Plant pathogen profiling with the EpiPv package

Ruairí Donnelly[1*], Israël Tankam[2] and Christopher A. Gilligan[1]

[1]Department of Plant Sciences, University of Cambridge, Cambridge, UK.

[2]Institut Agro Rennes-Angers

[*]rd501@cam.ac.uk

Abstract

1. This study introduces a flexible framework for epidemiological profiling of insect-borne plant pathogens (IBPPs), utilizing readily available experimental data. The framework is applicable to most IBPPs transmitted by insects feeding on plant veins, with particular relevance to whitefly-borne viruses that impact cassava production in sub-Saharan Africa. The goal of the study is to provide an approach to estimate critical parameters for IBPP epidemics and use these estimates to assess epidemic risk in the field.

2. The study employs analyses of access period experimental data to estimate three key parameters underlying IBPP epidemics: (i) the rate of pathogen acquisition by insects, (ii) the rate of plant inoculation by pathogen-carrying insects, and (iii) the rate of loss of infectiousness for pathogen-carrying insects. These parameters are incorporated into models that allow for the inference of epidemic risk following inoculum introduction in the field. The methods are packaged into the EpiPv R package, which facilitates rapid implementation and analysis.

3. The EpiPv R package was applied to analyze whitefly-transmitted cassava viruses. The results show that a critical whitefly density of approximately greater than 4 per plant is needed for sustained spread of the CBSI ipomovirus from infected planting material. In contrast, CBSI introductions in whitefly are liable to go extinct even in high-density whitefly populations. A different picture is uncovered for CMB begomovirus - whereby introductions in both plants and whitefly are found to be viable even at very low whitefly densities. This demonstrates significant, actionable, differences in the transmission attributes of these viruses - as uncovered by the EpiPv package.

4. These findings highlight the utility of the EpiPv framework for predicting the outcome of pathogen introductions and for guiding targeted disease management strategies. The ability to estimate key parameters and predict epidemic risk enables more informed decision-making for the control of insect-transmitted plant diseases, with broader applications for managing plant pests globally in both natural and cultivated systems.

1 INTRODUCTION

Global productivity from the cultivation of crops like cassava is severely limited by insect-borne plant pathogens (henceforth IBPPs) (Colvin et al., 2006). For instance, cassava provides more than half of the dietary calories for over 200 million people in East and Central Africa (Alene et al., 2013; FAO and IFAD., 2005; Reincke et al., 2018; Mwebaze et al., 2018), but African cassava production has been severely impacted by whitefly-borne viruses resulting in food insecurity (Mwebaze et al., 2018) and $>$ \$1.25 billion in annual crop losses (Legg et al., 2006; Macfadyen et al., 2021; Mwebaze et al., 2018). Researchers in insect-borne plant virology typically use a set of laboratory experiments referred to here as *access period assays* to confirm, and to investigate, virus transmission by putative insect vectors (Chant, 1958; Dubern, 1994; Maruthi et al., 2020). In this paper we introduce a framework for *epidemiological profiling* of IBPPs using access period data and collect the functions in the dedicated R package, EpiPV. By *epidemiological profiling* we mean estimation of virus transmission parameters and subsequent inference of epidemic risk. This paper describes the functions of the EpiPV package and applies them to profile two whitefly-transmitted viruses: cassava mosaic begomovirus (CMB) and cassava brown streak ipomovirus (CBSI).

The epidemiology of IBPP transmission is influenced by the rates of virus acquisition and inoculation and the retention period in the insect also plays an important role. Indeed, retention duration is a common means of classifying plant viruses as persistently transmitted (long retention, PT) and semi- or non-persistently transmitted (short retention, SPT or NPT) (Eigenbrode et al., 2018; Hogenhout et al., 2008) - with latent period relevant for PT but not for SPT or NPT viruses (Hogenhout et al., 2008). Note that true non-persistently transmitted viruses are acquired from epidermal plant cells by aphids (Carr et al., 2018) - it is important to note that at present our framework does not apply to these viruses which require separate treatment (Donnelly et al., 2019), and instead applies to the majority of IBPPs that are acquired from the phloem of plants (Carr et al., 2018). Among these phloem-restricted viruses are semi-persistently transmitted viruses like CBSI and persistently-transmitted viruses like CMB. The functions in *EpiPv* provide a user-friendly means to estimate the rates of virus acquisition and inoculation from access period data provided by the user.

Our approach is built around simple probability theory, which is central to modern epidemiology because of its dual utility in parameter estimation and statistical inference (Keeling and Rohani, 2011). In the first of these dual uses, probability models for epidemiological scenarios can be combined with epidemiological data to estimate parameters (Bolker, 2008). In the *EpiPv* package simple probability models for access period assays are combined with access period assay data to estimate acquisition and inoculation rates. In the second of these dual uses, probability models based upon parameter estimates can be used to make inferences. In the *EpiPv* package virus parameter estimates are used to infer the epidemic risk from inoculum introductions in the field.

For cassava viruses, and viruses of roots and tubers in general, the propagation of infected planting material for new growth constitutes an additional transmission mode in the field. This mode of transmission is influential in large-scale spatio-temporal dynamics of IBPPs, but is not a factor in the question of whether or not introduction of an infected plant or an infected insect vector into a population of host plants will result in an epidemic apart from as an epidemic seeding event. This is because in the absence of preferential selection of infected plant material propagation alone cannot lead to epidemic growth at a given location - but rather propagation plays an important role in the persistence of infection across growing seasons as well as virus movement to new locations. Note, however, that the initial introduction of an infected plant can be thought of as representing a single transmission event based on introduction of propagated infected material.

For insect-borne plant pathogens additional modes of transmission may exist. For instance, viruses of roots and tubers are transmitted when infected planting material is propagated for new growth. However once inoculum has been introduced - potentially through vegetative propagation - the risk of a local epidemic is a consequence of insect-borne transmission. This is because infected plant material propagation alone cannot lead to epidemic growth at a given location in the absence of preferential selection of infected material. This is why experiments in insect-borne transmission are important, yet there has been no way to translate laboratory data into epidemic risk. In this paper we provide this missing link. The paper is structured as follows: we first introduce the key epidemiological processes for phloem-limited IBPPs. We then describe the estimation of virus parameters using models based on these processes from access period data using Bayesian analysis. We next introduce the inference of local epidemic risk from the viral parameter estimates and several additional local parameters. We show how to apply the methods to published data as illustrated by whitefly-borne cassava viruses. The *EpiPv* package enables rapid adoption of the epidemiological profiling to analyse and identify effective strategies for disease management.

2 MATERIALS AND METHODS

Access period experiments are an essential means to study insect vector transmission of plant pathogens in the laboratory (Figure 1). In these experiments insect vector cohorts are provided feeding access to pathogen-infected plants, and are then transferred to healthy test plants. This article introduces a framework that makes use of access period experimental data to produce epidemiological profiles of plant pathogens. Two steps lie at the heart of the framework: parameter estimation from experimental data (step A) and the subsequent inference of field epidemic probabilities (step B) (together referred to as epidemiological profiling). Both steps (laboratory vs field model) are based on a quantitative description of the insect-borne plant pathogen (IBPP) interaction. In the first step we tailor a simple version of the quantitative model (referred to as a laboratory model) to generic access period assays for the estimation of virus transmission parameters. In the second step we use the parameter estimates with a probability model of field introductions (field model) to make inferences relating to epidemic risk.

How insects transmit the virus

For phloem-restricted plant viruses like the cassava viruses CMB and CBSI, the larger the period of feeding the greater the rate of virus acquisition and inoculation (Figure 2A). This means that the rates of acquisition and inoculation for a given host plant are proportional to the number of uninfected and infected insect vectors respectively feeding on the phloem of the plant,

$$\begin{aligned} &\text{Acquisition } (\textit{models: field; lab.}) \quad I_j^F \xrightarrow{(F-j)\alpha} I_{j+1}^F \qquad (1.1)\\ &\text{Inoculation } (\textit{models: field; lab.}) \quad S_j^F \xrightarrow{j\beta} E_j^F \qquad (1.2)\end{aligned}$$

where α and β in expressions 1.1-1.2 denote the acquisition and inoculation rates, and where X_j^F denotes a plant state (see Figure 1 and Table S1.1 for complete list of parameters) with

$X \in \{S, E, I\}$ (susceptible, S, virus expose, E, or infectious, I), with a number of phloem-feeding insect vectors on the plant (integer-valued superscript F), of which a number are virus-carrying (integer-valued subscript $j \leq F$). For instance, expression 1.1 represents the population transition from an infected plant with j infected insect vectors (i.e., I_j^F) to an infected plant with $j+1$ infected insect vectors (i.e., I_{j+1}^F) due to a virus acquisition event which occurs at rate $(F-j)\alpha$. Note that the event rates (expressions 1.1-1.2) are in units of per plant per day and are applicable when modelling both laboratory access period assays and field situations.

How insects disperse the virus

For phloem-restricted viruses like CMB and CBSI, the virus is acquired when an uninfected insect feeds on infected phloem (Figure 2B) and movement of an infected insect from 'birth' (i.e. acquisition) to 'death' (i.e., insect death or viral clearance) depends on the following life-history events with associated rates,

$$\begin{array}{lcr}
\text{Infected insect dispersal } (\textit{models: field}) & X_j^F \xrightarrow{j\theta} X_{j-1}^F,\ S_0^F \xrightarrow{j\theta} S_1^F & (1.3) \\
\text{Infected insect death } (\textit{models: field, } b_{y=l}\textit{; lab., } b_{y=f}) & X_j^F \xrightarrow{jb_y} X_{j-1}^F & (1.4) \\
\text{Infected insect recovery } (\textit{models: field; lab.}) & X_j^F \xrightarrow{j\mu} X_{j-1}^F & (1.5)
\end{array}$$

where θ, b_f, b_l and μ are the per-insect rates of dispersal, field mortality, laboratory mortality and virus clearance, respectively, and where $X \in \{S, E, I\}$. Note that infected insect dispersal results in changes to both source and destination plant states. At invasion all plants other than the inoculum, which itself originally arose from external means (i.e., through vegetative propagation or through infected insect migration into the field), are assumed to be free of infection. For this reason, the destination plant is assumed to have state S_0^F prior to infected insect dispersal. After dispersal has occurred the population may consist of two sets of inoculum (this occurs if the initial inoculum state was I_j^F or E_j^F with $j \geq 1$ or if it was S_j^F with $j > 1$). In addition, a number of events influence the infectiousness of plants, The event rates 1.3-1.5 are per plant per day, with the transitions as described in the previous section.

$$\begin{array}{lll}
\text{Exposed plant progression } (\textit{models: field}) & E_j^F \xrightarrow{\nu} I_j^F & (1.6) \\
\text{Exposed plant harvest } (\textit{models: field}) & E_j^F \xrightarrow{h} S_j^F & (1.7) \\
\text{Infected plant removal } (\textit{models: field}) & I_j^F \xrightarrow{r} S_j^F & (1.8) \\
\text{Infected plant harvest } (\textit{models: field}) & I_j^F \xrightarrow{h} S_j^F & (1.9)
\end{array}$$

where ν, r are the per-plant rates of onset of infection and infected plant removal, and h is the rate that plants are harvested (see Table S1.1 for complete list of parameters). The event rates (expressions 1.3-1.9) are in units of per plant per day. Note also that expressions 1.6-1.9 reflect an assumption of re-planting with susceptible material for simplicity. Note that a number of these events (1.3 and 1.6-1.9) apply only to field situations, i.e., they are not relevant to the laboratory context.

Step A, estimating viral transmission in the lab

Two parameters, α (acquisition rate) and β (inoculation rate), play a central role in IBPP transmission in the field as well as in laboratory access period assays, and a third, μ (insect virus clearance rate) is critical to IBPP epidemiology (Eqs 1.1-1.2). Estimating these parameters is a key task. It is also important to account for virus latent period in the insect where relevant in order to ensure reliable parameter estimation (through the rate with which virus-exposed insects become infectious, γ). A simplified representation of access period assays is shown in Figure 3A. For PT viruses the assay is frequently extended to include a latent access period (e.g., Dubern (1994)'s PT virus assay with - cf. Maruthi et al. (2020)'s SPT virus assay without - latent access period). As outlined in Figure 3A, parameter estimation from access period data requires calculation of the probability of the data given a probability model of the assay. In effect, the model calculates how likely the observed data is for different values of the underlying parameters given assay access durations (see Supporting Information S1-S2 for details). In this way sources of variation (assay variation and natural variation) that underlie access period data can be harnessed to estimate event rates such as α, β, μ and γ using Bayesian analysis.

Step B, inferring epidemic probability in the field

In step B the virus parameter estimates for α, β and μ from step A are used to produce estimates for the *epidemic probability* (when 1 infected plant or 1 infected insect vector is introduced into a population of susceptible host plants). Note that additional user-inputted local parameters are also required: the number of insects per plant (F), the insect dispersal (θ) and natural mortality (b_f) rates, the harvesting rate (h) and the rate that infected plants are removed (r), and, in addition, the rate that exposed plants become infectious (ν). The field model therefore assumes that a particular location or situation is associated with a constant insect-burden such that all plants have F phloem-feeding insects. This assumption is highly suitable when plant pathologists have associated a location or situation with a given insect burden - or where the aim is to investigate the impact of the magnitude of vector density on epidemic risk. We now briefly indicate how the *epidemic probability* is calculated for IBPPs.

The strategy is to condition the epidemic probability from a given inoculum state on possible future events (Keeling and Rohani, 2011) (Figure 3B, Supporting Information S3). By inoculum state we mean the infection state of a single plant unit of infection. This may correspond to infection or latent infection of the plant and/or infection of any of the insects feeding on the plant. Furthermore, the number of possible inoculum states depends on the local insect burden, F: there are $2 \times (F + 1) + F$ states ($I_0^F..I_F^F$, $E_0^F..E_F^F$ and $S_1^F..S_F^F$).

The calculation involves relating the extinction probability for a given inoculum state to the extinction probabilities for other inoculum states. This is achieved by examining the inoculum states produced by the events that may occur and accounting for the extinction probabilities from these new states. This process leads to a set of simultaneous equations that can be solved for the extinction probabilities associated with each inoculum state (and hence for epidemic probability i.e. $1-$ the probability of extinction).

In the following example, for illustration, we ask what is the fate of a single introduced infected plant $P(I_0^F)$? The possible future events for the inoculum state are as follows: the infected plant may be rogued (i.e., transition from I_0^F to S_0^F) with rate r, or it may be harvested (i.e., transition from I_0^F to S_0^F) with rate h, or a phloem-feeding insect on the plant may acquire the virus (i.e., a transition from I_0^F to I_1^F) with rate $F\alpha$. Conditioning on these possible events leads to $P(I_0^F) = (r/(r+h+F\alpha))P(S_0^F)+(h/(r+h+F\alpha))P(S_0^F)+(F\alpha/(r+h+F\alpha))P(I_1^F)$ in which the coefficients are the relative probabilities of a given event and the multiplicative terms are the probabilities of extinction for the ensuing inoculum states. Each of the inoculation states can be related to other inoculum states in the above manner. This leads to a set of simultaneous equations which can be solved for the extinction (hence epidemic) probabilities (Figure 3B).

The *EpiPv* R package

The *EpiPv* R package has two main uses. The first is the Bayesian estimation of virus transmission rates from access period data (step A in previous section). The second is the inference of epidemic probability based upon virus rate estimates and several local parameters (step B in previous section). In what follows we list the functions that are available in the *EpiPv* R package,

1. The *estimate_virus_parameters_PT* function receives AP data for a given vector-PT

virus-plant combination as user input together with assay configuration (i.e., AP feeding durations T_A T_L T_I and the number of insect vectors used X_0) and returns posterior parameter distributions for the transmission rates μ, α, β, and γ (see Supporting Information S1 for details).

2. The *estimate_virus_parameters_SPT* function receives AP data for a given vector-SPT virus-plant combination as user input together with assay configuration (i.e., AP feeding durations T_A T_I and the number of insect vectors used X_0) and returns posterior parameter distributions for the rates transmission rates μ, α, and β (see Supporting Information S2 for details).

3. The *calculate_epidemic_probability* function receives event rate parameters for virus transmission (μ, α, β) as well as local parameters F, θ, r, h and b_f as user input, and returns epidemic probability for different types of inoculum state (see Supporting Information S3 for details).

4. The *AP_data_simulator* function receives event rate parameters for virus transmission (μ, α, β) as well as assay feeding durations and returns simulated access period data (see Supporting Information S4 for details of the statistical simulation process).

3 RESULTS

In what follows we describe the epidemiological profiling of the whitefly-borne CMB and CBSI cassava viruses (cassava mosaic begomovirus; cassava brown streak ipomovirus). We first report the viral parameter estimates, and then the risk inferences, as described in the methods section.

Profiling whitefly-borne cassava viruses

When we applied the laboratory-scale model (materials and methods) to the Dubern (1994) CMB dataset (Supporting Information S1) and the Maruthi et al. (2020) CBSI dataset (Supporting Information S2), we obtained 95% credible intervals for CMB and CBSI acquisition, inoculation and insect clearance rates (Table 1 A-B). In addition, we estimated latent progression rate for the persistently-transmitted CMB virus - but note that use of an alternative plant for the LAP meant that we were unable to utilise the latent period varying sub-assay from Dubern (1994) in our model fitting exercise (see subsection 'The latent period varying sub-assay in Dubern (1994)' Supporting Information S1).

Model convergence was assessed by examining potential scale reduction factors ($Rhat \leq 1.01$ for all parameters for both CMB and CBSI), effective sample sizes ($n_{eff} >$ no. iterations for all parameters for both CMB and CBSI), and the absence of divergent transitions or tree depth exceedences. Absence of strong correlations among the estimated parameters was confirmed by examination of parameter pairs plots and mixing of chains confirmed by the earlier reported effective sample sizes. Excellent agreement between observed data and forward simulation from the estimated parameters is evident from figure S2.1 for CBSI and good agreement for CMB (figure S1.3). Excellent model fit, as evaluated through Bayesian R^2, was confirmed for CBSI (70% of variance explained by model) and good model fit for CMB ($\approx$ 50% of variance explained by model). Note that CMB modelling, though good, was not as strong as for CBSI - this likely relates to the use of an alternative plant (Chinese lantern rather than cassava) for the intermediate LAP phase in the CMB AP data of Dubern (1994), as discussed in Supporting Information S1.

