

RESEARCH ARTICLE

A method of determining where to target surveillance efforts in heterogeneous epidemiological systems

Alexander J. Mastin^{1*}, Frank van den Bosch², Timothy R. Gottwald³, Vasthi Alonso Chavez², Stephen R. Parnell¹

1 Ecosystems and Environment Research Centre, School of Environment and Life Sciences, University of Salford, Greater Manchester, United Kingdom, **2** Computational and Systems Biology, Rothamsted Research, Harpenden, Hertfordshire, United Kingdom, **3** USDA Agricultural Research Service, Fort Pierce, Florida, United States of America

* a.mastin@salford.ac.uk



OPEN ACCESS

Citation: Mastin AJ, van den Bosch F, Gottwald TR, Alonso Chavez V, Parnell SR (2017) A method of determining where to target surveillance efforts in heterogeneous epidemiological systems. *PLoS Comput Biol* 13(8): e1005712. <https://doi.org/10.1371/journal.pcbi.1005712>

Editor: Mark M. Tanaka, University of New South Wales, AUSTRALIA

Received: January 19, 2017

Accepted: August 2, 2017

Published: August 28, 2017

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by United States Department of Agriculture Farm Bill Section 10007 (Project 1A.0215.01) and Biotechnology and Biological Sciences Research Council (through funding to Rothamsted Research). The funders had no direct role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The spread of pathogens into new environments poses a considerable threat to human, animal, and plant health, and by extension, human and animal wellbeing, ecosystem function, and agricultural productivity, worldwide. Early detection through effective surveillance is a key strategy to reduce the risk of their establishment. Whilst it is well established that statistical and economic considerations are of vital importance when planning surveillance efforts, it is also important to consider epidemiological characteristics of the pathogen in question—including heterogeneities within the epidemiological system itself. One of the most pronounced realisations of this heterogeneity is seen in the case of vector-borne pathogens, which spread between ‘hosts’ and ‘vectors’—with each group possessing distinct epidemiological characteristics. As a result, an important question when planning surveillance for emerging vector-borne pathogens is where to place sampling resources in order to detect the pathogen as early as possible. We answer this question by developing a statistical function which describes the probability distributions of the prevalences of infection at first detection in both hosts and vectors. We also show how this method can be adapted in order to maximise the probability of early detection of an emerging pathogen within imposed sample size and/or cost constraints, and demonstrate its application using two simple models of vector-borne citrus pathogens. Under the assumption of a linear cost function, we find that sampling costs are generally minimised when either hosts or vectors, but not both, are sampled.

Author summary

Emerging pathogens are an increasing threat to human, animal, and plant health. In areas where these pathogens have not yet become established, surveillance is needed to detect incursions early enough to implement control measures. However, most epidemiological systems are heterogeneous in nature, and it is unclear how finite surveillance resources

Competing interests: The authors have declared that no competing interests exist.

should be divided between constituent groups (such as hosts and vectors in the case of vector-borne pathogens). We use mathematical and statistical methods to address this issue. Taking the example of vector-borne pathogens, we show how to estimate the proportion of infected hosts or vectors at the time of first detection for any combination of host and vector sampling rates, given some knowledge of the characteristics of pathogen spread within and between hosts and vectors. We predict that the required total sampling effort and cost for early detection will be lowest when either hosts or vectors are sampled, with the optimal group to sample being the one with the highest estimated prevalence during initial exponential growth (which has clear parallels with ‘targeted surveillance’). We demonstrate the use of our framework by applying it to two vector-borne diseases of citrus and evaluate its predictions using a simple simulation model of sampling.

Introduction

Human activities over the past 500 years have dramatically altered the distribution of organisms worldwide, through both purposeful and unintentional ‘invasions’ and ‘extinctions’ [1]. The global spread of plant pathogens, driven largely by the movement of people, plants, and products, as a result of globalisation [2–6], is an area of increasing concern, as these pathogens are a threat to natural ecosystems [7, 8] and horticultural industries [9–12] worldwide. The resilience of natural and managed ecosystems to new pathogens is further reduced by changes in land use and modern agricultural practices such as intensification, geographical consolidation, artificial selection, and genetic homogenisation [3, 13].

Plant disease control has historically been reactive in nature, but there is an increasing move towards proactive, risk-based, prevention strategies [14]. National and regional plant protection organisations therefore expend considerable effort in minimising the risk of emerging pathogens entering and establishing in new areas, through trade and movement restrictions/controls [15], border inspection and treatment [16], and ‘early detection surveillance’ activities [17]. Whilst movement restrictions and border checks help to minimise the risk of pathogen entry, early detection surveillance aims to detect pathogens following entry at a sufficiently early stage to allow control measures to be instigated. A failure of early detection may result in higher overall costs of control [16], or the loss of ability to control the pathogen altogether [18, 19]. It is well recognised that statistical and economic issues should be considered when planning early detection surveillance activities [20–23], but our previous work has shown that biological characteristics of the pathogen in question, in particular the rate of spread in a naive ecosystem, should also be considered [17, 24, 25]. This is particularly important in the case of emerging pathogens, where the prevalence of the invading epidemic will not be known until the time of first discovery, but can be approximated if the initial rate of transmission can be estimated.

Although many surveillance strategies are inherently founded on the assumption that the infection status of each individual is independent of all other individuals in the population (as is seen when simple random sampling is assumed to take place throughout the whole population), most epidemiological systems are characterised by marked heterogeneities [26, 27]. In these cases, pathogens tend to spread within and between distinct ‘groups’ of individuals: such as between different hosts, to and from environmental reservoirs, and between hosts and disease-carrying vectors. Although these groupings, or ‘heterogeneities’, may be implicitly acknowledged during surveillance planning for logistical reasons, sampling strategies are commonly driven by ease of sampling, availability, or perceived importance. For example,

surveillance strategies for plant diseases in particular have historically been largely based upon visual inspection of plants for signs of disease (despite this strategy delaying the timing of early detection [28, 29], reducing the ability to predict major disease outbreaks [30], and reducing the accuracy of prevalence estimation [12, 31]). Despite this, there has been increasing recognition in recent years of the potential to capitalise on heterogeneities in epidemiological systems by explicitly targeting early detection surveillance activities towards those groups which have a higher probability of infection (termed ‘risk-based’, or ‘targeted’, surveillance [26]) [32–34]. Although quantitative methods for targeting surveillance resources according to the risk of infection in a spatial context [35], or according to other epidemiological groupings [26] are available, these are often based on a largely ‘phenomenological’ interpretation of ‘risk groups’ (such as those obtained from statistical models). Few studies to date have attempted to develop a generic, biologically informed, framework for allocation of surveillance resources in heterogeneous systems based on a ‘mechanistic’ model of pathogen spread.

We focus here on vector-borne plant pathogens. These are responsible for a number of diseases of current concern, including huanglongbing (caused by bacteria of the genus *Candidatus Liberibacter*, and spread by hemipteran psyllid insects); olive quick decline syndrome (caused by the bacteria *Xylella fastidiosa* and spread predominantly by the Meadow spittlebug, *Philaenus spumarius*); and citrus tristeza syndromes (caused by the citrus tristeza virus and spread most effectively by the brown citrus aphid, *Toxoptera citricida* Kirkaldy). In the case of these pathogens, the aforementioned general focus on visual inspection for diagnosis means that commercially important host crops rather than insect vectors are often the primary focus of surveillance, as this is where symptoms and economic impacts are manifested, despite the recognised benefits of laboratory-based vector surveillance for diseases with long asymptomatic periods [28, 29].

In the current paper, we show how a mechanistic mathematical model of pathogen transmission can be linked with a statistical model of the timing of first detection during an ongoing surveillance campaign in order to estimate the mean prevalence at first detection in either group. As well as the total ‘sampling effort’ (the ‘rate’ of sample collection per unit time) from each group, the prevalence at first detection is affected by epidemiological characteristics of the pathogen in question (the rate of epidemic growth in the system as a whole and the relative prevalences of infection in hosts and vectors as the pathogen spreads through the system). We go on to show that the total sampling effort is minimised when either vectors or hosts, but not both, are sampled, and how the ‘costs’ of sampling can be incorporated into this framework.

Methods

Summary

The central question we wish to answer is what prevalence will a pathogen reach in hosts and in vectors when it is first detected (‘detection-prevalence’), given we know the sample size and the frequency of sampling from these two groups. In order to answer this, we develop (i) a statistically-based sampling model; and (ii) a mathematical model of pathogen population dynamics, which we combine in order to generate a heuristic (rule of thumb) for estimation of the prevalence at first detection and other useful outputs. Finally, we validate our heuristic by comparing its predictions with those obtained from a simulation model of host or vector sampling. In this section, we first outline the statistical sampling model, then go on to demonstrate how this can be parameterised using a compartmental mathematical model of the pathosystem in question, and finish by describing the two mathematical models we developed for this study.

Statistical sampling model

We start by describing how a simple binomial-based sampling model can be used to estimate the prevalence at first detection in a heterogeneous system comprised of ‘hosts’ and ‘vectors’. The reader is referred to our earlier reports [17, 24, 25] and S1 Text for additional information on the derivation described below.

