Predicting durability of resistance to crop pathogens: when is incorporating stochasticity necessary?

Elin Falla

Queens' College

Dr Nik Cunniffe, Benjamin Watkinson-Powell

Word count: 5,856

Abstract

Genetically engineering disease-resistance in crops is an effective way to reduce yield loss due to pathogens. Therefore, ensuring maximum disease-resistance durability in the face of pathogen evolution is an increasingly pressing issue in modern agriculture. The majority of mathematical models in this study area are deterministic rather than stochastic, despite many of them involving scenarios where pathogen extinction is likely, such as emergence of a resistance-breaking trait. This study compared a deterministic and stochastic version of the same 2-patch continuous model in order to find scenarios where the outputs of the two differed. The results will help inform future studies about the importance of stochasticity in certain contexts, and highlight the importance of an understanding of the dynamics of a biological system in order to ensure modelling most closely represents reality.

---

1 Introduction

There is an increasing need for effective agricultural practices and crop disease control to ensure sufficient levels of food production. The evolution of pathogen populations to overcome resistant crop cultivars is therefore an area of major concern. It's also an area of concern for environmental reasons, as the absence of other working methods often leads to a greater use of chemical pesticides, which can cause damage to agricultural and non-agricultural ecosystems. It's a spectacular case of disease emergence as resistances can break down in a matter of years or even months[1]. Therefore, one of the key goals of virulence management is to increase the durability of crop resistance, defined by Johnson[2] as the period of time that resistance remains effective with its widespread cultivation. However, the concept has different facets, and can be assessed using different measures[3]. Examples include the time to invasion of resistance-breaking genotypes, and the time during which the resistant cultivar still contributes to lowering disease overall[4].

Resistance or susceptibility of plants to pathogens (defined here as the qualitative ability of a pathogen to infect the host) results from the interaction between plant disease resistance and corresponding pathogen virulence. It is often caused by a molecular relationship governed by a gene-for-gene interaction[5]. This means a single resistance (R) gene in the host provides qualitative resistance to a corresponding avirulence gene in the pathogen. Often the avirulence gene is an effector in the mechanism of infection[6]. However, mutations in the avirulence gene may allow it to evade the R gene, meaning the pathogen can infect the resistant host, at which point it's referred to as 'resistance-breaking'[5].

The resistance-breaking trait often has a high reproductive fitness cost, meaning that the resistance-breaking pathogen will experience negative selection pressure on a susceptible crop when the wild-type pathogen is also present[7]. This loss of unnecessary virulence can be seen in field experimental evolution studies[8]. However, as mutations can occur quite frequently, and with as few as two nucleotide substitutions required to become resistance-breaking[9], resistance-breaking strains can persist at low background frequencies in the pathogen population. The level of this persistence is a representation of how robust the R gene is: ie. a combination of fitness costs and number of necessary mutations to determine how likely it is that mutation(s) will produce a viable resistance-breaking trait[1].

Mathematical modelling is often used to examine the factors affecting resistance durability. In recent years, population genetics has been combined with epidemiology to measure durability in terms of the overall intensity of the epidemic in a given time frame, rather than just considering gene frequencies as was done previously[10,11]. Van den Bosch and Gilligan[3] were the first to do this, by measuring the number of additional uninfected host growth days as a durability measure. This work was further developed by others such as Fabre et al.[1] and Iacono et al.[12]

However, the majority of models in this field are deterministic: Iacono et al.[12] were the first to introduce stochasticity to the resistance durability-measuring strategy proposed by van den Bosch

and Gilligan[3]. Demographic stochasticity is defined as the variability in population density arising from random differences among individuals in reproduction and survival[13]. Although its consideration in models is increasing, existing models of resistance durability rarely consider it, despite that fact that many consider pathogen emergence where stochasticity is likely to play an important role[3]. Adding stochasticity moves from the question of whether or not the pathogen will invade to the probability of it occurring[13]. The aim of this study, therefore, is to compare how different genetic and epidemiological processes affect resistance durability in a stochastic versus a deterministic model. The results of the two models I have formulated will be compared to identify scenarios where stochasticity gives a substantially different result for the epidemic intensity (EI) and hence should be considered in model creation in order to more accurately capture the true dynamics of the epidemic system.

All processes are being analysed in a scenario with susceptible and resistant crops, where a resistance-breaking pathogen has just been introduced to the system. The focus of my model on disease spread at the landscape scale, combined with the focus on starting at low initial pathogen frequencies, makes it applicable to any gene-for-gene system with a wind-dispersed pathogen, such as powdery mildews, rusts and Septoria[14].

