Asymmetrical over-infection as a process of plant virus emergence

F. Frédéric Fabre, Joël Chadœuf, Caroline Costa, Hervé Lecoq, Cecile Desbiez

► To cite this version:

F. Frédéric Fabre, Joël Chadœuf, Caroline Costa, Hervé Lecoq, Cecile Desbiez. Asymmetrical over-infection as a process of plant virus emergence. *Journal of Theoretical Biology*, 2010, 265 (3), pp.377. 10.1016/j.jtbi.2010.04.027 . hal-00608417

HAL Id: hal-00608417

<https://hal.science/hal-00608417>

Submitted on 13 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - ShareAlike| 4.0 International License

Author's Accepted Manuscript

Asymmetrical over-infection as a process of plant virus emergence

Frédéric Fabre, Joël Chadœuf, Caroline Costa, Hervé Lecoq, Cécile Desbiez

PII: S0022-5193(10)00218-3
DOI: doi:10.1016/j.jtbi.2010.04.027
Reference: YJTBI5976

www.elsevier.com/locate/jtbi

To appear in: *Journal of Theoretical Biology*

Received date: 23 December 2009
Revised date: 26 April 2010
Accepted date: 26 April 2010

Cite this article as: Frédéric Fabre, Joël Chadœuf, Caroline Costa, Hervé Lecoq and Cécile Desbiez, Asymmetrical over-infection as a process of plant virus emergence, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2010.04.027](https://doi.org/10.1016/j.jtbi.2010.04.027)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Asymmetrical over-infection as a process of plant virus emergence.

Frédéric Fabre^{a*}, Joël Chadœuf^b, Caroline Costa^a, Hervé Lecoq^a, Cécile Desbiez^a

^aINRA, UR 407 Unité Pathologie Végétale, F-84140 Avignon, France

^bINRA, UR 546 Biostatistique et Processus Spatiaux, F-84914 Avignon, France

*** Author for correspondence:**

Tel.: +33 4 32 72 28 47

Fax: +33 4 32 72 28 42

E-mail address: frederic.fabre@avignon.inra.fr

Postal address: INRA, UR 407 Unité Pathologie Végétale, F-84140 Avignon, France

Abstract

Disentangling the role of epidemiological factors in plant pathogen emergences is a prerequisite to identify the most likely future invaders. An example of emergence was recently observed in France: in 10 years, “classic” (CL) strains of *Watermelon mosaic virus* (WMV) were displaced at a regional scale by newly introduced “emerging” (EM) strains. Here we analyse a 3 years dataset describing the co-dynamics of CL and EM strains at field scale using state-space models estimating jointly (i) probabilities of primary and secondary infection and (ii) probabilities of over-infecting with a CL [EM] strain a plant already infected with an EM [CL] strain. Results especially indicate that it is more than 3 times less probable for a CL strain to over-infect an EM infected plant than for an EM strain to over-infect a CL infected plant. To investigate if these asymmetric interactions can explain the CL/EM shift observed at regional scale, an exploratory model describing WMV epidemiology over several years in a landscape composed of a reservoir and a cultivated compartment is introduced. In most simulations a shift is observed and both strains do coexist in the landscape, reaching an equilibrium that depends on the probabilities of over-infection.

Key words: Biological invasion, Epidemiology, Landscape, State-space model, *Watermelon mosaic virus*.

1. Introduction

Emergences of plant diseases often have detrimental consequences for food production and food security (Strange and Scott, 2005). Alone, plant virus emergences represent 47% of the cases of emerging plant infectious disease (Anderson et al., 2004) and are reported to be increasingly frequent (Jones 2009; Rojas and Gilbertson, 2008; Varma and Malathi, 2003). Factors driving the emergence of plant viruses are numerous. Anderson et al. (2004) reveal that, for viruses, pathogen introduction (mainly mediated by anthropogenic activities) and change in vector populations were the major factors of emergence. Many other factors including the intrinsic high capacity of viruses to evolve, changes in agricultural practices and altered pathosystem biology have been documented (Anderson et al., 2004; Elena et al., 2009; Jones, 2009; Rojas and Gilbertson, 2008; Woolhouse et al., 2005). A complete pathogen emergence is a 3-step process (Woolhouse et al., 2005): (i) the exposure of a new host species to a pathogen, (ii) the ability of the pathogen to infect an individual of this new host species (i.e. the pathogen and the host must be compatible) and (iii) the subsequent spread of the pathogen in its new host population. However, the term “emergence” often encompasses several phenomena: (i) newly recognised or newly evolved viruses that cause damaging epidemics, (ii) viruses that increase their geographical distribution and (iii) viruses that change their pathogenicity (host range, virulence) (Jones, 2009; Pulliam, 2008). The need to find new ways to restrict virus emergence requires additional insight into the ecological and evolutionary factors involved in these events (Jones 2009; Pulliam, 2008). Identifying these factors and understanding their relative roles (Anderson et al., 2004; Holmes and Drummond, 2007) is needed to predict which pathogens are most likely to become successful invaders in a given environment.

In France, *Watermelon mosaic virus* (WMV, genus *Potyvirus*) is currently undergoing a change in populations revealing a rapid replacement of local strains by “invasive”, recently introduced ones. WMV has a worldwide distribution, mostly in temperate and Mediterranean regions (Lecoq and Desbiez, 2008). It infects more than 170 crops and weed species, causes agronomic problems mostly in cucurbits and, like other potyviruses, it is non-persistently transmitted by at least 35 aphid species (Lecoq and Desbiez, 2008). At the world level, three molecular groups of WMV have been defined (Desbiez et al., 2007), one of which is widespread in France. WMV, severe in melon, only induced very mild symptoms on zucchini squash until the end of the 90’s. However, severe symptoms in zucchini squash leaves and fruits have been observed since 1999 in South-eastern (SE) France. Their appearance correlates with the introduction of new, “emerging” (EM) isolates distant at the molecular level from the “classic” (CL) isolates that have been present for more than 30 years. A survey performed between 2004 and 2007 at the regional scale showed that EM isolates did not spread over long distances, but rapidly displaced the pre-existing CL isolates in all sites where both groups occurred (Desbiez et al., 2009). This case constitutes a well documented example of viral emergence; not of the complete 3-step emergence processes, but rather a case of appearance in a new geographic region of a new strain of a pathogen. Indeed, before their introduction in France through a yet unknown route, EM WMV strains had already been reported to infect cucurbit crops in Eastern Asia. Moreover, as CL strains were pre-existing in France, investigating this case more deeply should provide interesting information on the factor involved in the “paradox of invasion” that is why an emerging viral species can displace a native viral species that should *a priori* be better adapted to the local conditions (Sax and Brown, 2000).

In addition to the regional survey, a field scale study of WMV epidemiology was performed from 2002 to 2004 in an experimental plot located in SE France. These data,

acquired in natural conditions, provided detailed spatio-temporal maps of the respective dynamics of CL and EM strains of WMV. In this paper, we will first examine the data collected at the field scale using models specifically built in order to elucidate the processes involved in CL strain displacement but hidden from direct observation. In a second step, using an exploratory model, we will investigate whether the processes identified at the field scale can explain the emergence of EM strains at the landscape scale where the whole annual cycle of WMV occurs.

2. Data description

A plot of 160 zucchini squash F1 Diamant consisting of 8 rows of 20 plants was planted each year from 2002 to 2004 at Montfavet, France. Each plant was surveyed weekly for virus infection until 100% of the plants became infected (week 3 in 2002 and 2004, week 4 in 2003) and a last survey done 4 or 5 weeks later (week 7 in 2002, week 8 in 2003 and 2004). Plants were classified into 4 compartments, (i) H (healthy), (ii) CL (infected by a classic strain), (iii) EM (infected by an emerging strain) and (iv) D (doubly infected by both a classic and an emerging strain), as follows: a young leaf was collected at each sampling date from each plant and tested by DAS-ELISA for WMV infection. If the sample is negative, the plant is classified H. For each positive sample, total RNA was extracted using TRI-Reagent (Molecular Research Center, Inc., Cincinnati, OH) and submitted to one-step RT-PCR according to the protocol used in the laboratory using primers WMV-5' (5'-GGCTTCTGAGCAAAGATG-3') and WMV-3' (5'-CCCACTCACTGTGGAAG-3') (Desbiez et al., 2007). PCR samples were sent to Cogenics (Grenoble, France) for direct sequencing with primer WMV-5'. Sequences were aligned with ClustalW included in DAMBE (Xia, 2000), adding reference sequences belonging to the main molecular groups

defined for WMV (Desbiez et al., 2007, 2009). Neighbour-joining and maximum-parsimony trees were built with MEGA 3.1 (Kumar et al., 2004) and bootstrap analysis performed to assess the robustness of the trees. According to the sequence data, the plant is classified as CL, EM or D.

Moreover, an epidemiological survey (Desbiez et al., 2009) was performed from 2004 to 2008 in the 3 neighbouring “départments” (administrative subdivision, 4-6000 km² each) Gard, Vaucluse and Bouches du Rhône in SE France where EM strains were first detected. Samples of cucurbit crops, received from farmers, farm advisers and seed companies, were tested as described above.

3. Models description

3.1. A model for the co-dynamics of viral strains at field scale

This section describes the model skeleton and underlying assumptions. The whole set of equations and additional mathematical details are provided in appendix A and B. In the model, $N(i,t)=[N_H(i,t), N_{CL}(i,t), N_{EM}(i,t), N_D(i,t)]$ is the number of plants in each compartments for a given year i and date t of observation. The model describes the dynamic of $N(i,t)$ at a time interval $\Delta t = 1$ week. Initially, all plants are healthy: $N(i,0)=[N_{tot}, 0, 0, 0]$ ($N_{tot}=160$). Plants can be infected through 8 categories of events resulting from the combination of 2 virus strains (CL and EM), 2 plant categories (when infected by a given virus strain a plant is either healthy or already infected by the other strain) and 2 infection processes (primary and secondary infection) (Fig. 1). Primary infection corresponds to the infection of a plant by aphids having acquired the virus from sources outside the considered field whereas sources are within this field in the case of secondary infection.

