, **102**(23), 8198–203.

47. Zavaliev, R., Ueki, S., Epel, B. L. & Citovsky, V. 2011 Biology of callose (beta-1,3-glucan) turnover at plasmodesmata. *Protoplasma*, **248**, 117–130. doi:10.1007/s00709-010-0247-0.

48. Latham, J. R. & Wilson, A. K. 2008 Transcomplementation and synergism in plants: implications for viral transgenes? *Molecular plant pathology*, **9**, 85–103. doi:10.1111/j.1364-3703.2007.00441.x.

49. Fondong, V. N. 2017 The search for resistance to cassava mosaic geminiviruses: How much we have accomplished, and what lies ahead. *Frontiers in plant science*, **8**, 408. doi:10.3389/fpls.2017.00408.

50. Gutiérrez, S., Yvon, M., Thébaud, G., Monsion, B., Michalakis, Y. & Blanc, S. 2010 Dynamics of the multiplicity of cellular infection in a plant virus. *PLoS pathogens*, **6**, e1001 113. doi:10.1371/journal.ppat.1001113.

51. Andreu-Moreno, I., Bou, J.-V. & Sanjuán, R. 2020 Cooperative nature of viral replication. *Science advances*, **6**. doi:10.1126/sciadv.abd4942.

52. Díaz-Munoz, S. L., Sanjuán, R. & West, S. 2017 Sociovirology: Conflict, cooperation, and communication among viruses. *Cell host & microbe*, **22**, 437–441. doi:10.1016/j.chom.2017.09.012.

53. Domingo, E. & Perales, C. 2019 Viral quasispecies. *PLoS genetics*, **15**, e1008 271. doi:10.1371/journal.pgen.1008271.

54. Sanjuán, R. 2021 The social life of viruses. *Annual review of virology*, **8**, 183–199. doi:10.1146/annurev-virology-091919-071712.

55. Graw, F., Balagopal, A., Kandathil, A. J., Ra y, S. C., Thomas, D. L., Ribeiro, R. M. & Perelson, A. S. 2014 Inferring viral dynamics in chronically HCV infected patients from the spatial distribution of infected hepatocytes. *PLoS computational biology*, **10**, e1003 934. doi:10.1371/journal.pcbi.1003934.

56. Graw, F. & Perelson, A. S. 2016 Modeling viral spread. *Annual review of virology*, **3**, 555–572. doi:10.1146/annurev-virology-110615-042249.

57. Durso-Cain, K., Kumberger, P., Schälte, Y., Fink, T., Dahari, H., Hasenauer, J., Uprichard, S. L. & Graw, F. 2021 HCV spread kinetics reveal varying contributions of transmission modes to infection dynamics. *Viruses*, **13**(1308), 1–22.

58. Gallagher, M. E., Brooke, C. B., Ke, R. & Koelle, K. 2018 Causes and consequences of spatial within-host viral spread. *Viruses*, **10**. doi:10.3390/v10110627.

59. Quirouette, C., Younis, N. P., Reddy, M. B. & Beauchemin, C. A. A. 2020 A mathematical model describing the localization and spread of influenza a virus infection within the human respiratory tract. *PLoS computational biology*, **16**, e1007 705. doi:10.1371/journal.pcbi.1007705.

60. Stiefel, P., Schmidt, F. I., Dörig, P., Behr, P., Zambelli, T., Vorholt, J. A. & Mercer, J. 2012 Cooperative vaccinia infection demonstrated at the single-cell level using fluidfm. *Nano letters*, **12**, 4219–4227. doi:10.1021/nl3018109.

61. Ito, Y., Tauzin, A., Remion, A., Ejima, K., Mammano, F. & Iwami, S. 2018 Dynamics of HIV-1 coinfection in different susceptible target cell populations during cell-free infection. *Journal of theoretical biology*, **455**, 39–46. doi:10.1016/j.jtbi.2018.06.025.

62. Deleage, C., Immonen, T. T., Fennessey, C. M., Reynaldi, A., Reid, C., Newman, L., Lipkey, L., Schlub, T. E., Camus, C. *et al.* 2019 Defining early SIV replication and dissemination dynamics following vaginal transmission. *Science advances*, **5**, eaav7116. doi:10.1126/sciadv.aav7116.

63. Kappagantu, M., Collum, T. D., Dardick, C. & Culver, J. N. 2020 Viral hacks of the plant vasculature: The role of phloem alterations in systemic virus infection. *Annual review of virology*, **7**, 351–370. doi:10.1146/annurev-virology-010320-072410.

64. Steinert, E. M., Schenkel, J. M., Fraser, K. A., Beura, L. K., Manlove, L. S., Igyrt, B. Z., Southern, P. J. & Masopust, D. 2015 Quantifying Memory CD8 T Cells Reveals Regionalization of Immunosurveillance. *Cell*, **161**(4), 737–749. doi:10.1016/j.cell.2015.03.031.

65. Gern, B. H., Adams, K. N., Plumlee, C. R., Stoltzfus, C. R., Shehata, L., Moguche, A. O., Busman-Sahay, K., Hansen, S. G., Axthelm, M. K. *et al.* 2021 TGFb restricts expansion, survival, and function of T cells within the tuberculous granuloma. *Cell host & microbe*, **29**, 594–606.e6. doi:10.1016/j.chom.2021.02.005.

66. Yang, X., Li, Y. & Wang, A. 2021 Research advances in Potyviruses: From the laboratory bench to the field. *Annual review of phytopathology*, **59**, 1–29. doi:10.1146/annurev-phyto-020620-114550.