The genetic processes I will be considering are the 'robustness' of the R gene and the reproductive fitness cost of the resistance-breaking trait. The former is defined as the background frequency of the resistance-breaking trait as it first emerges, determined by the difficulty of producing a viable resistance-breaking strain. These two factors are being considered as they have been identified as being likely to affect the degree of extinctions of the resistance-breaking pathogen[1,3]. The epidemiological process being considered is the degree of connectivity of the landscape, ie. the degree of coupling between the susceptible and resistant crop cultivar. Iacono et al.[12] assessed the impact of cropping ratio in stochastic versus deterministic models, but the effect of landscape connectivity has not yet been considered. Increased landscape connectivity (ie. increased spatial heterogeneity) creates dilution effects, where some of the force of infection from a given pathogen

strain is 'wasted' due to its reduced ability or inability to infect the other patch[15]. How this affects the level of extinctions and hence the difference between the two model types is an area of interest.

2 Materials and methods

2.1 Model description

Model overview

I have created a stochastic and deterministic version of the same epidemic model, which is a variation of the model used by Fabre et al.[1] As is seen in Fabre's model, there are two pathogen strains: a 'wild-type' genotype (wt) and a 'resistance-breaking' genotype (rb). These strains act on two host crop plant genotypes: 'susceptible' (S) and 'resistant' (R), in a 2-patch model. Resistance acts in a gene-for-gene system where the wt strain can only infect S fields, as resistance is assumed to be complete. The rb strain has equal infective ability on both host genotypes (R and S) but has a fitness cost on reproduction (δ) expressed on both host genotypes associated with its resistance-breaking trait. The model is continuous, describing in-season epidemic dynamics in the form of an SIS model. SIS refers to a system whereby a host plant can be either susceptible (uninfected) (S) or infected (I). Individual plants can both become infected, moving from the S to I compartment, and die and be replanted, passing back from the I to the S compartment (Figure 1). The total number of plants (N) is constant; each plant of a given cultivar is in one of the following states at any one time:

Resistant (R) cultivar

- S_R - uninfected resistant cultivar
- I_R^{rb} - resistant infected by resistance-breaking pathogen

Susceptible (S) cultivar

- S_S - uninfected susceptible cultivar
- I_S^{wt} - susceptible infected by wild-type pathogen

- I_S^{rb} - susceptible infected by resistance-breaking pathogen

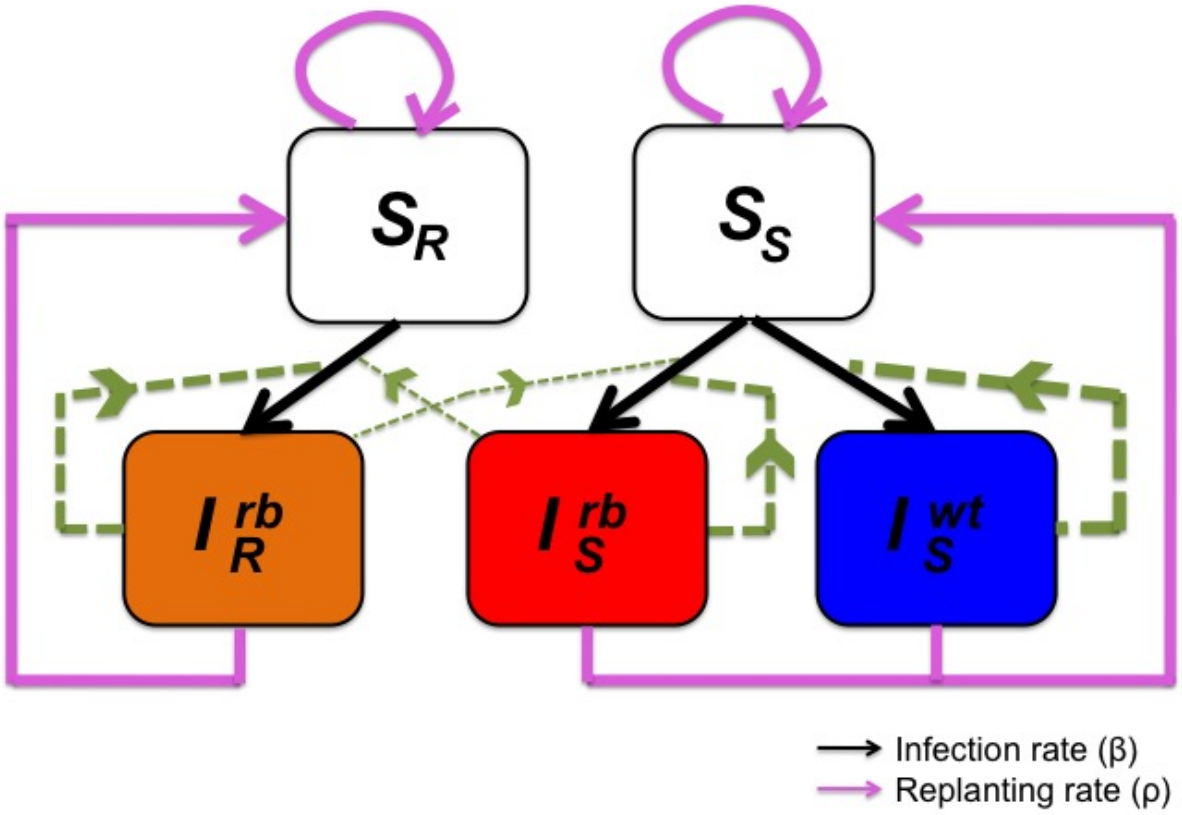

Figure 1: Model schematic showing compartments and transitions between them. Green dotted lines represent the effect of one state variable on the per capita rate of transition of another. Thickness of green lines represents the magnitude of effect. Black lines represent β, the underlying infection rate. Purple lines represent ρ, the replanting rate that moves individual plants from the infected back to susceptible classes.