3.1.1. Probabilities of infection of an H plant from $t-1$ to t

An H plant at time $t-1$ can become CL, EM or D or remain H at time t depending on the probabilities of 4 events (Fig. 1a): (i,ii) $p_{i,t}^{PI}(CL|H)$ [resp. $p_{i,t}^{PI}(EM|H)$], probability that “an H plant is infected by primary infection with a CL [resp. EM] strain during Δt ” and (iii,iv) $p_{i,t}^{SI}(CL|H)$ [resp. $p_{i,t}^{SI}(EM|H)$], probability that “an H plant is infected by secondary infection of a CL [resp. EM] strain during Δt ”.

Regarding primary infection, $p_{i,t}^{PI}(CL|H)$ is modelled as $\text{logit}[p_{i,t}^{PI}(CL|H)] = \alpha_{i,CL} + \gamma_{i,t}$ where (i) the parameter $\alpha_{i,CL}$ is, in the logit scale, the probability that an H plant is infected by primary infection with a CL strain during Δt in year i if no variation between weeks exists and (ii) $\gamma_{i,t}$ is a set of random effects mutually independent and identically normally distributed with unknown variance σ^2 accounting for the variability between weeks of the primary infection process. $p_{i,t}^{PI}(EM|H)$ is defined similarly (appendix A – equation A2).

Regarding secondary infection, deriving $p_{i,t}^{SI}(CL|H)$ required to consider the probability $q_i(SI_{CL}|H)$ of the elementary event “a plant infected in the field with a CL strain in year i infects an H plant by secondary infection during Δt ”. For a given year i , $q_i(SI_{CL}|H)$ was assumed constant and modelled as a pure fixed effect $\beta_{i,CL}$ as $\text{logit}[q_i(SI_{CL}|H)] = \beta_{i,CL}$.

Assuming that secondary infection does not depend on the distance between sources of CL strains and H plants, it comes that $p_{i,t}^{SI}(CL|H) = 1 - [1 - q_i(SI_{CL}|H)]^{N_{CL}(i,t-1) + (1-\pi)N_D(i,t-1)}$ where

(i) $N_{CL}(i,t-1) + (1-\pi)N_D(i,t-1)$ is the number of potential sources of CL strains in the field at time $t-1$ and (ii) π is a parameter describing the mean proportion of EM strains in D plants ($1-\pi$ is the proportion of CL strains). This equation means that an H plant at time $t-1$ become CL at t by secondary infection if at least one of the plants infected in the field with a CL strain at

date $t-1$ (either CL or D plants) was a source for this strain. Details on its derivation are provided in appendix B and in appendix A (equations A4, A6) for $p_{i,t}^{SI}(EM|H)$.

Finally, assuming that during Δt primary and secondary infection events occur independently, $p_{i,t}^M = (p_{i,t}(H_t | H_{t-1}), p_{i,t}(CL_t | H_{t-1}), p_{i,t}(EM_t | H_{t-1}), p_{i,t}(D_t | H_{t-1}))$, the vector of probabilities that, in year i , a plant is H [resp. CL, EM and D] at time t given that this plant was H at time $t-1$ is obtained (Appendix A - equations A7 to A10). No interaction is considered between infection events when they occurred during the same time step. But, as detailed now, such interactions are modelled when these events occurred sequentially during 2 different weeks.

3.1.2. Probabilities of infection of a singly infected plant from $t-1$ to t

A CL plant at time $t-1$ can become D or remain CL at time t depending on the probabilities of 2 events (Fig. 1b): (i) $p_{i,t}^{PI}(D|CL)$, probability that “a CL plant is infected by primary infection with a EM strain during Δt ” and (ii) $p_{i,t}^{SI}(D|CL)$, probability that “a CL plant is infected by secondary infection of an EM strain during Δt ”. As previously, deriving this probability required to consider the probability $q_i(SI_{EM}|CL)$ of “a plant infected in the field with an EM strain is source of infection of an CL plant during Δt ”.

$p_{i,t}^{PI}(D|CL)$ and $q_i(SI_{EM}|CL)$ are modelled with a parameter, γ_{CL} , defined as the odd ratio between the probability of infecting an H plant with a EM strain and the probability of infecting an CL plant with a EM strain. Values of $\gamma_{CL} > 1$ indicate that over-infecting an EM plant with a CL strain is less likely than infecting an H plant with a CL strain. The higher γ_{CL} , the less likely this event is. Values of $\gamma_{CL} < 1$ indicate that over-infecting is more likely. No difference arises when $\gamma_{CL} = 1$. For interpretation, odds ratios approximate relative risks for rare events ($p < 0.1$). γ_{CL} applies indifferently whatever the infection process leading to over-

infection (primary or secondary infection) and does not depends on the year and on the time lag between the first and the second infection events.

Finally assuming independence between primary and secondary infection leads to a simple expression of $p_{i,t}(D_t | CL_{t-1})$, the probability that, in year i , a plant becomes doubly infected (D) at time t given that this plant was CL at time $t-1$. All the equations necessary to derive $p_{i,t}(D_t | CL_{t-1})$ as well as their analogues for the over-infection of an EM plant with a CL strain (Fig. 1c) are presented in Appendix A (equations A11 to A18).

3.1.3. Dynamics of $N(i,t)$ during an epidemic

The dynamic of $N(i,t)$ from $t=0$ to t_{end} is modelled as a sequential 3 step stochastic process. During year i from $t-1$ to t , (i) the number of H plants remaining H or becoming CL, EM or D is the result of the multinomial process $\sim Multinomial(N_H(i, t-1), p_{i,t}^M)$, (ii) the number of CL plants becoming D is the result of the binomial process $\sim Binomial(N_{CL}(i, t-1), p_{i,t}(D_t | CL_{t-1}))$ and (iii) the number of EM plants becoming D is the result of the binomial process $\sim Binomial(N_{EM}(i, t-1), p_{i,t}(D_t | EM_{t-1}))$. Lastly, the number of plants in each compartment is updated from date $t-1$ to t (Appendix A – equation A22). These stochastic processes assume that: (i) all plants in a given compartment at $t-1$ have the same probability to be in another compartment at time t , (ii) changes between compartments are independent from one plant to another.

3.1.4. Model inferences

The vector of model parameters is $\theta = [\alpha_{CL,i}, \alpha_{EM,i}, \beta_{CL,i}, \beta_{EM,i}, \gamma_{EM}, \gamma_{CL}, \pi, \sigma]_{i=1..3}$. Statistical inferences were performed via Bayesian Monte Carlo Markov Chain (MCMC) methods

(Gelman et al. 2004) with OpenBUGS[®] 3.0.3 using uninformative prior probability density functions for θ . Consistency between the model and the data was checked using a χ^2 discrepancy test (Gelman et al. 2004) as well as linear regression of predicted versus observed values of $N_{CL}(\cdot, \cdot)$, $N_{EM}(\cdot, \cdot)$ and $N_D(\cdot, \cdot)$. Finally, to investigate more precisely the role of over-infection parameters (γ_{EM} and γ_{CL}), three models were compared using Bayes Factor (Kass and Raftery, 1995): (i) model M_0 , the model described above, (ii) model M_1 where $\gamma_{EM} = \gamma_{CL}$ (i.e. the probability of over-infection does not depend on the first strain inoculated) and (iii) model M_2 where $\gamma_{EM} = \gamma_{CL} = 1$ (i.e. the probability of over-infection are equal to the probability of infecting an H plant). Further details on model inferences are provided in appendix C.

3.2. A model of viral epidemiology at landscape scale.

3.2.1. Model description

A model of WMV epidemiology over several years in a landscape is described. It simulates $p_{EM}(i)_{\{i=0..y_{end}\}}$, the proportion of EM strains during y_{end} years found in an idealized landscape made of 2 compartments.

The reservoir compartment, available all year long, includes potential wild reservoir hosts of WMV. It is characterized by $p_{EM}(i)$, the proportion of EM strains among the reservoir hosts infected by WMV in year i ($1 - p_{EM}(i)$ is the proportion of CL strains).

The cultivated compartment, available during T_{max} weeks a year, is composed of N_{field} fields similar to our experimental plots (with N_{tot} plants). Using the field model, the dynamics of $N(i, f, t | \Omega_i) = [N_H(i, f, t), N_{CL}(i, f, t), N_{EM}(i, f, t), N_D(i, f, t)]$ are simulated, $N(i, f, t | \Omega_i)$ being the number of H, CL, EM and D plants during year i ($1 \leq i \leq y_{end}$), in the field f ($1 \leq f \leq N_{field}$) at date t ($1 \leq t \leq T_{max}$) simulated for a given model parameter Ω_i . $\Omega_i = [p^s_i(PI_{CL}|H), p^s_i(PI_{EM}|H), q^s_i(SP_{CL}|H),$

$q^s_i(SP_{EM}|H)$, γ^s_{EM} , γ^s_{CL} , π^s] (the upper script S is added to the notation defined in the “field scale” model to indicate that these parameters are those of the “landscape scale” model) is defined as follow:

(i) For the sake of simplicity, the probabilities of primary infection rates of CL [resp. EM] strains $p^s_i(PI_{CL}|H)$ [resp. $p^s_i(PI_{EM}|H)$] depend only on the year i and not on weeks. Primary infection events are inoculations to cultivated plants of viruses originating from the reservoir compartment (no between fields infections are considered). If p^s denotes the probability of having a primary infection event, and given that the relative proportion of EM strains in the reservoir compartment in the previous winter is $p_{EM}(i-1)$, it is assumed that $p^s_i(PI_{EM}|H)=p^s \cdot p_{EM}(i-1)$ and $p^s_i(PI_{CL}|H)=p^s \cdot (1-p_{EM}(i-1))$: the probability of primary infection by a given strain is proportional to the frequency of this strain in the reservoir compartment during the previous winter (*i.e.* there is an annual “reset” of WMV populations in the reservoir compartment).