67. Wisler, G. C. & Duffus, J. E. 2000 A century of plant virus management in the Salinas valley of California, 'East of Eden'. *Virus research*, **71**, 161–169. doi:10.1016/s0168-1702(00)00196-9.

68. Khatabi, B., Fajolu, O. L., Wen, R.-H. & Hajimorad, M. R. 2012 Evaluation of North American isolates of Soybean mosaic virus for gain of virulence on Rsv-genotype soybeans with special emphasis on resistance-breaking determinants on Rsv4. *Molecular plant pathology*, **13**, 1077–1088. doi:10.1111/j.1364-3703.2012.00817.x.

69. Ganusova, E. E. & Burch-Smith, T. M. 2019 Review: Plant-pathogen interactions through the plasmodesma prism. *Plant science : an international journal of experimental plant biology*, **279**, 70–80. doi:10.1016/j.plantsci.2018.05.017.

70. Lemoine, R., La Camera, S., Atanassova, R., Dédaldéchamp, F., Allario, T., Pourtau, N., Bonnemain, J.-L., Laloi, M., Coutos-Thévenot, P. *et al.* 2013 Source-to-sink transport of sugar and regulation by environmental factors. *Frontiers in plant science*, **4**, 272. doi:10.3389/fpls.2013.00272.
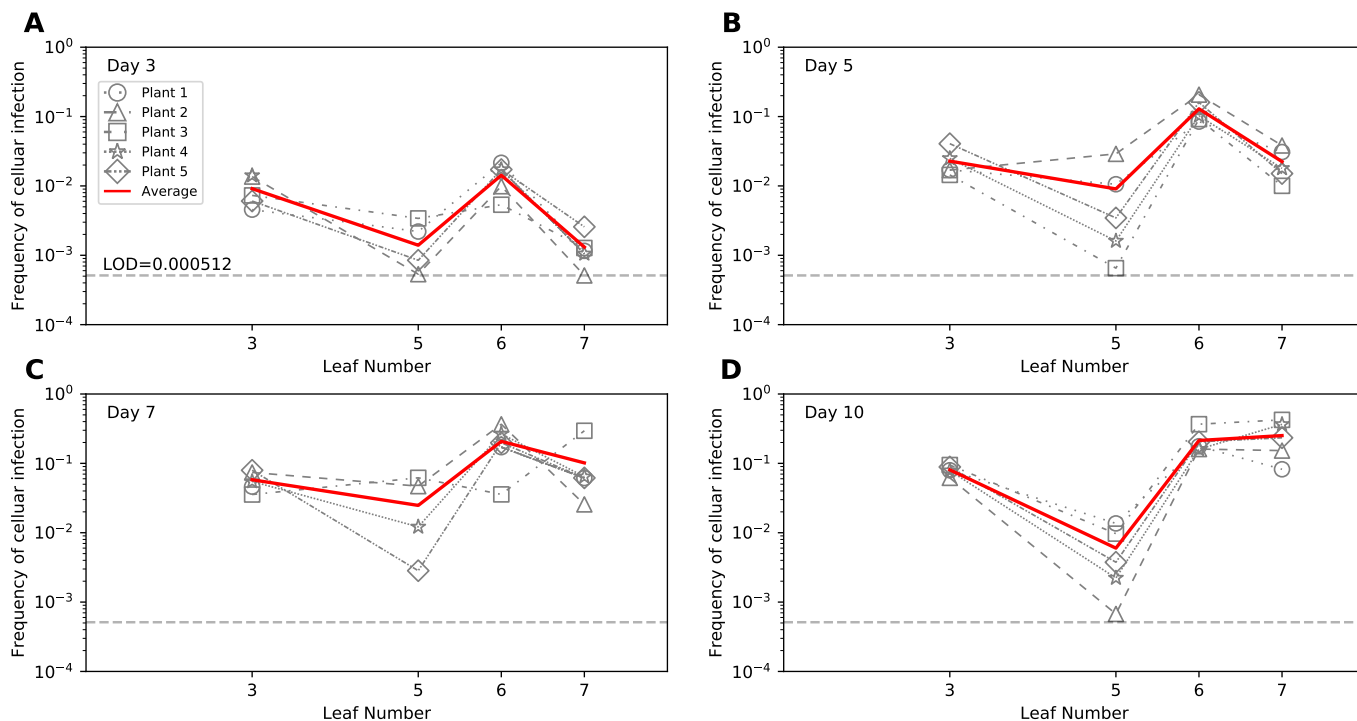
Supplemental Information

Mathematical modeling suggests cooperation of plant-infecting viruses

Joshua Miller, Tessa Burch-Smith, and Vitaly V. Ganusov

Additional figures and tables

Supplemental Figure S1: Kinetics of TEV dissemination in tobacco plants partitioned per individual plant show variability in leaf infection levels. We plot the data on infection of cells with either or both variants of TEV for individual leaves of a given plant for 3 (A), 5 (B), 7 (C), or 10 (D) days since infection. Symbols denote the frequency of infected cells in a leaf with lines connecting measurements in individual plants. Solid red line denotes average infection per leaf for a given time point.