Viral parameter estimates (full posterior distributions) were then taken forward for both viruses to calculate local epidemic probability (summarised in table 2) corresponding to set levels of insect burden. The epidemic probabilities that result are summarised in Table 2 (posterior credible intervals for epidemic probability vs. insect burden for viral introduction in plants, A, and insects in B). In addition, full posterior distributions are shown in Figure 4 focusing on plant vs insect inocula (plant and insect forms of inoculum for selected values of insect burden: F=1 v F=3 for CMB in A v C, and, F=4 v F=10 for CBSI in B v D). Full posterior distributions are also shown in Figure 5 focusing on insect burden variation. We collate these findings as epidemiological profiles for CBSI and then for CMB.

Cassava brown streak ipomovirus: a feeble inoculator that hides in plain sight

- *highly ephemeral retention* in *B. tabaci* insect vector. 95% credible interval for virus clearance per hour, $0.406 - 1.468h^{-1}$ (Table 1A iv); median virus retention $0.807^{-1} = 1.24h$.

- is *highly transmissible from infected cassava to uninfected B. tabaci.* 95% credible interval for acquisition rate per hour, $0.088 - 1.735h^{-1}$ (Table 1A i); median probability of acquisition per insect in a 24h feeding period ≈ 1.

- has *low transmissibility from infected B. tabaci to uninfected cassava.* 95% credible interval for inoculation rate per hour, $0.019 - 0.475h^{-1}$ (Table 1A ii). Thus, while median probability of inoculation per virus-bearing insect in a 24h feeding period is

≈ 0.7, median inoculation probability per insect infection is *only* ≈ 0.07 - when we take account of median retention time.

- *low epidemic risk from insect inocula, high epidemic risk from plant inocula requires high local insect burden.* High likelihood of local epidemics given infected plant introduction but only for moderate-high *B. tabaci* insect burden (95% epidemic probability credible interval $0.133 - 0.442$ for 4 whitefly per top 5 leaves, cf. $0.695 - 0.816$ for 10 whitefly, Table 2A iv cf, vi). Low likelihood of local epidemics given infected insect introduction (95% credible interval for epidemic probability, $0.027 - 0.238$ even for 10 whitefly per top 5 leaves, Table 2B vi);

- *low symptom detectability in plants* Beyond the scope of access period experiments - CBSI is notoriously difficult to identify in over-ground biomass. As such, there is a high tendency for human-mediated propagation to new seasons and new locations of cassava cultivation. Thus it has the ability to persist beyond the growth period of a cassava population.

Cassava mosaic begomovirus:
an all-rounder that is hard to miss

- is *weak- to moderate-ly transmissible from infected cassava to uninfected B. tabaci.* 95% credible interval for acquisition rate per hour, $0.012 - 0.016h^{-1}$ (Table 1B i); median probability of acquisition per uninfected insect in a 24h feeding period ≈ 0.285.

- is *highly transmissible from infected B. tabaci to uninfected cassava.* 95% credible interval for inoculation rate per hour, $2.219 - 3.995h^{-1}$ (Table 1B ii); median probability of at least one inoculation per infected insect in a 24h feeding period ≈ 1.

- *sustained retention* in the *B. tabaci* insect vector. 95% credible interval for loss of insect infectiousness per hour, $0.0001 - 0.019h^{-1}$ (Table 1B iv) (corresponding to between $2.2d$ and $> 100d$). This is supported by the finding of no evidence for whitefly clearance of CMB in Donnelly and Gilligan (2023).

- *very high risk from plant and insect introductions.* High likelihood of local epidemics given inoculum introductions in plant or in insect for even relatively low *B. tabaci* insect burdens. 95% credible interval for epidemic probability, $0.710 - 0.810$ (plant inoculum) and $0.758 - 0.865$ for 1 whitefly per top 5 leaves (Table 2A i, B i).

- *high symptom detectability in plants* Beyond the scope of access period experiments - CMB disease is highly visible in over-ground biomass. As such, there is greater scope for managing human-mediated propagation to new seasons and new locations of cassava cultivation. The release of CMB-tolerant varieties, however, in recent decades may have lead to chronic and less visible CMB disease incidence.

Computational validation of epidemic risk

For comprehensiveness, we also used individual-based simulation to compare predicted values of epidemic probability (calculate_epidemic_probability() function, EpiPv package) with the outcome of a large number of simulations. By individual-based simulation we mean the reproduction of the events that occur (on a per insect, per plant, basis) when inoculum is introduced into a field, with events simulated in proportion to event rates that are update each time an event occurs. Correspondence between prediction methods and simulation is achieved by introducing an infected host into a field of susceptible hosts at the start of a season - with fields evaluated at the end of the season for epidemic growth or extinction of the inoculum. This procedure provides a baseline for verifying the accuracy of our methods.

In brief, we found that the predicted epidemic probability matched the outcomes of simulations across a range of parameter value sets (Figure S5.1, blue crosses for predicted epidemic probabilities match circles for simulated epidemic probabilities by). The exercise also demonstrated that infrequent removal of infectious plants ($> 2weeks$ longevity of symptomatic plants) are associated with high epidemic probabilities, that low whitefly dispersal (< 1 dispersal per day per whitefly) lead to rapid declines in epidemic probability, that epidemic probability was less than 0.5 only for very rapid loss of insect infectiousness ($< 2h$ retention), and that epidemic probability was less than 0.5 only for very few whitefly per top 5 leaves (Figure S5.1 A,B,C,D respectively).

4 DISCUSSION

We have introduced a framework to estimate virus parameters and to predict the outcomes of virus introductions in the field. The framework utilises laboratory access period studies to estimate virus transmission parameters. Such studies are central to establishing insect transmission but say little about how transmissibility translates into epidemic risk. The framework that we introduce provides the missing link. Laboratory researchers can now extrapolate from their data to the consequences for insect-borne epidemics in field situations. We applied the methods to laboratory studies for CBSI and CMB showing that CBSI is characterised by high virus acquisition rates but is hindered by a combination of moderate inoculation rate and highly ephemeral retention in the insect, while CMB is characterised by a high virus inoculation rate and a relatively low acquisition rate.

The sustained retention of CMB in *B. tabaci* relative to CBSI makes its ability to cause epidemics at low whitefly abundances far more favorable than for CBSI. This is consistent with claims that high regional whitefly abundance was needed for CBSI expansion in sub-Saharan Africa to occur (Donnelly and Gilligan, 2020). The combination of moderate inoculation and ephemeral retention leads to strikingly lower risk associated with CBSI-infected vector introductions than from CBSI-infected plant introductions. This is a key finding emerging from this work that is important for epidemic management. For CMB epidemic risk is high from both types of introductions.

Plant pathologists use laboratory experiments in insect access to host plants to investigate IBPPs. The data from these experiments can be used to parameterise epidemic models. The wealth of published laboratory data from such experiments for a wide range of plant pathogens constitutes a valuable resource for epidemiologists - but they remain a relatively untapped resource. We encoded simple analyses of access period data - the proportion of infected test plants in acquisition and inoculation varying sub-assays - in a dedicated R package - *EpiPv*. The package estimates the main parameters of insect-borne transmission and then uses the estimates for an epidemiological profiling of the IBPP. As such the package offers a means for plant pathologists to complement experimental investigations with quantitative results that facilitate inter-species and inter-strain pathogen and insect vector comparisons of epidemic risk.

Cassava viruses, management, and future versions of the EpiPv R package

Cassava is produced by smallholder farmers whose average cultivated area is less than one hectare (Masamha et al., 2018) and is already mainly grown under inter-cropping systems (with crops such as maize, legumes, and bananas). It is important to understand the relationship of cassava inter-cropping and epidemics.

In general, epidemic growth is due to secondary transmission of insect-borne infection and this can be directly linked to the basic reproduction number, with sufficient insect-borne spread for epidemic growth occurring when the basic reproduction number $R_0 > 1$, which corresponds to non-zero epidemic probability. Our results show that the probability of CBSI epidemics is already relatively low in susceptible monocultures for infected insect introductions. This raises the possibility that CBSI epidemics arising from infected insect introductions could be entirely prevented using cultivar mixtures. A combination of strict control of the planting material that is propagated together with wide-spread intercropping of susceptible and resistant cultivars could entirely suppress landscape CBSI epidemics. A priority for future iterations of the *EpiPv* R package will be the calculation of epidemic probability for crop mixtures.

While the analyses of access period assays in the *EpiPv* package take account of the assay insect density, the inference of epidemic probability in the field requires user input of insect abundance per plant. For plant pathologists using the package this is likely to be an acceptable requirement. Field plant pathologists, for instance, are likely to have typical abundances per plant in mind for specific locations. We aim to extend the methods, however, in future package version to include functions that can predict insect abundance and hence epidemic probability based simply on temperature data for a given location - where this is the decisive environmental factor for the insect vector in question. In addition, the remaining non-cultural local parameters, b_f (field insect mortality rate) and θ (dispersal rate) that currently require user input could also be calculated from temperature data. This will require insect vector life-history data according to temperature which is already available for insect vectors of plant pathogens like whitefly (Aregbesola et al., 2019, 2020).

Uses of the EpiPv package

A critical facet of managing cassava virus epidemics at the landscape scale is the classification of cassava varieties in terms of their susceptibility to the virus in question. Such classification can be considered one of relative susceptibility i.e. classification will typically take account of a reference susceptible cultivar. The functions of the *EpiPv* package can estimate virus transmission parameters from AP assays for a reference susceptible cultivar, and also for a cultivar of interest. Comparison of these parameter estimates alone can provide a quantitative means of classifying cultivars in terms of susceptibility and resistance. In addition, MCMC methods can then provide a statistical test of relative susceptibility or resistance by accounting for variability in both sets of parameter estimates. Ultimately, relative susceptibility should be established in field trials - but such trials tend to be multi-year, resource intensive, and can often produce ambiguous results. Therefore, analysis of candidate cultivar AP assays with the *EpiPv* package could play a key role in streamlining and interpreting field trials, and crucially can show breeders the impact of a level of resistance of susceptibility in terms of epidemic risk.

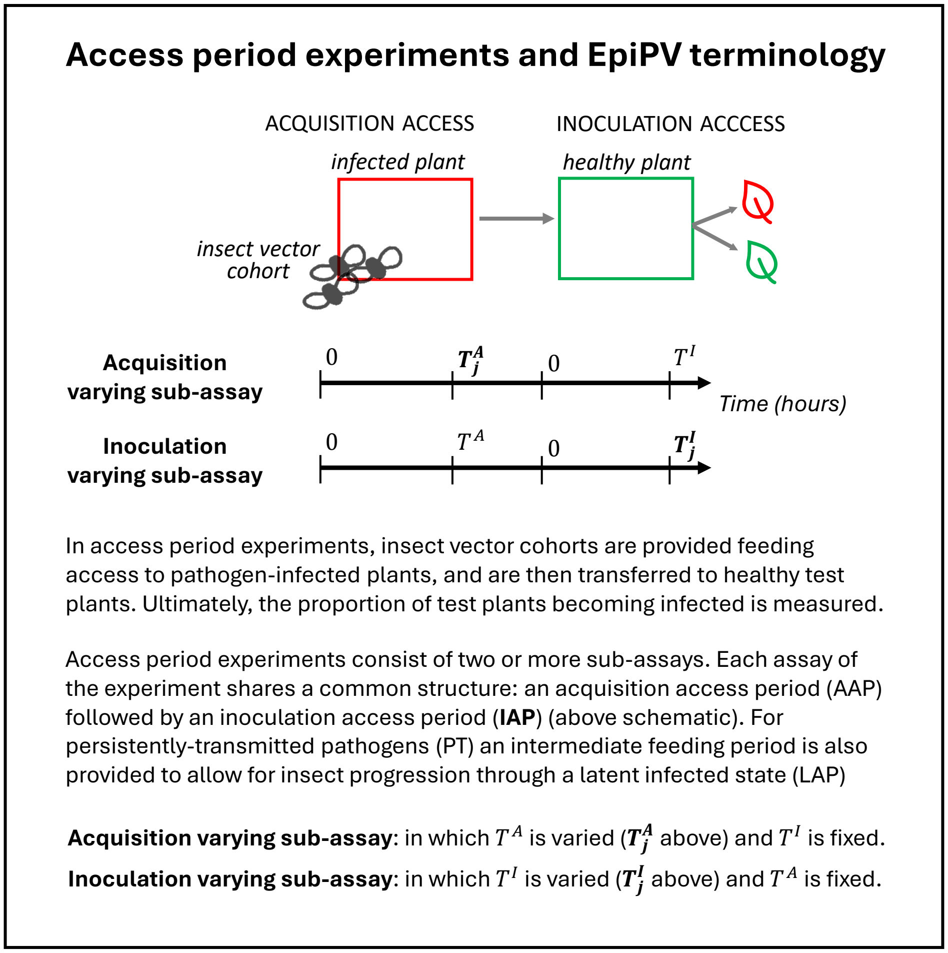
A second facet of managing cassava virus risk is the establishment of virus transmission for alternative virus and/or vector strains. While AP assays have historically been important for this objective - the functions of the *EpiPv* package in combination with these assays provides a means to quantify the parameters which was not previously possible and to translate these experiments into local epidemic risk. Therefore, where strain transmission has been demonstrated in the laboratory - but where it is unclear how this relates to effective risk in the field - the *EpiPv* package can be used to quantify the relative risk involved.

Modelling risk in a changing climate

Finally, where there is a risk of arrival of a plant virus that was not previous present in a region, the the *EpiPv* package provides a means to quantify local epidemic risk. This exercise is widely conducted by countries and regions as a part of pest risk assessment under WTO rules. The risk inference part of the *EpiPv* package is particularly suitable for modelling the risk of pest establishment - a key stage in pest risk assessments (Leung, 2012; World Trade Organization., 1995).

Conclusions

In summary, we provide a framework to harness the wealth of published access period data that is available for IBPPs. The methods are assembled in an R package for direct use by epidemiologists and by the plant pathologists who produce these datasets. The package estimates IBPP transmission rates from access period data and also uses inference of local epidemic probability upon introduction of infected material - which together we refer to as *epidemiological profiling.* The methods provide actionable findings relating to management of cassava virus invasions, for instance by calculating the risk of local epidemics for a given level of the insect vector. While other crops are projected to face significant adaptation challenges in Africa to changing climates, cassava is resilient to climate change because of its drought tolerance (Mwebaze et al., 2018) and may therefore undergo an increase in production in the decades ahead. *Bemisia tabaci* life-histories are also favoured by high temperature. For these reasons the problem of cassava virus disease in SSA may increase even further. The tools introduced in this paper and future extensions will provide a means to quantify changing epidemic threats to food security from cassava viruses, and for IBPPs in general, in a changing world.



Access period experiments and EpiPV terminology

In access period experiments, insect vector cohorts are provided feeding access to pathogen-infected plants, and are then transferred to healthy test plants. Ultimately, the proportion of test plants becoming infected is measured.

Access period experiments consist of two or more sub-assays. Each assay of the experiment shares a common structure: an acquisition access period (AAP) followed by an inoculation access period (**IAP**) (above schematic). For persistently-transmitted pathogens (PT) an intermediate feeding period is also provided to allow for insect progression through a latent infected state (LAP)

Acquisition varying sub-assay: in which T^A is varied ($\boldsymbol{T_j^A}$ above) and T^I is fixed.

Inoculation varying sub-assay: in which T^I is varied ($\boldsymbol{T_j^I}$ above) and T^A is fixed.