(ii) The probabilities of secondary infection, $q^s_i(SP_{CL}|H)$ and $q^s_i(SP_{EM}|H)$, are equal to q^s .

(iii) The probabilities of over-infection are assessed as previously assuming that $\gamma^s_{EM}=\gamma^s_{CL} \cdot \delta$ where $\delta > 0$.

Initially ($i=1$, $t=1$), only CL strains are in the reservoir hosts ($p_{EM}(i=0)=0$). Thus only CL strains can be introduced in the cultivated compartment through primary infection events. However during the first year a single plant infected with an EM strain is introduced at time t_{intro} in N_{intro} out of N_{field} . In these field(s), EM strains can only spread by secondary infection during this first year (*i.e.* between-fields events of primary infection are not modelled).

The dynamic of $p_{EM}(i)$ is simulated by iterating 2 steps for i in $1..y_{end}$:

(i) Step 1. Given the values of p^S , q^S , γ^S_{CL} , π^S , δ and $p_{EM}(i-1)$, the dynamics of $N(i, f, t | \Omega_i)$ are assessed. If $i=1$, a single plant infected with an EM strain is introduced at time t_{intro} in N_{intro} out of N_{field} .

(ii) Step 2. $p_{EM}(i)$ is the ratio between $S_{EM}(i)$ and $S_{Tot}(i)$, respectively defined as:

$$S_{EM}(i) = \sum_{f=1}^{N_{field}} \int_{t=1}^{T_{max}} (N_{EM}(i, f, t) + \pi^S N_D(i, f, t)) dt$$

$$S_{Tot}(i) = \sum_{f=1}^{N_{field}} \int_{t=1}^{T_{max}} (N_{CL}(i, f, t) + N_{EM}(i, f, t) + N_D(i, f, t)) dt$$

As such, $p_{EM}(i)$ is proportional to the cumulated dynamic of EM plants with respect to the cumulated dynamic of CL, EM and D infected plants over all the fields of the landscape. D plants are supposed to be source of EM strains proportionally to π^S .

3.2.2. Landscape model analysis

Analyses were focused on the effect of γ^S_{EM} and γ^S_{CL} . First, the effect of δ ($\delta = \gamma^S_{EM} / \gamma^S_{CL}$) on the temporal dynamics of $p_{EM}(i)$ for $i \in \{1.. y_{end}\}$ was studied. γ^S_{CL} was set to 2.95 and $p_{EM}(i)$ was assessed for values of δ ranging from 1 ($\gamma^S_{EM} = \gamma^S_{CL}$) to 40 ($\gamma^S_{EM} = 40. \gamma^S_{CL}$). Secondly, the final state reached after 40 years by $p_{EM}(y_{end})$ was studied in a $(\gamma^S_{EM}, \gamma^S_{CL})$ plane with $1 \leq \gamma^S_{EM} \leq \gamma^S_{CL} \leq 1000$. In practise, a log-regular grid with 900 points in this $(\gamma^S_{EM}, \gamma^S_{CL})$ plane was considered. Three final states were possible: (i) EM strains replaced CL strains (it exists $i \in \{2.. y_{end}\}$ such that $p_{EM}(j \geq i) = 1$), (ii) EM strains disappeared following their introduction (it exists $i \in \{2.. y_{end}\}$ such that $p_{EM}(j \geq i) = 0$), and (iii) EM and CL continued to coexist at y_{end} (for

all $i \in \{2.. y_{end}\}$, $0 < p_{EM}(i) < 1$). For each points of the grid in the plane $(\gamma_{EM}^S, \gamma_{CL}^S)$, the probabilities of the 3 possible final states were estimated by their proportions using 1000 independent replicates. In all simulations, primary and secondary infection probabilities were set to their overall mean from data analysis ($p^S=0.11$, $q^S=0.035$) as well as π^S set to 0.56. N_{field} was set to 25, t_{intro} and N_{intro} to 1, T_{max} to 10 weeks and y_{end} to 40.

4. Results

4.1. Epidemiology of CL and EM strains of WMV at the field scale

4.1.1. Experimental results at the field scale

The set of spatio-temporal maps of the disease dynamics observed are presented in online supplementary material (Fig. S1). During the 3 years, CL and EM strains were always first detected in the same plot as early as the second week after planting. Only one, EM1, of the four subgroups of EM strains observed in SE France (Desbiez et al., 2009), was found repeatedly in the experimental plot (data not shown). Complete infection of the plot was observed after 3 or 4 weeks and, at week 7 or 8, the number of D plants was close to 100%, indicating that cross protection between CL and EM strains is not fully efficient (Fig. S1).

4.1.2. Goodness of fit of the model

The baseline model (M_0) was consistent with the data. A good fit between observed and predicted values of the numbers of CL, EM and D plants ($N_{CL}(i,t)$, $N_{EM}(i,t)$, $N_D(i,t)$ respectively) was observed for the 3 years (Fig. 2). The overall r^2 value between observed and predicted values of $N_{CL}(i,t)$, $N_{EM}(i,t)$, $N_D(i,t)$ was 0.98 and 95 % of observed values were included in a 90% credibility interval ($n=74$). Moreover, the Bayesian p-value of the χ^2 discrepancy did not reveal any significant inconsistency between the model and the data: the

p-values were 0.87, 0.69 and 0.9 for N_{CL} , N_{EM} and N_D respectively. The statistical summary of the parameters of the model M_0 are provided in Table 1.

4.1.3. Mean annual probabilities of primary infection

Mean probabilities refer here to the probability of primary infection of an H plant averaged over the duration of the epidemic. Their values ranged from 0.05 to 0.31 depending on the year and the strain (Fig. 3). The overall mean was 0.11. From 2002 to 2004, the primary infection rate by EM strains was significantly higher than the one of CL strains in 2003 and 2004 (Bayesian p-value of 0.047 and 0 respectively) but not significantly different in 2002 (Bayesian p-value of 0.66). Finally, the hypotheses of a decrease or an increase through years of the mean probabilities of primary infection for CL and EM strains were tested. No significant trend was detected in mean probabilities of primary infection by a CL strain but the mean probability of primary infection by an EM strain was significantly higher in 2004 than in 2002 or 2003 (Bayesian p-value of 0.01 and 0.02 respectively).

4.1.4. Annual probabilities of secondary infection

The probabilities that a plant infected with a CL or an EM strain was a source of infection for an H plant during a week ranged from 0.002 to 0.07 depending on the year and the strain (Table 1). The overall mean was 0.036. The mean probabilities of secondary infection did not exhibit any clear annual trend. For a given year, the mean probabilities of secondary infection of a CL strain were significantly higher than the one of an EM strain in 2003 and 2004 (Bayesian p-values of 1 and 0.99 in 2003 and 2004 respectively) but not in 2002 (Bayesian p-values 0.89). The role of D plants in secondary infection events as sources of CL and EM

strains was modelled by the parameter π which is the mean frequency of EM strains in these plants (see appendix B for details). The mean value of π was estimated as 0.56 with a 90% Bayesian posterior credibility interval ranging from 0.34 to 0.79. Accordingly, within plant competition between CL and EM strains in D plants did not significantly favour the spread of one or another strain.

4.1.5. Odd ratios of over-infection

The mean value of the odd ratio for over-infecting an EM plant (γ_{EM}) and a CL plant (γ_{CL}) were respectively 8.92 and 2.95 (Table 1). Accordingly, over-infecting a CL plant with an EM strain is about 3 times less probable than infecting an H plant but over-infecting an EM plant with a CL strain is about 9 times less probable than infecting an H plant. Overall, it is 3 times less probable for a CL strain to infect an EM plant than for an EM strain to infect a CL plant. Moreover, when considering the 95% Bayesian posterior credibility interval and Bayesian p-values, it appears that γ_{EM} and γ_{CL} are significantly greater than one. Thus over-infecting a previously infected plant is significantly less probable than infecting a healthy plant. γ_{EM} is also significantly greater than γ_{CL} . The comparison of model M_0 (baseline model with $\gamma_{EM} \neq \gamma_{CL}$ *i.e.* the probability of over-infection depends on the first strain inoculated), M_1 (model with $\gamma_{EM} = \gamma_{CL}$ *i.e.* the probability of over-infecting does not depend on the first strain inoculated) and M_2 (model with $\gamma_{EM} = \gamma_{CL} = 1$ *i.e.* the probability of over-infection are equal to the probability of infecting an H plant) leads to the same conclusion with Bayes factors of respectively 4 for $2 \cdot \ln(BF_{1,0})$ (support for M_0) and 157 for $2 \cdot \ln(BF_{2,0})$ (strong evidence for M_0).

Back to the meaning of odd ratios, and taking the probabilities of secondary infection in 2003 as example, the value of the odd ratios for over-infection imply that, although the mean

probability of secondary infection of CL strains on a healthy plant was higher than the one of EM strains ($q_2(SI_{CL} | H) = 0.033$ and $q_2(SI_{EM} | H) = 0.011$), this is no longer the case when considering the infection by secondary spread of a previously infected plant ($q_2(SI_{CL} | EM) = 0.0038$ and $q_2(SI_{EM} | CL) = 0.0037$ according to equations A12 and A16 of Appendix A).