If we know the prevalence of infection at any specified timepoint, t_1 , we can use the binomial distribution to calculate the probability of failing to sample an infected individual during a single sampling round at time t_1 :

$$\left(1 - \left[\frac{I(t_1)}{\rho}\right]\right)^N \tag{1}$$

Where N is the number of samples collected, and the ratio of the number of infected individuals at time t_1 , $I(t_1)$, to the total number of individuals, ρ , is the prevalence of infection. We define ‘prevalence’ as the proportion of infected individuals, which is commonly referred to as the ‘incidence’ in the field of plant pathology. From this, we can estimate the probability of at least one detection during sampling at this time ($P(t_1)$):

$$P(t_1) = 1 - \left(1 - \left[\frac{I(t_1)}{\rho}\right]\right)^N \tag{2}$$

The probability of first detection at time t_1 in an ongoing sampling programme (where N samples are collected every Δ days) can be estimated as the product of eq 1 for each each of the K sampling points since initial entry of the pathogen (at time t_0), and Eq 2:

$$P(t_1 | t_0) = \left[1 - \left(1 - \left[\frac{I(t_1)}{\rho}\right]\right)^N\right] \cdot \prod_{k=1}^K \left(1 - \left[\frac{I(t_1 - k\Delta)}{\rho}\right]\right)^N \tag{3}$$

We can expand the framework in eq 3 in order to incorporate two groups of interest. Assuming a host-vector system where the number of infected hosts is I_h and the number of infected vectors is I_v , and the total numbers of hosts and vectors are given as ρ_h and ρ_v , we obtain the following:

$$P(t_1 | t_0) = \left[1 - \left(\left(1 - \left[\frac{I_h(t_1)}{\rho_h}\right]\right)^{N_h} \cdot \left(1 - \left[\frac{I_v(t_1)}{\rho_v}\right]\right)^{N_v}\right)\right] \cdot \prod_{k=1}^K \left(\left(1 - \left[\frac{I_h(t_1 - k\Delta)}{\rho_h}\right]\right)^{N_h} \cdot \left(1 - \left[\frac{I_v(t_1 - k\Delta)}{\rho_v}\right]\right)^{N_v}\right) \tag{4}$$

Where N_h is the number of hosts sampled at each sampling point, and N_v is the number of vectors sampled at each sampling point.

In reality, t_0 is not known, but it is possible to estimate t_0 given that detection occurs [17, 24]. First, we can simplify eq 4 if we assume that the initial increase in the prevalence is exponential in nature, that prevalences are low, and that sampling occurs as a continuous process rather than at discrete intervals (with a sampling rate of $\frac{N}{\Delta} = \theta$). Given that there is pathogen transmission between hosts and vectors in both directions, we can assume a single rate of exponential growth, r , for the system as a whole. We can then use Bayes’ theorem to represent

the probability of first entry at time t_0 given the pathogen was detected at time t_1 [17, 24]:

$$P(t_0 | t_1) \approx \left[\left(\theta_h \left[\frac{v_h}{\rho_h} \right] + \theta_v \left[\frac{v_v}{\rho_v} \right] \right) e^{r(t_1-t_0)} \right] \cdot \exp \left(- \left(\frac{1}{r} \right) \left[\left(\theta_h \left[\frac{v_h}{\rho_h} \right] + \theta_v \left[\frac{v_v}{\rho_v} \right] \right) e^{r(t_1-t_0)} \right] \right) \quad (5)$$

The two new parameters v_h and v_v can be interpreted as the relative numbers of infected hosts and vectors as exponential growth within the system as a whole is first achieved. Given that the initial increase in the number of infected hosts and vectors is exponential in nature, these estimates can be obtained from analysis of a system of ordinary differential equations (ODEs) (see the Sampling model parameterisation section below and S2 Text). If we assume deterministic growth in the prevalence over time, we can adjust eq 5 in order to calculate the expected prevalence, q , in each group at the time of first detection. Using the approach described in our previous work [17, 24] and in S1 Text, we find that the prevalence at first detection in each group ($\kappa = h;v$) follows an exponential distribution:

$$P(q_\kappa^* | t_1) \lambda_\kappa e^{-\lambda_\kappa q_\kappa^*} \quad (6)$$

The exponential rate parameter λ_κ in eq 6 will vary depending upon whether the prevalence in hosts or vectors is desired. For the host prevalence at first detection, λ_h is used, and is calculated as:

$$\lambda_h = \left(\left(\frac{1}{r} \right) \left(\theta_h + \theta_v \left[\frac{\left(\frac{v_v}{\rho_v} \right)}{\left(\frac{v_h}{\rho_h} \right)} \right] \right) \right) \quad (7)$$

For the vector prevalence at first detection, λ_v is used, and is calculated as:

$$\lambda_v = \left(\left(\frac{1}{r} \right) \left(\theta_h \left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right] + \theta_v \right) \right) \quad (8)$$

The values of r , $\left(\frac{v_h}{\rho_h} \right)$, and $\left(\frac{v_v}{\rho_v} \right)$ are estimated from a mathematical model of the pathosystem under study as described in the Sampling model parameterisation section below.

Estimating sampling effort

The mean prevalences at first detection ($E(q_h^*)$ and $E(q_v^*)$) can be estimated as the inverse of the rate (λ_κ) parameters of the exponential distributions in eqs 6 to 8:

$$E(q_h^*) = \frac{r}{\left(\theta_h + \theta_v \left[\frac{\left(\frac{v_v}{\rho_v} \right)}{\left(\frac{v_h}{\rho_h} \right)} \right] \right)} \quad (9)$$

$$E(q_v^*) = \frac{r}{\left(\theta_h \left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right] + \theta_v \right)} \quad (10)$$

When there is only one group of interest (i.e. only one group is sampled, and the mean prevalence in that group is to be estimated), these formulae reduce down to our original rule of thumb [17, 24]:

$$E(q^*) = \frac{r}{\theta} = \frac{r\Delta}{N} \tag{11}$$

Eqs 9 and 10 can also be rearranged in order to estimate the rate of sampling required from each group in order to first detect the pathogen at a specified mean prevalence in either group, which we define here as the ‘sampling effort’. This gives four separate linear equations which represent the sampling effort required for first detection in hosts or vectors at any specified

mean prevalence as a function of r , the ratio $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right]$, and the sampling effort from the other

group (shown in S1 Text). Interestingly, we found that these are linear functions, meaning that the sampling rate will not be minimised by sampling from both groups and indicating that a single group alone should be sampled in order to minimise the total sampling effort. Representing the sampling effort when only one group is sampled as Θ , we can manipulate eqs 9 and 10 to show how to calculate the relative required rate of exclusive vector sampling, Θ_v (as compared to exclusive host sampling, Θ_h) for detection at any specified mean prevalence (more details on this derivation can be found in S1 Text):

$$\left(\frac{\Theta_v}{\Theta_h} \right) = \left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right] \tag{12}$$

Estimating sampling costs

In many cases, the main constraint to planned surveillance activities will be the resources available for sampling. If we assume that sampling will be conducted from either hosts or vectors, and that the total ‘cost’ of sampling from either group during each sampling round (which may be purely financial cost, or some other metric), Z_h or Z_v , can be calculated as the sum of the ‘fixed’ costs of sampling from the group in question per sampling round (ζ_{0h} and ζ_{0v}) and the product of the sampling effort and the cost of sampling a single individual from the group (ζ_h or ζ_v):

$$Z_h = \zeta_{0h} + \zeta_h \Theta_h \tag{13}$$

$$Z_v = \zeta_{0v} + \zeta_v \Theta_v \tag{14}$$

This allows us to reformulate eq 12 as:

$$\left[\frac{\left(\frac{Z_v - \zeta_{0v}}{\zeta_v} \right)}{\left(\frac{Z_h - \zeta_{0h}}{\zeta_h} \right)} \right] = \left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right] \tag{15}$$

If we now assume that the total cost is constant and therefore equal regardless of which group is sampled ($Z_h = Z_v$), and that the fixed costs of surveillance are also equal for either group ($\zeta_{0h} = \zeta_{0v}$), then the left side of eq 15 reduces down to the ratio $\left(\frac{\zeta_h}{\zeta_v} \right)$. Under these

constraints, the ratio $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right]$ therefore indicates the ratio of individual unit sampling costs at which the total cost of sampling exclusively from hosts would be equal to that when sampling exclusively from vectors. This can be also treated as a ‘threshold quantity’ which indicates whether to sample from hosts or vectors in order to minimise the total sampling cost:

- If $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right] > \left[\frac{c_h}{c_v} \right]$ then sample from hosts only.
- If $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right] = \left[\frac{c_h}{c_v} \right]$ then sample from hosts and/or vectors.
- If $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right] < \left[\frac{c_h}{c_v} \right]$ then sample from vectors only.

Sampling model parameterisation

The rate of exponential increase for both groups, r , and the ratio $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right]$ can be estimated using techniques from conventional model stability analysis. If we create an ODE model of our epidemiological system and represent the number of infected individuals in the form of a matrix equation, we can extract the Jacobian matrix (the 2x2 matrix of partial differential equations describing the change in the number of infected individuals in each group):

$$\begin{pmatrix} \dot{I}_h \\ \dot{I}_v \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} I_h \\ I_v \end{pmatrix} \tag{16}$$

The left side of eq 16 represents the derivative of the infected categories (represented here using dot notation rather than the Leibniz notation used earlier, for ease of visualisation): \dot{I}_h represents the rate of change in the number of infected hosts, and \dot{I}_v represents the rate of change in the number of infected vectors. The first term on the right of eq 16 is the Jacobian matrix, and the second term describes the current state of the infected hosts (I_h) and vectors (I_v).