The epidemic runs over a specified time frame (t_f) whereby integer values represent seasons, in a continuous time dynamic. Over the time frame, plants can become infected through one of two routes: either by inoculum from the same type of host (p) or the other host type ($1-p$). These are expressed as a function of a coupling coefficient (ϵ) in equations 1 and 2 respectively. ϵ provides implicit spatial structure to the model by determining the degree to which the two cultivar types are mixed (in terms of pathogen dispersal), ie. the degree of spatial heterogeneity/landscape connectivity. An ϵ value of 0 means the S and R cultivar are completely separate and a value of

1 means the two cultivar types are completely mixed.

$$p = \frac{1}{1+\epsilon} \tag{1}$$

$$1 - p = \frac{\epsilon}{1+\epsilon} \tag{2}$$

The rate of infection of each genotype combination is given by the following equations:

$$\frac{dI_S^{wt}}{dt} = \beta S_S(pI_S^{wt}) - \rho I_S^{wt} \tag{3}$$

$$\frac{dI_S^{rb}}{dt} = \beta S_S(1-\delta)(pI_S^{rb} + (1-p)I_R^{rb}) - \rho I_S^{rb} \tag{4}$$

$$\frac{dI_R^{rb}}{dt} = \beta S_R(1-\delta)(pI_R^{rb} + (1-p)I_S^{rb}) - \rho I_R^{rb} \tag{5}$$

Equations for the rate of change of uninfected plants are not explicitly characterised, as due to the constant population size (N), the number of uninfected S and R crops at any point is given by equations 6 and 7 respectively, shown below.

$$S_S = N - I_S^{wt} - I_S^{rb} \tag{6}$$

$$S_R = N - I_S^{wt} \tag{7}$$

In equations 3 to 5, the 'replanting rate' is represented by ρ, which determines the rate at which individuals from all classes die and are replanted (Figure 1). The proportion of N plants that are resistant (R) is determined by the cropping ratio (ϕ), which represents the proportion of resistant cultivar. The treatment of coupling between the two patches is appropriate for a system with an even cropping ratio ($\rho = 0.5$). β is the underlying infection rate. The parameters and state variables used in the model and their reference values are summarized in Table 1.

The reference values (Table 1) define the baseline epidemiological context, that is both responsive to parameter changes and represents a realistic scenario. In this scenario, the initial proportion

Symbol	State variable	Reference value (constraints)
I_S^{wt}	Number of *wt*-infected Susceptible plants	-
I_S^{rb}	Number of *rb*-infected Susceptible plants	-
I_R^{rb}	Number of *rb*-infected Resistant plants	-
S_S	Number of uninfected Susceptible plants	-
S_R	Number of uninfected Resistant plants	-
	Parameter (units)	
ϵ	Coupling coefficient - degree of spatial heterogeneity	$0.5 \quad (0 \leq \epsilon \leq 1)$
p	Proportion of inoculum landing on the same type of host	$\frac{2}{3} \quad (\frac{1}{2} \leq p \leq 1$ corresponding to $0 \leq \epsilon \leq 1)$
β	Infection rate (per infected plant per year)	9×10^{-6}
ρ	Replanting rate - rate at which individuals from all classes die and are replanted (per year)	1
δ	Fitness cost of the *rb* trait	$0.3 \quad (0 \leq \delta \leq 1)$
ϕ	Cropping ratio - proportion of resistant cultivar	0.5
N	Total number of crop plants	10^6
t_f	Time frame - number of seasons/years	$40 \quad (0 \leq t_f \leq 40)$
θ	Initial proportion of infected (I) individuals	$\frac{1}{3}$
γ	Initial proportion of I individuals that are *rb*	3×10^{-6} $(3 \times 10^{-6} \leq \gamma \leq 4.5 \times 10^{-5})$
A_S^{wt}	Epidemic intensity (EI) of I_S^{wt}	$\int_{t=0}^{t_f} \frac{I_S^{wt}(t)}{N} dt$
A_S^{rb}	Epidemic intensity (EI) of I_S^{rb}	$\int_{t=0}^{t_f} \frac{I_S^{rb}(t)}{N} dt$
A_R^{rb}	Epidemic intensity (EI) of I_R^{rb}	$\int_{t=0}^{t_f} \frac{I_R^{rb}(t)}{N} dt$
A_{total}	Total epidemic intensity	$A_S^{wt} + A_S^{rb} + A_R^{rb}$