Figure 1. Important EpiPv terminology

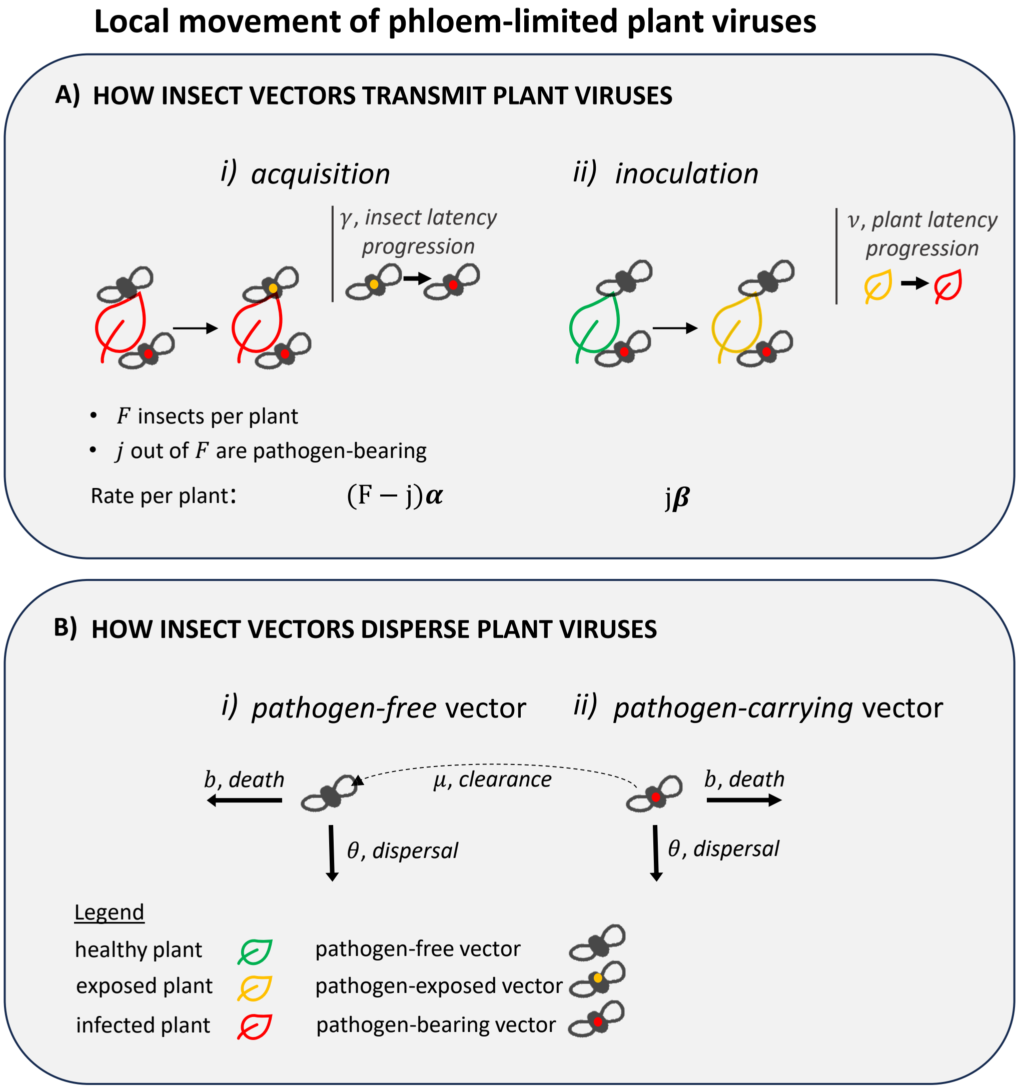
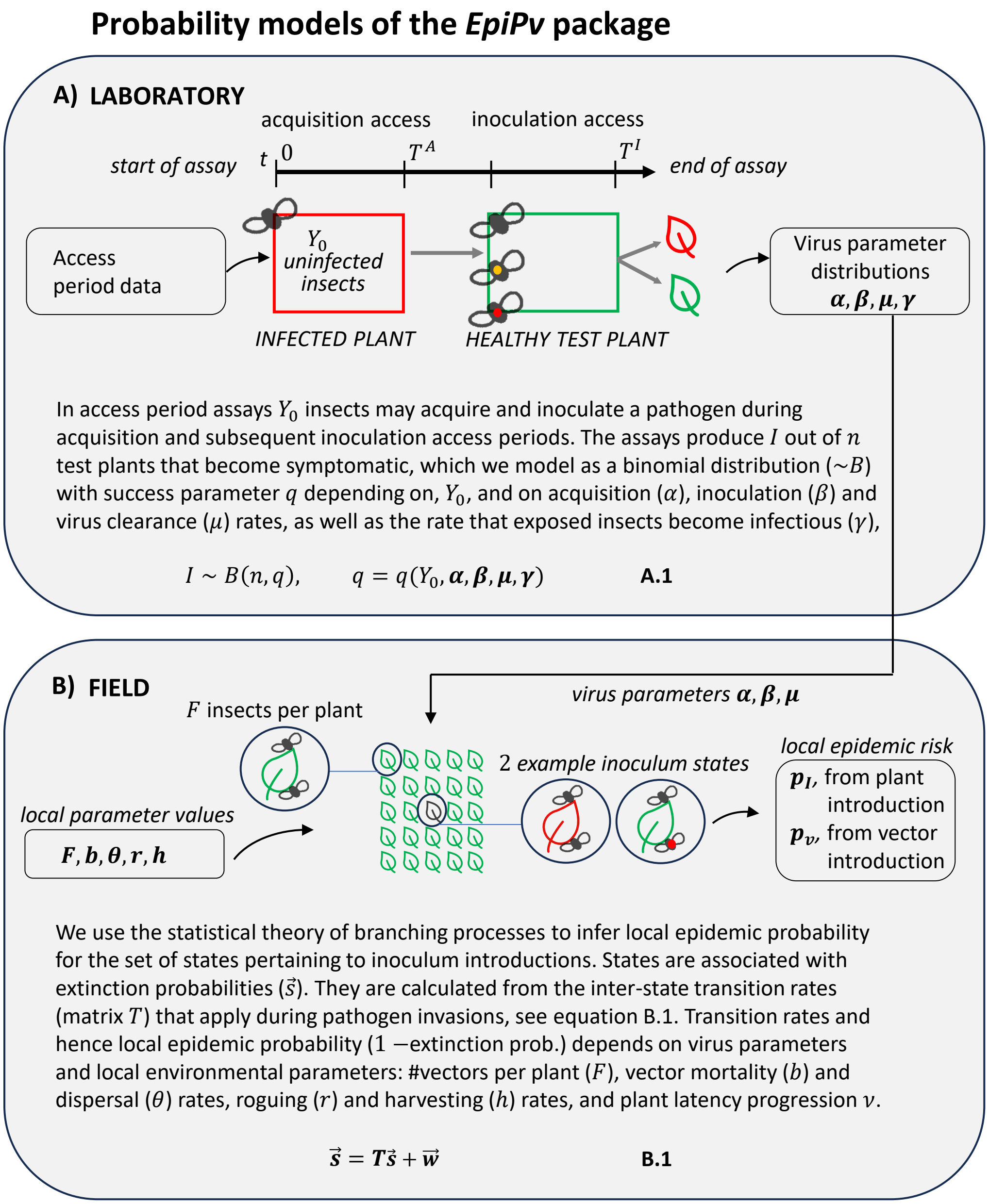


Figure 2. The processes by which phloem-restricted insect-borne plant pathogens undergo transmission and dispersal. In A, virus acquisition (by virus-free insects from virus-infected plants) and virus inoculation (of healthy plants by virus-bearing insects) occur when insects feed on host plant phloem. The overall acquisition rate (see Ai) is proportional to the number of virus-free insects that are feeding on virus-infected plants and the per-insect rate of virus acquisition, α. The overall inoculation rate (see Aii) is proportional to the number of virus-bearing insects that are feeding on healthy plants and the per-insect rate of virus inoculation, β. Note that virus-exposed insects become infectious at rate γ (see Ai inset) and virus-exposed plants become infectious at rate ν (see Aii inset). In B, insect life-history events alter the distribution of phloem-restricted insect-borne plant pathogens with virus-free and virus-bearing insects dispersing to new host plants at rate θ and with insect mortality loss occurring at rate b. In addition, virus-infected insects can clear the virus at rate μ becoming virus-free.


A) LABORATORY
acquisition access
inoculation access
start of assay
t
0
T^A
T^I
end of assay
Access period data
Y_0 uninfected insects
INFECTED PLANT
HEALTHY TEST PLANT
Virus parameter distributions $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\gamma}$
B) FIELD
F insects per plant
virus parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}$
2 example inoculum states
local epidemic risk
$\boldsymbol{p_I}$, from plant introduction
$\boldsymbol{p_v}$, from vector introduction
local parameter values
$\boldsymbol{F}, \boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{r}, \boldsymbol{h}$

Probability models of the *EpiPv* package

In access period assays Y_0 insects may acquire and inoculate a pathogen during acquisition and subsequent inoculation access periods. The assays produce I out of n test plants that become symptomatic, which we model as a binomial distribution ($\sim B$) with success parameter q depending on, Y_0, and on acquisition (α), inoculation (β) and virus clearance (μ) rates, as well as the rate that exposed insects become infectious (γ),

$$I \sim B(n, q), \qquad q = q(Y_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\gamma}) \qquad \textbf{A.1}$$

We use the statistical theory of branching processes to infer local epidemic probability for the set of states pertaining to inoculum introductions. States are associated with extinction probabilities ($\vec{s}$). They are calculated from the inter-state transition rates (matrix T) that apply during pathogen invasions, see equation B.1. Transition rates and hence local epidemic probability (1 −extinction prob.) depends on virus parameters and local environmental parameters: #vectors per plant (F), vector mortality (b) and dispersal (θ) rates, roguing (r) and harvesting (h) rates, and plant latency progression ν.

$$\vec{\boldsymbol{s}} = \boldsymbol{T}\vec{\boldsymbol{s}} + \vec{\boldsymbol{w}} \qquad \textbf{B.1}$$

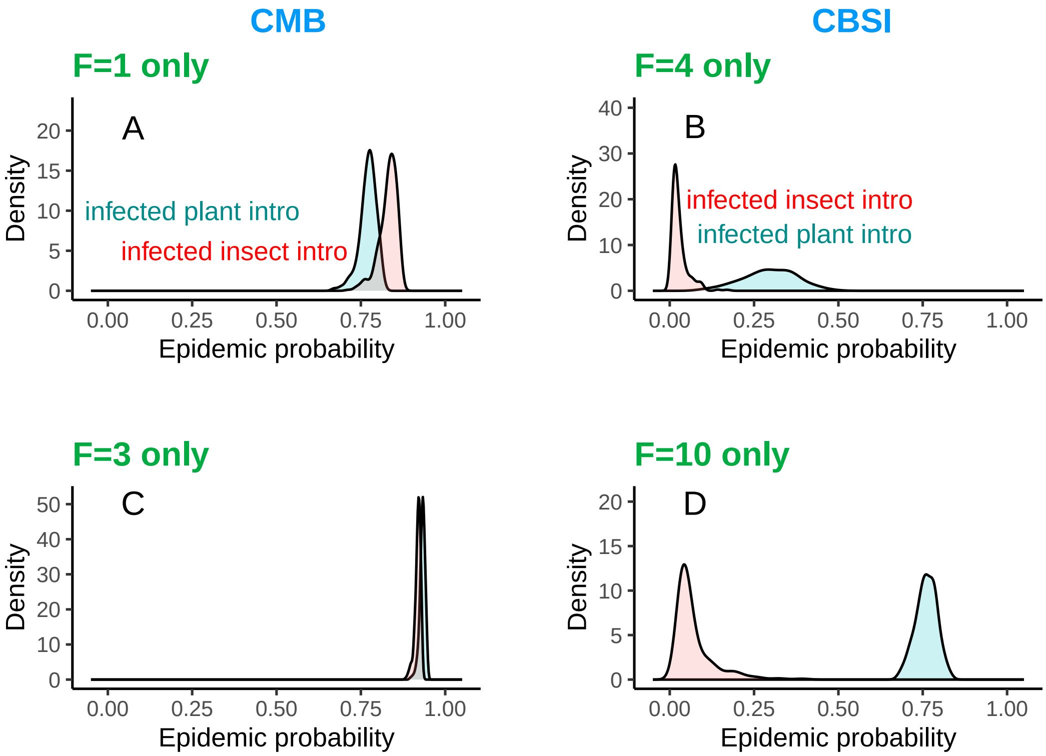
Figure 3. Probability models of the EpiPV package.

			A) CBSI[a]			*B) CMB*		
	rate parameter		*median*	*2.5%*	*97.5%*	*median*	*2.5%*	*97.5%*
i)	*acquisition*	α	$0.638h^{-1}$	$0.088h^{-1}$	$1.735h^{-1}$	$0.014h^{-1}$	$0.012h^{-1}$	$0.016h^{-1}$
ii)	*inoculation*	β	$0.056h^{-1}$	$0.019h^{-1}$	$0.475h^{-1}$	$3.01h^{-1}$	$2.219h^{-1}$	$3.995h^{-1}$
iii)	*latency*	γ	–	–	–	$0.85h^{-1}$	$0.317h^{-1}$	$2.053h^{-1}$
iv)	*insect clearance*	μ	$0.807h^{-1}$	$0.406h^{-1}$	$1.468h^{-1}$	$0.005h^{-1}$	$0.0001h^{-1}$	$0.019h^{-1}$

Table 1. Parameter estimates for viral transmission rates using the *estimate_virus_parameters_SPT* (A) and *estimate_virus_parameters_PT* (B) functions. Parameter estimates are shown for: whitefly-borne CBSI as calculated from the acquisition and inoculation access period assay datasets of Maruthi et al. (2020) (A), and for whitefly-borne CBSI as calculated from the acquisition and inoculation access period assay datasets of Dubern (1994) (B) (see Supporting Information S1-S2 for further details).

			A) plant inoculum			*B) insect inoculum*		
			median	*2.5%*	*97.5%*	*median*	*2.5%*	*97.5%*
CMB	*i)*	$F = 1$	0.774	0.710	0.810	0.835	0.758	0.865
	ii)	$F = 2$	0.916	0.896	0.929	0.927	0.904	0.936
	iii)	$F = 3$	0.933	0.912	0.945	0.920	0.896	0.930
CBSI	*iv)*	$F = 4$	0.297	0.133	0.442	0.020	0.006	0.092
	v)	$F = 7$	0.629	0.534	0.710	0.040	0.021	0.197
	vi)	$F = 10$	0.761	0.695	0.816	0.048	0.027	0.238

Table 2. Epidemiological field inferences for whitefly-borne CMB and CBSI using the *epidemic_probability* function (materials and methods, step B), illustrated with CBSI (top) and CMB (bottom). See Table 1 for parameter estimates representing virus transmission that were passed as arguments to the *epidemic_probability* function together with the following local parameter values: $\theta = 0.45\ d^{-1}$, $r = 1/28\ d^{-1}$, $h = 1/365\ d^{-1}$, $b = 1/14\ d^{-1}$, $\nu = 1/14\ d^{-1}$ (see Table S1.1 for a list of parameters).



Source: Dubern '94 AP dataset

Source: Maruthi et al '17 AP dataset

Figure 4. Risk inference for cassava viruses depending on form of inoculum. Posterior distributions for the epidemic probability parameter for CMB (left) and CBSI (right) when inoculum arrives in the form of an infected plant or insect for different background levels of insect burden, F. Posterior distributions are presented for a single level of insect abundance ($F = 1$, top; $F = 2$, bottom). See Supporting Information S1, S2 and S3 for further details of the model fitting and see Table S1.1 for a list of parameters. Figure data-points were obtained using a combination of the function calls $estimate\_virus\_parameters\_SPT$, $estimate\_virus\_parameters\_PT$ and $calculate\_epidemic\_probability$ functions.

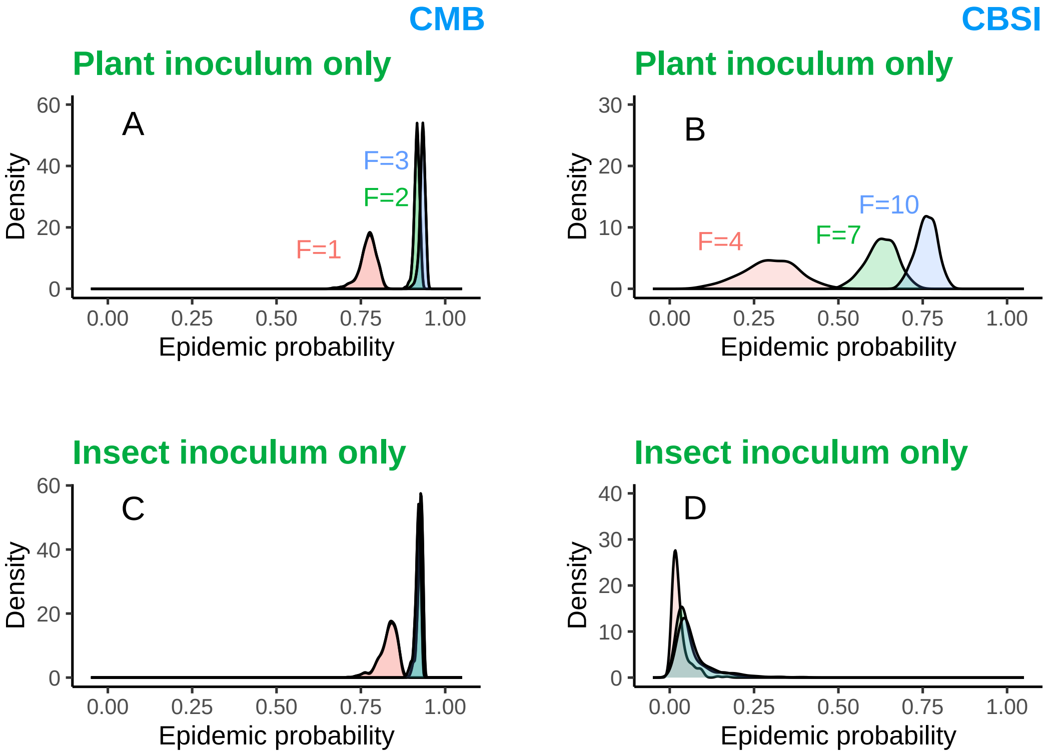


Figure 5. Risk inference for cassava viruses depending on insect burden. Posterior distributions for the epidemic probability parameter for CMB (left) and CBSI (right) for various background levels of insect burden, $F \in 1, 3, 5$. Posterior distributions are presented for inoculum arriving in the form of an infected plant (top) or in the form of an infected insect (bottom). See Supporting Information S1, S2 and S3 for further details of the model fitting and see Table S1.1 for a list of parameters. Figure data-points were obtained using a combination of the function calls *estimate_virus_parameters_SPT*, *estimate_virus_parameters_PT* and *calculate_epidemic_probability* functions.

REFERENCES

Aregbesola OZ, Legg JP, Sigsgaard L, Lund OS, Rapisarda C. (2019) Potential impact of climate change on whiteflies and implications for the spread of vectored viruses. *Journal of Pest Science*, 15;92:381-92.

Aregbesola OZ, Legg JP, Lund OS, Sigsgaard L, Sporleder M, Carhuapoma P, Rapisarda C. (2020) Life history and temperature-dependence of cassava-colonising populations of Bemisia tabaci. *Journal of Pest Science*, 93:1225-41.

Alene, A.D., Abdoulaye, T., Rusike, J., Labarta, R., Creamer, B., Del Río, M., Ceballos, H. and Becerra, L.A. (2013). Identifying crop research priorities based on potential economic and poverty reduction impacts: The case of cassava in Africa, Asia, and Latin America. *PLoS One*, 13(8), p.e0201803.

Bolker BM. (2008) Ecological models and data in R. *Princeton university press.*

Carr JP, Donnelly R, Tungadi T, Murphy AM, Jiang S, Bravo-Cazar A, Yoon JY, Cunniffe NJ, Glover BJ, Gilligan CA. (2018) Viral manipulation of plant stress responses and host interactions with insects. *Advances in Virus Research*, 102:177-97.

Caswell H. 2000 *Matrix population models.* Sunderland, MA: Sinauer.

Chant S.R. (1958) Studies on the transmission of cassava mosaic virus by Bemisia spp.(Aleyrodidae). *Annals of Applied Biology*, 46(2):210-5.

Colvin, J., Omongo, C.A., Govindappa, M.R., Stevenson, P.C., Maruthi, M.N., Gibson, G., *et al.* (2006) Host-plant viral infection effects on arthropod-insect population growth, development and behavior: Management and epidemiological implications. *Advances in Virus Research*, 67, 419-452.

Core team. (2014) *R: A language and environment for statistical computing*, in: R Foundation for Statistical Computing, Vienna, Austria, Available at: http://www.R-project.org/

Donnelly, R., & Gilligan, C.A. (2020) What is pathogen-mediated insect superabundance? *Journal of the Royal Society Interface.*.