4.2. Landscape scale model

The landscape model simulates the dynamic of the proportion of EM strains in a landscape during 40 years ($p_{EM}(i)$ with $i \in \{1..40\}$) following the accidental introduction of a single cucurbit seedling ($N_{intro}=1$ out of $N_{field}=25$ fields with 160 plants) infected with an EM strain. The effect of the parameter, $\delta = \gamma_{EM}^S / \gamma_{CL}^S$, measuring how many times it is less likely for a CL strain to infect an EM plant than for an EM strain to infect a CL plant, on the dynamic of the proportion of EM strains in a landscape was particularly investigated. A value of $\delta > 1$ indicates that it is δ times more probable for an EM strain to infect a CL plant than for a CL strain to infect an EM plant. Inversely, a value of $\delta < 1$ indicates that it is $1/\delta$ times more probable for a CL strain to infect an EM plant than for an EM strain to infect a CL plant. In Fig. 4a, the odd ratio for over-infecting an CL plant, γ_{CL}^S , was set to its reference value 2.95 and the dynamic of the proportion of EM strains in the landscape was plotted for 4 values of δ (corresponding to 4 values of the odd ratio for over-infecting an CL plant, γ_{EM}^S) [$\delta=1$ ($\gamma_{EM}^S=2.95$), $\delta=3$ ($\gamma_{EM}^S \approx 8.8$), $\delta=10$ ($\gamma_{EM}^S \approx 29$) and $\delta=40$ ($\gamma_{EM}^S \approx 118$)]. For most simulations, the proportion of EM strains in the landscape rapidly increased during the first 15 years and then reached a noisy steady state where both CL and EM strains of WMV coexisted in the landscape. In simulations where δ ranged between 0.001 and 1000, coexistence between EM

and CL strains almost always occurred for $0.05 \leq \delta \leq 10$ (Fig. 5a). When coexistence occurred, the level of the steady state between CL and EM strains mostly depended on δ (Fig. 5b). For $\delta=1$, the proportion reached by EM strains in the landscape, 0.57, was close to their proportion in D plants, π^S . However, this correspondence was not observed for higher values of π^S (not illustrated).

Coexistence of EM and CL strains was not always observed. The probability that EM strains disappeared following their introduction is higher than 0.25 for values of $\delta \leq 0.05$ and even higher than 0.5 for values of $\delta \leq 0.01$ (Fig. 5c). Similarly, EM strains could replace CL strains (Fig. 5d). However this event was infrequent in the range of the plane $(\gamma_{EM}^S, \gamma_{CL}^S)$ explored and only occurred with probabilities higher than 0.5 for values of $\delta \geq 50$. Finally, for a given value of δ (e.g. $\gamma_{EM}^S=100, \gamma_{CL}^S - \delta=100$) it was somewhat more likely for EM strains to replace CL strains than for EM strains to go extinct in the inverse case ($\gamma_{CL}^S=100, \gamma_{EM}^S - \delta=0.01$). This asymmetry resulted from a trade-off between the endemic and native origins of CL and EM strains respectively and their relative fitness in competition in doubly infected plant.

5. Discussion

The shift between CL and EM strains of WMV observed in France since 2000 is a striking example of the invasion paradox, locally adapted strains being rapidly displaced by invading ones originating possibly from Asia (Desbiez et al., 2009). A similar observation was recently reported by Gómez et al. (2009) in Spain where, after their recent introduction, new isolates of *Pepino mosaic virus* (PepMV-CH2) have spread in a few years in an epidemic fashion into a niche previously occupied by PepMV-EU isolates. As a first step in this study, we developed stochastic models to disentangle the field co-dynamics of CL and EM strains with regard to several epidemiological mechanisms (primary and secondary infection and a

wide range of host-virus interactions in co-infected host). These models also explicitly deal with the uncertainty in the observation process. They are examples of state-space model, a class of model now widely used in various fields of ecology (*e.g.* Clark, 2005; Cressie et al., 2009; Soubeyrand et al., 2009) notably due to its propensity to reconcile the point of view of statisticians and mechanism based modellers.

The main originality of the model was to define probabilities of over-infection that can handle the different kinds of plant-virus interaction arising in mixed infections. A virus may stimulate the replication and/or the movement of another virus (Atabekov and Dorokhov, 1984) (Karyieja et al., 2000). A plant resistant to a virus can even become infected with the help of another virus (Dodds and Hamilton, 1972) (Wang et al., 2004) or another strain of the same virus (Gomez et al., 2009). Diverse mechanisms are involved in these synergistic interactions, including efficient inhibition of posttranscriptional gene silencing by at least one of the viruses (for a review see Latham and Wilson, 2008; Cuellar et al., 2009). However, one virus can also prevent the infection of a second, a phenomenon known as cross-protection (Fraser, 1998) and used to control some major virus diseases (Lecoq, 1998). To our knowledge, very few papers deal with the mathematical modelling of interactions between plant viruses (Zhang et al., 2000; Zhang and Holt, 2001). Only Zhang and Holt (2001) modelled the effect of cross protection on plant-virus epidemiology assuming that infection by the protective virus precludes completely infection by the severe one, but not vice-versa. More generally for plant pathogens, many model-based studies used the concept of the basic reproduction number (R_0) to analyse disease invasion and persistence from the microscopic to the regional scale. However, most of these studies rely on the assumption that the invading pathogen is not competing with resident pathotypes (see Gilligan and van den Bosh (2008) for a review and references for exceptions to this trend).

In our case, inferences on the odd ratio for over-infecting an EM plant (γ_{EM}) and a CL plant (γ_{CL}) firstly reveal that over-infecting an already infected plant is less probable than infecting a healthy plant but nevertheless occurs as revealed by both experimental and field observations. Normally, over-infection between strains of the same potyvirus is supposed to be suppressed by the mechanisms of cross-protection, possibly related to RNA silencing (Ratcliff et al., 1999); cross-protection was once considered as a taxonomic criterion for species/strain demarcation (Ward and Shukla, 1991). The efficiency of cross-protection among potyviruses depends on the molecular relatedness between strains, highly divergent isolates being poorly cross-protective (Nakazono-Nagaoka et al., 2009; Valkonen et al., 2002; Wang et al., 1991; You et al., 2005). CL and EM isolates have 4 to 6% divergence in their coat protein (CP) nucleotide sequence (Desbiez et al., 2009). At this level of divergence, cross-protection between potyvirus isolates is usually effective (Nakazono-Nagaoka et al., 2009; Lecoq, 1998). However some exceptions were already observed and the relation between CP sequence relatedness and protection efficiency has been questioned (Valkonen et al., 2002).

Inferences on the odd ratio for over-infecting an EM plant (γ_{EM}) and a CL plant (γ_{CL}) also indicate that it is 3 times less probable for a CL strain to infect an EM plant than for an EM strain to infect a CL plant. Examples of such asymmetrical efficiency are quite rare in the literature (e.g. between strains of *Sugarcane mosaic virus* (Krstic, 1995), *Potato virus A* (Valkonen et al., 2002), *Cucumber mosaic virus* (Tian et al., 2009) or between *Oilseed rape mosaic virus* and *Tobacco mild green mosaic virus* (Aguilar et al., 2000)). This asymmetry might be related to a more efficient accumulation, migration and/or silencing suppression in EM strains, thus overcoming the cross-protection mechanisms. Whereas strains of some potyviruses rarely infect the same cell (Dietrich and Maiss 2003; Takahashi et al., 2007), the high frequency of natural recombinants in WMV (Desbiez and Lecoq, 2008) suggests that cell

co-infection is frequent in this virus; the asymmetry between EM and CL strains might happen at this level. These hypotheses remain to be investigated. Asymmetrical interactions have also been recently revealed in animal pathology. Wolfe et al. (2007) showed that mice previously infected with *Bordetella pertussis* were not protected against a later infection with *B. parapertussis* (the main agents of whooping cough), while primary infection with *B. parapertussis* conferred cross-protection. Much attention has been given to the mathematical modelling of cross-protection in animal pathology and especially virology as it is a key factor of vaccination success (e.g. human influenza A (Ferguson et al., 2003)). However, with a few exception (Restif et al., 2008 and references therein), asymmetric cross-immunity has also been overlooked in most mathematical models.

The model goodness of fit was satisfactory. However, improvements in several ways are possible by taking advantage of the flexibility of the state-space models. Modelling the spatiotemporal dynamics of CL and EM strains on a daily time step is particularly appealing. Such space-time data analysis often improves our understanding of the mechanisms governing both animal and plant epidemics (e.g. Keeling et al., 2001; Otten et al., 2003; Soubeyrand et al., 2009). Here it will provide inferences about the distance of dispersion of secondary infection events for example, a process seldom estimated *in natura* for non-persistently transmitted viruses. Bayesian inferences on spatiotemporal process are of special interest as they account coherently for parameter variability (Cressie et al., 2009). If classical Bayesian inferences are difficult to handle, statistical methods like approximate Bayesian calculations (Beaumont et al., 2002) could be an alternative.

The second part of this work investigated the consequences, at the landscape scale, of the asymmetrical probabilities of over-infection revealed at the field scale during the course of the annual epidemiological cycle. An explanatory model is proposed in order to fit the scales

at which hosts and aphid-borne viruses naturally interact. Simulations addressed the case of an accidental introduction of a cucurbit seedling infected with an EM strain. The results indicated that δ , the parameter measuring how many times it is less likely for a CL strain to infect an EM plant than for an EM strain to infect a CL plant, deeply impacted the equilibrium between CL and EM strains in the landscape throughout the following mechanism. Assuming for example $\delta > 1$ implies that plants remained solely EM infected longer than solely CL infected. In other words, the higher δ is, the longer is the lag time necessary for a EM infected plant to become D infected relatively to the lag time necessary for a CL plant to become D. In turn, this tends to increase the cumulated dynamics of plants solely infected with EM strains in the fields of the landscape and thus the proportion of reservoir hosts infected with EM strains. A consequence of this mechanism is that, for some value of δ , CL strains mostly persist in the environment in D plants. The persistence of viral isolates in mixed infected plants was proposed by Gómez et al. (2009) to explain PepMV-EU isolates persistence in Spain after PepMV-CH2 spread.