Table 1: List of all parameters and state variables used in model. Constraints are given for parameters that are varied.

of infected hosts (θ) is set at $\frac{1}{3}$, the wt equilibrium when both S and R groups are present. This is the value that I_S^{wt} tends towards over time in the event that there is no invasion by the rb pathogen. This is evidenced by Figure 2a and b, as the wt proportion infected remains unchanged until the rb invades ˜8 seasons in. However, of the initially infected hosts, only one is infected by the rb pathogen (ie. $\gamma = 3 \times 10^{-6}$ - explained in Section 2.2). Note that the model assumes that a proportion of initially infected individuals are infected by the rb pathogen: emergence of the rb pathogen is not considered. This is analogous to a situation where the R crop has been in use for some time and the model starts from the time point where the rb has just mutated into existence or immigrated in from elsewhere. This also aligns with the assumption that the number of hosts infected with rb starts off at a very low frequency (ie. low γ). The infection rate β was optimised to give a wt epidemic of $\frac{1}{3}$ at equilibrium. The cropping ratio (ϕ) and degree of coupling (ϵ) are both set at an intermediate value of 0.5. Throughout the current study the replanting rate ρ is fixed at 1 (per year), which is realistic as this represents seasonal harvesting and replanting of the crop.

Deterministic versus stochastic implementation

For the deterministic version of the model, the equations were differentiated numerically using the 'deSolve' R package (function: 'ode()').

In order to introduce stochasticity to the model, Gillespie's Direct Method was used[16]. The 'adaptivetau' R package (function: 'ssa.exact()') was used to implement this. For this package, all rate equations must be positive, as the Gillespie algorithm is event-driven. This means it must have separate consideration of all possible events and the state transitions associated with each.[17] Therefore, each of equations 3-6 was split into two separate equations, because as they were they represented two possible transitions each. Equations 8 and 9 show an example of this for I_S^{wt}.

$$Forward\ \frac{dI_S^{wt}}{dt} = \beta S_S(pI_S^{wt}) \tag{8}$$

$$Reverse\ \frac{dI_S^{wt}}{dt} = \rho I_S^{wt} \tag{9}$$

2.2 Model analysis

General: Effect of stochasticity

The aim of the analysis was to quantify the difference between the results given by the stochastic and deterministic versions of the model, in order to compare them. This was done by running the model to produce the infection dynamics over the time frame (t_f) (Figure 2), then calculating the Area Under Disease Progress Curve (AUDPC), which acts as a measure of the epidemic intensity (EI). This also acts as a proxy for pathogen population size. This was based on the proportion of individuals infected not the absolute number (ie. divided by N). Each calculation was done twice: once for the stochastic and once for the deterministic model. For the stochastic model, the AUDPC was calculated using the mean of all of the runs done. The AUDPC calculation for each group is given by equation 10. Only the infected compartments were considered.

$$A_S^{wt} = \int_{t=0}^{t_f} \frac{I_S^{wt}(t)}{N} dt \qquad A_S^{rb} = \int_{t=0}^{t_f} \frac{I_S^{rb}(t)}{N} dt \qquad A_R^{rb} = \int_{t=0}^{t_f} \frac{I_R^{rb}(t)}{N} dt \tag{10}$$

These values were then summed to give the total AUDPC for all infection classes (genotype combinations) as seen in equation 11. This resulted in two A_{total} values: one for the stochastic and one for the deterministic model (A_{total}^{St} and A_{total}^{De} respectively).

$$A_{total} = A_S^{wt} + A_S^{rb} + A_R^{rb} \tag{11}$$

Quantifying the difference between the two A_{total} values was done in two ways: a calculation of absolute difference (D_{Abs}) (equation 12) and of proportional difference (D_{Prop}) (equation 13). In both cases, the calculation is the degree to which the deterministic EI is larger than the stochastic EI. Therefore, in a scenario where the EI is larger in the stochastic model $D_{Abs} < 0$ and $D_{Prop} < 1$.

$$D_{Abs} = A_{total}^{De} - A_{total}^{St} \tag{12}$$

$$D_{Prop} = \frac{A_{total}^{De}}{A_{total}^{St}} \tag{13}$$

In cases where infection groups were considered separately, equation 10 was used rather than A_{total} and the differences were calculated in the same way.

Characteristics of the R gene

The current study analysed two characteristics of the resistance gene. The first is its robustness, meaning the ease with which mutations in the *wt* pathogen can produce a viable *rb* strain. This was parameterised using the assumption that a more robust R gene would result in a lower background frequency of the new *rb* strain on the S crop at the point when it first arose. This parameter (γ) is therefore the initial proportion of I plants that are I_S^{rb}, ie. the initial proportion that are *rb* as there is no initial infection of R cultivar ($I_R^{rb} = 0$). This is a proxy for the emergence frequency of the *rb* pathogen.