Donnelly R, Cunniffe NJ, Carr JP, Gilligan CA. (2019) Pathogenic modification of plants enhances long-distance dispersal of nonpersistently transmitted viruses to new hosts. *Ecology*, 100(7):e02725.

Donnelly R, Sikazwe GW, Gilligan CA. (2020) Estimating epidemiological parameters from experiments in insect access to host plants, the method of matching gradients. *PLoS computational biology*, 16(3):e1007724.

Donnelly R, Gilligan CA. (2023) A new method for the analysis of access period experiments, illustrated with whitefly-borne cassava mosaic begomovirus. *PLoS computational biology*, 19(8):e1011291.

Dubern J. (1994) Transmission of African cassava mosaic geminivirus by the whitefly (Bemisia tabaci). *Tropical Science*, 34(1):82-91.

Eigenbrode, S.D., Bosque-Pérez, N.A., Davis, T.S. (2018). Insect-borne plant pathogens and their insects: ecology, evolution, and complex interactions. *Annual Review of Entomology*, 63 (2018) 169-191.

FAO & IFAD. (2005). A Review of Cassava in Africa with Country Case Studies on Nigeria, Ghana, the United Republic of Tanzania, Uganda and Benin. *Proceedings of the Validation Forum on the Global Cassava Development Strategy*, Vol. 2.

Ferris AC, Stutt RO, Godding D, Gilligan CA. (2020) Computational models to improve surveillance for cassava brown streak disease and minimize yield loss. *PLoS computational biology*, 16(7):e1007823.

Hogenhout, S.A., Ammar, E.D., Whitfield, A.E. and Redinbaugh, M.G. (2008) Insect vector interactions with persistently transmitted viruses. *Annu. Rev. Phytopathol*,46(1), pp.327-359.

Keeling, M.J., Rohani, P. (2011) Modeling infectious diseases in humans and animals. *Princeton university press.*

Legg, J.P., Owor, B., Sseruwagi, P. and Ndunguru, J. (2006). Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Advances in virus research*, 67, pp.355-418.

Leung B, Roura-Pascual N, Bacher S, Heikkilä J, Brotons L, Burgman MA, Dehnen-Schmutz K, Essl F, Hulme PE, Richardson DM, Sol D. 2012 TEASIng apart alien species risk assessments: a framework for best practices. *Ecology Letters*, 15(12), 1475-93.

Macfadyen, S., Tay, W.T., Hulthen, A.D., Paull, C., Kalyebi, A., Jacomb, F., Parry, H., Sseruwagi, P., Seguni, Z., Omongo, C.A. and Kachigamba, D. (2021). Landscape factors and how they influence whitefly pests in cassava fields across East Africa. *Landscape Ecology*, 36, pp.45-67.

Maruthi MN, Jeremiah SC, Mohammed IU, Legg JP. (2016) The role of the whitefly, Bemisia tabaci (Gennadius), and farmer practices in the spread of cassava brown streak ipomoviruses. *J Phytopathol.* 31: 1–11.

Masamha, B., Thebe, V. and Uzokwe, V.N. (2018) Mapping cassava food value chains in Tanzania's smallholder farming sector: The implications of intra-household gender dynamics. *Journal of Rural Studies* 58, pp.82-92.

Mwebaze, P., Macfadyen, S., De Barro, P., Bua, A., Kalyebi, A., Bayiyana, I., Tairo, F. and Colvin, J. (2024). Adoption determinants of improved cassava varieties and intercropping among East and Central African smallholder farmers. *Journal of the Agricultural and Applied Economics Association.*

Reincke, K., Vilvert, E., Fasse, A., Graef, F., Sieber, S. and Lana, M.A. (2018). Key factors influencing food security of smallholder farmers in Tanzania and the role of cassava as a strategic crop. *Food security*, 10, pp.911-924.

Stan Development Team. (2022) *R: A language and environment for statistical computing*, in: R Foundation for Statistical Computing, Vienna, Austria, *RStan: the R interface to Stan*, R package version 2.21.7 Available at: https://mc-stan.org/

World Trade Organization. 1995 *The WTO Agreement on the Application of Sanitary and Phytosanitary Measures (SPS Agreement).* Geneva: World Trade Organization.

Data and Code Availability Statement:

Data and code underlying this manuscript are currently not publicly available in order to comply with journal conditions. They will be made openly available on github upon journal decision.

Supporting Information S1, the estimate_virus_parameters_PT function

In the following sets of Supporting Information, we describe and document the *EpiPV* R package. The first 3 Supporting Information sections are derivations of the probability models used in the main *EpiPV* functions. Subsequent Supporting Information then further describe package functions and datasets and we also include package documentation.

1. In Supporting Information S1 we derive the probability model that is the basis of the *estimate_virus_parameters_PT* function.

2. In Supporting Information S2 we derive the probability model that is the basis of the *estimate_virus_parameters_SPT* function.

3. In Supporting Information S3 we derive the probability model that is the basis of the *calculate_epidemic_probability* function.

4. In Supporting Information S4 we describe the statistical simulation process that underlies the *AP_data_simulator* function.

5. In Supporting Information S5 we provide brief validation of *calculate_epidemic_probability.*

6. In Supporting Information S6 we briefly describe the obligatory structure of AP datasets in the *EpiPv* package (as required in the arguments of estimate virus parameters functions), and we show how to produce this from *AP_data_simulator.*

7. In Supporting Information S7 we include the *EpiPv* package manual and vignettes.

The estimate_virus_parameters_PT function

In access period (AP) experiments, individuals of a cohort of insects may acquire a pathogen when the cohort is provided feeding access to a donor infected plant. When the cohort moves to a healthy test plant, the cohort individuals that acquired the pathogen can now inoculate the test plant. This procedure is followed in a variety of assay sets in order to produce data that reveal the length of access periods required for transmission to occur. We focus on the most common assay structure which is based upon two sets: the acquisition varying and the inoculation varying sub-assays. When the pathogen is question is persistently-transmitted, this is likely to also be accompanied by an assay in which the duration of latent access period on an intermediate plant is varied, i.e., where a latent period is expected for the given pathogen.

We derive a probability model that is tailored to the structure of this set of assays. *When actual access period data are combined with the model using Bayesian analysis, investigators can estimate the following epidemiological rates: α, the rate of acquisition of the pathogen by insects, γ, the rate of progress of insects from virus exposed to virus infectious, μ, the rate that infectiousness is lost, and, β, the rate of pathogen inoculation of plants by insects.* In addition, the mortality rate of insects, b, influences the effective duration of the experiment - and this is also estimated by conditioning the model on insect survival duration.

In this Supporting Information,

- We describe the probability model underlying parameter estimation.
- We demonstrate the accuracy of the parameter estimation.
- We reproduce data and summarise results for two plant virus assays.
- we list the relevant function calls from the package *EpiPv*.

Probability model underlying parameter estimation

Parameter estimation is applied here to the most commonly performed assay within access period experiments - which performs replications of infected plant to healthy plant transfers of insect cohorts for variable durations of the acquisition access period (AAP), and then separately for variable durations of the latent access period (LAP) and the inoculation access period (IAP) (while in each case the alternative periods are held fixed). We will henceforth refer to this trio of assays as the AP experiment. In addition, we denote the three phases of plant transfer by A (acquire), L (latency), I (inoculate).

While there is a trio of sub-assays (as defined by which phase in the sequence of 3 phases is varied) each assay replicate has the same structure based on access durations in each of the 3 phases: $\delta T^A, \delta T^L, \delta T^I$. For clarity we represent the acquisition varying sub-assay therefore as having j subgroups with N replicates each having the phase durations $\delta T^A_j, \delta T^L, \delta T^I$ (and similarly the latent period and inoculation varying sub-assay have the durations $\delta T^A, \delta T^L_j, \delta T^I$ and $\delta T^A, \delta T^L, \delta T^I_j$, respectively). In terms of formulating the model, however, for the moment we simply drop the j subscript and consider the general assay replicate with durations $\delta T^A, \delta T^L, \delta T^I$. Therefore the probability of ultimate test plant infection for a single insect after the durations in each phase $\delta T^A, \delta T^L, \delta T^I$, is denoted by $P(\delta T^A, \delta T^L, \delta T^I)$. The probability can be expressed using integrals that condition the overall probability of test plant infection on the time that acquisition, latent progression, and recovery events occur - denoted by t^A, t^L, t^I,

There is an additional factor that may play a role in these experiments: if insect survival is shorter than the length of the IAP, then it may be important to additionally account for mortality. It is important also to note that in AP experiments, the insect number in the cohort is the number which are placed on the healthy test plants: therefore the AAP and LAP are already conditioned on insect survival, so it is only in the IAP that mortality is a potential factor. Our approach to this is to omit insect survival from the main part of the model - but then to condition the probability of test plant infection on values of IAP duration that are in part determined by discrete levels of insect survival (which depends on the mortality rate). Specifically we assume that $T_I = T_L + min(X, \delta T_I)$ where X is insect survival, so that the effective IAP an insect experiences is bounded period given by the experimenter or the insect's mortality time - whichever comes first. Thus, insect mortality rate is estimated in the model as a nuisance parameter that is unlikely to significantly influence the dynamics, but can be ignored in the initial stages of model derivation.

$$\begin{aligned} p(\delta T^A, \delta T^L, \delta T^I | X) = \int\limits_{t^A=0}^{T^A} \alpha e^{-\alpha t^A} dt^A \Bigg[& \overbrace{\int\limits_{t^L=0}^{T^L-t^A} \gamma e^{-\gamma t^L} dt^L e^{-\mu(T^L - t^{A+L})} \psi(\tau = \delta T^I)}^{\text{early latent progression: full potential IAP}} \\ & + \underbrace{\int\limits_{t^L=T^L-t^A}^{T^I-t^A} \gamma e^{-\gamma t^L} dt^L \psi(\tau = T^I - t^{L+A})}_{\text{late latent progression: reduced potential IAP}} \Bigg] \end{aligned} \tag{S1.1}$$

in which,

$$\psi(\tau) = \int\limits_{t^R=0}^{\tau} \mu e^{-\mu t^R}(1 - e^{-\beta t^R}) dt^R, \tag{S1.2}$$

$$T^I = T^L + min(X, \delta T^I), \tag{S1.3}$$

$$T^L = T^A + \delta T^L,$$

$$T^A = \delta T^A,$$

$$X \sim exp(b) \tag{S1.4}$$

in which $p(\delta T^A, \delta T^L, \delta T^I | X)$ (equation S1.1) is the probability of test plant infection from 1 insect given the insect survival duration X, where $X \sim exp(b)$ (depending on the rate of insect mortality, b). Equation S1.1 calculates the overall probability of test plant infection for a single insect - as determined by the time of the acquisition, latent progression and recovery events and then all possible such times and their probabilities are accounted for through the integrals. We do all this in order to calculate the full probability of inoculation by the insect (final part in parentheses, equation S1.2). See subsection below (sequential events in AP assays) for further details of how these events are constrained in the AP experiment and in equation S1.1.

Finally, then the unconditioned expression for the probability of ultimate test plant infection for a single insect is,

$$\begin{aligned} P(\delta T^A, \delta T^L, \delta T^I) &= \int_{X=0}^{X=LS} p(\delta T^A, \delta T^L, \delta T^I | X) P(X) dX \\ &\approx \sum_j p(\delta T^A, \delta T^L, \delta T^I | X_j) P(X_j) \qquad \text{(S1.5)} \\ &\approx \sum_j p\left(\delta T^A, \delta T^L, \delta T^I \middle| X_j = \frac{T_{j-1} + T_j}{2}\right) \left(e^{-bT_j} - e^{-bT_{j-1}}\right), \qquad \text{(S1.6)} \end{aligned}$$

for $j = 1..N$ such that probability with $X_1 = 0$ and $X_N = LS$, where LS denotes assumed *maximum* insect lifespan. Since we are allowing for a distribution for the parameter b, however, LS simply defines an upper bound for the discretisation. In addition, $P(X_j) = e^{-bT_j} - e^{-bT_{j-1}}$ is probability of insect mortality between time T_{j-1} and T_j, or equivalently, insect survival up until the above time window. Note also that $X_j = (T_{j-1} + T_j)/2$ assigns a survival value that is mid-point with respect to the time window - that is survival is discretised such that the survival at the mid-point of a time window has corresponding probability related to the probability of mortality at any point in the window with mortality assumed to follow an exponential distribution with rate b. Finally, note that at the upper bound $X_N = T_N$, $P(X_N) = 1 - e^{-bT_N}$. In the estimate_virus_parameters_PT function the user supplies LS in days and the number of time points per day tp, so that $N = LS * tp$, and by default $tp = 1$.

So far we have derived the probability of test plant infection from 1 insect vector. We now need to combine this expression with the number of insect vectors in the cohort and the number of experimental replicates to describe the distribution of the data. The first step is to take account of the W_0 insect vectors in each replicate: the operation $1-(1-P)^{W_0}$ transforms the probability of test plant infection from 1 insect into probability of test plant infection from any insect in the cohort. The ensuing quantity is in fact binomial probability when we consider the number of replicates N as corresponding to the number of Bernoulli trials, and therefore the distribution of the data i.e. the probability of test plant infection, TPI, is,

$$TPI \sim Bin\Bigg(N,\ 1-\big(1-p(\delta T^A, \delta T^L, \delta T^I)\big)^{W_0}\Bigg) \tag{S1.7}$$

In summary, equation S1.1 represents the probability of test plant infection from a single insect - as conditioned by the time of insect survival X. Equation S2.34 then represents the unconditioned probability of test plant infection from 1 insect. Equation S2.35 represents the approximation of the unconditioned probability by evaluating the insect survival probabilities for discrete windows rather than integrating over all possible survival values. Equation S1.7 represents the probability of test plant infection for AP experiment replicates.

Sequential events in AP assays

Figure S1.1 is a visual depiction of the events represented in the integral equation S1.1. It can be summarised as follows, acquisition can occur at any time (t^A) between 0 and T^A (left most integral equation S1.1, first part of timeline arrow and blue example event time line #1, Figure S1.1). Latent progression can occur at any time (t^L) between t^A and the end of

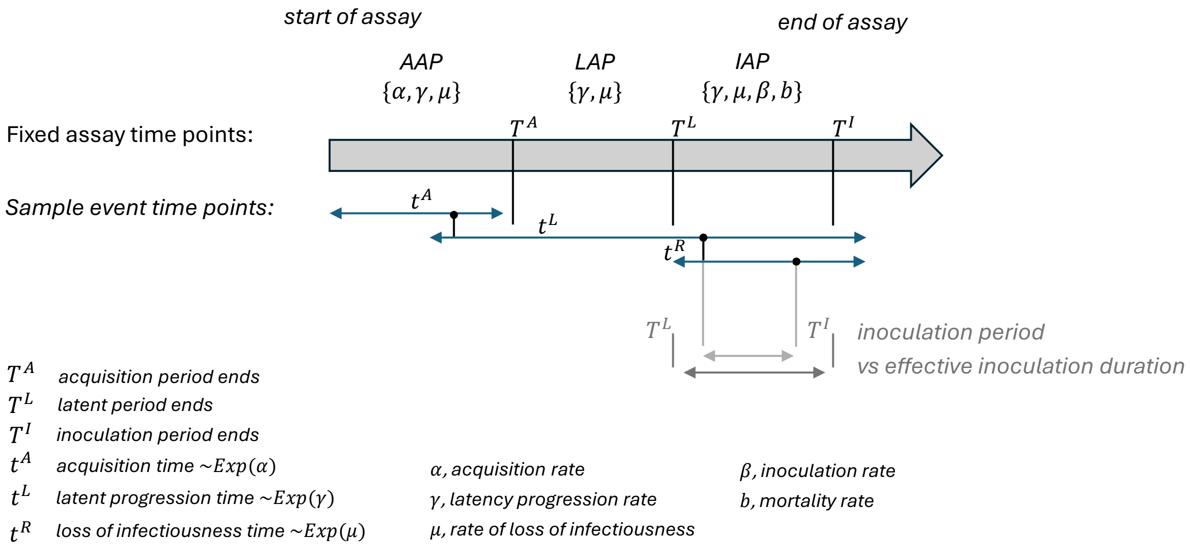

start of assay
end of assay
AAP
{α, γ, μ}
LAP
{γ, μ}
IAP
{γ, μ, β, b}
Fixed assay time points:
T^A
T^L
T^I
Sample event time points:
t^A
t^L
t^R
T^L
T^I
inoculation period
vs effective inoculation duration
T^A acquisition period ends
T^L latent period ends
T^I inoculation period ends
t^A acquisition time ~Exp(α)
t^L latent progression time ~Exp(γ)
t^R loss of infectiousness time ~Exp(μ)
α, acquisition rate
γ, latency progression rate
μ, rate of loss of infectiousness
β, inoculation rate
b, mortality rate

Figure S1.1. Structure of access period assay experiments in relation to the acquisition, inoculation, latency progression and recovery of insect vectors for persistently transmitted plant viruses.

the experiment T^I (up to end of third part of timeline arrow and blue example event time line #2, Figure S1.1). Once an insect passes the latent stage it may lose infectiousness. Loss of infectiousness can occur at any time (t^R) between t^L and the end of the experiment T^I (up to end of third part of timeline arrow and blue example event time line #3, Figure S1.1). In addition, in the AP experiments that we have seen insect mortality is a factor in the IAP phase only - this is because a set number of alive insects are transferred at the end of the acquisition and latent phases while in the inoculation phase it is by no means certain that all insects will be alive at the end of the period and moreover experimenters may define the IAP phase as lasting until the death of all insects. As such mortality in the IAP phase must also be incorporated as we have done here through the probability of insect survival expression $P(X_j)$ (equation S2.34 and IAP event set, Figure S1.1).

Finally, if $t^A + t^L < T^L$ then the potential period for plant inoculation is the full IAP $T^I - T^L$. On the other hand, if $t^A + t^L > T^L$ then the potential period for plant inoculation is the $T^I - T^L - (t^A + t^L - T^L) = T^I - (t^A + t^L)$. These 2 cases are the main expressions in equation S2.34 and the effective inoculation periods are reflected in the respective probability inoculation terms $\psi(T^I - T^L)$ and $\psi(T^I - (t^A + t^L))$ therein.