In the simulation with the parameters closest to the field observation (Fig. 4a, $\delta=3$ - black curve), CL and EM strains almost always coexist in the landscape at a ratio of about 3:7. Although this is not a formal validation, these figures should be compared to the data of the epidemiological survey in the Gard and Bouches du Rhône regions where the proportion of EM strains seems to reach an equilibrium near 70-80% (Fig. 4b) in less than 10 years after their introduction. In Vaucluse, the situation is different as the proportion of CL strains decreased steadily and reached only 1% in 2008. Although it is tempting to bring nearer these data, the epidemiological situation of WMV is much more complex than the simulation scenario with in particular 4 subgroups of EM isolates probably originating from as many introduction events. Only one subgroup of isolates was detected in the field data (EM1). The value of δ derived thus characterizes the interactions between CL and EM1 strain in D plants.

The landscape model describes a disease with discontinuous crop-host availability typical of the annual crops in temperate climates. During the crop-free season, viruses and vectors are hosted by “reservoirs”, wild hosts playing then a major role. Up to date, viruses in the wild compartment have received little attention. Consequently, in the model, the reservoir compartment is simply seen as a mirror reflecting, from year n to year $n+1$, the proportion of EM strain according to its evolution in the sole cultivated compartment. Despite this simplification, one should notice that the trend detected in the mean probability of primary infection by CL and EM strains (Fig. 3) corroborates this hypothesis. Nevertheless, further research on virus-plant pathosystems at the interface between cultivated and natural vegetation is needed (Jones, 2009). Firstly, this need arises from the importance of studying epidemics at the scale where they naturally occur as well as the need to match this scale with the one at which control strategy must be deployed (Dybiec et al. 2004; Gilligan 2008). In this regard, the recent focus on the landscape epidemiology of plant disease (Plantegenest et al. 2007) is a promising field of research. A global sensitivity analysis of the landscape model aiming to rank the relative role of model parameters on the probabilities of extinction and coexistence as well as on the proportion of EM strains at equilibrium could help to point out the major processes that should be studied in priority at the landscape scale. Secondly this need arises because alternation between hosts in a heterogeneous landscape tends to select for generalist viruses which in turn are more likely to jump over the species boundary (Elena et al., 2009), and are thus more likely to successfully emerge.

Acknowledgments

We are grateful to Catherine Wipf-Scheibel and Charlotte Chandeysson for their excellent technical assistance. We also thank Amandine Lê Van for her enthusiasm with early

522 modelling work on this dataset and Denis Fargette and Cindy E. Morris for their comments
523 and help with the English.

524

Accepted manuscript

References

Aguilar, I., Sánchez, F., Ponz, F., 2000. Different forms of interference between two tobamoviruses in two different hosts. *Plant Pathol.* 49, 659-665.

Anderson, P., Cunningham, A., Patel, N., Morales, F., Epstein, P., Daszak, P., 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 19, 535–44.

Atabekov, J.G., Dorokhov, Y.L., 1984. Plant virus-specific transport function and resistance of plants to viruses. *Adv. Vir. Res.* 29, 313-364.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.

Brooks, S.P., Gelman, A., 1998. General Methods for Monitoring Convergence of Iterative Simulations. *J. Comp. Graph. Stat.* 7, 434-55.

Clark, J.S., 2005. Why environmental scientists are becoming Bayesians. *Ecol. Lett.* 8, 2-14.

Cressie, N., Calder, C.A., Clark, J.S., Ver Hoeff, J.M., Wikle, C.K., 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modelling. *Ecol. Appl.* 19, 553-570.

Desbiez, C., Costa, C., Wipf-Scheibel, C., Girard, M., Lecoq, H., 2007. Serological and molecular variability of Watermelon mosaic virus (genus Potyvirus). Arch. Virol. 152, 775-781.

Desbiez, C., Lecoq, H. 2008. Evidence for multiple intraspecific recombinants in natural populations of Watermelon mosaic virus (WMV). Arch. Virol. 153, 1749-1754.

Desbiez, C., Joannon, B., Wipf-Scheibel, C., Chandeysson, C., Lecoq, H. 2009. Emergence of new strains of Watermelon mosaic virus in South-eastern France: Evidence for limited spread but rapid local population shift. Vir. Res. 141, 201-208.

Dietrich, C., Maiss, E., 2003. Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected *Nicotiana benthamiana* plants. J. Gen. Virol. 84, 2871-2876.

Dodds, J.A., Hamilton, R.I., 1972. The influence of Barley stripe mosaic virus on the replication of TMV in *Hordeum vulgare* L. Virology 50, 404-411.

Dybiec, B., Kleczkowski, A., Gilligan, C.A., 2004. Controlling disease spread on networks with incomplete knowledge. Phys. Rev. E 70, 1-5.

Elena, S.F., Agudelo-Romero, P., Lalić, J., 2009. The evolution of viruses in multi-host fitness landscapes. Open Virol. J. 3, 1-6.

Ferguson, N.M., Galvani, A.P., Bush, R.M., 2003. Ecological and immunological determinants of influenza evolution. *Nature* 422, 428–433.

Fraser, R.S.S., 1998. Introduction to classical cross protection. In: Foster, G.D., Taylor, S.C. (Eds), *Plant Virology Protocols*. Humana Press, Totowa, New Jersey, pp 13-24.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. *Bayesian Data Analysis*, second edn. Chapman & Hall, London, UK.

Gilligan, C.A., 2008. Sustainable agriculture and plant diseases: an epidemiological perspective. *Phil. Trans. R. Soc. B* 363, 741-759.

Gilligan, C.A., van den Bosch, F., 2008. Epidemiological models for invasion and persistence of pathogens. *Ann. Rev. Phytopath.* 46, 385-418.

Gómez, P., Sempere, R.N., Elena, S.F., Aranda, M.A., 2009. Mixed infections of Pepino Mosaic Virus strains modulate the evolutionary dynamics of this emergent virus. *J. Virol.* 83, 12378-12387.

Holmes, E.C., Drummond, A., 2007. The evolutionary genetics of viral emergence. In: Childs, J.E., Mackenzie, J.S., Richt, J.A. (Eds), *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission*. Springer, Berlin Heidelberg, Germany, pp 51-66.

Jones, R.A.C., 2009. Plant virus emergence and evolution: Origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Res.* 141, 113-130.

Karyeija, R.F., Kreuze, J.F., Gibson, R.W. and Valkonen, J.P.T. (2000) Synergistic interactions of a Potyvirus and a phloem-limited Crinivirus in sweet potato plants. *Virology* 269, 26-36.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. *J. Am. Stat. Asso.* 90, 773-795.

Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J., Grenfell, B.T., 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* 294, 813–817.

Krstic, B., 1995. Cross-protection between strains of Sugarcane mosaic, Maize dwarf mosaic, Johnsongrass mosaic, and Sorghum mosaic Potyviruses. *Plant Dis.* 79, 135-138.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetic Analysis and sequence alignment. *Brief. Bioinform.* 5, 150-163.

Latham, J.R. and Wilson, A.K. (2008) Transcomplementation and synergism in plants: implications for viral transgenes? *Molecular Plant Pathology* 9, 85-103.

Lecoq, H., 1998. Control of plant virus diseases by cross protection. In: Hadidi, A., Khetarpal, R.K., Koganezawa, H. (Eds), *Plant virus disease control*. The American Phytopathological Society, St Paul, Minnesota, pp 33-40.

Lecoq, H., Desbiez, C., 2008. Watermelon mosaic virus and Zucchini yellow mosaic virus. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds), Encyclopedia of Virology, Third edn. Elsevier, Oxford, UK, pp 433-440.

Nakazono-Nagaoka, E., Takahashi, T., Shimizu, T., Kosaka, Y., Natsuaki, T., Omura, T., Sasaya, T., 2009. Cross-protection against Bean yellow mosaic virus (BYMV) and Clover yellow vein virus by attenuated BYMV isolate M11. *Phytopathology* 99, 251-257.

Otten, W., Filipe, J.A.N., Bailey, D.J., Gilligan, C.A., 2003. Quantification and analysis of transmission rates for soilborne epidemics. *Ecology* 84, 3232–3239.

Plantegenest, M., Le May, C., Fabre, F., 2007. Landscape epidemiology of plant diseases. *J. R. Soc. Interface* 4, 963-972.

Pulliam, J.R.C., 2008. Viral Host Jumps: Moving toward a Predictive Framework. *EcoHealth* 5, 80-91.

Ratcliff, F.G., MacFarlane, S.A., Baulcombe, D.C., 1999. Gene silencing without DNA: RNA-mediated cross-protection between viruses. *Plant Cell* 11, 1207-1215.

Restif, O., Wolfe, D.N., Goebel, E.M., Bjornstad, O.N., Harvill, E.T., 2008. Of mice and men: asymmetric interactions between *Bordetella* pathogen species. *Parasitology* 135, 1517-1529.

Rivot, E., Prévost, E., Cuzol, A., Baglinière, J., Parent, E., 2008. Hierarchical Bayesian modelling with habitat and time covariates for estimating riverine fish population size by successive removal method. *Can. J. Fish. Aquat. Sci.* 65, 117–133.

Rojas, M.R., Gilbertson, R.L., 2008. Emerging Plant Viruses: a Diversity of Mechanisms and Opportunities. In: Roossinck, M.J. (Ed), *Plant Virus Evolution*. Springer-Verlag Berlin Heidelberg, Germany, pp 28-48.

Sax, D.F., Brown, J.H., 2000. The paradox of invasion. *Global Ecol. Biogeo.* 9, 363-371.

Soubeyrand, S., Laine, A.L., Hanski, I., Penttinen, A., 2009. Spatio-temporal structure of host-pathogen interactions in a metapopulation. *Am. Nat.* 174, 308-320.

Strange, R.N., Scott, P.R., 2005. Plant disease: a threat to global food security. *Ann. Rev. Phytopath.* 43, 83–116.