The initial proportion of I individuals that are *rb* (γ) was used to calculate the number of plants in the I_S^{rb} and I_S^{wt} classes at $t = 0$. Equations 14 and 15 show how this was calculated. The total proportion of infected plants at $t = 0$, θ, remained constant throughout the analyses, set at the *wt* equilibrium when both cultivar types are present ($\frac{1}{3}$ of all plants infected). Note that $\theta * N$, seen in equations 14 and 15, is equal to the total number of infected plants at $t = 0$. This can be represented as $I_{S(t=0)}$ due to the lack of initial infection of the R cultivar.

$$I_{S(t=0)}^{rb} = \gamma\, \theta\, N \tag{14}$$

$$I_{S(t=0)}^{wt} = \theta\, N\ - I_{S(t=0)}^{rb} \tag{15}$$

The range of values of γ selected (Table 1) corresponds to between 1 and 15 I_S^{rb} plants initially.

The second R gene characteristic tested was the reproductive fitness cost of the *rb* trait (δ), which is dependent on the R gene used to create the R crop. The value ranges from 0, meaning no fitness cost, to 1.

γ and δ were first tested individually to analyse their effect on the magnitude of the EI in deterministic compared to stochastic models. The dynamic between the two parameters was then tested to see their combinatorial effect on the difference between the epidemic sizes of the two model types.

Characteristics of the landscape

The second area of analysis was at the ecological scale. The parameter ϵ is the connectivity of the landscape, also known as the spatial heterogeneity or degree of coupling/mixing of the S and R crops. This parameter, as for γ and δ, was first tested in isolation to quantify its effects on the epidemic intensities of the stochastic and deterministic models. It was then analysed in combination with the characteristics of the R gene tested (γ and δ) to quantify their interactions.

For the interaction between ϵ and δ, the ϵ measure used was the 'benefit of spatial heterogeneity'. This was calculated as the absolute difference between the value of A_{total} at a high level of coupling ($\epsilon = 0.95$) and at a low level of coupling ($\epsilon = 0.05$). A positive value signifies that increased spatial heterogeneity decreases the EI and vice versa for a negative value. In addition, further analysis was done on the stochastic model to try to identify the causes of the trends seen. Firstly, the extinction proportion was calculated: the proportion of runs of the stochastic model in which the rb pathogen goes extinct. Secondly, the local extinctions were calculated: the mean number of extinction and reinfection events per run per group of the model.