Solution

The integral equation equation S1.1 has the solution $p(T^A, T^L, T^I)$ (see the additional supplementary document S5 for workings),

$$
\begin{aligned}
&p(T^A, T^L, T^I|X) \\
&= \alpha\beta\frac{\gamma}{\gamma-\mu}\Big(\frac{e^{(\mu+\beta)(T^L-T^I)}-1}{\mu+\beta}\Big)\Bigg(e^{-\gamma T^L}\Big(\frac{e^{(\gamma-\alpha)T^A}-1}{\gamma-\alpha}\Big) - e^{-\mu T^L}\Big(\frac{e^{(\mu-\alpha)T^A}-1}{\mu-\alpha}\Big)\Bigg) \qquad \text{(S1.8)} \\
&\qquad\qquad + \alpha\gamma\frac{\beta}{\mu+\beta}\frac{(e^{(\gamma-\alpha)T^A}-1)}{\gamma-\alpha}e^{-\gamma T^I}\Big(\frac{e^{\gamma(T^I-T^L)}-1}{\gamma} - \frac{e^{(\gamma-(\mu+\beta))(T^I-T^L)}-1}{\gamma-(\mu+\beta)}\Big)\Bigg) \\
&= \alpha\beta F\Big(\frac{\mu}{\gamma}\Big)H(\mu+\beta, T^L-T^I)\Big(e^{-\gamma T^L}H(\gamma-\alpha, T^A) - e^{-\mu T^L}H(\mu-\alpha, T^A)\Big) \qquad \text{(S1.9)} \\
&\qquad\qquad + \frac{\alpha\gamma\beta}{\mu+\beta}e^{-\gamma T^I}H(\gamma-\alpha, T^A)\Big(H(\gamma, T^I-T^L) - H(\gamma-(\mu+\beta), T^I-T^L)\Big)
\end{aligned}
$$

Note that terms such as $(exp((\gamma-\alpha)T^A)-1)/(\gamma-\alpha)$ correspond to a well-known probability distribution (the two parameter hypo-exponential e.g., $hypo(\alpha, \gamma; T^A)$) and as such are equal to values from the unit interval ($[0, 1]$). In Bayesian model-fitting, however, there is a risk that if the sampled points for α and γ are equal then the computational expression will be undefined due to the singularity that would be present. To circumvent this, in the implementation of the probability model, we replace all such terms with their Maclaurin expansion to the nth degree, denoted by H in equation S1.9, e.g.,

$$H(\gamma - \alpha, T^A) = \frac{(e^{(\gamma-\alpha)T^A} - 1)}{\gamma - \alpha} \tag{S1.10}$$

$$\approx T^A \sum_{j=0} \frac{((\gamma - \alpha)T^A)^j}{j!}. \tag{S1.11}$$

In addition, another expression in equation S1.8 represents the convolution of 2 hypo-exponential distributions and in principal could also be expressed in terms of H functions. For simplicity, however, we simply replace the coefficient term $\gamma/(\gamma-\mu)$ with its geometric expansion (equation S1.9) denoted here by F,

$$F\Big(\frac{\mu}{\gamma}\Big) = \frac{1}{1 - \frac{\mu}{\gamma}} \tag{S1.12}$$

$$\approx \sum_{j=0} \Big(\frac{\mu}{\gamma}\Big)^j. \tag{S1.13}$$

Summary

In summary, the statistical model of the AP experiment is given by equations S1.7 and this forms the basis of the R function 'estimate_AP_parameters_PT'. In typical acquisition and inoculation varying sub-assay there is sufficient data to estimate posterior distributions for each $\alpha, \beta, \gamma, \mu$, using Bayesian analysis. We now demonstrate this capability using first simulated data where the task is to recover the original parameters from which the simulated data was produced. We then present results for CMB based upon AP experimental data. Finally, insect mortality rate b is also estimated for completeness - we set an uninformative uniform prior for mortality rate, between $0.1h$ and D_{LS}, where D_{LS} can be provided by the user but otherwise defaults to $50d$. In fact, D_{LS} merely helps to structure the discretisation of insect survival and is not expected to impact the outcome. For *B. tabaci*, for example, we used the natural survival for D_{LS}, since laboratory survival is expected to be far less, so it is therefore a reasonable upper bound. Note that in our experience b, has little influence on AP dynamics, and is effectively a nuisance parameter that is estimated in the modelling for thoroughness - the prior therefore is uninformative on a reasonable interval and in the model-fitting exercise it tends to remain uninformative.

A note on the latent period varying sub-assay in Dubern (1994)

The latent period varying sub-assay was conducted with *Physalis alkekinge* (Chinese lantern) as the intermediate plant (LAP) rather than the cassava host plant which was used in all other parts of the assay (AAP and IAP) in Dubern (1994). It is evident, however, on close inspection of the AP data that the intermediate plant has exerted an influence on whitefly behavior. To see this, we compare the acquisition varying sub-assay (values of AAP ranging from 2-8h followed by IAP on healthy cassava until insect death) and the latent period varying sub-assay (values of LAP ranging from 0.5-8h, preceded by 5h AAP and followed by IAP on healthy cassava until insect death). As such, it is clear that in the limit of LAP approaching 0 hours (latent period varying sub-assay) the assay becomes identical to that of 5h AAP in the acquisition varying sub-assay. Yet, while in the acquisition varying sub-assay when the AAP was 5h there was already 16/30 test plants that became infected, nevertheless in the latent period varying sub-assay no infection of test plants at all were recorded until there was at least 4h LAP (i.e., 0/12 test plants became infected for 0.5 and 1h LAP, 0/22 for 2 and 3 h LAP - and then 3/34 became infected for 4h LAP rising to 23/34 infected after 8h LAP). In other words, the intermediate plant in the latent period varying sub-assay has exerted an influence on insect behavior. For this reason, we omitted the latent period varying sub-assay from our analyses of the Dubern (1994) CMB AP dataset. Future experimenters should bear in mind that LAP on intermediate plants that are different from the host plant, may exert unexpected transient effects on insect behavior.

A)	**Notation**	***Parameters & principal variables***	***Units/labelling***
	S	susceptible plant	*plant type*
	E	latent-infected plant	*plant type*
	I	infectious plant	*plant type*
	α	pathogen acquisition rate	*rate* h^{-1}
	β	pathogen inoculation rate	*rate* h^{-1}
	μ	rate of loss of insect infectiousness	*rate* h^{-1}
	$P(X_F^j)$	extinction prob plant type X_F^j	*probability*
	X_F^j	plant type X with j from F infected insects	*inoculum state*
B)		***Laboratory model only***	
	W_0	# insects in cohort	*integer*
	n	# experimental replicates	*integer*
	γ	rate of onset of insect infectiousness	*rate* h^{-1}
C)		***Field model only***	
	F	#insect vectors per plant	*integer*
	$\nu = 1/14$	rate of onset of plant infectiousness[a]	*rate* d^{-1}
	r=1/21	rate of infected plant removal[b]	*rate* d^{-1}
	$\theta = 0.45$	rate of insect dispersal[c]	*rate* d^{-1}
	$b = 1/21$	rate of insect mortality[d]	*rate* d^{-1}
	$h = 1/365$	harvesting rate[e]	*rate* d^{-1}

Table S1.1. Parameter definitions for the laboratory and field models. C are representative choices of local parameters for *calculate_epidemic_probability* and $\theta = 0.45$ is the median value from the posterior dispersal distribution in Ferris et al. (2020) – but note that all values to some extent will vary with locality. Note that all results in the main text that relate to epidemic probability use these values. Note also that whatever the appropriate choice of values for these parameters these should not differ for CMB and CBSI except for the rate of onset of plant infectiousness (ν), which is likely to be similar for CMB and CBSI but depends on the cultivars in question.

Parameter estimation from simulated and experimental datasets

In this section of the Supporting Information we present AP data for a PT virus that was simulated with the AP_data_simulator (Supporting Information S4) and the results of analysing that data with the estimate_virus_parameters_PT function (table S1.3). The benefit of analysing data is that the estimated parameters can be compared with the original parameters (see original, median and credible interval columns of S1.3 B). In addition we present the published AP data and the parameter estimates (from analysing the data with the estimate_virus_parameters_PT function) for the PT CMB virus in table S1.4.

A) Model 1	***Parameter to be fitted***	***Prior distribution***
α	acquisition rate per hour	$\sim half-normal(0,1)$
β	inoculation rate per hour	$\sim half-normal(0,1)$
γ	latency progression rate per hour	$\sim half-normal(0,1)$
μ/γ	rate per hour of loss of infectiousness (μ)	$\sim beta(1,5)$
B)	***Data variables***	
$nReps_j^A$	# experimental reps j^{th} AAP	assay data
$nReps_j^L$	# experimental reps j^{th} LAP	assay data
$nReps_j^I$	# experimental reps j^{th} IAP	assay data
δT_j^A	access period length j^{th} AAP	assay data
δT_j^L	access period length j^{th} LAP	assay data
δT_j^I	access period length j^{th} IAP	assay data
TPI_j^A	# infected test plants j^{th} AAP	assay data
TPI_j^L	# infected test plants j^{th} LAP	assay data
TPI_j^I	# infected test plants j^{th} IAP	assay data

Table S1.2. Parameter definitions, and prior distributions, for the model-based Bayesian analysis. Derived parameters are combinations of fitted parameters in A). All prior distributions were chosen to be non-informative - smooth model-fitting was aided by estimating the ratio parameter μ/γ with a beta prior as is appropriate for the ratio of rate events - and with prior parameterisation ($beta(1,5)$) to reflect the magnitude difference expected for these parameters for PT viruses ($1/\mu$ is expected to be on the order of days and $1/\gamma$ on the order of hours - see Table 2 of Hogenhout et al. (2008)

Simulated AP dataset, PT virus

A)	*AAP length*	*2h*	*3h*	*3.5h*	*4h*	*4.5h*	*5h*	*6h*	*8h*	
i)	*no. reps*	30	30	30	30	30	30	30	30	
ii)	*test plant infections*	0	0	9	13	11	16	21	16	
	LAP length	*0.5h*	*1h*	*2h*	*3h*	*4h*	*5h*	*6h*	*7h*	*8h*
iii)	*no. reps*	30	30	30	30	30	30	30	30	30
iv)	*test plant infections*	30	30	30	30	30	30	30	30	30
	IAP length	*5m*	*10m*	*15m*	*20m*	*25m*	*30m*	*40m*	*50m*	*60m*
v)	*no. reps*	30	30	30	30	30	30	30	30	30
vi)	*test plant infections*	13	17	17	23	27	28	30	30	30

B			*mean*	*2.5%*	*97.5%*	*original simulation values*
i)	*acquisition*	α	$0.139h^{-1}$	$0.098h^{-1}$	$0.207h^{-1}$	$0.1h^{-1}$
ii)	*latency*	γ	$0.652h^{-1}$	$0.169h^{-1}$	$1.949h^{-1}$	$0.5h^{-1}$
iii)	*inoculation*	β	$0.882h^{-1}$	$0.547h^{-1}$	$1.464h^{-1}$	$1.0h^{-1}$
iii)	*duration of infectiousness*	μ	$0.044h^{-1}$	$0.003h^{-1}$	$0.099h^{-1}$	$0.01h^{-1}$

Table S1.3. Simulated PT virus AP data and analysis. A: Simulated acquisition varying sub-assay (A), latent period varying sub-assay (B) and inoculation varying sub-assay (C) assay results for a representative plant virus. Simulations consisted of sampling random times of acquisition, progression through latency and loss of pathogen for individual whitefly from exponential distributions with intensity values based on the 'original' parameter values given in B. B: Parameter estimates for insect-borne transmission resulting from analysis of the simulated access period assay data in A using the *estimate_virus_parameters_PT()* function. Posterior parameter distributions were obtained using Hamiltonian Monte Carlo, RStan version 2.21.0 (Stan Development Team, 2022), R version 3.6.3 (Core team, 2014). Additional rStan settings were: $warmup = 1000$, $chains = 4$, $iterations = 2000$, $max(treedepth) = 10$, $adapt\_delta = 0.95$).

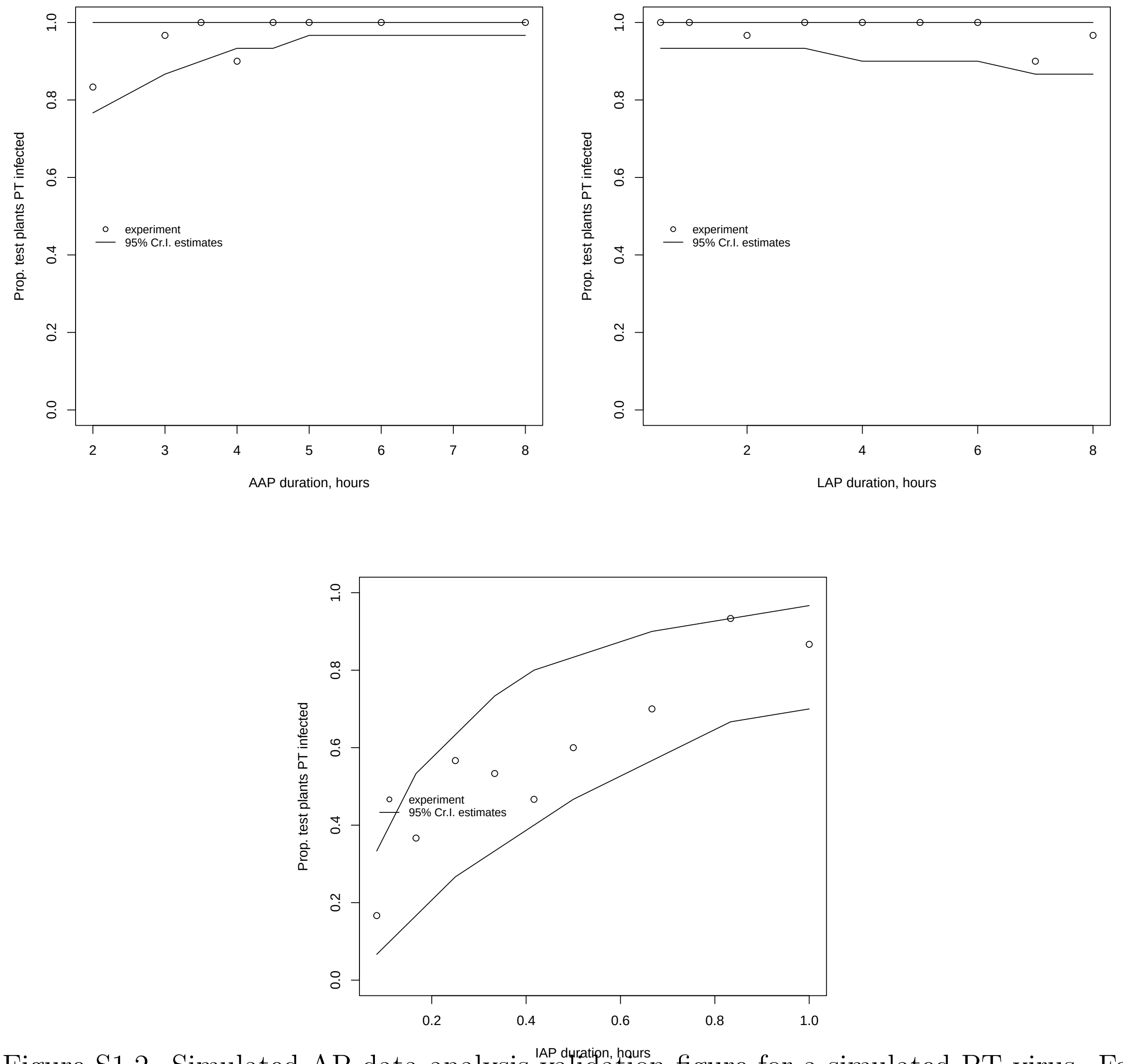


Figure S1.2. Simulated AP data analysis validation figure for a simulated PT virus. Forward simulation of the access period dataset (Table S1.3 A) using the estimate parameters (Table S1.3 B). We show curves for the 95% forward simulated credible interval and we superimpose the orginal simulated dataset (empty circles). This process is repeated for each of the three sub-assays in the experiment (acquisition varying sub-assay, latent access varying sub-assay and inoculation varying sub-assay).

Empirical AP dataset, CMB

A)	*AAP length*	*2h*	*3h*	*3.5h*	*4h*	*4.5h*	*5h*	*6h*	*8h*	
i)	*test plant infections*	0	0	9	13	11	16	21	16	
ii)	*no. reps*	30	30	30	30	30	30	30	30	
	IAP length	*5m*	*10m*	*15m*	*20m*	*25m*	*30m*	*40m*	*50m*	*60m*
iii)	*test plant infections*	0	8	11	10	16	17	19	8	9
iv)	*no. reps*	12	36	36	36	36	24	24	12	12

B			*mean*	*2.5%*	*97.5%*
i)	*acquisition*	α	$0.014h^{-1}$	$0.012h^{-1}$	$0.016h^{-1}$
ii)	*inoculation*	β	$3.01h^{-1}$	$2.219h^{-1}$	$3.995h^{-1}$
iii)	*latency*	γ	$0.85h^{-1}$	$0.317h^{-1}$	$2.053h^{-1}$
iii)	*duration of infectiousness*	μ	$0.005h^{-1}$	$0.0001h^{-1}$	$0.019h^{-1}$

Table S1.4. CMB AP data and analysis. A: The acquisition varying sub-assay (A), and inoculation varying sub-assay period (B) data from Dubern (1994), in which *B. tabaci* transmission of CMB in cassava was studied. In the acquisition varying sub-assay insect cohorts consisting of 10 insects were moved from infected source plants (for variable duration) to healthy test plants (for the remainder of their lives) with no intermediate plant supplied. In the inoculation varying sub-assay insect cohorts consisting of 10 insects were moved from infected source plants (5h feeding access), to healthy test plants (for variable duration) via a feeding period of 6h on an uninfected intermediate plant. B: CMB parameter estimates for insect-borne transmission generated from the access period assay data in A using the *estimate_virus_parameters_SPT()* function. Posterior parameter distributions were obtained using Hamiltonian Monte Carlo, RStan version 2.21.0 (Stan Development Team, 2022), R version 3.6.3 (Core team, 2014). Additional rStan settings were: $warmup = 1000$, $chains = 4$, $iterations = 2000$, $max(treedepth) = 10$, $adapt\_delta = 0.95$.