We describe in S2 Text how eq 16 can be solved in order to estimate the number (and the proportion) of infected individuals at any time point during exponential growth, and how this

relates to the ratio $\left[\frac{\left(\frac{v_v}{\rho_v} \right)}{\left(\frac{v_h}{\rho_h} \right)} \right]$. We therefore need to calculate the eigenvector (v) associated with the dominant eigenvalue. We can do this by first calculating the trace (T) of the Jacobian matrix in eq 16 as $(a + d)$, and the determinant (D) of the matrix as $(ad - bc)$, which we can use to calculate the eigenvalues of the system. When we linearise our system around the disease-free steady state, the largest eigenvalue will approximate the initial exponential growth rate (r) for the system as a whole (i.e. the rate of increase in the number of both infected hosts

and vectors):

$$r \approx \left(\frac{T}{2}\right) + \sqrt{\left(\frac{T^2}{4}\right) - D} \tag{17}$$

The ratio of the values of the eigenvector associated with this eigenvalue will describe the relative numbers of infected hosts and vectors $\left(\frac{v_h}{v_v}\right)$ as exponential growth proceeds. Since r is fixed for the system as a whole, this ratio captures the heterogeneities between host and vector infection as the pathogen spreads through the system. Assuming that there is some transmission between the two groups, the ratio of eigenvectors can be calculated using the following formula:

$$\left(\frac{v_h}{v_v}\right) \approx \left(\frac{r-d}{c}\right) = \left(\frac{b}{r-a}\right) \tag{18}$$

Multiplying eq 18 with the ratio of vector to host numbers gives us an estimate of the ratio $\left[\frac{\left(\frac{v_h}{\rho_h}\right)}{\left(\frac{v_v}{\rho_v}\right)}\right]$ (which can be interpreted as an estimate of the relative proportions of infected hosts and vectors—the relative prevalences—as exponential growth proceeds):

$$\left(\frac{\left(\frac{v_h}{\rho_h}\right)}{\left(\frac{v_v}{\rho_v}\right)}\right) \approx \left(\frac{r-d}{c}\right) \left(\frac{\rho_v}{\rho_h}\right) = \left(\frac{b}{r-a}\right) \left(\frac{\rho_v}{\rho_h}\right) \tag{19}$$

Epidemiological models

To demonstrate our approach, we used a simple SI-type compartmental host-vector model framework [36] (described in S2 Text) to simulate the epidemiological dynamics of two important citrus pathogens. We used Southern Gardens Citrus, a commercial citrus plantation in south Florida, as the conceptual setting for our model, and parameterised the models as shown in Tables 1 and 2.

Table 1. Parameter values used in the estimation of the transmission parameters (β) for the two models in the current study.

Par	Interpretation	HLB model	Tristeza model
T	Duration of each visit (h)	5	6
ϕ	Rate of host visits by vector	1	1
b_v	Rate of acquisition by vectors	$\frac{0.8}{32d} = 0.02$	$\frac{0.8}{1d} = 0.8$
b_h	Rate of host inoculation	0.05	0.2

$\left(\beta_\kappa = \left(\frac{\phi}{\rho_h}\right)(1 - \exp(-b_\kappa T))\right)$, where κ refers to the receiving group. Most estimates are taken from [29, 37] for HLB and [38, 39] for tristeza. The duration of feeding per visit for the HLB model was taken from [40] and for the tristeza model was adjusted according to the total efficiency of CTV transmission described in [41]. All rates (ϕ , b_v , b_h) are per day.

<https://doi.org/10.1371/journal.pcbi.1005712.t001>

Table 2. Parameter values for different models used in the current study.

Par	Interpretation	HLB model	Tristeza model
ρ_h	Number of hosts	250,000	250,000
β_{hv}	Host-vector transmission rate	$2e-8$	$7e-7$
μ_h	Rate of infected host removal	$\left(\frac{1}{365 \times 5}\right) = 5e-4$	$\left(\frac{1}{365 \times 10}\right) = 3e-4$
τ_h	Rate of host recovery	0.0	0.0
π_h	Prob of graft transmission	0.0	0.0
ρ_v	Number of vectors	3,924,040	802,426
β_{vh}	Vector-host transmission rate	$4e-8$	$2e-7$
μ_v	Rate of infected vector removal	$\frac{1}{82}$	$\frac{1}{28}$
τ_v	Rate of vector recovery	0.0	$\frac{1}{2}$
π_v	Prob of transovarial transmission	0.036	0.0

All rates (β , μ , τ) are per day.

<https://doi.org/10.1371/journal.pcbi.1005712.t002>

The full system of ODEs for the model framework are given in [S2 Text](#), but the ODEs describing the numbers of infected hosts and vectors are as follows:

$$\frac{dI_h}{dt} = S_h I_v \beta_{vh} + (\pi_h - 1) \mu_h I_h - \tau_h I_h \tag{20}$$

$$\frac{dI_v}{dt} = S_v I_h \beta_{hv} + (\pi_v - 1) \mu_v I_v - \tau_v I_v \tag{21}$$

Linearising around the disease-free steady state, the components of the Jacobian matrix in [eq 16](#) for this system can be calculated:

$$a = \frac{\partial \dot{I}_h}{\partial I_h} \approx \mu_h (\pi_h - 1) - \tau_h \tag{22}$$

$$b = \frac{\partial \dot{I}_h}{\partial I_v} \approx \rho_h \beta_{vh} \tag{23}$$

$$c = \frac{\partial \dot{I}_v}{\partial I_h} \approx \rho_v \beta_{hv} \tag{24}$$

$$d = \frac{\partial \dot{I}_v}{\partial I_v} \approx \mu_v (\pi_v - 1) - \tau_v \tag{25}$$

We used this model framework to create models of two insect-vector citrus diseases of economic importance to the global citrus industry: huanglongbing (HLB) and tristeza diseases. The epidemiological unit in each model was an individual sweet orange tree (*Citrus × sinensis*) host, or single insect vector (Asian citrus psyllid, *Diaphorina citri* Kuwayama, or brown citrus aphid, *Toxoptera citricida* (Kirkaldy), respectively). Parameter values and sources for the two models are shown in [Tables 1 and 2](#). We assume that there is no differential immigration and emigration of infected vectors [\[42\]](#), that the total number of hosts and vectors does not change over time, and that there was no vertical transmission amongst hosts due to certification and testing of budwood source trees [\[43\]](#). We estimated transmission parameters using the approach described by Jeger and others [\[42, 44\]](#), and selected a suitable number of vectors to

achieve an overall R_0 of 100 when the number of hosts is fixed at 250,000 (see [S2 Text](#)). Our decision to fix the R_0 for each pathosystem at 100 and use this to calculate the relative densities of hosts and vectors was primarily intended to account for the lack of data on vector abundance, and to allow comparison of different pathogen types [45, 46].

Huanglongbing (also known as citrus greening) is a fatal disease of citrus and related plants caused by phloem-restricted gram negative Alphaproteobacteria of the genus *Candidatus Liberibacter* [47]. The most common species of *Liberibacter* worldwide is *Ca. L. asiaticus* (Las), which is the cause of ‘Asian citrus greening’, and is spread by the phloem-feeding Asian citrus psyllid. Las can be considered a ‘persistently transmitted, circulative pathogen’ [45, 46]. These pathogens enter the haemolymph of the vector and have the potential for transovarial transmission (although this is disputed in the particular case of Las [37, 48]).

Unlike many plant viruses, the citrus tristeza virus (CTV) complex comprises a number of strains which are responsible for a wide range of syndromes in citrus and their relatives [49–51]. Although strains were traditionally differentiated according to disease phenotype [49, 52, 53], this relationship remains unclear [50, 54], and we therefore focus on the spread of an undefined ‘novel’ CTV strain by the brown citrus aphid (considered the most efficient vector of CTV [55]). CTV is considered a ‘semipersistently transmitted, foregut-borne’ pathogen [45, 46], which does not spread systemically and therefore is characterised by rapid acquisition [49, 52, 56] and short persistence [39, 57].

Sampling simulation

In order to assess how well our sampling models (eqs 6 to 8) performed, we created a model to simulate the sampling process using a Monte Carlo approach [58] with 1000 iterations. For each iteration, we used the output of the full ODE transmission model to indicate the spread of our pathogen through a susceptible population, and simulated a sampling process during the resultant epidemic by randomly selecting a series of timepoints from the model output, accounting for the probability of detection at each. We used the specified sampling interval to estimate the timing of first sampling and the interval between subsequent samples, and we calculated the prevalence in hosts and vectors at each of these points, along with the probability of detection given the sample size (using the binomial sampling strategy described in eq 2). In order to convert these probabilistic estimates into a dichotomous classification of whether the pathogen was successfully detected or not, we generated a pseudorandom number between 0 and 1 for each sampling point and classified detection as ‘successful’ if this number was less than or equal to our estimated detection probability. We then recorded the earliest time of first detection for the iteration in question, and identified the associated host and vector prevalences at this point. We assumed that a total of 800 samples were collected and tested per month (based upon data provided by the United States Sugar Corporation to the Citrus Greening Symposium in 2009, detailing laboratory testing instigated in Southern Gardens Citrus during 2006 and 2007 [59]).

For ease of interpretation, we conducted most analysis assuming a cost ratio at the ‘threshold’ of $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right]$ (indicating the cost ratio at which point the total costs and the prevalences at first detection would be expected to be equal regardless of which stratum was sampled).

Sensitivity analysis

We first investigated the effect of varying sampling effort on the mean prevalence at first detection amongst hosts and vectors. Since vector sampling was not routinely performed in

Southern Gardens, we assumed similar parameters to host sampling (i.e. 800 vectors per month, individually tested). We then investigated the prevalence at first detection when either hosts or vectors are exclusively sampled within fixed cost constraints, and conducted a brief sensitivity analysis of our model parameter estimates on the model outputs, focussing on transmission rates (β). Centering on a ‘threshold’ cost ratio (as described above), we adjusted the parameter values by a factor of ten in order to investigate the effect on the value of the ratio

$$\left[\frac{\left(\frac{v_h}{\rho_h}\right)}{\left(\frac{v_v}{\rho_v}\right)} \right]$$

and the mean prevalence in each stratum at first detection.

Analyses were conducted using R (version 3.3.1) [79] and the Anaconda distribution (version 2.4.0; Continuum Analytics, <https://continuum.io/>) of the Python programming language (version 3.5.0; Python Software Foundation, <https://www.python.org/>). Full code is provided in S3, S4, S5 and S6 Text.

Results

The number of vectors per host required to achieve an R_0 of 100 was 16 for the HLB model, and 3 for the tristeza model (reflecting the higher transmission rates in this model). The transmission dynamics of the two models over a period of three years (including the exponential growth approximation upon which the rule of thumb is based) are shown in S1 and S2 Figs.