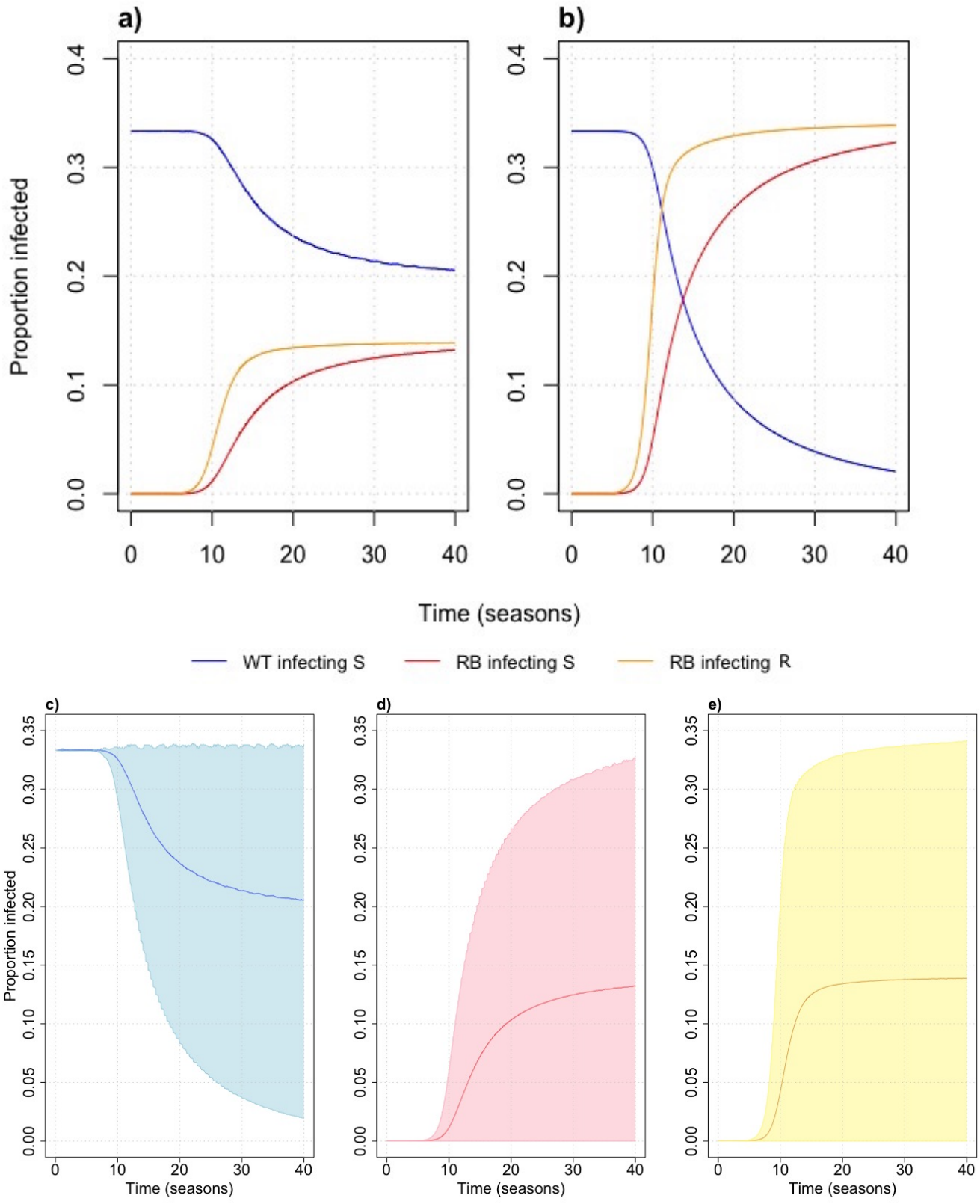

Figure 2: Deterministic and stochastic epidemic trajectories. (a) Mean trajectory of 100 runs of stochastic model over time (b) Deterministic model trajectory over time (c-e) I_S^{wt},I_S^{rb}, I_R^{rb} stochastic trajectory (respectively), including 5th and 95th percentiles of runs. AUDPC is calculated for lines in (a)+(b). All parameters are at reference values.

3 Results

3.1 Broad differences between stochastic and deterministic trajectories

There is a considerable difference between the epidemic trajectories of the stochastic and deterministic models when run at the reference values (Figure 2a,b) (Table 1). The *rb* pathogen doesn't invade in the stochastic model to the same extent it does in the deterministic, although the qualitative shape of the trajectories are largely the same over time. The increase in the proportion infected by the *rb* inoculum over time (ie. I_S^{rb} and I_R^{rb}) is mirrored by a decrease in the *wt* inoculum (I_S^{wt}). This may be due to the fact that N is constant, and the rate of change of I_S^{wt}, as for the other infection groups, is dependent on the availability of host tissue. The amount of available uninfected host tissue in the S crop is affected by I_S^{rb} and I_R^{rb} (Figure 1 – see green dashed lines). Therefore, although stochasticity is likely driving the *rb* strain, the effect is reflected in the *wt* strain.

The reason that the stochasticity is likely affecting the *rb* strain is that in the stochastic epidemic model, unlike in the deterministic, infection classes are affected by a degree of randomness and have the ability to go extinct. Moreover, the stochastic trajectory in Figure 2a is a mean of 100 runs of the model. This is shown by Figure 3, where the proportion of the runs resulting in extinction decreases with increasing γ. The variability that is inherent to stochasticity is clear in Figure 2c,d,e, where the 5th percentile of runs go extinct whereas the 95th percentile shows the same trajectory as the deterministic model.

This has important implications for interpretation of the results throughout the study. Despite what Figure 2a+b might first appear to show, it isn't the case that stochasticity lowers the *rb* equilibrium under these conditions. The fact the stochastic model has a lower proportion infected at t_f represents the probability of extinction at the given parameters. This means that a lower EI in the stochastic model signifies a reduced probability that the *rb* pathogen will successfully colonise

the landscape. If the *rb* pathogen is successful, it would likely show a very similar equilibrium and disease trajectory to the deterministic model, as in Figure 3e where there are no extinctions the stochastic EI is very similar to the deterministic (Figure 3a). Therefore the closer the stochastic result is to the deterministic, the lower the probability of extinction of the *rb* pathogen and hence the lower the likelihood that adding stochasticity is necessary to represent the true dynamics of the system. In addition, a lower *rb* EI means a lower total EI as although the rise in *rb* infecteds is countered by a decrease in *wt* infecteds, it cannot compensate for the loss as there are two *rb* groups and only one *wt* group. Therefore a lower value of A_{total} means a lower probability of *rb* invasion due to higher extinction rates. EI is also a proxy for resistance durability[1], as a smaller epidemic suggests the resistance is more robust.

The mean rather than the median stochastic run is being used due to the fact the latter doesn't convey the extent of extinctions. Whereas a higher number of extinctions lowers the mean incrementally, the median just shows a typical (middle) trajectory, which either colonises or doesn't (Figure 3c vs d). It therefore conveys much less information about the extent of extinctions. As extinction events are the main difference between the deterministic and stochastic models, using the mean is therefore more interesting.