| Parameter	Original	Alt. Model 1	Alt. Model 2	Alt. Model 3	Alt. Model 4	Alt. Model 5	Alt. Model 6	Alt. Model 7
Fitted with binomial distribution-based likelihood method								
I_0	.0003	.0005	.0004	.00006	.0001	.0006	.0001	.0001
β, 1/day	.950	.837	.887	1.289	1.040	1.000	1.037	1.032
χ_3, 1/day	N/A	N/A	N/A	N/A	.391	.042	.385	.064
χ_5, 1/day	.167	.120	.135	.775	N/A	1.876	.239	.047
χ_6, 1/day	1.046	5.489	4.964	8.101	2.275	N/A	2.269	2.251
χ_7, 1/day	.029	.059	.465	.972	.193	.022	N/A	N/A
ψ_3	.080	.083	.081	.075	.074	.075	.072	.072
ψ_5	.016	.017	.016	.016	.016	.011	.016	.016
ψ_6	.224	.223	.223	.217	.225	.235	.219	.219
ψ_7	.269	.276	.276	.590	.265	.268	.269	.269
nll	378317	378212	378172	378343	378495	379377	378442	378524
AIC_{Lik}	756652	756442	756362	756704	757008	758772	756902	757066
Δ_{AIC}	290	80	0	342	646	2410	540	704

Supplemental Table S1: Alternative models for viral dissemination fitted to data using binomial distribution-based likelihood describe the data with different quality based on AIC values. We performed the same analysis as **Table 2** except that models were fitted to data using binomial distribution-based likelihood.

| Param	I_0	β, 1/day	χ_5, 1/day	χ_6, 1/day	χ_7, 1/day						SSR_{Log}	$AIC_{SSR_{Log}}$	Δ_{AIC}
Fixed n						T_3	T_5	T_6	T_7	n			
	0.0005	0.829	0.061	0.724	0.032	5.762	4.194	5.192	7.654	7.122	52.960	-13	3
Fixed T						n_3	n_5	n_6	n_7	T			
	0.00008	2.518	0.049	0.835	0.055	3.524	6.017	3.659	1.82E+00	3.2	53.073	-13	3
Fixed n, T						T	n						
	0.001	0.673	0.025	0.602	0.068	6.127	9.316				67.555	0	16