Using the stability analysis technique described above, the rate of exponential growth (r) for the HLB system was 0.02, and the ratio $\left[\frac{\left(\frac{v_h}{\rho_h}\right)}{\left(\frac{v_v}{\rho_v}\right)} \right]$ (see eq 15) was 8. Assuming that the ratio of

sampling costs was equal to this would mean that sufficient resources would be available to sample either 800 hosts or 6,382 vectors per month. As expected, when either hosts or vectors alone were sampled at this rate, the mean time of first detection from the simulation model was 266 days for exclusive host sampling, and 267 days for exclusive vector sampling. The simulation model and the heuristic both predicted a mean host prevalence of 0.0007 and a mean vector prevalence of 0.0001 at first detection, regardless of which group was sampled.

The estimate of the exponential growth parameter (r) obtained from the tristeza model was 0.03, and the ratio $\left[\frac{\left(\frac{v_h}{\rho_h}\right)}{\left(\frac{v_v}{\rho_v}\right)} \right]$ (eq 15) was 6. When the ratio of sampling costs was set to this, the

available resources would allow sampling of either 800 hosts or 4,687 vectors every 28 days, which both gave a simulated time of first detection of 164 days. The heuristic predicted a mean prevalence in hosts at first detection of 0.0009 and the simulation model predicted a mean prevalence of 0.0010 at first detection, regardless of which group was sampled. Similarly, amongst vectors, both methods predicted a mean prevalence of 0.0002, regardless of which group was sampled. Graphs of the distribution of the timing of first detection are shown in S3 and S4 Figs, and graphs of the prevalence distribution at first detection are shown in S3 and S4 Figs. Fig 1 shows the effect of varying the sampling effort (regardless of cost) on the mean prevalence at first detection using the heuristic described in eqs 9 and 10.

Fig 2 shows the effect of varying the transmission parameters on the ratio $\left[\frac{\left(\frac{v_h}{\rho_h}\right)}{\left(\frac{v_v}{\rho_v}\right)} \right]$ (shown on the log scale to assist visualisation). Similar graphs for the host and vector longevity parameters, along with the effect of varying transmission and longevity parameters on the mean prevalence at first detection, are shown in S9–S12 Figs.

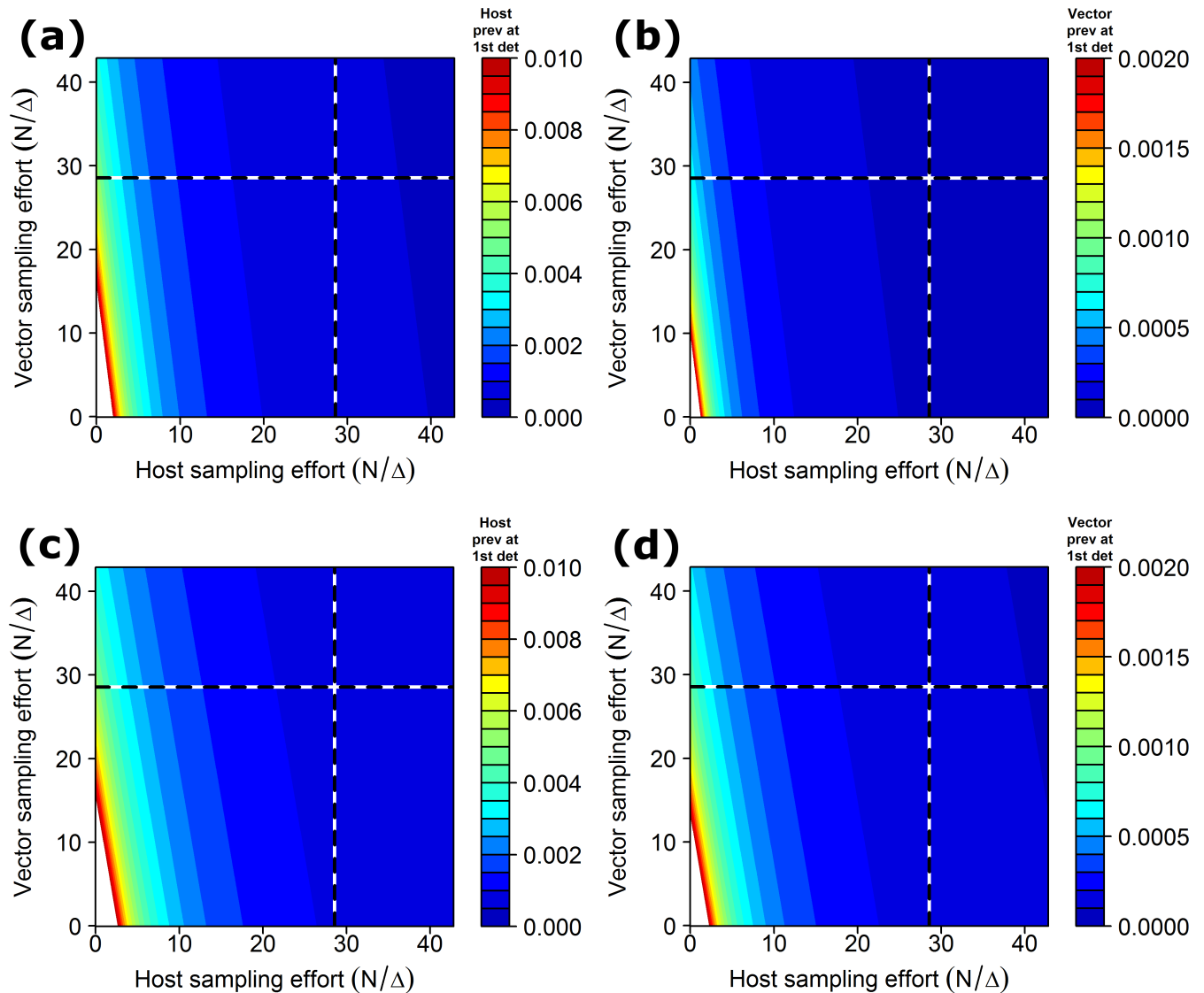


Fig 1. Effect of varying sampling effort ($\theta = \frac{N}{\Delta}$) on the mean prevalence at first detection for the HLB model (panels (a) and (b) and the tristeza model (panels (c) and (d)). The estimated prevalence at first detection in hosts is shown in the graphs on the left, and that in vectors is shown in the graphs on the right. The dashed line indicates a host (vertical line) and a vector (horizontal line) sampling effort of 800 samples per 28 days, with the intersection of these dashed lines indicating a theoretical scenario in which a total of 800 hosts and 800 vectors were sampled.

<https://doi.org/10.1371/journal.pcbi.1005712.g001>

Discussion

The protection of natural and managed ecosystems against the incursion of emerging pathogens increasingly relies upon the use of planned surveillance activities to detect pathogen entry at a suitably early time to allow control measures to be implemented. Failure of early detection can have catastrophic consequences, as was observed in the UK in 2001 following entry of the foot and mouth disease virus [60], and has been predicted due to chalara dieback of ash trees (caused by *Hymenoscyphus fraxineus*) throughout Europe [61, 62]. Although the risk of infection generally varies between different epidemiological groups within a single pathosystem, this heterogeneity has been commonly overlooked when planning surveillance activities. One

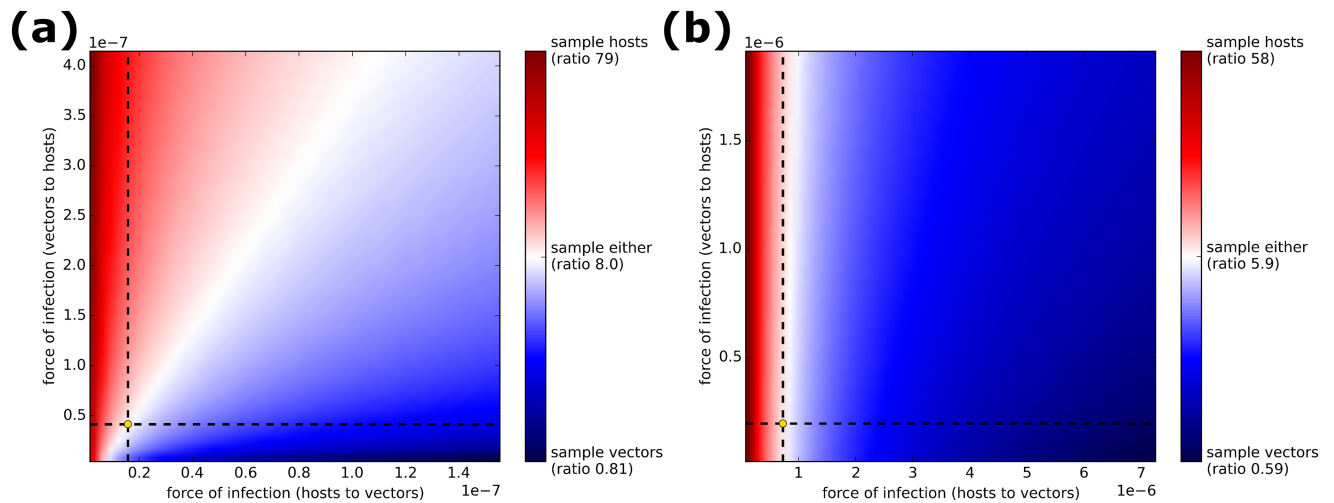


Fig 2. Effect of varying transmission parameters (β) on the suggested group of sampling for the HLB model (panel (a)) and the tristeza model (panel (b)). We estimate the relative sampling efforts required from vectors compared to that from hosts when using the current model parameters (located at the intersection of the dashed lines) using the ratio $\left[\frac{y_h}{y_v}\right]$, and assume that the relative cost of sampling hosts compared to vectors is equal to this threshold (8 for HLB, 6 for Tristeza)—indicating the ‘equivalence point’ as described in the text. The numbers in the key on the right describe the relative vector sampling effort $\left[\frac{y_h}{y_v}\right]$ for different transmission rates, but the colour gradient relates to the ratio of the relative vector sampling effort to the relative host sampling cost $\left[\frac{y_h}{y_v}\right]$, and is shown on the log scale in order to better discriminate values less than 1. Regions shown in red have a sampling effort ratio greater than the cost ratio (suggesting that sampling hosts would minimise the total cost) and those in blue have a ratio less than the cost ratio (suggesting that sampling vectors would minimise the total cost). The frontier between these two (indicating a ratio equal to the cost ratio) is shown in white.