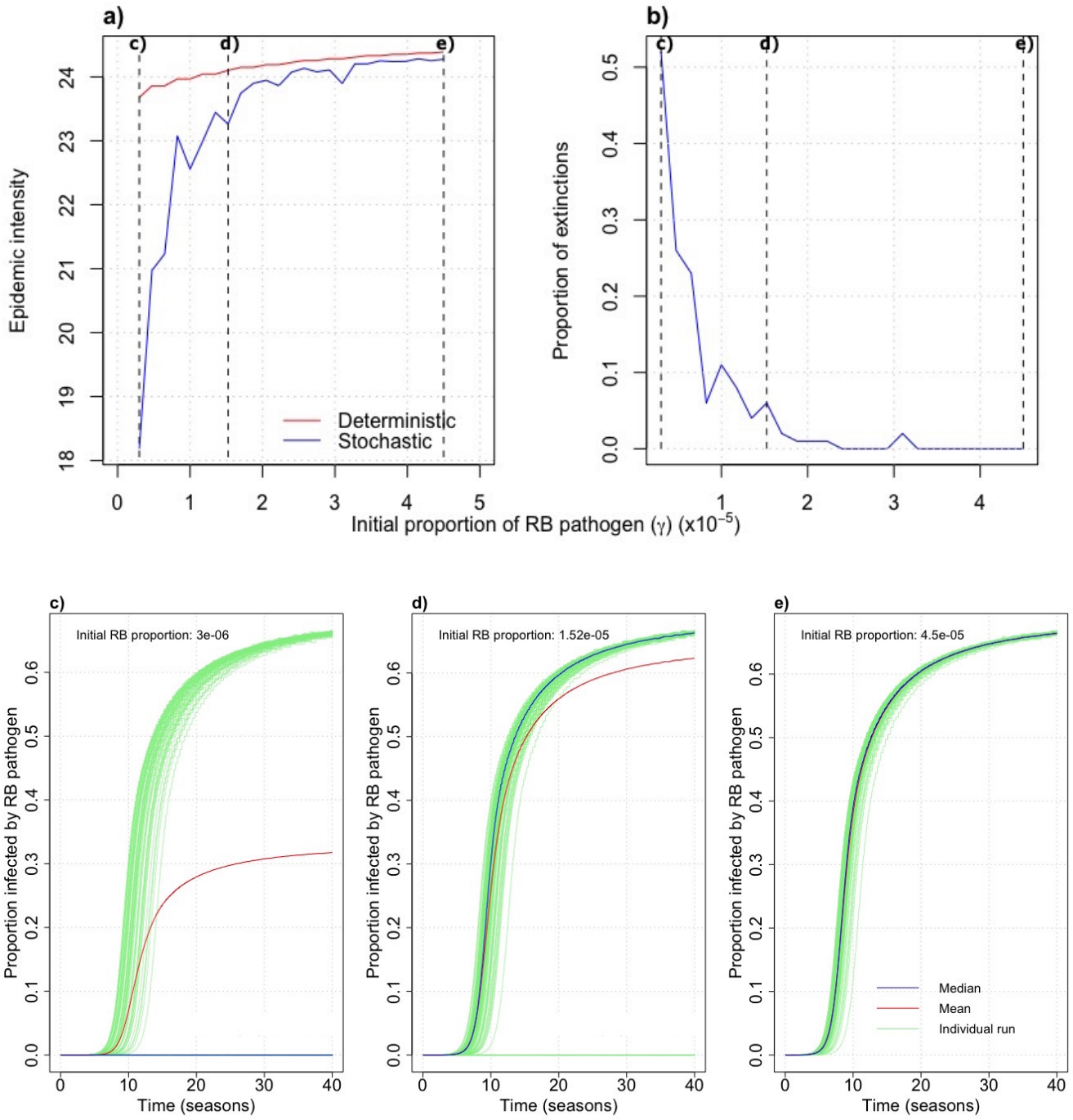

Figure 3: Demonstration of extinction rates in stochastic model. (a) Initial proportion of infected plants infected by rb (γ) against the EI for both stochastic and deterministic models. (b) The proportion of runs of the stochastic model that resulted in extinction of the rb strain across values of γ. (c-e) The proportion of N plants that are infected by the rb strain (ie. I_S^{rb} and I_R^{rb}) across time for three of values of γ. The values of γ correspond to the black vertical dotted lines on (a) and (b). All other parameters are at their reference values.

3.2 Effects of R gene characteristics

Robustness: Initial levels of rb strain (γ)

The stochastic model has a substantially lower EI than the deterministic at low γ (Figure 4a,b), with the deterministic EI nearly 25% higher at $\gamma = 3 \times 10^{-6}$ (Figure 4c). This difference declines as γ increases (Figure 4c,d). This is probably due to the fact that at low initial levels of rb pathogen, extinction is very likely which will lower the EI. As γ increases, the proportion of extinctions occurring decreases and hence the EI rises. The deterministic model is fairly constant across all γ values but shows a slight decline with decreasing γ, likely due to the fact that when there are initially very few rb-infected plants, the epidemic takes longer to establish itself as the initial propagation rate of the rb pathogen is low. which results in slightly lower EI at low γ.