Supplemental Table S2: A mathematical model assuming that virus dissemination is influenced by the leaf-specific and systemic immunity can describe the experimental data. We changed the original, Tromas *et al.* [14] model by assuming the time-dependent and leaf-dependent S_k function (Alternative model 9, see eqn. (17)) and fitted the model to the data using least squares with a logarithmic transform eqn. (35). In fits we either varied the time (T_k) at which S_k declines to zero, the Hill coefficient (n_k) which determines the speed at which S_k declines to zero, or both parameters being independent of the leaf number (k). The resulting SSR_{Log} and $AIC_{SSR_{Log}}$ values for different model fits are shown (AICs are rounded to the nearest whole number). We found the following values of the alternative model 8 (eqn. (16)) fits of the data: $SSR_{Log} = 52.285$ and $AIC_{SSR_{Log}} = -10$.
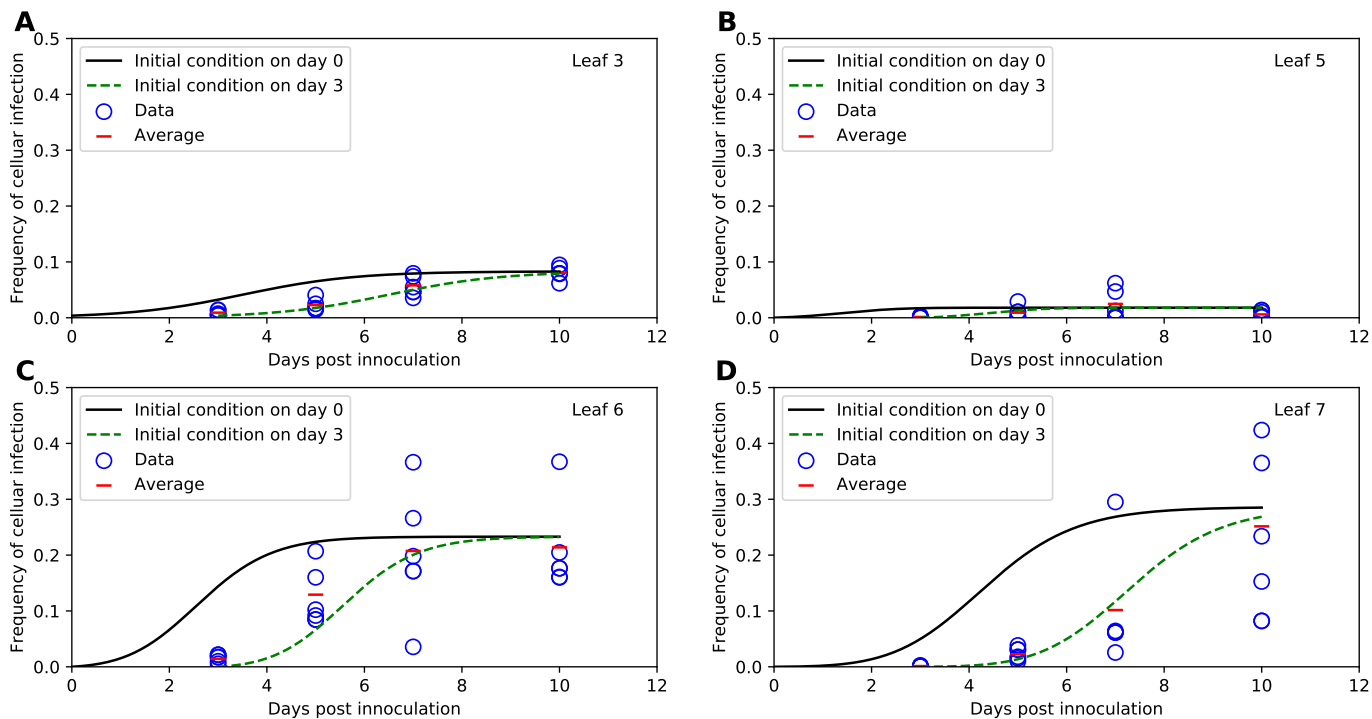
Supplemental Figure S2: Shifting original model predictions (found in Tromas *et al.* [14]) by three days reasonably well matches experimental data. We integrated the original model (given in eqns. (1)–(3)) using an ODE solver in python either assuming that infection starts at day 0 (solid lines) or infection starts at day 3 (dashed lines); data are shown by markers for leaf 3 (A), leaf 5 (B), leaf 6 (C), and leaf 7 (D). By default, ODE solver in python is initialized by the first time point provided in the data which is day 3 in the data set. We overrode the default by forcing the solver to start infection at day 0.
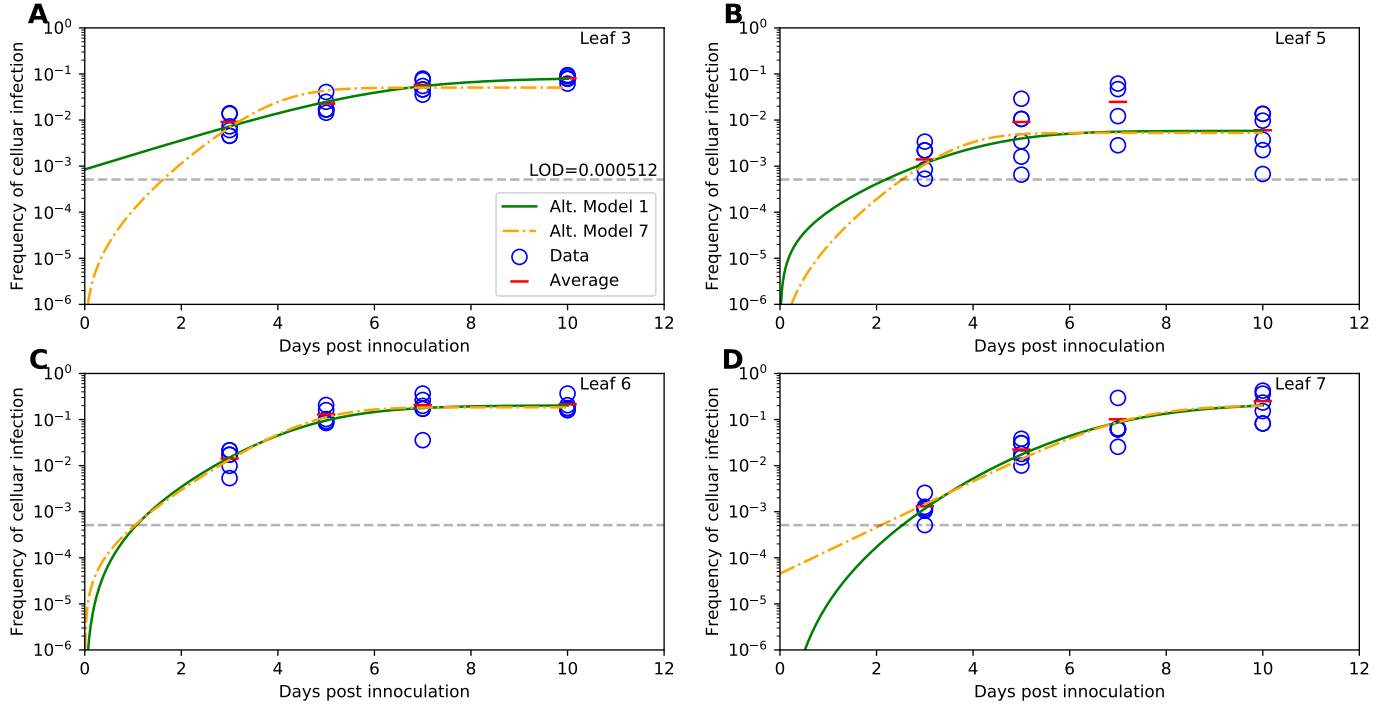