<https://doi.org/10.1371/journal.pcbi.1005712.g002>

particular example of this heterogeneity is seen in the case of pathogens spread by insect vectors, which are increasingly identified as ‘emerging pathogens’ and are a current source of considerable concern due to their potential impact upon animal and plant health [63]. Despite the clear epidemiological differences between ‘hosts’ and ‘vectors’, relatively little work has been conducted to date on how best to distribute surveillance resources between these groups in order to ensure that incursion of these emerging pathogens is rapidly detected by ongoing surveillance activities (‘early detection surveillance’).

In the current paper, we use vector-borne pathogens as an example of a ‘heterogeneous epidemiological system’, and describe how the prevalence at first detection in both hosts and vectors is related to the rate of sampling from these groups (eqs 6 to 8). We have developed a

heuristic, which is parameterised using the rate of exponential growth, r , and a ratio, $\left[\frac{y_h}{y_v}\right]$,

which describes the relative prevalences in each group during early exponential growth (and can also be interpreted as the relative sampling effort required from vectors when they are exclusively sampled, compared to that required from hosts). Both of these parameters can easily be obtained from a simple mathematical model of the system in question (as described in the Sampling model parameterisation section). Our heuristic is based upon an assumption of exponential growth in the prevalence of infection. Although this is unlikely to be epidemiologically realistic beyond the initial stages of epidemic growth, we find that our output (the probability of first detection), is constrained by the rapidly decreasing probability of having failed to

detect the pathogen earlier, as reported in our previous work [17, 24]. Therefore, given a suitable sampling interval, the importance of being able to accurately estimate the prevalence decreases as the prevalence grows.

An important output of our method is the heuristic described in eqs 6 to 8. We can use this heuristic directly to evaluate ongoing or planned surveillance activities, in particular by predicting the distribution of prevalences (or mean prevalence) at first detection in either group, assuming a particular rate of sampling from each group. Alternatively, we can reformulate it in order to assist in surveillance planning, by estimating the sampling rate required in order to detect a specified mean prevalence (or specified prevalence percentile) in either group.

Another useful output of our work is shown in eq 15. This simple heuristic is focussed on

direct interpretation of the ratio $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right]$, in order to determine whether hosts or vectors

should be sampled in order to minimise both the prevalence at first detection and the total ‘cost’ of sampling, and can be used in two main ways:

- By explicitly specifying the sampling costs and adopting a dichotomised ‘threshold’ interpretation based upon the ratio $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right]$, such as that described above.
- By using the ratio $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right]$ to quantify the ratio of sampling costs at which the suggested sampling strategy would change (possibly in combination with sensitivity analysis, such as that shown in Fig 2). This strategy may be useful when the true ratio of sampling costs is less well known.

With the exception of our earlier work [17, 24, 25], the only other study we know of which attempted to develop a heuristic for evaluating early detection surveillance focussed on the estimation of the probability of detection before a specified prevalence was reached [64]. As our methods are able to estimate the whole probability distribution of the prevalence at the time of first detection, we are also able to estimate this probability if desired, along with measures such as the average prevalence at first detection.

The concept of surveillance within a host-vector system has been previously studied by Ferguson and others [65], who found that the relative prevalences and the costs of sampling determined the probability of detection in any group at any single sampling point. This shares

similarities with our own formulation, since the ratio $\left[\frac{\left(\frac{v_h}{\rho_h} \right)}{\left(\frac{v_v}{\rho_v} \right)} \right]$ can be considered the ratio of

prevalences amongst hosts and vectors during initial exponential growth. As with our approach, the Ferguson model had threshold-like behaviour in which the optimal sampling strategy suddenly changed, with the optimal sampling strategy generally being to focus on a single group of interest [65].

Finally, we considered similarities between our work and the body of literature on early detection surveillance within the more general field of invasion biology. Although a number of studies have investigated how to improve the early detection and control of invasive species, these generally considered the issue as an optimisation problem—using complex simulation models to determine the optimal strategy for surveillance and control, often in conjunction with economic modelling [22, 66–70]. Our method differs from these in that it does not

require the creation of a complex model, but is still able to account for important biological properties of the pathosystem in question, including epidemiological groupings. Indeed, it has been argued that simple heuristics such as ours can be particularly useful for decision makers, since they can reduce a complex system down into a more manageable and understandable form [71].

Our framework assumes that the surveillance strategy in place is able to detect asymptomatic infection, and that the diagnostic test used is applied regardless of perceived infection status. As mentioned earlier, visual detection is the most commonly used method of first line diagnosis of infection status for plant pathogens, which is likely to be a problem for effective early detection surveillance [29], and is not compatible with our methodology in the case of most plant pathogens. Another repercussion of basing early detection surveillance strategies on visual detection is that consideration is rarely given to first line detection of infection in insect vectors (which generally do not show clinical signs following infection). Whilst we therefore suggest the use of highly sensitive tests able to detect asymptomatic infection, the additional immediate costs of applying these diagnostic methods means that extra consideration must be given to targeting those individuals most likely to be infected [30]. We achieve this in our framework by considering how to minimise total cost or effort by sampling exclusively from either hosts or vectors. However, our framework is equally capable of evaluating surveillance activities in which a combination of hosts and vectors are sampled, and could therefore be used by growers or regulatory agencies to help plan and evaluate ongoing surveillance activities.

We demonstrate the application of the current framework by developing two simple models of important vector-borne citrus pathogens: Las (the cause of HLB, which is a current emerging threat to the Californian citrus industry [72]) and CTV (the cause of citrus tristeza syndromes which have historically shaped the global citrus industry [49]) and base these models on a large plantation in south Florida. Despite arbitrarily setting R_0 at 100, our estimates of r are comparable to those reported in the literature (r for Las has been estimated as between 0.002 and 0.01 [24], and that for CTV around 0.008 [73]). Our analysis of both the HLB and tristeza pathosystems suggested that sampling exclusively from hosts would minimise the total sampling effort, but that if the cost of sampling an individual host is more than eight times (sampling for Las) or six times (sampling for CTV) that of a single vector, vectors should instead be sampled in order to minimise total sampling costs. We do not attempt to estimate sampling costs, since it could be argued that pooled testing of multiple vectors together would raise the ratio of host to vector sample testing costs, whereas the additional effort required to capture motile vectors compared to sampling sessile hosts would lower the ratio of host to vector sample collection costs. Instead, we identify the ratio of sampling costs at which the sug-

gested sampling strategy would change, using the ratio $\left[\frac{\left(\frac{\nu_h}{\rho_h} \right)}{\left(\frac{\nu_v}{\rho_v} \right)} \right]$ as described in eq 15 and the

associated text. As well as correctly identifying this ‘equivalence point’ at which either hosts or vectors could reasonably be sampled in order to minimise sampling costs, our heuristic agreed well with results obtained from a simulated sampling model.

We investigated the effect of varying the transmission parameters on the suggested stratum of sampling. Fig 2 shows that an increase in host to vector transmission favoured vector sampling for both pathogens (associated with an increase in the relative prevalence amongst vectors), but that varying the rate of vector to host transmission only affected the suggested group of sampling in the HLB model (with higher values favouring host sampling). The lack of an effect of vector to host transmission in the tristeza model likely represents the constraining effect of the short duration of virus persistence in vectors.

Both the Las and CTV pathosystems are characterised by latent and incubation periods [12, 29, 74], and irregular distribution of the pathogen within the host [75–77], meaning that current available tests are imperfect (although work is currently in place to improve these tests). These characteristics would be expected to impact upon the optimal sampling strategy but are not explicitly captured in our current model. The issue of latency is a particularly important one for emerging plant pathogens [30], and has previously been used as an argument for sampling vectors instead of hosts for detection of Las [28]. It may be possible to adjust the statistical framework underlying our framework in order to capture latency [25] and imperfect test sensitivity, but incubation (where an individual is infected but not infectious) cannot be easily captured since the underlying mathematical model must be fully identifiable from the numbers of infected hosts and vectors at any time point. Also, we have purposefully selected a simple model for the costs of sampling and testing, with equal fixed costs and a linear cost function for variable costs. Further work is therefore needed to investigate the impact of these epidemiological and economic assumptions on model predictions, and to incorporate characteristics of importance into the framework.

Although we have described our approach using examples of host-vector systems, our framework should be applicable to any ‘heterogeneous’ epidemiological or ecological system—given that there is some transmission between the two groups (if this is not the case, each group should be sampled independently using our earlier frameworks [17, 24, 25]). As well as incorporating imperfect test performance and latency, further work will focus on investigation of the effect of nonlinear cost functions (since the per-sample collection cost would be expected to decrease as the surveillance intensity increases [78]), differences in fixed costs, generalisation to systems containing more than two linked epidemiological groups (offering the potential for investigating multiple hosts and/or vectors), ‘temporal targeting’ of surveillance effort by accounting for seasonality in the epidemiological system, and evaluation using more realistic, spatially explicit, transmission models.

Conclusion

We propose an epidemiologically-informed approach to help answer the question of where best to place sampling resources for early detection of emerging pathogens in a system comprised of two epidemiologically distinct, but connected, groups (such as hosts and vectors). We show that the prevalence at first detection in each group can be estimated using a simple heuristic which, although novel, can be considered a generalisation of that from our own previous work [17, 24]. We demonstrate how to parameterise this heuristic using two epidemiological parameters which can be extracted from a system of ordinary differential equations: these are the initial rate of exponential growth of the pathogen in the system (r), and a ratio,

$\left[\frac{\left(\frac{y_H}{p_H} \right)}{\left(\frac{y_V}{p_V} \right)} \right]$, which describes the relative prevalences in each group during exponential growth.