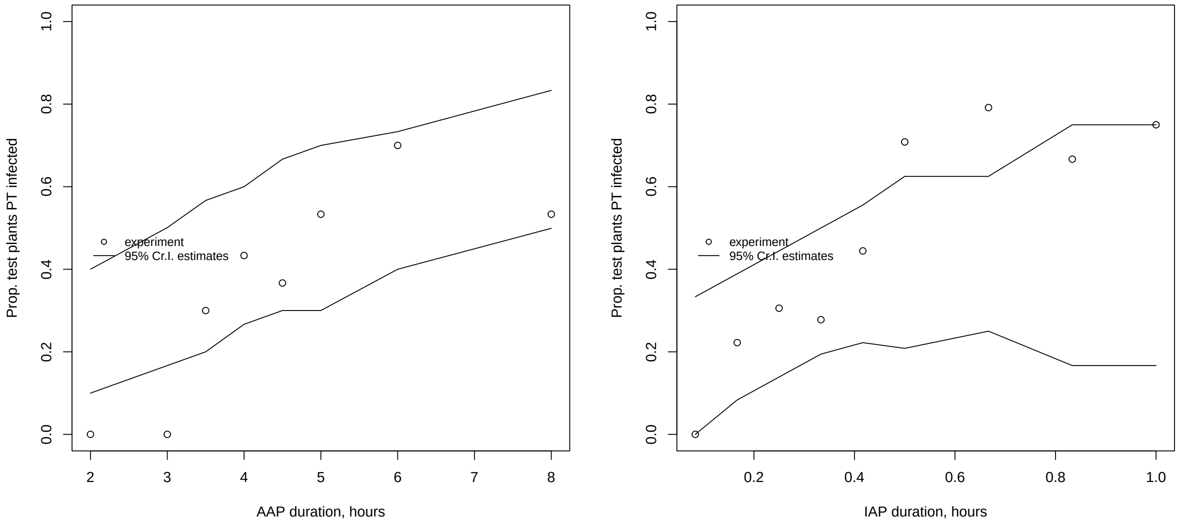


Figure S1.3. Simulated AP data analysis validation figure for the CMB virus. Forward simulation of the CMB access period dataset (Table S1.4 A) using the estimated parameters (Table S1.5 B). We show curves for the 95% forward simulated credible interval and we superimpose the orginal CMB dataset (empty circles). This process is repeated for the two sub-assays (acquisition varying sub-assay and inoculation varying sub-assay).

Supporting Information S2, the *estimate_virus_parameters_SPT* function

Probability model underlying parameter estimation

For the SPT variant of the assay we take a different approach to deriving the probability model. This is because the probability model has already been derived in Donnelly and Gilligan (2020) where it was used to produce point estimates (the method of matching gradients). In this work, in contrast, the probability model is used in a Bayesian analysis to estimate posterior parameter distributions. In what follows we repeat the steps taken in the derivation (Supporting Information S1-S2 in Donnelly et al. (2020); note that several typos are corrected here), and we add additional steps relating to parameter estimation. In addition, note that for SPT viruses latent periods are assumed to be insignificant - see Table 2 of Hogenhout et al. (2008) - and therefore SPTs assays are not expected to feature LAPs nor latent period varying sub-assays, and accordingly in our modelling we do not attempt to estimate a rate of progression through latency (γ).

Note, as per Supporting Information S1, insect mortality is an additional factor that may play a role in these experiments. We omit insect survival from the main model derivation - but then condition the final probability model on discrete levels of insect survival, i.e., for comprehensiveness we estimate insect mortality rate as a nuisance parameter that is unlikely to significantly influence the dynamics.

Model of acquisition access period

The equations governing the feeding dynamics of insect vectors in an acquisition access period (AAP) are,

Joint probability dynamics of virus-free and virus-bearing vectors $P_{X,Y}(t+\delta t)$

$$= P_{X,Y}(t) + \tag{S2.1}$$

$$\left(\overset{\text{Virus loss}}{\underbrace{\mu(Y+1)P_{X-1,Y+1}(t) - \mu Y P_{X,Y}(t)}} + \overset{\text{Acquisition loss}}{\underbrace{\alpha(X+1)P_{X+1,Y-1}(t) - \alpha X P_{X,Y}(t)}} \right)\delta t,$$

where X and Y represent the number of virus-free and virus-bearing insect vectors. System S2.1 has the initial condition: $X(0) = X_0, Y(0) = 0$ (i.e., $P_{X_0,0}(0) = 1$). Insects acquire the virus at rate α per hour and lose the virus at per capita rate μ per hour (as per main text).

System S2.1 can be rewritten as a partial differential equation (PDE) in which the dependent variable is the probability generating function, denoted $g(z_1, z_2, t)$, of the variables X and Y from system S2.1, i.e., $g(z_1, z_2, t) = \sum\limits_{X,Y} z_1^X z_2^Y P_{X,Y}(t)$. The strategy is to solve the PDE and hence to recover the distribution of either population variable by manipulating the system's probability generating function. Multiplying both the left hand side and right hand terms of the stochastic process in system S2.1 by $\sum\limits_{X,Y} z_1^X z_2^Y$ and rearranging, produces the PDE:

$$\frac{\partial g}{\partial t} = \frac{\partial g}{\partial z_1}(\alpha z_2 - \alpha)z_1 + \frac{\partial g}{\partial z_2}(\mu z_1 - \mu z_2) \tag{S2.2}$$

The PDE in equation S2.2 is linear and can be solved along characteristic curves (curves on which the solution $g(z_1, z_2, t)$ is constant). This involves forming a linear system of ODEs from the PDE. They are given by,

$$\begin{aligned}\frac{dz_1}{dt} &= -(\alpha z_2 - \alpha z_1)\\ \frac{dz_2}{dt} &= -(\mu z_1 - \mu z_2)\end{aligned} \tag{S2.3}$$

This linear ODE system can be solved by first solving it as a homogeneous system with coefficient matrix,

$$A = \begin{pmatrix} \alpha & -\alpha \\ -\mu & \mu \end{pmatrix}$$

which has the eigenvalues: $\lambda_1 = 0$, $\lambda_2 = \alpha + \mu$. The problem is homogeneous and the homogeneous solution is a linear combination based on the e-values, i.e.,

$$\hat{z}(t) = c_1 \hat{v}_1 e^{\lambda_1 t} + d_1 \hat{v}_2 e^{\lambda_2 t}, \tag{S2.4}$$

where the vectors $\hat{v}_1$ and $\hat{v}_2$ are the eigen-vectors for the corresponding eigen-values. Calculating the eigen-vectors leads to $\hat{v}_1 = (1 \;\; 1)^T$ and $\hat{v}_2 = (-\alpha \;\; \mu)^T$. Which can be written as,

$$\begin{aligned} z_1(t) &= c_1 - c_2 \alpha e^{(\alpha+\mu)t}\\ z_2(t) &= c_1 + c_2 \mu e^{(\alpha+\mu)t}. \end{aligned} \tag{S2.5}$$

Then letting $z_1(0) = z_1^0$ and $z_2(0) = z_2^0$ we find that $c_1 = z_1^0 + (\alpha/(\alpha+\mu))(z_2^0 - z_1^0)$ and $c_2 = (z_2^0 - z_1^0)/(\alpha+\mu))$. Then using the relation,

$$z_1^0 = (z_1(t) - z_2(t))e^{-(\alpha+\mu)t} + z_2^0, \tag{S2.6}$$

we can express the solutions with the constant terms on the left hand side in accordance with the method of characteristics,

$$z_1^0 = (z_1(t) - z_2(t))e^{-(\alpha+\mu)t}(\frac{\alpha}{\mu+\alpha} + \frac{\mu}{\mu+\alpha}e^{-(\alpha+\mu)t}) + z_2(t) \tag{S2.7}$$

$$z_2^0 = z_2(t) - \frac{\mu}{\mu+\alpha}(z_1(t) - z_2(t))(e^{-(\alpha+\mu)t} - 1) \tag{S2.8}$$

By the method of characteristics (which finds solutions of the PDE along curves where solutions are constant), the PDE solution, and hence the generating function, is some function of the right hand sides, i.e.,

$$g(z_1, z_2, t) = F(z_0^1, z_0^2) \tag{S2.9}$$

All that remains is to find the function F and this is achieved by using the generating function's initial condition, i.e. $g(z_1, z_2, 0) = z_1^{X_0}$ (to see this note that at $t = 0$: $X = X_0$ and $Y = 0$) so that,

$$F(z_0^1, z_0^2)\Big|_{t=0} = z_1^{X_0}\Big|_{t=0} \tag{S2.10}$$

$$\implies g(z_1, z_2, t) = \Big((z_1(t) - z_2(t))e^{-(\alpha+\mu)t}(\frac{\alpha}{\mu+\alpha} + \frac{\mu}{\mu+\alpha}e^{-(\alpha+\mu)t}) + (z_2(t) - 1) + 1\Big)^{X_0}. \tag{S2.11}$$

Next we recall that we are interested chiefly in Y (i.e., the number of virus-bearing insects), and hence we reduce equation S2.11 to a generating function in Y only (that is marginalising over the number without virus, X), i.e.,

$$\begin{aligned} G(z,t) = g(1, z_2, t) &= \Big(1 + (z(t) - 1 + \frac{\mu}{\mu+\alpha}(1 - z(t)) + \frac{\alpha}{\mu+\alpha}(1 - z(t))e^{-(\mu+\alpha)t}\Big)^{X_0} \\ &= \Big(1 + (z(t) - 1)(\frac{\alpha}{\mu+\alpha})(1 - e^{-(\mu+\alpha)t})\Big)^{X_0} \end{aligned} \tag{S2.12}$$

Finally, taking the k^{th} derivative over z of $G(z,t)$ and evaluating at $z = 0$ (which produces the probability that $Y = k$), we see that the underlying distribution for the variable Y after t hours of acquisition access has the binomial form,

$$\begin{aligned} \frac{dG^{(k)}(0,t)}{dz} &= \frac{X_0(X_0-1)...(X_0-k+1)}{k!}p(t)^k(1-p(t))^{x_0-k} \\ &= \sum_k \binom{X_0}{k} p(t)^k(1-p(t))^{X_0-k} \end{aligned} \tag{S2.13}$$

where $p(t) = \frac{\alpha}{\alpha+\mu}(1 - e^{-(\alpha+\mu)t})$ and $1 - p(t)$ equal to the term within the large parantheses in equation S2.12.

Model of inoculation access period

The equations describing vector dynamics in an inoculation access period (IAP) are governed by,

Joint dynamics of infected plants and virus-bearing vectors $Q_{M,N}(t+\delta t) = Q_{M,N}(t)$

$$+\left(\overbrace{\mu(N+1)Q_{M,N+1}(t) - \mu N Q_{M,N}(t)}^{\text{Virus loss}} + \overbrace{\beta N Q_{M-1,N}(t) - \beta N Q_{M,N}(t)}^{\text{Inoculation}}\right)\delta t \tag{S2.14}$$

with initial conditions: $N(0) = 0, M(0) = y_0$ (i.e., $Q_{0,y_0}(0) = 1$). The system can be rewritten as a partial differential equation (PDE) in which the dependent variable is the probability generating function, denoted $w(s_1, s_2, t)$, of the variables M and N from equation S2.14, i.e., $w(s_1, s_2, t) = \sum\limits_{M,N} s_1^M s_2^N Q_{M,N}(t)$. Multiplying both the left hand side and right hand terms of the process in equation S2.14 by $\sum\limits_{M,N} s_1^M s_2^N$ and rearranging, produces the PDE,

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial s_1}0 + \frac{\partial w}{\partial s_2}(\mu(1-s_2) - \beta s_2(1-s_1)) \tag{S2.15}$$

The PDE given by equation S2.15 is linear and can be solved along characteristic curves (curves on which the solution $w(s_1, s_2, t)$ is constant). This involves forming a linear system of ODEs from the PDE. They are given by,

$$\begin{aligned}\frac{ds_1}{dt} &= 0\\ \frac{ds_2}{dt} &= (\beta s_2(1-s_1)) - (\mu(1-s_2)\end{aligned} \tag{S2.16}$$

Thus s_1 is constant with respect to time - henceforth $s_1 = \sigma$ for clarity, and s_2 is governed by a linear ODE which can be solved for homogeneous and inhomogeneous parts, and letting $s_2(0) = s_0^2$, leads to,

$$s_0^2 = s_2 e^{-(\beta(1-\sigma)+\mu)t} + \frac{\mu}{\beta(1-\sigma)+\mu}(1 - e^{-(\beta(1-\sigma)+\mu)t}) \tag{S2.17}$$

and hence,

$$\begin{aligned} w(s_1, s_2, t) &= H(s_0^1, s_0^2) \\ &= H(s_1, s_2 e^{-(\beta(1-\sigma)+\mu)t} + \frac{\mu}{\beta(1-\sigma)+\mu}(1 - e^{-(\beta(1-\sigma)+\mu)t})) \end{aligned} \tag{S2.18}$$

All that remains is to find the function H and this is achieved by using the generating function's initial condition, i.e. $w(s_1, s_2, 0) = s_2^{y_0}$ so that,

$$H(s_0^1, s_0^2)\Big|_{t=0} = s_2^{y_0}\Big|_{t=0} \tag{S2.19}$$

$$\Leftrightarrow H(s_0^1, s_0^2) = s_0^{2y_0} \tag{S2.20}$$

leading to,

$$w(s_1 = \sigma, s_2, t) = \left(s_2 e^{-(\beta(1-\sigma)+\mu)t} + \frac{\mu}{\beta(1-\sigma)+\mu}(1 - e^{-(\beta(1-\sigma)+\mu)t})\right)^{y_0} \tag{S2.21}$$

Next we recall that we are interested chiefly in N (i.e., the number of plant inoculations), and hence we reduce equation S2.21 to a generating function in N only, i.e.,

$$W(s,t) = w(s_1, 1, t) = \left(\frac{\mu + \beta(1-s_1)e^{-\beta(1-s_1)+\mu)t}}{\beta(1-s_1)+\mu} \right)^{y_0}. \tag{S2.22}$$

since $w(1, s_2, t) = \sum_{M,N} 1^M s_2^N P_{M,N}(t) = w(s_2, t)$ by the definition of generating functions. Since we are interested in the probability of plant infection, denoted $S(t)$, we can finally convert equation S2.22 to a simpler form by calculating the probability that $N \geq 1$. This leads to,

$$S(t) = 1 - W(0,t) = 1 - \left(\frac{\mu + \beta e^{-(\beta+\mu)t}}{\beta + \mu} \right)^{y_0}. \tag{S2.23}$$

Parameter estimation from the combined access period model

The preceding models are combined by noting that equation S2.23 has the exponent y_0 which is the number of infected insects at the end of the AAP (equation S2.13). Together they correspond to the probability of test plant infection, P_{TPI}, at the end of the IAP,

$$P_{TPI} = \sum_k \binom{X_0}{k} q(\delta T_A)^k (1 - q(\delta T_A))^{X_0 - k} \left(1 - \left(\frac{\mu + \beta e^{-(\beta+\mu)\delta T_I}}{\beta + \mu} \right)^k \right) \tag{S2.24}$$

where $q(t) = \frac{\alpha}{\alpha+\mu}(e^{-bt} - e^{-(\alpha+\mu)t})$ and where δT_A and δT_I are the acquisition and inoculation durations respectively.

Using binomial expansion this can be written as,

$$\begin{aligned} P_{TPI} &= 1 - (1 - \frac{\alpha}{\alpha+\mu}\frac{\beta}{\beta+\mu}(1 - e^{-(\alpha+\mu)\delta T_A})(1 - e^{-(\beta+\mu)\delta T_I}))^{X_0} && \text{(S2.25)} \\ &= 1 - (1 - c_1(1 - e^{-c_2 \delta T_A})(1 - e^{-c_3 \delta T_I}))^{X_0}, && \text{(S2.26)} \end{aligned}$$

where $c_1 = \frac{\alpha}{\alpha+\mu}\frac{\beta}{\beta+\mu}$, $c_2 = \alpha + \mu$ and $c_3 = \beta + \mu$.

Finally,

$$TPI_j^A \sim Bin(nReps_j^A, 1 - (1 - c_1(1 - e^{-c_2 \delta T_A})(1 - e^{-c_3 \delta T_I}))^{X_0}) \tag{S2.27}$$

$$TPI_j^I \sim Bin(nReps_j^I, 1 - (1 - c_1(1 - e^{-c_2 \delta T_A})(1 - e^{-c_3 \delta T_I}))^{X_0}). \tag{S2.28}$$

where δT^A and δT^I denote the fixed acquisition and inoculation access duration portions of the inoculation and acquisition access assays respectively. In addition, δT_j^A and δT_j^I denote the particular durations from the portions of the inoculation and acquisition access assays with variable duration.