Takahashi, T., Sugawara, T., Yamatsuta, T., Isogai, M., Natsuaki, T., Yoshikawa, N., 2007. Analysis of the spatial distribution of identical and two distinct virus populations differently labelled with cyan and yellow fluorescent proteins in coinfecting plants. *Phytopathology* 97, 1200-1207.

Thomas, A., O Hara, B., Ligges, U., Sturtz, S., 2006. Making BUGS Open. *R News* 6, 12-17.

Tian, Z., Qiu, J., Yu, J., Han, C., Liu, W., 2009. Competition between Cucumber mosaic virus subgroup I and II isolates in Tobacco. *J. Phytopath.* 157, 457-464.

Valkonen, J.P.T., Rajamäki, M.L., Kekarainen, T., 2002. Mapping of viral genomic regions important in cross-protection between strains of a potyvirus. *Mol. Plant Pathol.* 15, 683-692.

Varma, A., Malathi, V.G., 2003. Emerging geminivirus problems: a serious threat to crop production. *Ann. Appl. Biol.* 142, 145–164.

Wang, H.L., Gonsalves, D., Provvidenti, R., Lecoq, H., 1991. Effectiveness of cross protection by a mild strain of Zucchini yellow mosaic virus in cucumber, melon, and squash. *Plant Dis.* 75, 203-207.

Wang, Y., Lee, K.-C., Wong, S.-M., Palukaitis, P. and Gal-On, A. (2004) Breakage of resistance to Cucumber mosaic virus by co-infection with Zucchini yellow mosaic virus: enhancement of CMV accumulation independent of symptom expression. *Arch. Virol.* 149, 379-396.

Ward, C.W., Shukla, D.D., 1991. Taxonomy of potyviruses: current problems and some solutions. *Intervirology* 32, 269-296.

Wolfe, D.N., Goebel, E.M., Bjørnstad, O.N., Restif, O., Harvill, E.T., 2007. The O antigen enables *Bordetella parapertussis* to avoid *Bordetella pertussis*-induced immunity. *Infect. Immun.* 75, 4972–4979.

Woolhouse, M.E.J., Haydon, D., Antia, R., 2005. Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.* 20, 238–244.

Xia, X., 2000. Data analysis in molecular biology and evolution. Kluwer Academic Publishers, Boston.

You, B.J., Chiang, C.H., Chen, L.F., Su, W.C., Yeh, S.D., 2005. Engineered Mild Strains of Papaya ringspot virus for Broader Cross Protection in Cucurbits. *Phytopathology* 95, 533-540.

Zhang, X.S., Holt, J., Colvin, J., 2000. Mathematical models of host plant infection by helper-dependent virus complexes: Why are helper viruses always avirulent ? *Phytopathology* 90, 85-93.

Zhang, X.S., Holt, J., 2001. Mathematical models of cross protection in the epidemiology of plant-virus diseases. *Phytopathology* 91, 924-934.

Table 1. Statistical summary of the inference on parameters of model M_0 . Main statistics (mean, q-5%: percentile 5%, q-25%: quartile 25%, median, q-75%: quartile 75%, q-95: percentile 95%) of the marginal posterior distribution are derived from a MCMC sample of size 10^5 . For the secondary infection process, the values provided for the β 's parameters (transformed in the logit^{-1} scale), estimate directly the probability that a plant infected with a CL or an EM strain is the source of infection of an H plant by secondary infection during a Δt time step ($q_i(SI_{CL}|H)$ and $q_i(SI_{EM}|H)$ with $1 \leq i \leq 3$).

Process	Parameter	mean	q-5%	q-25%	median	q-75%	q-95%
Primary infection	$\alpha_{1,CL}$	-3.49	-7.20	-4.90	-3.41	-2.09	-0.10
	$\alpha_{2,CL}$	-4.85	-7.99	-6.32	-4.99	-3.58	-0.79
	$\alpha_{3,CL}$	-4.54	-7.59	-5.56	-4.40	-3.47	-1.87
	$\alpha_{1,EM}$	-3.69	-7.32	-5.04	-3.62	-2.34	-0.51
	$\alpha_{2,EM}$	-4.36	-7.52	-5.82	-4.50	-3.06	-0.27
	$\alpha_{3,EM}$	-2.36	-5.40	-3.35	-2.25	-1.32	0.33
	σ	3.28	1.80	2.41	3.01	3.84	5.64
Secondary infection	$\text{logit}^{-1}(\beta_{1,CL})$	0.050	0.031	0.046	0.051	0.057	0.064
	$\text{logit}^{-1}(\beta_{2,CL})$	0.033	0.025	0.030	0.033	0.037	0.042
	$\text{logit}^{-1}(\beta_{3,CL})$	0.071	0.039	0.054	0.067	0.084	0.119
	$\text{logit}^{-1}(\beta_{1,EM})$	0.039	0.022	0.034	0.040	0.044	0.051
	$\text{logit}^{-1}(\beta_{2,EM})$	0.011	0.007	0.009	0.011	0.013	0.015
	$\text{logit}^{-1}(\beta_{3,EM})$	0.009	10^{-8}	10^{-5}	0.001	0.013	0.042
	π	0.56	0.34	0.46	0.56	0.66	0.79
Over Infection	γ_{CL}	2.95	2.08	2.54	2.90	3.31	3.98
	γ_{EM}	8.92	6.09	7.51	8.70	10.09	12.54

Figure legends

Fig. 1. Possible transitions from time $t-1$ to time t in year i between the 4 plant compartments (H: healthy, CL: infected by a classic strain, EM: infected by an emerging strain and D: infected by both a CL and an EM strain). **a)** The possible transitions of an H plant depend on the probabilities of 4 events: (i,ii) $p_{i,t}^{PI}(CL|H)$ [resp. $p_{i,t}^{PI}(EM|H)$], probability that “an H plant is infected by primary infection with a CL [resp. EM] strain during Δt ” and (iii,iv) $p_{i,t}^{SI}(CL|H)$ [resp. $p_{i,t}^{SI}(EM|H)$], probability that “an H plant is infected by secondary infection of a CL [resp. EM] strain during Δt ”. **b,c)** The possible transitions of a CL [resp. EM] plant depend on the probabilities of 2 events: (i) $p_{i,t}^{PI}(D|CL)$ [resp. $p_{i,t}^{PI}(D|EM)$], probability that “a CL [resp. EM] plant is infected by primary infection with a EM [resp. CL] strain during Δt ” and (ii) $p_{i,t}^{SI}(D|CL)$ [resp. $p_{i,t}^{SI}(D|EM)$] probability that “a CL [resp. EM] plant is infected by secondary infection of an EM [resp. CL] strain during Δt ”.

Fig. 2. Observed and predicted temporal dynamics of WMV epidemics in cucurbit plots from 2002 to 2004. The field contained 160 cucurbit plants. Blue, red and green dots indicate the observation of the number of plants infected with a classical strain (CL), infected with an emerging strain (EM) and infected with both strains (D) respectively. Lines indicate the mean number of CL, EM and D plants respectively as predicted by the model. For each prediction, the extent of a 90% posterior confidence interval is indicated.

Fig. 3. Mean annual probabilities of primary infection by CL and EM strains from 2002 to 2004. Boxes indicate the lower (25%) and upper (75%) quartiles, and symbols indicate the mean values (stars for CL strain, dots for EM strain). Dotted lines extending from each end of the box show the extent of the 90% posterior confidence interval.

Fig. 4. Simulated and observed dynamics of the proportion of EM strains of WMV following its introduction in a landscape. **a)** Dynamics of the proportion of EM strains of WMV during the 40 first years ($p_{EM}(i)$ with $i \in \{1..40\}$) after their introduction in a landscape composed of 25 fields, each field having 160 cucurbit plants. The introduction consists in putting, at the onset ($i=1$, $t_{intro}=1$) and in 1 field, a single plant infected with an EM strain. The dynamics of $p_{EM}(i)$ is indicated for 4 values of δ , the ratio between γ_{EM}^s and γ_{CL}^s ($\delta=1$ in red, $\delta=3$ in black, $\delta=10$ in green and $\delta=40$ in blue). δ measures how many times it is less likely for a CL strain to infect an EM plant than for an EM strain to infect a CL plant. For each value of δ , a sample of 20 independent realizations of the model are plotted (dotted lines), and the mean of $p_{EM}(i)$ indicated in solid line (the mean is assessed with the 1000 simulation realized). Others model parameters were set to the values detailed in the model description section. **b)** Annual dynamics of the proportion of EM strains detected in the sample collected during an epidemiological survey performed from 2004 to 2008 in the 3 neighbouring regions where EM strains were first detected in France.

Fig. 5. Effect of γ_{EM}^s , γ_{CL}^s and their ratio δ on the final state of the landscape model 40 years after the introduction in 1 field out of 25 a single plant infected with an EM strain of WMV. γ_{EM}^s [resp. γ_{CL}^s] measures how many times over-infecting an EM [resp. CL] plant with a CL [resp. EM] strain is less likely than infecting an H plant with a CL [resp. EM] strain. Their ratio $\delta = \gamma_{EM}^s / \gamma_{CL}^s$ is then the differential ability for over-infection between EM and CL strains. **a)** Probabilities that EM and CL still coexist in the landscape 40 years after the initial introduction as a function of γ_{EM}^s and γ_{CL}^s . **b)** When EM and CL strains coexist at $y_{end}=40$, the mean value of $p_{EM}(y_{end})$ is plotted as a function of γ_{EM}^s and γ_{CL}^s . **c)** Probabilities that EM strains disappeared in the landscape as a function of γ_{EM}^s and γ_{CL}^s . **d)** Probabilities that EM

775 strains replaced CL strains in the landscape as a function of γ_{EM}^s and γ_{CL}^s . For all graphs,
776 model parameters other than γ_{EM}^s and γ_{CL}^s were set to the values detailed in the model
777 description section. Probabilities were estimated based on 1000 independent realizations of
778 the model. In graphs **a)** and **b)**, the black dot is positioned at the value of γ_{EM}^s and γ_{CL}^s
779 estimated from the field data set ($\gamma_{EM}^s = 9.19$ and $\gamma_{CL}^s = 2.89$).