We also show that the optimal strategy for minimising the total sample size (or the total sampling cost, if a linear cost function is assumed) will generally be to sample from a single group rather than both. Although this is contrary to many surveillance strategies, it is conceptually related to the idea of ‘risk-based’ surveillance, which is increasingly used for early detection surveillance. We have validated our approach using simple transmission models, but further work is needed to evaluate how well it performs in the face of more realistic, spatially explicit, transmission models.

Supporting information

S1 Text. Details on the full derivation of the sampling model.

(PDF)

S2 Text. Additional details on the transmission model.

(PDF)

S3 Text. Python 3.x code for sampling simulation model.

(PY)

S4 Text. Python 3.x code for single parameter sensitivity analysis model.

(PY)

S5 Text. Python 3.x code for dual parameter sensitivity analysis model.

(PY)

S6 Text. Simple R code for estimating the heuristic.

(R)

S1 Fig. HLB model transmission dynamics. Host and vector transmission dynamics in HLB model over the course of two years. Hosts are shown in panel (a) and vectors in panel (b). The relative densities of hosts and vectors for both models was fixed in order to give an R_0 estimate of 100.

(TIF)

S2 Fig. Tristeza model transmission dynamics. Host and vector transmission dynamics in the tristeza model over the course of two years. Hosts are shown in panel (a) and vectors in panel (b). The relative densities of hosts and vectors for both models was fixed in order to give an R_0 estimate of 100.

(TIF)

S3 Fig. Simulated distribution of time of first detection for the HLB model. Simulated distribution of time of first detection at the cost ratio threshold with a sampling ‘cost’ equivalent to that of 800 hosts every 28 days in the HLB model (i.e. either 800 hosts or 6,382 vectors). Panel (a) shows the results predicted when sampling 800 hosts and no vectors, and Panel (b) shows those predicted when sampling 6,382 vectors and no hosts. The dotted lines show the average time at first detection.

(TIF)

S4 Fig. Simulated distribution of time of first detection for the tristeza model. Predicted distribution of time of first detection at the cost ratio threshold with a sampling ‘cost’ equivalent to that of 800 hosts every 28 days in the tristeza model (i.e. either 800 hosts or 4,687 vectors). Panel (a) shows the results predicted when sampling 800 hosts and no vectors, and Panel (b) shows those predicted when sampling 4,687 vectors and no hosts. The dotted lines show the average time at first detection.

(TIF)

S5 Fig. Predicted distribution of prevalence at first detection for the HLB model. Predicted distribution of prevalence at first detection at the cost ratio threshold with a sampling ‘cost’ equivalent to that of 800 hosts every 28 days in the HLB model (i.e. either 800 hosts or 6,382 vectors), using both model simulation and the heuristic (‘rule of thumb’). Host prevalence at first detection is shown in panels (a) and (c), and vector prevalence in panels (b) and (d). Panels (a) and (b) show the results when sampling only from hosts, and panels (c) and (d) show

those predicted when vectors alone are sampled. Dotted lines show the mean prevalence at first detection.

(TIF)

S6 Fig. Predicted distribution of prevalence at first detection for the tristeza model. Predicted distribution of prevalence at first detection at the cost ratio threshold with a sampling 'cost' equivalent to that of 800 hosts every 28 days in the tristeza model (i.e. either 800 hosts or 4,687 vectors), using both model simulation and the heuristic ('rule of thumb'). Host prevalence at first detection is shown in panels (a) and (c), and vector prevalence in panels (b) and (d). Panels (a) and (b) show the results when sampling only from hosts, and panels (c) and (d) show those predicted when vectors alone are sampled. Dotted lines show the mean prevalence at first detection.

(TIF)

S7 Fig. Effect of varying longevity parameters (μ) on the suggested group of sampling.

Effect of varying longevity parameters (μ) on the suggested group of sampling for the HLB model (panel (a)) and the tristeza model (panel (b)), assuming a sampling cost ratio at the threshold (8 for HLB, 6 for Tristeza). The intersection of the dashed lines shows the current parameter values. The colour gradient relates to the ratio $\left[\frac{(\frac{\nu_h}{\rho_h})}{(\frac{\nu_v}{\rho_v})} \right]$, and is shown on the log scale.

Red indicates a ratio greater than the cost ratio (suggesting host sampling) and blue indicates a ratio less than the cost ratio (suggesting vector sampling).

(TIF)

S8 Fig. Effect of varying numbers of hosts and vectors (ρ parameters) on the suggested group of sampling. Effect of varying numbers of hosts and vectors (ρ parameters) on the suggested stratum of sampling for the HLB model (panel (a)) and the tristeza model (panel (b)), assuming a sampling cost ratio at the threshold (8 for HLB, 6 for Tristeza). The intersection of the dashed lines shows the current parameter values. The colour gradient relates to the ratio

$\left[\frac{(\frac{\nu_h}{\rho_h})}{(\frac{\nu_v}{\rho_v})} \right]$, and is shown on the log scale. Red indicates a ratio greater than the cost ratio (suggesting host sampling) and blue indicates a ratio less than the cost ratio (suggesting vector sampling).

(TIF)

S9 Fig. Effect of varying transmission rates (β parameters) on the mean prevalence at first detection for the HLB model. Effect of varying transmission rates (β parameters) on the mean prevalence at first detection for the HLB model (host prevalence shown in panels (a) and (c) and vector prevalence in panels (b) and (d)). Red lines show the estimated prevalence when 800 hosts are sampled every 28 days, and blue lines show the estimated prevalence when 6,382 vectors are sampled every 28 days. Plots in panels (a) and (b) show the effect of varying host to vector transmission, and those in panels (c) and (d) show the effect of varying vector to host transmission. The dashed line shows the parameter value used in the model. The transmission parameters have units of 'infections per host per vector per day'

(TIF)

S10 Fig. Effect of varying transmission rates (β parameters) on the mean prevalence at first detection for the tristeza model. Effect of varying transmission rates (β parameters) on the mean prevalence at first detection for the tristeza model (host prevalence shown on the left and vector prevalence on the right). Red lines show the estimated prevalence when 800 hosts

are sampled every 28 days, and blue lines show the estimated prevalence when 4,687 vectors are sampled every 28 days. Plots in panels (a) and (b) show the effect of varying host to vector transmission, and those in panels (c) and (d) show the effect of varying vector to host transmission. The dashed line shows the parameter value used in the model. The transmission parameters have units of ‘infections per host per vector per day’ (TIF)

S11 Fig. Effect of varying longevity (μ parameters) on the mean prevalence at first detection for the HLB model. Effect of varying longevity (μ parameters) on the mean prevalence at first detection for the HLB model (host prevalence shown on the left and vector prevalence on the right). Red lines show the estimated prevalence when 800 hosts are sampled every 28 days, and blue lines show the estimated prevalence when 6,382 vectors are sampled every 28 days. Plots in panels (a) and (b) show the effect of varying host longevity, and those in panels (c) and (d) show the effect of varying vector longevity. The dashed line shows the parameter value used in the model. (TIF)

S12 Fig. Effect of varying longevity (μ parameters) on the mean prevalence at first detection for the tristeza model. Effect of varying longevity (μ parameters) on the mean prevalence at first detection for the tristeza model (host prevalence shown on the left and vector prevalence on the right). Red lines show the estimated prevalence when 800 hosts are sampled every 28 days, and blue lines show the estimated prevalence when 4,687 vectors are sampled every 28 days. Plots in panels (a) and (b) show the effect of varying host longevity, and those in panels (c) and (d) show the effect of varying vector longevity. The dashed line shows the parameter value used in the model. (TIF)

Author Contributions

Conceptualization: Alexander J. Mastin, Frank van den Bosch, Timothy R. Gottwald, Stephen R. Parnell.

Formal analysis: Alexander J. Mastin, Frank van den Bosch, Timothy R. Gottwald, Vasthi Alonso Chavez, Stephen R. Parnell.

Funding acquisition: Frank van den Bosch, Timothy R. Gottwald, Stephen R. Parnell.

Investigation: Alexander J. Mastin, Timothy R. Gottwald, Stephen R. Parnell.

Methodology: Alexander J. Mastin, Frank van den Bosch, Vasthi Alonso Chavez, Stephen R. Parnell.

Resources: Stephen R. Parnell.

Software: Alexander J. Mastin, Stephen R. Parnell.

Supervision: Frank van den Bosch, Timothy R. Gottwald, Stephen R. Parnell.

Validation: Alexander J. Mastin, Stephen R. Parnell.

Visualization: Alexander J. Mastin, Stephen R. Parnell.

Writing – original draft: Alexander J. Mastin, Frank van den Bosch, Timothy R. Gottwald, Vasthi Alonso Chavez, Stephen R. Parnell.

Writing – review & editing: Alexander J. Mastin, Stephen R. Parnell.