Supplemental Figure S3: The difference between the best performing alternative model 1 (eqn. (4)) and the worst-performing, alternative model 7 (eqn. (15)) is visually small. We fitted the two models using the least-squares method with a log transformation where zeros were replaced with a limit of detection model (LOD) value, in this case $5.12 \cdot 10^{-4}$. The parameters for Alt Model 1 with 95% confidence intervals are: $I_0 = 0.0008\ (0.0002, 0.0017)$, $\beta = 0.744\ (0.554, 1.308)$, $\chi_5 = 0.059\ (0.033, 0.119)$, $\chi_6 = 8.201\ (4.151, 13.206)$, $\chi_7 = 0.073\ (0.052, 0.152)$, $\psi_3 = 0.083\ (0.059, 0.117)$, $\psi_5 = 0.006\ (0.005, 0.013)$, $\psi_6 = 0.204\ (0.186, 0.227)$, $\psi_7 = 0.228\ (0.081, 0.699)$. The parameters for Alt. Model 7 with 95% confidence intervals are: $I_0 = 0.0005\ (2 \cdot 10^{-6}, 1 \cdot 10^{-5})$, $\beta = 1.159\ (0.973, 1.898)$, $\chi_3 = 0.286\ (0.117, 0.645)$, $\chi_5 = 0.050\ (0.023, 20)$, $\chi_6 = 3.263\ (2.365, 5.890)$, $\psi_3 = 0.051\ (0.041, 0.069)$, $\psi_5 = 0.005\ (0.003, 0.008)$, $\psi_6 = 0.184\ (0.143, 0.226)$, $\psi_7 = 0.209\ (0.081, 0.415)$. The goodness of fit metrics for Alt. Model 1 are: $SSR_{Log} = 51.991$ and $AIC_{SSR_{Log}} = -16$, and for comparison the Alt. Model 7's metrics are: $SSR_{Log} = 56.309$ and $AIC_{SSR_{Log}} = -10$, giving $\Delta_{AIC} = 6$.
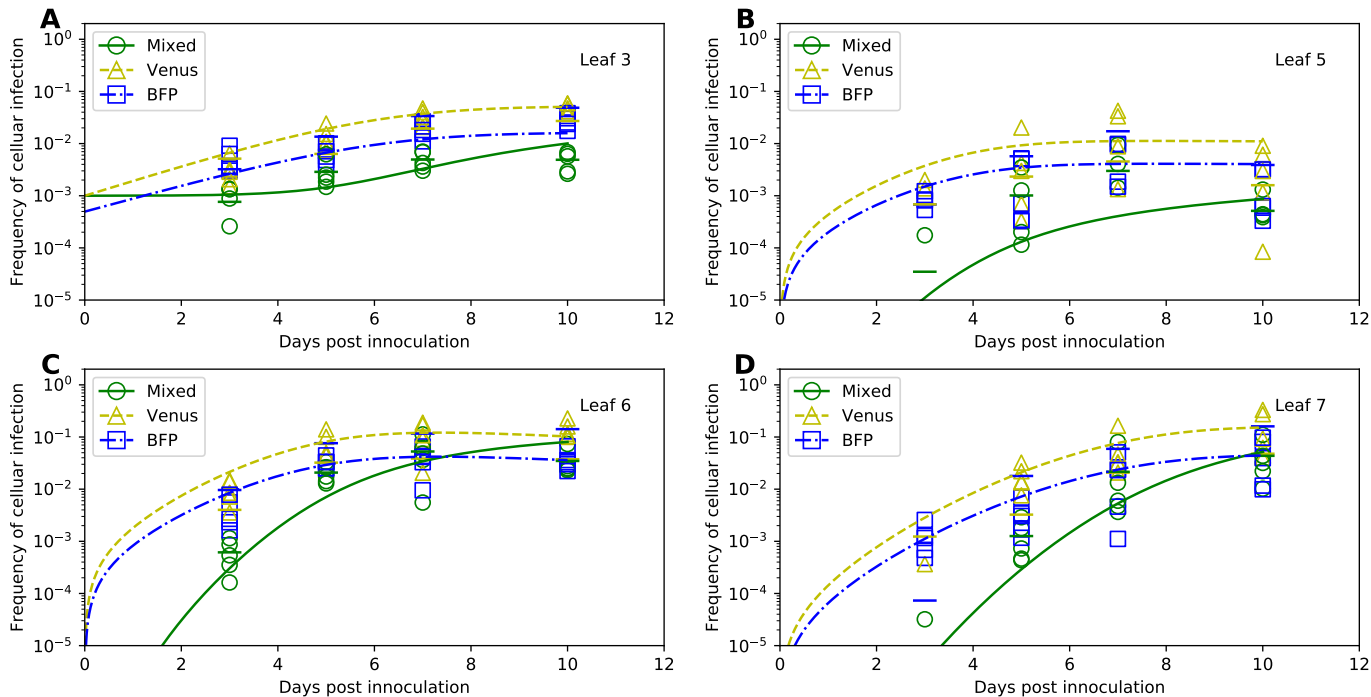
Supplemental Figure S4: The simplest 1-alpha coinfection model does not adequately describe coinfection data. The 1-alpha coinfection model (given by eqns. (18)–(19) and eqn. (21)) assumes that coinfection of individual cells by two different strains occurs independently ($\alpha = 1$) or coinfection may be more ($\alpha > 1$) or less ($0 < \alpha < 1$) likely that infection of an uninfected cell (**Figure 2**A). Other graph details are similar to those given in **Figure 6**. The parameters and 95% confidence intervals for this model are: $V_0 = .0006$ $(.0003, .001)$, $B_0 = .0003$ $(0.3 \times 10^{-5}, .001)$, $M_0 = .0001$ $(.0004, .001)$, $\beta_V = .744$ $(.443, 6.178)/\text{day}$, $\beta_B = .666$ $(.260, 6.580)/\text{day}$, $\chi_5 = .269$ $(.034, 8.76)/\text{day}$, $\chi_6 = .939$ $(.395, 2.292)/\text{day}$, $\chi_7 = .044$ $(.015, 2.107)/\text{day}$, $\psi_3 = .078$ $(.044, .954)$, $\psi_5 = .016$ $(.006, .028)$, $\psi_6 = .201$ $(.156, .267)$, $\psi_7 = .254$ $(.100, .449)$, $\alpha = 2.332$ $(.103, 5.772)$.
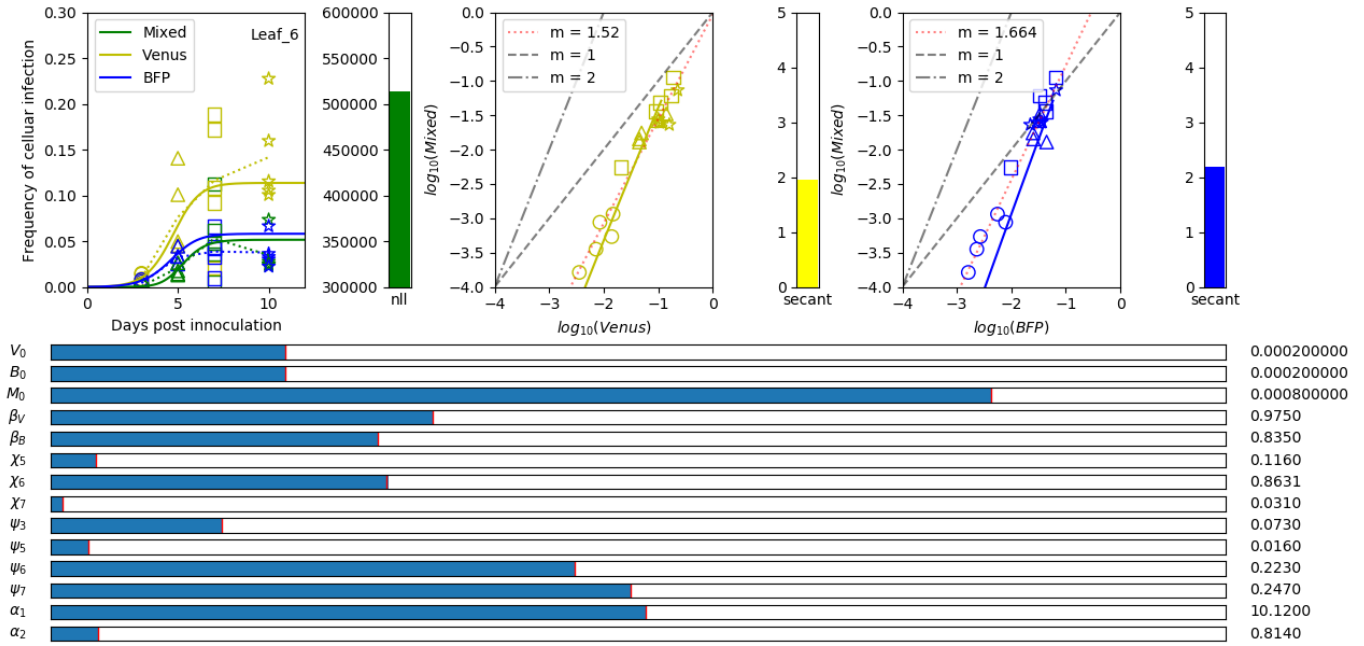
Supplemental Figure S5: A python-based tool to study the impact of parameters of a given mathematical model on the infection rate of a given leaf. By using a function `Slider` from `pylab` in python we visualized the dynamics of cell infection by individual viruses (Venus and BFP) and coinfection of cells by the two viruses according to the Probabilistic Analytic Model. Parameters of the model can be changed using sliders resulting in the changed kinetics of virus infection (shown in the left panel), or changes in the predicted relationships between the degree of coinfection of cells by two viruses (denoted as "Mixed") and singly infected cells (Venus or BFP for middle and right panels, respectively). The example shown is for infection of leaf 6; the code allows to chose any individual leaf for visualization. In all panels data are shown by markers and predictions of the model by lines. Additional parameters shown are i) the negative log-likelihood (nll, see eqns. (30)–(33)); ii) the average ratio of the frequency of coinfected cells to singly-infected cells (secant); iii) the values of the expressions $m_{\frac{M}{V}}$ and $m_{\frac{M}{B}}$ evaluated at the rightmost timepoint in the leftmost panel, in this case t=12 ($m_{\frac{M}{V}}$ and $m_{\frac{M}{B}}$).