780

Accepted manuscript

Appendix A. The model describing the co-dynamics of viral strains at field scale

In the model, plants were classified into 4 compartments: (i) H (healthy), (ii) CL (infected by a classic strain), (iii) EM (infected by an emerging strain) and (iv) D (doubly infected by both a CL and an EM strain). For a given year i and date t of observation, $N(i, t) = [N_H(i, t), N_{CL}(i, t), N_{EM}(i, t), N_D(i, t)]$ is the vector of the number of plants in each compartment. The model describes the dynamic of $N(i, t)$ at a Δt time interval equal to 1 week given that initially all plants are healthy: $N(i, 0) = [N_{tot}, 0, 0, 0]$ ($N_{tot} = 160$).

Probabilities of infection of an H plant from $t-1$ to t

From time $t-1$ to time t , an H plant can remain H or become CL, EM or D with probabilities $p_{i,t}^M = (p_{i,t}(H_t | H_{t-1}), p_{i,t}(CL_t | H_{t-1}), p_{i,t}(EM_t | H_{t-1}), p_{i,t}(D_t | H_{t-1}))$. These transitions between compartments depend on the probabilities of 4 events (Fig. 1a).

$p_{i,t}^{PI}(CL | H)$ [resp. $p_{i,t}^{PI}(EM | H)$], are the probabilities that “an H plant is infected by primary infection with a CL [resp. EM] strain during Δt ”. They are modelled as:

$$\text{logit}[p_{i,t}^{PI}(CL | H)] = \alpha_{i,CL} + \gamma_{i,t} \quad (A1)$$

$$\text{logit}[p_{i,t}^{PI}(EM | H)] = \alpha_{i,EM} + \gamma_{i,t} \quad (A2)$$

where (i) $\alpha_{i,CL}$ [resp. $\alpha_{i,EM}$] is, in the logit scale, a parameter describing the probability that an H plant is infected by primary infection with a CL [resp. EM] strain during Δt in year i if no variation between weeks exists and (ii) $\gamma_{i,t}$ is a set of random effects mutually independent and identically normally distributed with unknown variance σ^2 .

$p_{i,t}^{SI}(CL | H)$ [resp. $p_{i,t}^{SI}(EM | H)$] are probabilities that “an H plant is infected by secondary infection of a CL [resp. EM] strain during Δt ”. They depend on the probabilities $q_i(SI_{CL} | H)$

[resp. $q_i(SI_{EM} | H)$] that “a plant infected in the field with a CL [resp. EM] strain in year i infects an H plant by secondary infection during Δt ”. In the model, it is assumed that

$$\text{logit}[q_i(SI_{CL} | H)] = \beta_{i,CL} \quad (A3)$$

$$\text{logit}[q_i(SI_{EM} | H)] = \beta_{i,EM} \quad (A4)$$

where $\beta_{i,CL}$ [resp. $\beta_{i,EM}$] is, in the logit scale, a parameter modelling the probability that a plant infected with a CL [resp. EM] strain is a source of infection for an H plant by secondary spread during Δt in year i .

Following the calculations detailed in appendix B, it comes that

$$p_{i,t}^{SI}(CL | H) = 1 - [1 - q_i(SI_{CL} | H)]^{N_{CL}(i,t-1) + (1-\pi)N_D(i,t-1)} \quad (A5)$$

$$p_{i,t}^{SI}(EM | H) = 1 - [1 - q_i(SI_{EM} | H)]^{N_{EM}(i,t-1) + \pi N_D(i,t-1)} \quad (A6)$$

where π is a parameter modelling the mean proportion of EM strains in D plants ($1-\pi$ is the proportion of CL strains).

Finally, assuming that during Δt primary and secondary infection events occur independently, the probability that, in year i , a plant is H [resp. CL, EM and D] at time t given that this plant was H at time $t-1$ can be derived:

$$p_{i,t}(H_t | H_{t-1}) = [1 - p_{i,t}^{PI}(CL | H)] \cdot [1 - p_{i,t}^{SI}(CL | H)] \cdot [1 - p_{i,t}^{PI}(EM | H)] \cdot [1 - p_{i,t}^{SI}(EM | H)] \quad (A7)$$

$$p_{i,t}(CL_t | H_{t-1}) = [1 - p_{i,t}^{PI}(EM | H)] \cdot [1 - p_{i,t}^{SI}(EM | H)] \cdot [p_{i,t}^{PI}(CL | H) \cdot p_{i,t}^{SI}(CL | H) + p_{i,t}^{PI}(CL | H) \cdot (1 - p_{i,t}^{SI}(CL | H)) + (1 - p_{i,t}^{PI}(CL | H)) \cdot p_{i,t}^{SI}(CL | H)] \quad (A8)$$

$$p_{i,t}(EM_t | H_{t-1}) = [1 - p_{i,t}^{PI}(CL | H)] \cdot [1 - p_{i,t}^{SI}(CL | H)] \cdot [p_{i,t}^{PI}(EM | H) \cdot p_{i,t}^{SI}(EM | H) + p_{i,t}^{PI}(EM | H) \cdot (1 - p_{i,t}^{SI}(EM | H)) + (1 - p_{i,t}^{PI}(EM | H)) \cdot p_{i,t}^{SI}(EM | H)] \quad (A9)$$

$$p_{i,t}(D_t | H_{t-1}) = 1 - p_{i,t}(H_t | H_{t-1}) - p_{i,t}(CL_t | H_{t-1}) - p_{i,t}(EM_t | H_{t-1}) \quad (A10)$$

Probabilities of over-infection of a singly infected plant from $t-1$ to t

From time $t-1$ to time t , a CL plant can remain CL or become D with probability $p_{i,t}(D_t | CL_{t-1})$ that depends on the probabilities of 2 events (Fig. 1b): (i) $p_{i,t}^{PI}(D | CL)$, the probability that “a CL plant is infected by primary infection with a EM strain during Δt ” and (ii) $p_{i,t}^{SI}(D | CL)$, the probability that “a CL plant is infected by secondary infection of an EM strain during Δt ”. As previously, deriving the latter probability required to introduce the probability $q_i(SI_{EM} | CL)$ that “a given plant infected in the field with an EM strain is the source of infection of a CL plant by secondary infection during Δt ”.

Defining the parameter γ_{CL} as the odd ratio between the probability of infecting an H plant with a EM strain and the probability of infecting an CL plant with a EM strain implies that

$$p_{i,t}^{PI}(D | CL) = \frac{\text{odds}[p_{i,t}^{PI}(EM | H)]}{\text{odds}[p_{i,t}^{PI}(EM | H)] + \gamma_{CL}} \quad (A11)$$

$$q_i(SI_{EM} | CL) = \frac{\text{odds}[q_i(SI_{EM} | H)]}{\text{odds}[q_i(SI_{EM} | H)] + \gamma_{CL}} \quad (A12)$$

As previously (Appendix B), it comes that

$$p_{i,t}^{SI}(D | CL) = 1 - [1 - q_i(SI_{EM} | CL)]^{N_{EM}(i,t-1) + \pi N_D(i,t-1)} \quad (A13)$$

Finally, assuming independence between primary and secondary infections, leads to

$$p_{i,t}(D_t | CL_{t-1}) = 1 - [1 - p_{i,t}^{PI}(D | CL)] \cdot [1 - p_{i,t}^{SI}(D | CL)] \quad (A14)$$

Similarly, an EM plant at time $t-1$ can remain EM or become D at time t (Fig. 1c) with probability $p_{i,t}(D_t | EM_{t-1})$. The same developments as previously lead successively to

$$p_{i,t}^{PI}(D | EM) = \frac{\text{odds}[p_{i,t}^{PI}(CL | H)]}{\text{odds}[p_{i,t}^{PI}(CL | H)] + \gamma_{EM}} \quad (A15)$$

$$q_i(SI_{CL} | EM) = \frac{odds[q_i(SI_{CL} | H)]}{odds[q_i(SI_{CL} | H)] + \gamma_{EM}} \quad (A16)$$

$$p_{i,t}^{SI}(D | EM) = 1 - [1 - q_i(SI_{CL} | EM)]^{N_{CL}(i,t-1) + (1-\pi)N_D(i,t-1)} \quad (A17)$$

And

$$p_{i,t}(D_t | EM_{t-1}) = 1 - [1 - p_{i,t}^{PI}(D | EM)] \cdot [1 - p_{i,t}^{SI}(D | EM)] \quad (A18)$$

850 ***Dynamics of $N(i,t)$ during an epidemic***

851 The dynamic of $N(i,t)$ from $t=0$ to t_{end} is modelled as a sequential 3 step stochastic processes:

852 (i) $\Delta H(i,t) = [\Delta H_H(i,t), \Delta H_{CL}(i,t), \Delta H_{EM}(i,t), \Delta H_D(i,t)]$, the vector of the number of H
853 plants in year i at date $t-1$ remaining H at date t or becoming respectively CL, EM or D at date
854 t , is drawn from a multinomial distribution:

$$\Delta H(i,t) \sim Multinomial(N_H(i,t-1), p_{i,t}^M) \quad (A19)$$

856 (ii) $\Delta CL(i,t)$, the number of CL plants in year i at date $t-1$ becoming D at date t is the
857 result of a binomial process:

$$\Delta CL(i,t) \sim Binomial(N_{CL}(i,t-1), p_{i,t}(D_t | CL_{t-1})) \quad (A20)$$

859 (iii) $\Delta EM(i,t)$, the number of EM plants in year i at date $t-1$ becoming D at date t is the
860 result of a binomial process:

$$\Delta EM(i,t) \sim Binomial(N_{EM}(i,t-1), p_{i,t}(D_t | EM_{t-1})) \quad (A21)$$

862 Then $N(i,t)$ is updated from time $t-1$ to time t ,

$$N(i,t) = \begin{pmatrix} N_H(i,t-1) - \Delta H_{CL}(i,t) - \Delta H_{EM}(i,t) - \Delta H_D(i,t) \\ N_{CL}(i,t-1) + \Delta H_{CL}(i,t) - \Delta CL(i,t) \\ N_{EM}(i,t-1) + \Delta H_{EM}(i,t) - \Delta EM(i,t) \\ N_D(i,t-1) + \Delta H_D(i,t) + \Delta CL(i,t) + \Delta EM(i,t) \end{pmatrix} \quad (A22)$$

Appendix B: derivation of the probabilities of secondary infection

Suppose that each infected plant sends on another plant (through aphid population dynamics) a Poisson number of viral particles with mean λ independent of the strain of the viral particles. Suppose also that the frequency of viral particles of EM strain is π in doubly infected plants. Thus a plant infected by the two viral strain sends a Poisson number of viral particles of EM strain with mean $\pi.\lambda$ and a Poisson number of particles of CL strain with mean $(1 - \pi).\lambda$.