References

1. Chapin FS 3rd, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds HL, Hooper DU, Lavorel S, Sala OE, Hobbie SE, Mack M, Diaz S. Consequences of changing biodiversity. *Nature*. 2000; 405(6783):234–242. <https://doi.org/10.1038/35012241> PMID: 10821284
2. Brasier CM. The biosecurity threat to the UK and global environment from international trade in plants. *Plant Pathology*. 2008; 57(5):792–808. <https://doi.org/10.1111/j.1365-3059.2008.01886.x>
3. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution*. 2004; 19(10):535–544. <https://doi.org/10.1016/j.tree.2004.07.021> PMID: 16701319
4. Dehnen-Schmutz K, Holdenrieder O, Jeger MJ, Pautasso M. Structural change in the international horticultural industry: some implications for plant health. *Scientia Horticulturae*. 2010; 125(1):1–15. <https://doi.org/10.1016/j.scienta.2010.02.017>
5. Wilkinson K, Grant WP, Green LE, Hunter S, Jeger MJ, Lowe P, Medley GF, Mills P, Phillipson J, Poppy GM, Waage J. Infectious diseases of animals and plants: an interdisciplinary approach. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2011; 366(1573):1933–1942. <https://doi.org/10.1098/rstb.2010.0415>
6. Waage JK, Mumford JD. Agricultural biosecurity. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2008; 363(1492):863–876. <https://doi.org/10.1098/rstb.2007.2188>
7. Rizzo DM, Garbelotto M, Hansen EM. *Phytophthora ramorum*: integrative research and management of an emerging pathogen in California and Oregon forests. *Annual Review of Phytopathology*. 2005; 43:309–335. <https://doi.org/10.1146/annurev.phyto.42.040803.140418> PMID: 16078887
8. Mitchell RJ, Beaton JK, Bellamy PE, Broome A, Chetcuti J, Eaton S, Ellis CJ, Gimona A, Harmer R, Hester AJ, Hewison RL, Hodgetts NG, Iason GR, Kerr G, Littlewood NA, Newey S, Potts JM, Pozsgai G, Ray D, Sim DA, Stockan JA, Taylor AFS, Woodward S. Ash dieback in the UK: a review of the ecological and conservation implications and potential management options. *Biological Conservation*. 2014; 175:95–109. <https://doi.org/10.1016/j.biocon.2014.04.019>
9. Ordonez N, Seidl MF, Waalwijk C, Drenth A, Kilian A, Thomma BPHJ, Ploetz RC, Kema GHJ. Worse comes to worst: bananas and Panama disease—when plant and pathogen clones meet. *PLOS Pathogens*. 2015; 11(11):e1005197. <https://doi.org/10.1371/journal.ppat.1005197> PMID: 26584184
10. Singh RP, Hodson DP, Huerta-Espino J, Jin Y, Njau P, Wanyera R, Herrera-Foessel SA, Ward RW. Will stem rust destroy the world's wheat crop? *Advances in Agronomy*. 2008; 98:271–309. [https://doi.org/10.1016/S0065-2113\(08\)00205-8](https://doi.org/10.1016/S0065-2113(08)00205-8)
11. Martelli GP, Boscia D, Porcelli F, Saponari M. The olive quick decline syndrome in south-east Italy: a threatening phytosanitary emergency. *European Journal of Plant Pathology*. 2015; 144(2):235–243. <https://doi.org/10.1007/s10658-015-0784-7>
12. Gottwald TR. Current epidemiological understanding of citrus huanglongbing. *Annual Review of Phytopathology*. 2010; 48(1):119–139. <https://doi.org/10.1146/annurev-phyto-073009-114418> PMID: 20415578
13. Bebbier DP, Holmes T, Gurr SJ. The global spread of crop pests and pathogens. *Global Ecology and Biogeography*. 2014; 23(12):1398–1407. <https://doi.org/10.1111/geb.12214>
14. Mills P, Dehnen-Schmutz K, Ilbery B, Jeger M, Jones G, Little R, MacLeod A, Parker S, Pautasso M, Pietravalle S, Maye D. Integrating natural and social science perspectives on plant disease risk, management and policy formulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2011; 366(1573):2035–2044. <https://doi.org/10.1098/rstb.2010.0411>
15. WTO. Agreement on the Application of Sanitary and Phytosanitary Measures. World Trade Organization; 1995. Available from: https://www.wto.org/english/docs_e/legal_e/15-sps.pdf.
16. Magarey RD, Colunga-Garcia M, Fieselman DA. Plant biosecurity in the United States: roles, responsibilities, and information needs. *Bioscience*. 2009; 59(10):875–884. <https://doi.org/10.1525/bio.2009.59.10.9>
17. Parnell S, Gottwald TR, Gilks WR, van den Bosch F. Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. *Journal of Theoretical Biology*. 2012; 305:30–36. <https://doi.org/10.1016/j.jtbi.2012.03.009> PMID: 22480434
18. Cunniffe NJ, Cobb RC, Meentemeyer RK, Rizzo DM, Gilligan CA. Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. *Proceedings of the National Academy of Sciences*. 2016; <https://doi.org/10.1073/pnas.1602153113>
19. Thompson RN, Cobb RC, Gilligan CA, Cunniffe NJ. Management of invading pathogens should be informed by epidemiology rather than administrative boundaries. *Ecological Modelling*. 2016; 324: 28–32. <https://doi.org/10.1016/j.ecolmodel.2015.12.014> PMID: 27019546

20. Burnett KM. Introductions of invasive species: failure of the weaker link. *Agricultural and Resource Economics Review*. 2006; 35(1):21–28. <https://doi.org/10.1017/S1068280500010029>
21. Perrings C, Williamson M, Barbier EB, Delfino D, Dalmazzone S, Shogren J, Simmons P, Watkinson A. Biological invasion risks and the public good: an economic perspective. *Conservation Ecology*. 2002; 6(1). <https://doi.org/10.5751/ES-00396-060101>
22. Mehta SV, Haight RG, Homans FR, Polasky S, Venette RC. Optimal detection and control strategies for invasive species management. *Ecological Economics*. 2007; 61(2–3):237–245. <https://doi.org/10.1016/j.ecolecon.2006.10.024>
23. Venette RC, Moon RD, Hutchison WD. Strategies and statistics of sampling for rare individuals. *Annual Review of Entomology*. 2002; 47(1):143–174. <https://doi.org/10.1146/annurev.ento.47.091201.145147> PMID: 11729072
24. Parnell S, Gottwald TR, Cunniffe NJ, Alonso Chavez V, van den Bosch F. Early detection surveillance for an emerging plant pathogen: a rule of thumb to predict prevalence at first discovery. *Proceedings of the Royal Society B: Biological Sciences*. 2015; 282(1814):20151478. <https://doi.org/10.1098/rspb.2015.1478> PMID: 26336177
25. Alonso Chavez V, Parnell S, van den Bosch F. Monitoring invasive pathogens in plant nurseries for early-detection and to minimise the probability of escape. *Journal of Theoretical Biology*. 2016; 407:290–302. <https://doi.org/10.1016/j.jtbi.2016.07.041> PMID: 27477202
26. Stärk KDC, Regula G, Hernandez J, Knopf L, Fuchs K, Morris RS, Davies P. Concepts for risk-based surveillance in the field of veterinary medicine and veterinary public health: Review of current approaches. *BMC Health Services Research*. 2006; 6:20. <https://doi.org/10.1186/1472-6963-6-20> PMID: 16507106
27. Martin PAJ, Cameron AR, Barford K, Sergeant ESG, Greiner M. Demonstrating freedom from disease using multiple complex data sources 2: case study—classical swine fever in Denmark. *Preventive Veterinary Medicine*. 2007; 79(2-4):98–115. PMID: 17239459
28. Manjunath KL, Halbert SE, Ramadugu C, Webb S, Lee RF. Detection of *Candidatus Liberibacter asiaticus* in *Diaphorina citri* and its importance in the management of citrus huanglongbing in Florida. *Phytopathology*. 2008; 98(4):387–396. <https://doi.org/10.1094/PHYTO-98-4-0387> PMID: 18944186
29. Lee JA, Halbert SE, Dawson WO, Robertson CJ, Keesling JE, Singer BH. Asymptomatic spread of huanglongbing and implications for disease control. *Proceedings of the National Academy of Sciences*. 2015; 112(24):7605–7610. <https://doi.org/10.1073/pnas.1508253112>
30. Thompson RN, Gilligan CA, Cunniffe NJ. Detecting presymptomatic infection is necessary to forecast major epidemics in the earliest stages of infectious disease outbreaks. *PLOS Computational Biology*. 2016; 12(4):1–18. <https://doi.org/10.1371/journal.pcbi.1004836>
31. Irey MS, Gast T, Gottwald TR. Comparison of visual assessment and polymerase chain reaction assay testing to estimate the incidence of the huanglongbing pathogen in commercial Florida citrus. *Proceedings of the Florida State Horticultural Society*. 2006; 119:89–93.
32. RISKSUR Consortium. RISKSUR (Providing a new generation of methodologies and tools for cost-effective risk-based animal health surveillance systems for the benefit of livestock producers, decision makers and consumers); 2016. Available from: http://cordis.europa.eu/result/rcn/186796_en.html.
33. DEFRA. Protecting Plant Health—A Plant Biosecurity Strategy for Great Britain; 2014.
34. Doherr MG, Audigé L. Monitoring and surveillance for rare health-related events: a review from the veterinary perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2001; 356(1411):1097–1106. <https://doi.org/10.1098/rstb.2001.0898>
35. Parnell S, Gottwald TR, Riley T, van den Bosch F. A generic risk-based surveying method for invading plant pathogens. *Ecological Applications*. 2014; 24(4):779–790. <https://doi.org/10.1890/13-0704.1> PMID: 24988776
36. Keeling MJ, Rohani P. *Modeling infectious diseases in humans and animals*. Princeton University Press; 2008.
37. Pelz-Stelinski KS, Brlansky RH, Ebert TA, Rogers ME. Transmission parameters for *Candidatus Liberibacter asiaticus* by Asian citrus psyllid (Hemiptera: Psyllidae). *Journal of Economic Entomology*. 2010; 103(5):1531–1541. <https://doi.org/10.1603/EC10123> PMID: 21061950
38. Roistacher CN, Bar-Joseph M. Aphid transmission of citrus tristeza virus: a review. *Phytophylactica*. 1987; 19(2):163–167.
39. Costa AS, Grant TJ. Studies on transmission of the tristeza virus by the vector, *Aphis citricidus*. *Phytopathology*. 1951; 41(2):105–113.
40. Cen Y, Yang C, Holford P, Beattie GAC, Spooner-Hart RN, Liang G, Deng X. Feeding behaviour of the Asiatic citrus psyllid, *Diaphorina citri*, on healthy and huanglongbing-infected citrus. *Entomologia Experimentalis et Applicata*. 2012; 143(1):13–22. <https://doi.org/10.1111/j.1570-7458.2012.01222.x>