Expansion of the Tromas *et al.* [14] model

$$\frac{dI_3}{dt} = \beta I_3 S_3 \tag{S.1}$$

$$\frac{dI_5}{dt} = \beta I_5 S_5 + \chi_5 S_5 I_3 \tag{S.2}$$

$$\frac{dI_6}{dt} = \beta I_6 S_6 + \chi_6 S_6 (I_3 + I_5) \tag{S.3}$$

$$\frac{dI_7}{dt} = \beta I_7 S_7 + \chi_7 S_7 (I_3 + I_5 + I_6) \tag{S.4}$$

Alternative formulations of coinfection models

Expansion of the 1-alpha coinfection model (eqns. (18)–(20) and eqn. (21)):

$$\frac{dV_3}{dt} = \beta_V V_3 S_3 \tag{S.5}$$

$$\frac{dB_3}{dt} = \beta_B B_3 S_3 \tag{S.6}$$

$$\frac{dM_3}{dt} = \alpha(\beta_B V_3 B_3 + \beta_V V_3 B_3) \tag{S.7}$$

$$\frac{dV_5}{dt} = \beta_V V_5 S_5 + \chi_5 S_5 V_3 \tag{S.8}$$

$$\frac{dB_5}{dt} = \beta_B B_5 S_5 + \chi_5 S_5 B_3 \tag{S.9}$$

$$\frac{dM_5}{dt} = \alpha(\beta_B V_5 B_5 + \beta_V V_5 B_5) \tag{S.10}$$

$$\frac{dV_6}{dt} = \beta_V V_6 S_6 + \chi_6 S_6 (V_3 + V_5) \tag{S.11}$$

$$\frac{dB_6}{dt} = \beta_B B_6 S_6 + \chi_6 S_6 (B_3 + V_5) \tag{S.12}$$

$$\frac{dM_6}{dt} = \alpha(\beta_B V_6 B_6 + \beta_V V_6 B_6) \tag{S.13}$$

$$\frac{dV_7}{dt} = \beta_V V_7 S_7 + \chi_7 S_7 (V_3 + V_5 + V_6) \tag{S.14}$$

$$\frac{dB_7}{dt} = \beta_B B_7 S_7 + \chi_7 S_7 (B_3 + B_5 + B_6) \tag{S.15}$$

$$\frac{dM_7}{dt} = \alpha(\beta_B V_7 B_7 + \beta_V V_7 B_7). \tag{S.16}$$

Expansion of the 2-alpha coinfection model (eqns. (18)–(20) and eqn. (22)):

$$\frac{dV_3}{dt} = \beta_V V_3 S_3, \tag{S.17}$$

$$\frac{dB_3}{dt} = \beta_B B_3 S_3, \tag{S.18}$$

$$\frac{dM_3}{dt} = \alpha_B \beta_B V_3 B_3 + \alpha_V \beta_V V_3 B_3, \tag{S.19}$$

$$\frac{dV_5}{dt} = \beta_V V_5 S_5 + \chi_5 S_5 V_3 \tag{S.20}$$

$$\frac{dB_5}{dt} = \beta_B B_5 S_5 + \chi_5 S_5 B_3 \tag{S.21}$$

$$\frac{dM_5}{dt} = \alpha_B \beta_B V_5 B_5 + \alpha_V \beta_V V_5 B_5 \tag{S.22}$$

$$\frac{dV_6}{dt} = \beta_V V_6 S_6 + \chi_6 S_6 (V_3 + V_5) \tag{S.23}$$

$$\frac{dB_6}{dt} = \beta_B B_6 S_6 + \chi_6 S_6 (B_3 + V_5) \tag{S.24}$$

$$\frac{dM_6}{dt} = \alpha_B \beta_B V_6 B_6 + \alpha_V \beta_V V_6 B_6 \tag{S.25}$$

$$\frac{dV_7}{dt} = \beta_V V_7 S_7 + \chi_7 S_7 (V_3 + V_5 + V_6) \tag{S.26}$$

$$\frac{dB_7}{dt} = \beta_B B_7 S_7 + \chi_7 S_7 (B_3 + B_5 + B_6) \tag{S.27}$$

$$\frac{dM_7}{dt} = \alpha_B \beta_B V_7 B_7 + \alpha_V \beta_V V_7 B_7. \tag{S.28}$$

949   Expansion of the 1-alpha probabilistic model (eqns. (18)–(20) and eqn. (24)):

$$\frac{dV_3}{dt} = \beta_V V_3 S_3, \tag{S.29}$$

$$\frac{dB_3}{dt} = \beta_B B_3 S_3 \tag{S.30}$$

$$\frac{dM_3}{dt} = \alpha[S_3 V_3 B_3 (\beta_B + \beta_V)] \tag{S.31}$$

$$\frac{dV_5}{dt} = \beta_V V_5 S_5 + \chi_5 S_5 V_3 \tag{S.32}$$

$$\frac{dB_5}{dt} = \beta_B B_5 S_5 + \chi_5 S_5 B_3 \tag{S.33}$$

$$\frac{dM_5}{dt} = \alpha[S_5 B_5 V_5 (\beta_B + \beta_V) + \chi_5 S_5 (B_5 V_3 + V_5 B_3)] \tag{S.34}$$

$$\frac{dV_6}{dt} = \beta_V V_6 S_6 + \chi_6 S_6 (V_3 + V_5) \tag{S.35}$$

$$\frac{dB_6}{dt} = \beta_B B_6 S_6 + \chi_6 S_6 (B_3 + B_5) \tag{S.36}$$

$$\frac{dM_6}{dt} = \alpha[(S_6 V_6 B_6 (\beta_V + \beta_B) + \chi_6 S_6 (B_6 (V_3 + V_5) + V_6 (B_3 + B_5)))] \tag{S.37}$$

$$\frac{dV_7}{dt} = \beta_V V_7 S_7 + \chi_7 S_7 (V_3 + V_5 + V_6) \tag{S.38}$$

$$\frac{dB_7}{dt} = \beta_B B_7 S_7 + \chi_7 S_7 (B_3 + B_5 + B_6) \tag{S.39}$$

$$\frac{dM_7}{dt} = \alpha[(S_7 V_7 B_7 (\beta_V + \beta_B) + \chi_7 S_7 (B_7 (V_3 + V_5 + V_6) + V_7 (B_3 + B_5 + B_6)))]. \tag{S.40}$$

Deriving relationship between coinfected and single-infected cells

We found that the relationship between the frequency of coinfected cells and of singly infected cells is approximately linear (e.g., **Figure 7**). To understand this we performed the following analyses. Specifically, we aim at calculating asymptotic behavior of $\frac{dM_k}{dV_k}$ and $\frac{dM_k}{dB_k}$.