Note that in practice the access period data may often specify X_0 as the number of insects at the start of the IAP rather than at the start of the AAP (as was done for the CBSI dataset of Maruthi et al. (2020)). This means that insect mortality is only relevant during the IAP (effectively the acquisition varying sub-assay data already conditions on insect survival because only living insects are taken forward to form the X_0 insects at the start of the IAP). This appears to be the most common situation and therefore, henceforth, for simplicity, taking a similar approach to Supporting Information S1, we set $b = 0$ throughout the model (zero insect mortality) - but we set $\delta T_I^{eff} = min(\delta T_I, X)$ (effective duration of the IAP for a given insect is the smaller of the length of the IAP and insect survival, X). As with Supporting Information S1, we condition the probability of test plant infection for a single insect on the discretised length of survival, $X(b)_j$. This situation corresponds to,

$$\begin{aligned}
TPI_j^A &\sim Bin\left(nReps_j^A, 1-\left(1-P(\delta T_{Aj}, \delta T_I^{eff})\right)^{X_0}\right) && \text{(S2.29)}\\
TPI_j^I &\sim Bin\left(nReps_j^I, 1-\left(1-P(\delta T_A, \delta T_{I_j}^{eff})\right)^{X_0}\right), && \text{(S2.30)}\\
p(\delta T_A, \delta T_I^{eff}|X) &= c_1(1-e^{-c_2\delta T_A})(1-e^{-c_3\delta T_I}) && \text{(S2.31)}\\
\delta T_I^{eff} &= min(\delta T_I, X) && \text{(S2.32)}\\
X &\sim exp(b). && \text{(S2.33)}
\end{aligned}$$

where $c_1 = \frac{\alpha}{\alpha+\mu}\frac{\beta}{\beta+\mu}$, $c_2 = \alpha+\mu$ and $c_3 = \beta+\mu$. Finally, then the unconditioned expression for the probability of ultimate test plant infection for a single insect is,

$$\begin{aligned}
P(\delta T^A, \delta T^I) &= \int_{X=0}^{X=LS} p(\delta T^A, \delta T^I|X)P(X)dX\\
&\approx \sum_j p(\delta T^A, \delta T^I|X_j)P(X_j) && \text{(S2.34)}\\
&\approx \sum_j p\left(\delta T^A, \delta T^I \middle| X_j = \frac{T_{j-1}+T_j}{2}\right)\left(e^{-bT_j} - e^{-bT_{j-1}}\right), && \text{(S2.35)}
\end{aligned}$$

for $j = 1..N$ such that probability with $X_1 = 0$ and $X_N = LS$, where LS denotes assumed *maximum* insect lifespan. Since we are allowing for a distribution for the parameter b, however, LS simply defines an upper bound for the discretisation. In addition, $P(X_j) = e^{-bT_j} - e^{-bT_{j-1}}$ is probability of insect mortality between time T_{j-1} and T_j, or equivalently, insect survival up until the above time window. Note also that $X_j = (T_{j-1}+T_j)/2$ assigns a survival value that is mid-point with respect to the time window - that is survival is discretised such that the survival at the mid-point of a time window has corresponding probability related to the probability of mortality at any point in the window with mortality assumed to follow an exponential distribution with rate b. Finally, note that at the upper bound $X_N = T_N$, $P(X_N) = 1 - e^{-bT_N}$. In the estimate_virus_parameters_SPT function the user supplies LS in days and the number of time points per day tp, so that $N = LS * tp$, and by default $tp = 1$.

Parameter estimation from simulated and experimental datasets

Firstly, note that we can unpack the estimated composite parameters $c_1,\ c_2,\ c_3$ (equations S2.29-S2.33) as follows. Since $c_1 = \frac{\alpha}{\alpha+\mu}\frac{\beta}{\beta+\mu}$, $c_2 = \alpha + \mu$ and $c_3 = \beta + \mu$,

$$c_2 - c_3 = XX = \alpha - \beta \tag{S2.36}$$

$$c_1 c_2 c_2 = YY = \alpha\beta \tag{S2.37}$$

$$\implies \beta^2 + XX\beta - YY = 0 \tag{S2.38}$$

$$\implies \beta_+ = \frac{-XX + \sqrt{XX^2 + 4YY}}{2}, \tag{S2.39}$$

where it is straightforward to see that there will always be a negative and a positive root β. Therefore, proceeding with the positive root only (β_+), it then follows that,

$$\alpha = XX + beta_+ \tag{S2.40}$$

$$\mu = c_3 - beta_+, \tag{S2.41}$$

and posterior parameter distributions can accordingly be produced using MCMC for α, β_+ and μ from the main estimated composite parameters c_1, c_2 and c_3 (this is achieved in rstan using the generated quantities block of the stan file).

In this Supporting Information section we present simulated AP data and parameter estimates that emerge from analysing the simulated data with estimate_virus_parameters_SPT function (Table S2.2 A-B). We also present the published empirical AP data and the parameter estimates (from analysing the data with the estimate_virus_parameters_SPT function) for the SPT CBSI virus in tables S2.3. For comparison we show the point estimates for CBSI reported in Donnelly et al. (2020) (tables S2.3 'math estimate' column). In general the correspondence is good, but that the rate of virus acquisition is somewhat overestimated in Donnelly et al. (2020). Finally, as per Supporting Information 1, insect mortality rate b is also estimated for completeness - we set an uninformative uniform prior for mortality rate, between $0.1h$ and D_{LS}, where D_{LS} can be provided by the user but otherwise defaults to $50d$. In fact, D_{LS} merely helps to structure the discretisation of insect survival and is not expected to impact the outcome. Note that we find that b, has little influence on AP dynamics, and is effectively a nuisance parameter that is estimated in the modelling for thoroughness.

Simulated test dataset

A) Model 1	***Parameter to be fitted***	***Prior distribution***
$c_1 = \frac{\alpha}{\alpha+\mu}\frac{\beta}{\beta+\mu}$	composite parameter	$\sim half-normal(0,1)$
$c_2 = \alpha + \mu$	composite parameter	$\sim half-normal(0,1)$
$c_3 = \beta + \mu$	composite parameter	$\sim half-normal(0,1)$
B)	***Data variables***	
$nReps_j^A$	# experimental reps j^{th} AAP	assay data
$nReps_j^L$	# experimental reps j^{th} LAP	assay data
$nReps_j^I$	# experimental reps j^{th} IAP	assay data
δT_j^A	access period length j^{th} AAP	assay data
δT_j^L	access period length j^{th} LAP	assay data
δT_j^I	access period length j^{th} IAP	assay data
TPI_j^A	# infected test plants j^{th} AAP	assay data
TPI_j^L	# infected test plants j^{th} LAP	assay data
TPI_j^I	# infected test plants j^{th} IAP	assay data

Table S2.1. Parameter definitions, and prior distributions, for the model-based Bayesian analysis. Composite parameters involve parameters defined in Table S1.1. All prior distributions were chosen to be non-informative.

Simulated AP dataset, SPT virus

A)	*AAP length*	*2h*	*3h*	*3.5h*	*4h*	*4.5h*	*5h*	*6h*	*8h*	
i)	*no. test plant infections*	16	19	16	23	15	15	18	18	
ii)	*no. reps*	30	30	30	30	30	30	30	30	
	IAP length	*5m*	*10m*	*15m*	*20m*	*25m*	*30m*	*40m*	*50m*	*60m*
iii)	*test plant infections*	3	8	14	11	12	13	13	18	19
iv)	*no. reps*	30	30	30	30	30	30	30	30	30

B			*median*	*2.5%*	*97.5%*	*original simulation values*
i)	*acquisition*	α	$0.085h^{-1}$	$0.027h^{-1}$	$2.194h^{-1}$	$0.1\ h^{-1}$
ii)	*inoculation*	β	$1.149h^{-1}$	$0.267h^{-1}$	$2.194h^{-1}$	$1.0\ h^{-1}$
iii)	*latency*	γ	–	–	–	
iii)	*duration of infectiousness*	μ	$0.920h^{-1}$	$0.378h^{-1}$	$1.750h^{-1}$	$1.0\ h^{-1}$

Table S2.2. Simulated SPT virus AP data and analysis. A: Simulated acquisition varying sub-assay and inoculation varying sub-assay assay results for a representative SPT plant virus. See B 'original' column for the underlying viral parameter values. Simulations consisted of sampling random times of acquisition and loss of pathogen for individual whitefly from exponential distributions with intensity values based on the 'original' parameter values given in B. B: SPT viral parameter estimates for insect-borne transmission generated from the simulated access period assay data in A using the *estimate_virus_parameters_SPT()* function. Posterior parameter distributions were obtained using Hamiltonian Monte Carlo, RStan version 2.21.0 (Stan Development Team, 2022), R version 3.6.3 (Core team, 2014). Additional rStan settings were: $warmup = 4500$, $chains = 4$, $iterations = 6000$, $max(treedepth) = 10$, $adapt\_delta = 0.95$.

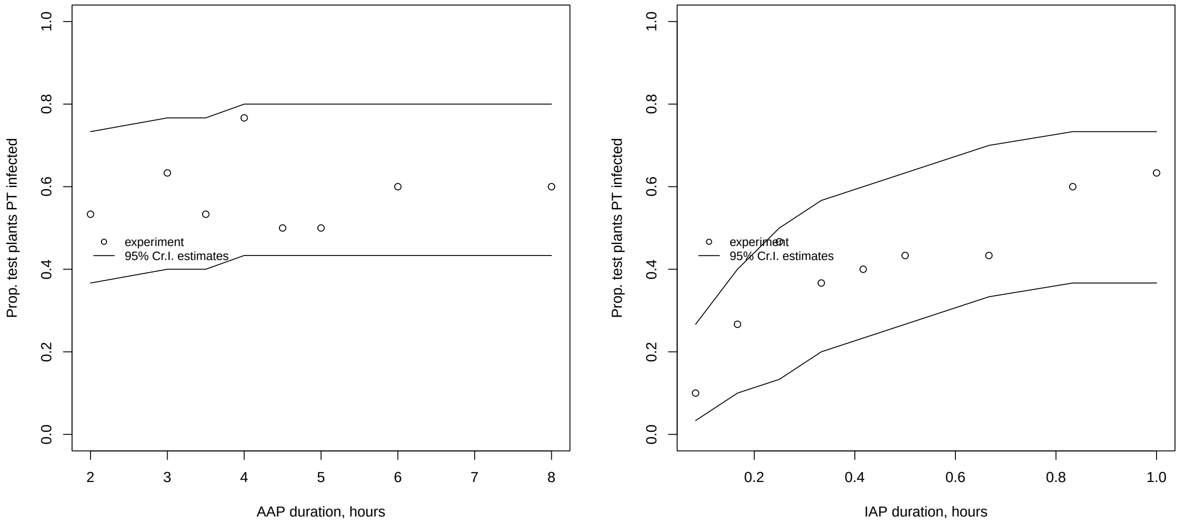


Figure S2.1. Simulated AP data analysis validation figure for a simulated SPT virus. Forward simulation of the access period dataset (Table S2.2 A) using the estimate parameters (Table S2.2 B). We show curves for the 95% forward simulated credible interval and we superimpose the original simulated dataset (empty circles). This process is repeated for each of the two sub-assays in the experiment (acquisition and inoculation varying sub-assays).

Empirical AP dataset, CBSI

A)	*AAP length*	*0.13h*	*0.5h*	*1h*	*4h*	*24h*	*48h*
i)	*test plant infections*	4	8	10	6	9	6
ii)	*no. reps*	25	25	25	15	20	15
	iAP length	*0.13h*	*0.5h*	*1h*	*4h*	*24h*	*48h*
iii)	*test plant infections*	6	7	8	13	29	6
iv)	*no. reps*	31	33	39	35	48	15

B			*median*	*2.5%*	*97.5%*	*Donnelly et al. (2020) estimates*
i)	*acquisition*	α	$0.638h^{-1}$	$0.088h^{-1}$	$1.735h^{-1}$	$1.818\ h^{-1}$
ii)	*inoculation*	β	$0.056h^{-1}$	$0.019h^{-1}$	$0.475h^{-1}$	$0.021\ h^{-1}$
iii)	*latency*	γ	–	–	–	
iii)	*duration of infectiousness*	μ	$0.807h^{-1}$	$0.406h^{-1}$	$1.468h^{-1}$	$0.632\ h^{-1}$

Table S2.3. CBSI AP data and analysis. A: The acquisition varying sub-assay (A) and inoculation varying sub-assay (B) assay results from Maruthi et al. (2020). In the acquisition varying sub-assay insect cohorts consisting of 20-25 insects were moved from infected source plants (for variable duration) to healthy test plants (for 48h). In the inoculation varying sub-assay insect cohorts consisting of 20-25 insects were moved from infected source plants (48h feeding access), to healthy test plants (for variable duration). Note that for out analyses we assumed a value of 23 insects in place of the reported 20-25 insects. B: CBSI parameter estimates for insect-borne transmission generated from access period assay data using the *estimate_virus_parameters_SPT()* function. Posterior parameter distributions were obtained using Hamiltonian Monte Carlo, RStan version 2.21.0 (Stan Development Team, 2022), R version 3.6.3 (Core team, 2014). Additional rStan settings were: $warmup = 4500$, $chains = 4$, $iterations = 6000$, $max(treedepth) = 10$, $adapt\_delta = 0.95$.

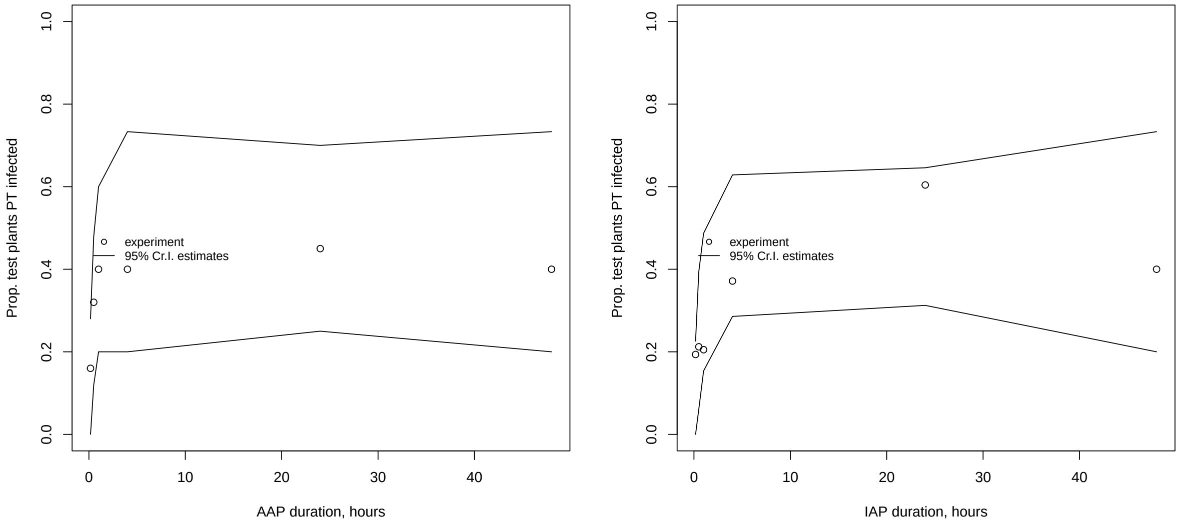


Figure S2.2. Simulated AP data analysis validation figure for the CBSI virus. Forward simulation of the CBSI access period dataset (Table S2.3 A) using the estimated parameters (Table S2.3 B). We show curves for the 95% forward simulated credible interval and we superimpose the original CBSI dataset (empty circles). This process is repeated for the two sub-assays (acquisition and inoculation varying sub-assays).

Supporting Information S3, the *calculate_epidemic_probability* function

4.0.1 The processes determining inoculum extinction

We begin by asking what is the fate of one introduced infected plant i.e., what is the extinction probability, denoted $P(I_0^F)$ (with notation as listed in Table S3.1)? This then requires consideration of all the possible future events for one infected plant with 0 infected insects. Note that the set of inoculum states depends on the number of insects per plant which is assumed constant and is denoted F. For a simple example, when $F = 2$ there are 8 inoculum states, they are $I_0^2, I_1^2, I_2^2, E_0^2, E_1^2, E_2^2, S_1^2, S_2^2$. More generally, there are exactly $3 \times F - 1$ inoculum states: the -1 accounts for the state S_0^F which corresponds to extinction.

For instance, the infected plant may be rogued (i.e., a transition from I_0^F to S_0^F) with rate r, or it may be harvested (i.e., a transition from I_0^F to S_0^F) with rate h, or, alternatively, one of the phloem-feeding insects on the plant may acquire the virus (i.e., a transition from I_0^F to I_1^F) with rate $F\alpha$. Conditioning extinction probabilities on future events one can write,

$$P(I_0) = \frac{r}{r + (F-0)\alpha} P(S_0) + \frac{(F-0)\alpha}{r + (F-0)\alpha} P(I_1). \tag{S3.1}$$

Conditioning equations can be written too for $P(I_1^F)$, and in turn for the extinction probabilities downstream of $P(I_1^F)$, noting, however, that $P(S_0^F) = 1$. When the extinction probabilities in each inoculum state have been related to each other in this way together they form semi-linear systems of equations. Writing these now using general notation the systems are of the form,

$$\vec{s_i} = \left(\sum_{j\neq i} \delta_{ij}\vec{s_j}\right) + \vec{w}, \tag{S3.2}$$

$$\Leftrightarrow T\vec{s} + \vec{w} = 0, \tag{S3.3}$$

where $\vec{s}$ is a vector of length $L = 3 \times (F + 1) - 1$ containing the extinction probabilities for each state, and T is a square matrix of dimension L consisting of transition rates between the states - with $T_{i=j} = -1$. Finally $\vec{w}$ is a vector of length L containing the rates of transition that result in direct extinction from each state.

4.0.2 The calculate_epidemic_probability function

The transition rates for the matrix T are shown in Table S3.2 for the simple case $F = 1$. The general case, however, can be constructed in a similar way. The elements of the vector $\vec{w}$ are also shown in Table S3.2 - they are the events that result in a post-transition extinction probability of 1. The semi-linear system 4.0.3 has the fixed point solution vector $\vec{\hat{s}}$ which represents the extinction probability for each possible inoculum state (recalling that we equate inoculum with a single plant and its resident insects). The calculate_epidemic_probability function constructs the semi-linear system based on the number of insects per plant and it returns 1 minus the fixed points of the system (i.e., the epidemic probability from each starting inoculum state). This involves numerical solving of system with the standard base R function.

Propagation of inoculum foci

It is important to note that system takes account of the generation of new inocula plants implicitly - this occurs through insect dispersal which results in a product of extinction probabilities for the relevant inoculum state (see e.g. event 2.13 in Table S3.2). In other words, when there is dispersal of an infected insect the extinction probability in the first instance increases to that of the same inoculum state but with 1 fewer infected insect (for instance, from $P(I^1_1)$ to $P(I^F_0)$). But this is not the end of the story, the assumption that the dispersing infected insect alights elsewhere on an uninfected plant (since all other plants are assumed uninfected) means that the extinction probability is instead the product of the first inoculum plant state and the new one, i.e., $P(I^F_0)P(S^F_1)$, so that the probability of extinction has in fact decreased significantly after the infected insect dispersal.

Finally, note that for convenience the algorithm models insect dispersal as an insect exchange between two plants. This allows the number of insects per plant to remain constant across all plants. Note that in practice this means that rate of insect dispersal is exactly doubled in the algorithm and therefore when we report the results we report them as the natural dispersal rate (i.e., twice the algorthim dispersal rate). Note that future versions of the EpiPv package will allow for dynamic insect abundance.