41. Komazaki S. Biology and virus transmission of citrus aphids. Food and Fertilizer Technology Center; 1993.
42. Madden LV, Jeger MJ, van den Bosch F. A theoretical assessment of the effects of vector-virus transmission mechanism on plant virus disease epidemics. *Phytopathology*. 2000; 90(6):576–594. <https://doi.org/10.1094/PHTO.2000.90.6.576> PMID: 18944537
43. Sieburth PJ, Nolan KG, Alderman SM, Dexter RJ. Increased efficiency and sensitivity for identifying citrus greening and citrus tristeza virus using real-time PCR testing. *Proceedings of the Florida State Horticultural Society*. 2009; 122:141–146.
44. Jeger MJ, van den Bosch F, Madden LV, Holt J. A model for analysing plant-virus transmission characteristics and epidemic development. *Mathematical Medicine and Biology*. 1998; 15(1):1–18. <https://doi.org/10.1093/imammb/15.1.1>
45. Nault LR. Arthropod transmission of plant viruses: a new synthesis. *Annals of the Entomological Society of America*. 1997; 90(5):521–541. <https://doi.org/10.1093/aesa/90.5.521>
46. Nault LR, Ammar ED. Leafhopper and planthopper transmission of plant viruses. *Annual Review of Entomology*. 1989; 34(1):503–529. <https://doi.org/10.1146/annurev.en.34.010189.002443>
47. Jagoueix S, Bove JM, Garnier M. The phloem-limited bacterium of greening disease of citrus is a member of the alpha subdivision of the Proteobacteria. *International Journal of Systematic and Evolutionary Microbiology*. 1994; 44(3):379–386.
48. Hung TH, Hung SC, Chen CN, Hsu MH, Su HJ. Detection by PCR of *Candidatus Liberibacter asiaticus*, the bacterium causing citrus huanglongbing in vector psyllids: application to the study of vector–pathogen relationships. *Plant Pathology*. 2004; 53(1):96–102. <https://doi.org/10.1111/j.1365-3059.2004.00948.x>
49. Moreno P, Ambrós S, Albiach-Martí MR, Guerri J, Peña L. Citrus tristeza virus: a pathogen that changed the course of the citrus industry. *Molecular Plant Pathology*. 2008; 9(2):251–268. <https://doi.org/10.1111/j.1364-3703.2007.00455.x> PMID: 18705856
50. Harper SJ. Citrus tristeza virus: evolution of complex and varied genotypic groups. *Frontiers in Microbiology*. 2013; 4:93. <https://doi.org/10.3389/fmicb.2013.00093> PMID: 23630519
51. Dawson WO, Bar-Joseph M, Garnsey SM, Moreno P. Citrus tristeza virus: making an ally from an enemy. *Annual Review of Phytopathology*. 2015; 53:137–155. <https://doi.org/10.1146/annurev-phyto-080614-120012> PMID: 25973695
52. Bar-Joseph M, Garnsey SM, Gonsalves D. The closteroviruses: a distinct group of elongated plant viruses. *Advances in Virus Research*. 1979; 25:93–168. [https://doi.org/10.1016/S0065-3527\(08\)60569-2](https://doi.org/10.1016/S0065-3527(08)60569-2) PMID: 393100
53. Garnsey SM, Gumpf DJ, Roistacher CN, Civerolo EL, Lee RF, Yokomi RK, Bar-Joseph M. Toward a standardized evaluation of the biological properties of citrus tristeza virus. *Phytophylactica*. 1987; 19:151–157.
54. Hilf ME, Garnsey SM, Robertson CJ, Gowda S, Satyanarayana T, Irely M, Sieburth P, Dawson W. Characterization of recently introduced HLB and CTV isolates. *Proceedings of the Florida State Horticultural Society*. 2007; 120:138–141.
55. Brlansky RH, Damsteegt VD, Howd DS, Roy A. Molecular analyses of citrus tristeza virus subisolates separated by aphid transmission. *Plant Disease*. 2003; 87(4):397–401. <https://doi.org/10.1094/PDIS.2003.87.4.397>
56. Bar-Joseph M, Marcus R, Lee RF. The continuous challenge of citrus tristeza virus control. *Annual Review of Phytopathology*. 1989; 27(1):291–316. <https://doi.org/10.1146/annurev.py.27.090189.001451>
57. Michaud JP. A review of the literature on *Toxoptera citricida* (Kirkaldy) (Homoptera: Aphididae). *Florida Entomologist*. 1998; 81(1):37–61. <https://doi.org/10.2307/3495995>
58. Metropolis N. The beginning of the Monte Carlo method. *Los Alamos Science*. 1987; 15(584):125–130.
59. Irely M, Sieburth P, Brlansky R, DaGraça J, Graham J, Gottwald T, Hartung J, Hilf M, Kunta M, Manjunath K, Ling H, Ramdugu C, Roberts P, Rogers M, Shatters R, Sun X, Wang N. PCR results From multiple HLB testing laboratories; 2009.
60. Keeling MJ, Woolhouse ME, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, Cornell SJ, Kappey J, Wilesmith J, Grenfell BT. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*. 2001; 294(5543):813–817. <https://doi.org/10.1126/science.1065973> PMID: 11679661
61. Alonso Chavez V, Parnell S, van den Bosch F. Designing strategies for epidemic control in a tree nursery: the case of ash dieback in the UK. *Forests*. 2015; 6(11):4135–4145. <https://doi.org/10.3390/f6114135>
62. Thomas PA. Biological flora of the British isles: *Fraxinus excelsior*. *Journal of Ecology*. 2016;

63. Woolhouse MEJ, Brierley L, McCaffery C, Lycett S. Assessing the epidemic potential of RNA and DNA viruses. *Emerging Infectious Disease*. 2016; 22(12):2037. <https://doi.org/10.3201/eid2212.160123>
64. Metz JA, Wedel M, Angulo AF. Discovering an epidemic before it has reached a certain level of prevalence. *Biometrics*. 1983; 39(3):765–770. <https://doi.org/10.2307/2531106> PMID: 6360226
65. Ferguson JM, Langebrake JB, Cannataro VL, Garcia AJ, Hamman EA, Martcheva M, Osenberg CW. Optimal sampling strategies for detecting zoonotic disease epidemics. *PLOS Computational Biology*. 2014; 10(6):e1003668. Jake M. Ferguson¹ *, Jessica B. Langebrake², Vincent L. Cannataro¹, Andres J. Garcia^{3,4}, Elizabeth A. Hamman¹, Maia Martcheva², Craig W. Osenberg¹ <https://doi.org/10.1371/journal.pcbi.1003668> PMID: 24968100
66. Olson LJ, Roy S. On prevention and control of an uncertain biological invasion. *Review of Agricultural Economics*. 2005; 27(3):491–497. <https://doi.org/10.1111/j.1467-9353.2005.00249.x>
67. Hauser CE, McCarthy MA. Streamlining 'search and destroy': cost-effective surveillance for invasive species management. *Ecology Letters*. 2009; 12(7):683–692. <https://doi.org/10.1111/j.1461-0248.2009.01323.x> PMID: 19453617
68. Leung B, Lodge DM, Finnoff D, Shogren JF, Lewis MA, Lamberti G. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. *Proceedings of the Royal Society B: Biological Sciences*. 2002; 269(1508):2407–2413. <https://doi.org/10.1098/rspb.2002.2179> PMID: 12495482
69. Bogich TL, Liebhold AM, Shea K. To sample or eradicate? A cost minimization model for monitoring and managing an invasive species. *Journal of Applied Ecology*. 2008; 45(4):1134–1142. <https://doi.org/10.1111/j.1365-2664.2008.01494.x>
70. Epanchin-Niell RS, Haight RG, Berec L, Kean JM, Liebhold AM. Optimal surveillance and eradication of invasive species in heterogeneous landscapes. *Ecology Letters*. 2012; 15(8):803–812. <https://doi.org/10.1111/j.1461-0248.2012.01800.x> PMID: 22642613
71. Leung B, Finnoff D, Shogren JF, Lodge D. Managing invasive species: Rules of thumb for rapid assessment. *Ecological Economics*. 2005; 55(1):24–36. <https://doi.org/10.1016/j.ecolecon.2005.04.017>
72. Warnert J, Others. Research news: Asian citrus psyllid and Huanglongbing disease threaten California citrus. *California Agriculture*. 2012; 66(4):127–130. <https://doi.org/10.3733/ca.v066n04p127>
73. Gottwald TR, Garnsey SM, Borbón J. Increase and patterns of spread of citrus tristeza virus infections in Costa Rica and the Dominican Republic in the presence of the brown citrus aphid, *Toxoptera citricida*. *Phytopathology*. 1998; 88(7):621–636. <https://doi.org/10.1094/PHTO.1998.88.7.621> PMID: 18944934
74. Coletta-Filho HD, Daugherty MP, Ferreira C, Lopes JRS. Temporal progression of *Candidatus Liberibacter asiaticus* infection in citrus and acquisition efficiency by *Diaphorina citri*. *Phytopathology*. 2014; 104(4):416–421. <https://doi.org/10.1094/PHTO-06-13-0157-R> PMID: 24620723
75. Gottwald TR, Garnsey SM, Riley TD. Latency of systemic infection in young field-grown sweet orange trees following graft-inoculation with citrus tristeza virus. In: *Proceedings of the 15th Conference of International Organization of Citrus Virologists*. IOCV, Riverside, CA; 2002. p. 48–53.
76. Gottwald T, Parnell S, Taylor E, Poole K, Hodge J. Within-tree spatial distribution of *Candidatus Liberibacter asiaticus*. In: Gottwald TR, Graham JH, editors. *Proceedings of the 1st International Research Conference on Huanglongbing*; 2008. p. 310–313.
77. Tatineni S, Sagaram US, Gowda S, Robertson CJ, Dawson WO, Iwanami T, Wang N. In planta distribution of *Candidatus Liberibacter asiaticus* as revealed by polymerase chain reaction (PCR) and real-time PCR. *Phytopathology*. 2008; 98(5):592–599. <https://doi.org/10.1094/PHTO-98-5-0592> PMID: 18943228
78. Blackburn L, Epanchin-Niell R, Thompson A, Liebhold A. Predicting costs of alien species surveillance across varying transportation networks. *Journal of Applied Ecology*. 2016;
79. R Core TeamR: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. 2016. <https://www.R-project.org/>