Derivation of the "V_k" case

Using basic calculus and eqn. (18), eqn. (19), and eqn. (23) we find:

$$\frac{dM_k}{dV_k} = \frac{\frac{dM_k}{dt}}{\frac{dV_k}{dt}} = \frac{\alpha[B_k V_k' + V_k B_k']}{V_k'} = \alpha\left[B_k + \frac{V_k B_k'}{V_k'}\right]. \tag{S.41}$$

where $'$ denotes derivative in time. The key to the behavior of the relationship between coinfected and single-infected cells thus lies in understanding the behavior of

$$\frac{V_k B_k'}{V_k'}. \tag{S.42}$$

Leaf 3. We first consider the leaf 3 as it is the simplest and provides a method we can use to understand patterns for higher leaves. Simplifying eqn. (S.42) gives:

$$\frac{V_3 B_3'}{V_3'} = \frac{V_3(\beta_B S_3 B_3)}{\beta_V S_3 V_3} = \frac{\beta_B}{\beta_V} B_3 \tag{S.43}$$

Using this, we can find the expression for the original equation.

$$\frac{dM_k}{dV_k} = \alpha \left[B_3 + \frac{\beta_B}{\beta_V} B_3 \right] = \alpha \left(1 + \frac{\beta_B}{\beta_V} \right) B_3 = cB_3 \tag{S.44}$$

Further leaves. In the cases where $k > 3$ we have:

$$\frac{V_k B_k'}{V_k'} = \frac{V_k(\beta_B S_k B_k + \chi_k S_k \sum_{i=3}^{k-1} B_i)}{\beta_V S_k V_k + \chi_k S_k \sum_{i=3}^{k-1} V_i} \tag{S.45}$$

This expression is much more difficult to simply than the $k = 3$ case. However, if we take a linear combination between V_3, V_5, etc. we can proceed. For simplicity, we can use the average:

$$\overline{V} = \frac{1}{n+1} \left(V_k + \sum_{i=3}^{k-1} V_i \right), \qquad \overline{B} = \frac{1}{n+1} \left(B_k + \sum_{i=3}^{k-1} B_i \right) \tag{S.46}$$

where n is the number of proper leaves below the kth leaf. Using this eqn. (S.45) becomes:

$$\frac{\overline{V}(\beta_B S_k \overline{B} + \chi_k S_k \overline{B})}{\beta_V S_k \overline{V} + \chi_k S_k \overline{V}} = \overline{B} \frac{\beta_B + \chi_k}{\beta_V + \chi_k} \tag{S.47}$$

And thus we have:

$$\frac{dM_k}{dV_k} = \alpha \left[\overline{B} + \overline{B} \frac{\beta_B + \chi_k}{\beta_V + \chi_k} \right] = \alpha \left(1 + \frac{\beta_B + \chi_k}{\beta_V + \chi_k} \right) \overline{B} = c\overline{B} \tag{S.48}$$

Because $\overline{V}$ and $\overline{B}$ are linear functions of V_k and B_k, we can conclude that indeed $\frac{dM_k}{dV_k}$ is proportional to B_k, and by inference, $\frac{dM_k}{dB_k}$ is proportional to V_k.

Derivation of the "B_k" case

Proceeding similarly as with eqn. (S.41) we find

$$\frac{\frac{dM_k}{dt}}{\frac{dB_k}{dt}} = \frac{\alpha[B_k V_k' + V_k B_k']}{B_k'} = \alpha \left[\frac{B_k V_k'}{B_k'} + V_k \right]. \tag{S.49}$$

Leaf 3.

$$\frac{B_3 V_3'}{B_3'} = \frac{B_3(\beta_V S_3 V_3)}{\beta_B S_3 B_3} = \frac{\beta_V}{\beta_B} V_3 \implies \frac{\frac{dM_k}{dt}}{\frac{dV_k}{dt}} = \alpha \left[\frac{\beta_V}{\beta_B} V_3 + V_3 \right] = \alpha \left(\frac{\beta_V}{\beta_B} + 1 \right) V_3 = c V_3 \qquad \text{(S.50)}$$

Further leaves. In the cases where $k > 3$ we have:

$$\frac{B_k V_k'}{B_k'} = \frac{V_k(\beta_V S_k V_k + \chi_k S_k \sum_{i=3}^{k-1} V_i)}{\beta_B S_k B_k + \chi_k S_k \sum_{i=3}^{k-1} B_i} \qquad \text{(S.51)}$$

Let

$$\overline{V} = \frac{1}{n+1} \left(V_k + \sum_{i=3}^{k-1} V_i \right), \qquad \overline{B} = \frac{1}{n+1} \left(B_k + \sum_{i=3}^{k-1} B_i \right) \qquad \text{(S.52)}$$

where n is the number of proper leaves below the kth leaf.

$$\frac{B_k V_k'}{B_k'} \rightarrow \frac{\overline{B}(\beta_V S_k \overline{V} + \chi_k S_k \overline{V})}{\beta_B S_k \overline{B} + \chi_k S_k \overline{B}} = \overline{V} \frac{\beta_V + \chi_k}{\beta_B + \chi_k} \qquad \text{(S.53)}$$

And thus we have:

$$\frac{dM_k}{dB_k} = \alpha \left[\overline{V} \frac{\beta_V + \chi_k}{\beta_B + \chi_k} + \overline{V} \right] = \alpha \left(\frac{\beta_B + \chi_k}{\beta_V + \chi_k} + 1 \right) \overline{V} = c \overline{V} \qquad \text{(S.54)}$$