4.0.3 Calculation of epidemic probability using an iterative approach

System has multiple solutions because it contains a (weak) non-linearity - though there is typically only one viable solution. In fact, constraints inherent in the branching process mean that one can simply take the smallest fixed point for each state across the multiple extinction probability solutions.

Nevertheless, an alternative approach works by iterating extinction probability over time so that the extinction probabilities depend on the starting point and time t $\vec{s} = \vec{s}(t_0, t)$. When run to an asymptote in $\vec{s}$ this is equivalent to the smallest extinction probability solutions across the fixed points of the semi-linear system . The iterative approach, however, has several advantages: it is a natural way to calculate extinction probability with respect to a finite season, and it is straightforward to incorporate time-dependent rates (see final subsection in this Supporting Information).

We iterate the equivalent of system 4.0.3 from initial conditions representing initial invasion (i.e., $\vec{s}(t_0) = 0$ since at the point where inoculum has been introduced the local epidemic is definitely not extinct) until a solution (asymptote) has been reached $\vec{s}(t_{asymp})$ or the season-end has been reached. We now briefly summarise the iterative algorithm which is encoded in the *calculate_epidemic_probability* function of the EpiPv package.

Just as T and $\vec{w}$ determined the solution of system 4.0.3, the iterative extinction probability vector $\vec{s}(t)$ is updated using transition and fertility matrices $\tilde{t}$ and f (Caswell, 2000). The entries of $\tilde{t}$ are the probabilities that a transition occurs in a given time step.

In addition, $\tilde{t}$ includes an $(L+1)^{th}$ row for direct extinction (i.e, equivalent to $\vec{w}$ in system 4.0.3), making $\tilde{t}$ of dimension $L+1 \times L$. We must also allow for no event occurrence in the fixed time step: $\tilde{t}_{i,i} = 1 - \sum_{i \neq j} \tilde{t}_{i,j}$. In addition, *calculate_epidemic_probability* selects the biggest interval δt such that $\sum_i \tilde{t}_{i,j} < 1$ for all j. The fertility matrix f (dimension $L \times L$) is zero apart from the row corresponding to the inoculum state S_1^F (the $L-(F-1)^{th}$ row. This row of f contains the probabilities of infected insect dispersal from the various inoculum states to a new plant in the given time step (positive only for the states X_j^F with $j \geq 1$ and $X \in S, E, I$). These values are the same as the entries corresponding to infected insect dispersal in $\tilde{t}$.

The *EpiPv* R package function *calculate_epidemic_probability* outputs the epidemic probability for each inoculum state for a location with user-input F insect burden per plant and local parameters ν, r, θ, b, h. In addition the function takes the virus parameters α, β, μ as arguments and these can be estimated from access period experiment data using the estimate_virus_parameters_SPT and estimate_virus_parameters_PT functions. A precision value which is used to identify when an asymptote has been reached in the iterative process or alternatively, the user can specify when a growing season ends. Users of the function will typically be interested in the solution elements $\vec{s}(t_{asymp})_1$ (probability of extinction for a single infected plant introduction) and $\vec{s}(t_{asymp})_{L-(F-1)}$ (probability of extinction for a single infected insect vector introduction). Note that $1-$ the extinction probability in question results in the *epidemic probability* output.

A single iteration of the algorithm updates $\vec{s}(t)$ to $\vec{s}(t+1)$ for each inoculum state i using the set of calculations,

$$\begin{aligned} gb &= 1 - f(\nu,:) + f(\nu,:)\vec{s}_\nu(t) \\ gt &= \tilde{t}' * [\vec{s}(t);1] \\ \vec{s}(t+1) &= gt.^* gb \end{aligned}$$

We now provide explanation for each step of the iteration, note also that in the above $*$ denotes matrix multiplication, and where $.^*$ denotes element-wise vector multiplication.

First, gb scales extinction probability by the probability that a new foci is produced. This is needed because spread to new foci leads to a proportional reduction in extinction probability. This corresponds to a multiplication of extinction probability from the current state with extinction probability for the new expanded state. The algorithm is written above to reflect that new foci correspond to a the S_1 inoculum state only (i.e., $\nu = L - (F-1)$, the 'birth state' corresponding to S_1^F). Note that gb is simply 1 when there is no foci spread from an inoculum state (scaled by the probability of no foci spread), otherwise it is $P(S_1^F)$ (scaled by the probability of foci spread). In this way the gb calculation produces a container column vector of length n.

Second, gt computes the transitions between the inoculum states. Note that $\tilde{t}$ is an $(n+1)xn$ matrix where the final row is the probability of direct extinction. The transpose of this matrix is multiplied by the vector of inoculum states with the value 1 appended, i.e., an column vector of length $n+1$. In this way the gt calculation produces a container column vector of length n.

Finally, the element-wise multiplication of the gb and gt vectors produces updated extinction probabilities from each inoculum state.

4.0.4 Incorporating dynamic insect abundance using an iterative approach

As a final note, though intuitive to field pathologists and convenient given frequent associations of whitefly burden with location, assuming a single level of whitefly abundance per plant at a location is unrealistic. The iterative approach, however, means that one can readily incorporate dynamic insect abundance. One way to achieve this is for the algorithm to be constructed for the maximum number of insects per plant in a season. The iteration would then commence as usual with extinction probability equal to 1 for all inoculum states except the one corresponding to inoculum introduction. The dynamic level of abundance $F'(t)$ is then incorporated only in the rate of virus acquisition. Where previously this was $\alpha(F - j)$ for the inoculum state I_j^F this would now become $\alpha(F'(t) - j)$ - but note that the notation for the inoculum states retain their previous form e.g. I_j^F with F here referring to the maximum insect abundance per plant over the season (cf. $F'(t)$ in the total acquisition rate which denotes actual dynamic abundance at time t).

A)	Notation	***Parameters & principal variables***	***Units/labelling***
	S	susceptible plant	*plant type*
	E	latent-infected plants	*plant type*
	I	infectious plants	*plant type*
	$P(X_F^j)$	extinction prob plant type X_F^j	*probability*
	X_F^j	plant type X with j from F infected insects	*inoculum state*
	α	pathogen acquisition rate	*rate* h^{-1}
	β	pathogen inoculation rate	*rate* h^{-1}
B)		***Field model***	
	ν	rate of onset of plant infectiousness	*rate* h^{-1}
	μ	rate of loss of insect infectiousness	*rate* h^{-1}
	r	rate of infected plant removal	*rate* h^{-1}
	F	#insect insects per plant	*integer*
	θ	rate of insect dispersal	*rate* h^{-1}
	h	harvesting rate	*rate* h^{-1}
	b	rate of insect mortality	*rate* h^{-1}

Table S3.1. Parameter definitions for the laboratory and field models.

pre-transition probability					*post-transition probability*
$P(I_0^1)$	Infected plant rogue	$I_0^1 \to S_0^1$	$\frac{r}{r+h+\alpha}$	(S3.4)	1
	Infected plant harvest	$I_0^1 \to S_0^1$	$\frac{h}{r+h+\alpha}$	(S3.5)	1
	Insect acquisition	$I_0^1 \to I_1^1$	$\frac{\alpha}{r+h+\alpha}$	(S3.6)	$P(I_1^1)$
$P(I_1^1)$	Infected insect dispersal	$I_1^1 \to I_0^1$	$\frac{\theta}{\theta+\mu+b+r+h}$	(S3.7)	$P(I_0^1)P(S_1^1)$
	Infected insect recovery	$I_1^1 \to I_0^1$	$\frac{\mu}{\theta+\mu+b+r+h}$	(S3.8)	$P(I_0^1)$
	Infected insect death	$I_1^1 \to I_0^1$	$\frac{b}{\theta+\mu+b+r+h}$	(S3.9)	$P(I_0^1)$
	Infected plant rogue	$I_1^1 \to S_1^1$	$\frac{r}{\theta+\mu+b+r+h}$	(S3.10)	$P(S_1^1)$
	Infected plant harvest	$I_1^1 \to S_1^1$	$\frac{h}{\theta+\mu+b+r+h}$	(S3.11)	$P(S_1^1)$
$P(E_0^1)$	Exposed plant harvest	$E_0^1 \to S_0^1$	$\frac{h}{\nu+h}$	(S3.12)	1
	Exposed plant progression	$E_0^1 \to I_0^1$	$\frac{\nu}{\nu+h}$	(S3.13)	$P(I_0^1)$
$P(E_1^1)$	Infected insect dispersal	$E_1^1 \to E_0^1$	$\frac{\theta}{\theta+\mu+b+\nu+h}$	(S3.14)	$P(E_0^1)P(S_1^1)$
	Infected insect recovery	$E_1^1 \to E_1^1$	$\frac{\mu}{\theta+\mu+b+\nu+h}$	(S3.15)	$P(E_0^1)$
	Infected insect death	$E_1^1 \to E_1^1$	$\frac{b}{\theta+\mu+b+\nu+h}$	(S3.16)	$P(E_0^1)$
	Exposed plant harvest	$E_1^1 \to S_1^1$	$\frac{h}{\theta+\mu+b+\nu+h}$	(S3.17)	$P(S_1^1)$
	Exposed plant progression	$E_1^1 \to I_1^1$	$\frac{\nu}{\theta+\mu+b+\nu+h}$	(S3.18)	$P(I_1^1)$
$P(S_1^1)$	Infected insect dispersal	$S_1^1 \to S_0^1$	$\frac{\theta}{\theta+\mu+b+h}$	(S3.19)	$P(S_1^1)$
	Infected insect recovery	$S_1^1 \to S_0^1$	$\frac{\mu}{\theta+\mu+b+h}$	(S3.20)	1
	Infected insect death	$S_1^1 \to S_0^1$	$\frac{b}{\theta+\mu+b+h}$	(S3.21)	1
	Susceptible plant harvest	$S_1^1 \to S_1^1$	$\frac{h}{\theta+\mu+b+h}$	(S3.22)	$P(S_1^1)$

Table S3.2. Table of transition rates, with pre- and post-transition extinction probabilities, for the case of $F = 1$, which can be generalised for any insect burden F.

Supporting Information S4, the *AP_data_simulator* function

We additionally include a function for simulating AP data given underlying assay structure (T_A, T_L and T_I, number of replicates n and number of insect vectors used W_0). It is also necessary to specify virus transmission parameters from which to simulate virus acquisition, inoculation, latency progression and clearance (i.e. the user must provide values for α, β, γ and μ).

The *AP_assay_simulator* R function (EpiPv package) then draws n samples from exponential distributions for each the following virus transmission parameters, α, β and γ. The *AP_assay_simulator* function calls *AP_insect_simulator* (to simulate random times of acquisition, inoculation, loss of infectiousness, progression through latency on a per insect basis). This in turns calls the *inoc_durtn_calculator* function which calculates the effective duration of inoculation for each set of samples, T_I^{eff}: this is the intersection of the IAP period and the period when a given insect was infectious. Finally the probability of test plant infection is calculated by sampling from an exponential distribution with rate β and testing whether the sample is less than T_I^{eff} (i.e., whether or not an inoculation event occurred within the effective inoculation window).

The *AP_assay_simulator* R function performs these calculations for each insect of W_0 initial insects, to produce n experimental replicates of an AP assay.

Supporting Information S5, Validation of virus parameter estimates

We here describe the exercise of comparing individual-based simulation of epidemic data with predicted epidemic risk. By individual-based simulation we mean the reproduction of the events that occur when inoculum is introduced into a field. The events are simulated in proportion to expected statistical distributions which are updated each time an event occurs. We perform this computationally complex procedure as a baseline with which to compare our calculations of epidemic probability (i.e., if they agree then accuracy of the calculations are verified). When we ran the *calculate epidemic probability* function for a range of values for four focal parameters (with remaining values held constant), we found that the predicted epidemic probability matched the outcomes of individual-based computer simulations for the same underlying sets of parameter values. An infected plant introduction at the start of a season was simulated with fields evaluated at the end of the season to assess whether or not a local epidemic had occurred (a value of 1 was assigned if there was > 1 infected plant after $1yr$ of simulated events, otherwise 0).

In figure S5.1 stochastic simulations were conducted in batches of 50 (i.e. an epidemic probability sensitivity of 1/50 per batch) for a representative plant virus. For each parameter value there were 20 batches (i.e. 20 data points per parameter value). Grey circles denote individual batch outcomes, black circles denote the mean over the batches and blue crosses denote the predicted value (A-D). This process was repeated for four underlying epidemiological parameters: the rate of roguing (A), the dispersal rate (B), the rate of pathogen clearance from the insect (C) and the number of insects per plant (per top 5 leaves) (D). Figure data-points for predicted epidemic probability were obtained using the *calculate_epidemic_probability* function. Note that the individual-based simulation is for manuscript validation only and is not provided in the EpiPv function.

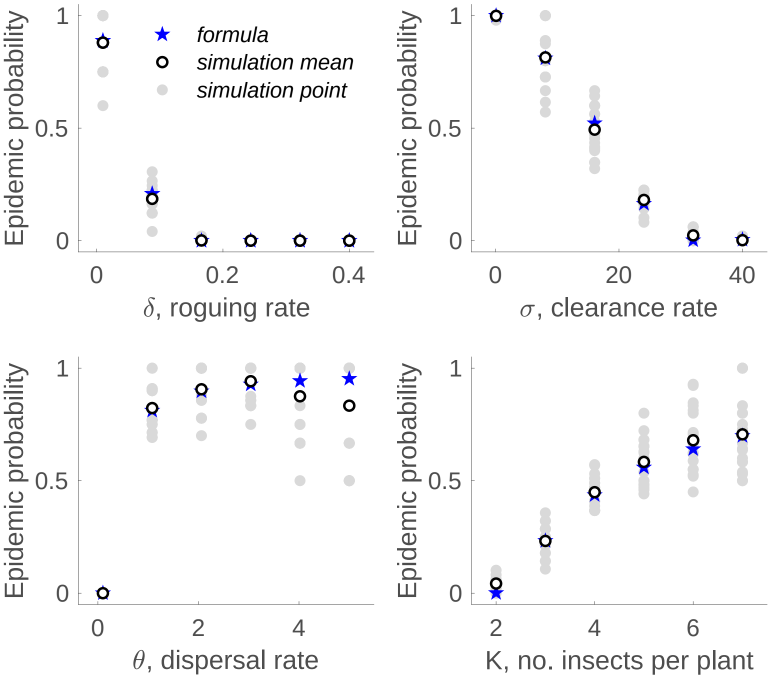


Figure S5.1. Validation of predicted epidemic probabilities using individual-based simulations. In A-D we plot the epidemic outcome of stochastic simulations of infected plant introductions against the predicted values.

Supporting Information S6, AP dataset structure in the *EpiPv* package

AP datasets have an obligatory structure in the *EpiPv* package. This structure is required in the arguments of the estimate virus parameters package functions. In vignette B, assay simulation script - we show how to produce this structure using the AP data simulator.

Experimenters with empirical data should first determine whether they are analysing an SPT virus or a PT virus (see e.g. Carr et al. (2018); Hogenhout et al. (2008) and note that non-persistently aphid-transmitted viruses should not be analysed with this package due to their unique transmission mode see e.g.).

They are then advised to load the corresponding example dataset that is available in the package: either *load(ap_data_sim_PT)* or *load(ap_data_sim_SPT)* as appropriate. They can then edit the ensuing list (i.e. dataset object in r) and assign it to a dedicated list: *ap_data_myVirus_PT=ap_data_sim_PT*. They should then replace the elements of the list with their data.

We now show the list elements using *ap_data_sim_PT* for reference:

```
$d_AAP
      [,1] [,2]  [,3] [,4]  [,5] [,6] [,7] [,8]
T_vec    1  1.5  1.75    2  2.25  2.5    3    4
R_vec   30 30.0 30.00   30 30.00 30.0   30   30
I_vec   15 17.0 21.00   27 25.00 27.0   24   30

$d_LAP
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
T_vec  0.5    1    2    3    4    5    6    7    8
R_vec 30.0   30   30   30   30   30   30   30   30
```

```
I_vec 26.0   28   28   29   24   27   26   25   28

$d_IAP
             [,1]       [,2]  [,3]       [,4]       [,5] [,6]       [,7]
T_vec  0.08333333  0.1666667  0.25  0.3333333  0.4166667  0.5  0.6666667
R_vec 30.00000000 30.0000000 30.00 30.0000000 30.0000000 30.0 30.0000000
I_vec  4.00000000  7.0000000 12.00 17.0000000 14.0000000 18.0 18.0000000
            [,8] [,9]
T_vec  0.8333333    1
R_vec 30.0000000   30
I_vec 21.0000000   25

$d_durations
                  [,1] [,2] [,3]
AAPfixedComponent   -1  0.5    1
LAPfixedComponent    2 -1.0    1
IAPfixedComponent    2  0.5   -1

$d_vectorspp
[1] 20

$d_virusType
[1] "PT"

attr(,"alpha")
[1] 0.1
```

```
attr(,"beta")
[1] 1
attr(,"gamma")
[1] 0.5
attr(,"mu")
[1] 0.01
```

And the list elements using *ap_data_sim_SPT* for reference:

```
$d_AAP
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
T_vec    2    3  3.5    4  4.5    5    6    8
R_vec   30   30 30.0   30 30.0   30   30   30
I_vec   19   15 20.0   17 20.0   23   18   20

$d_IAP
             [,1]       [,2]  [,3]       [,4]       [,5] [,6]       [,7]
T_vec  0.08333333  0.1666667  0.25  0.3333333  0.4166667  0.5  0.6666667
R_vec 30.00000000 30.0000000 30.00 30.0000000 30.0000000 30.0 30.0000000
I_vec  3.00000000  7.0000000  5.00 12.0000000 14.0000000 12.0 16.0000000
            [,8] [,9]
T_vec  0.8333333    1
R_vec 30.0000000   30
I_vec 15.0000000   16

$d_durations
     [,1] [,2]
```

```
[1,]   -1    6
[2,]    4   -1

$d_vectorspp
[1] 20

$d_virusType
[1] "SPT"

attr(,"alpha")
[1] 0.1
attr(,"beta")
[1] 1
attr(,"mu")
[1] 1
```