A given plant then receives a Poisson number of viral particles of EM strain with mean $\Lambda = (N_{EM} + \pi N_D).\lambda$ where N_{EM} is the number of plants infected with EM strain only at the preceding date and N_D the number of plants doubly infected at the same date.

Let $p(EM, H)$ be the probability that a viral particle of EM strain is transmitted to an healthy plant at a given date.

Suppose that this healthy plant receives n particles of strain EM, and that each particle can be transmitted independently from another particle. The probability that this plant is not infected by these particles is $[1 - p(EM, H)]^n$.

A healthy plant receiving a Poisson number of viral particles of EM strain with mean $\Lambda = (N_{EM} + \pi N_D).\lambda$ remains uninfected by an EM strain with probability

$$\begin{aligned} E\left[(1 - p(EM, H))^n\right] &= \sum_{n=0}^{\infty} \frac{(1 - p(EM, H))^n}{n!} \Lambda^n \exp(-\Lambda) \\ &= \exp(-p(EM, H).\Lambda) \\ &= \exp[-p(EM, H).(N_{EM} + \pi.N_D).\lambda] \end{aligned}$$

Similarly, an healthy plant is infected by a given EM plant (so that $N_{EM} = 1$ and $N_D = 0$ in the preceding formula) with probability

$$q(SI_{EM} | H) = 1 - \exp(-p(EM, H).\lambda)$$

888 Thus a healthy plant receiving a Poisson number of viral particles of EM strain with mean

889 $\Lambda = (N_{EM} + \pi N_D) \cdot \lambda$ remains uninfected by an EM strain with probability

890
$$\exp[-p(EM, H) \cdot (N_{EM} + \pi N_D) \lambda] = [1 - q(SI_{EM} | H)]^{N_{EM} + \pi N_D}$$

891 and becomes EM infected with probability

892
$$1 - [1 - q(SI_{EM} | H)]^{N_{EM} + \pi N_D}$$

893

Accepted manuscript

Appendix C: Model inferences

Link between the observed data and the model

During the first weeks until 100% of the plants became infected by WMV (week 3 in 2002 and 2004, and week 4 in 2003), maps of the epidemics were established at a weekly interval. This is precisely the Δt time step adopted in the model. Thus, assuming that the DAS-ELISA and RT-PCR tests used to detect viruses have high levels of sensitivity and specificity, both $\Delta H(i, t)$, $\Delta CL(i, t)$ and $\Delta EM(i, t)$ are directly observed. After this period of weekly data collection, a last map of the epidemic is observed 4 weeks later (week 7 in 2002, week 8 in 2003 and 2004 denoted thereafter t_{end}). Here, only $N(i, t_{end})$ is observed. Given that no map was observed in week $t_{end} - 1$ it is impossible to derive $\Delta H(i, t_{end})$, $\Delta CL(i, t_{end})$ and $\Delta EM(i, t_{end})$. However, $N(i, t_{end})$ can be related to the model output $N(i, t_{end}-1)$ by noting that

$$N(i, t_{end}) \sim Multinomial \left(N_{tot}, \begin{pmatrix} p_i(H_{t_{end}}) \\ p_i(CL_{t_{end}}) \\ p_i(EM_{t_{end}}) \\ p_i(D_{t_{end}}) \end{pmatrix} \right)$$

where the probabilities of being respectively H, CL, EM and D at t_{end} during year i can be derived using classical conditional probability rules given for example that to be H at time t a plant must be H at time $t-1$ and not be infected from times $t-1$ to t :

$$p_i(H_{t_{end}}) = p_i(H_{t_{end}-1}) \cdot p_i(H_{t_{end}} | H_{t_{end}-1})$$

$$\text{And, because } p(H_{t_{end}-1}) = \frac{N_H(i, t_{end}-1)}{N_{tot}}, \text{ it comes}$$

$$p_i(H_{t_{end}}) = \frac{N_H(i, t_{end}-1)}{N_{tot}} \cdot p_i(H_{t_{end}} | H_{t_{end}-1})$$

Similarly, it results that

$$p_i(CL_{t_{end}}) = \frac{N_{CL}(i, t_{end}-1)}{N_{tot}} \cdot [1 - p_i(D_{t_{end}} | CL_{t_{end}-1})] + \frac{N_H(i, t_{end}-1)}{N_{tot}} \cdot p_i(CL_{t_{end}} | H_{t_{end}-1})$$

$$p_i(EM_{t_{end}}) = \frac{N_{EM}(i, t_{end}-1)}{N_{tot}} \cdot [1 - p_i(D_{t_{end}} | EM_{t_{end}-1})] + \frac{N_H(i, t_{end}-1)}{N_{tot}} \cdot p_i(EM_{t_{end}} | H_{t_{end}-1})$$

$$p_i(D_{t_{end}}) = 1 - p_i(H_{t_{end}}) - p_i(CL_{t_{end}}) - p_i(EM_{t_{end}})$$

916

917 These equations defined a Markov chain linking dates $t-1$ to t . The likelihood at t_{end} , given
 918 the last observation realized in the field, is obtained by numerical integration of all the
 919 possible states of $N(i, t_{end}-1)$ by simulation of this Markov chain.

920

921 Bayesian statistical inferences

922 Statistical inferences were performed via Bayesian Monte Carlo Markov Chain (MCMC)
 923 methods (Gelman et al. 2004) with OpenBUGS[®] 3.0.3 (Bayesian inference Using Gibbs
 924 Sampling) software (available from <http://www.mrc-bsu.cam.ac.uk/bugs/>) (Thomas et al.
 925 2006).

926 The vector of model parameters is $\theta = [\alpha_{CL,i}, \alpha_{EM,i}, \beta_{CL,i}, \beta_{EM,i}, \gamma_{EM}, \gamma_{CL}, \pi, \sigma]_{i=1..3}$. To reflect our
 927 initial lack of knowledge we used uninformative prior probability density function (PDF) for
 928 θ . Log-uniform $[10^{-3}, 10^3]$ probability distributions were assigned to γ_{CL} and γ_{EM} .
 929 Normal(0,100) probability distributions were assigned to α 's and β 's parameters, a Beta(1,1)
 930 probability distribution assigned to π and a half positive Normal(0,100)I(0, $+\infty$) assigned to σ .

931 To check for convergence of the MCMC chain to its ergodic target distribution, we ran 3
 932 chains and used the Gelman-Rubin diagnostics implemented in OpenBUGS[®] 3.0.3 (Brooks
 933 and Gelman 1998). Marginal posterior PDFs were summarised by statistics computed from
 934 the MCMC samples: mean, median, 2.5th and 97.5th percentile defining the 95% Bayesian
 935 posterior credibility interval. From the joint posterior distribution of model parameter, derived
 936 quantities such as the mean probability that an H plant is infected by primary or secondary
 937 infection are easy to obtain in OpenBUGS.

To check for consistency between the model and the data, the χ^2 discrepancy test proposed by Gelman (2004, p 175) was used (see also Rivot et al. (2008) for details on its practical implementation). χ^2 discrepancy statistics were computed for the observation of $N_{CL}(i,t)$, $N_{EM}(i,t)$, $N_D(i,t)$ summed over the available observation units i and t . Extreme Bayesian p values (<0.05 or >0.95) reveal inconsistencies between model predictions and the actual data. Simple predicted versus observed values of these variables were also regressed and the coefficient of determination (r^2), which provides a measure of the proportion of the total variance explained by the model, was estimated.

In order to investigate more precisely the role of the process of over-infection, a model comparison approach was used. Let's denote M_0 the previous model parameterisation. Two alternative formulations of M_0 were fitted to the dataset. In the model M_1 , it is assumed that $\gamma_{EM} = \gamma_{CL}$ (i.e. the probability of over-infecting an already infected plant differs from the probability of infecting an H plant but does not depend on the strain (CL or EM) first inoculated). In the model M_2 , it is assumed that $\gamma_{EM} = \gamma_{CL} = 1$ (i.e. the probability of over-infecting an already infected plant is the same as the probability of infecting an H plant). Models M_1 , M_2 and M_0 were compared using Bayes factors. The Bayes factor of 2 alternative models M_0 and M_1 , $BF_{1,0}$, is defined as the ratio of the marginal likelihood of the data under model M_1 , $[y|M_1]$, to that obtained under model M_0 , $[y|M_0]$. It is a measure of the relative credibility of M_1 compared with M_0 for given dataset. In practise, $[y|M_i]$ can be approximated as the harmonic mean of the likelihood function (Kass and Raftery, 1995). These approximations of Bayes factor are accurate enough to be interpreted in a logarithm scale according to guidelines provided by Kass and Raftery (1995).

4. Figure

4. Figure