The responses of the individual infection classes (Figure 5) corroborates this result. The rb groups start with much lower epidemic intensities in the stochastic model than the deterministic but tend towards the same values seen in the deterministic, while the absolute differences decrease towards 0 (Figure 5c).

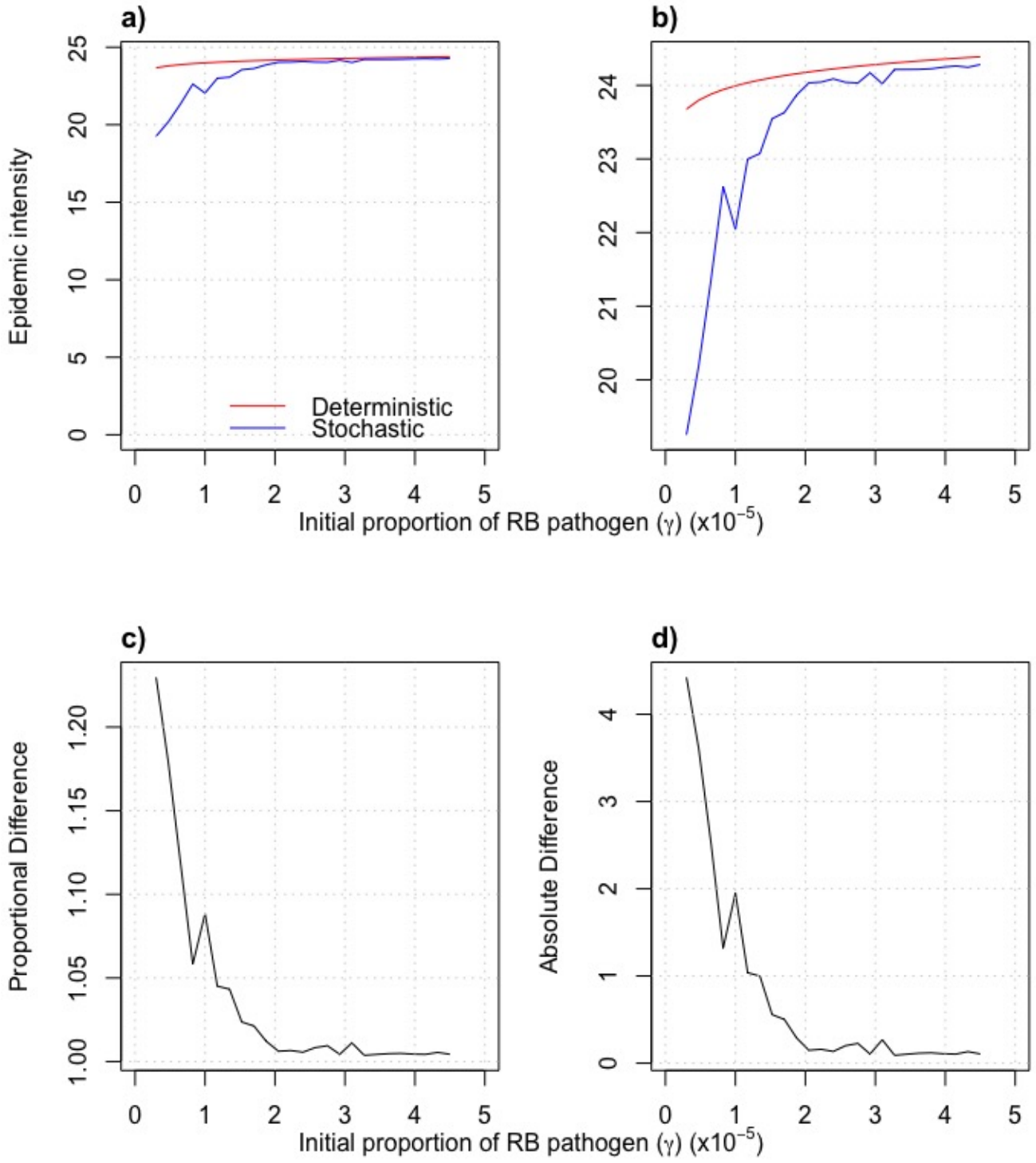

Figure 4: Comparison of EI for deterministic and stochastic models across values of initial *rb* strain proportion (γ). (a) Epidemic intensity (A_{total}) in stochastic and deterministic systems over a range of values of γ. Stochastic line is the mean of 100 runs. (b) Enhanced x-axis of (a). (c-d) Proportional and absolute difference (respectively) between the mean stochastic and deterministic result. All other parameters are at their reference values.

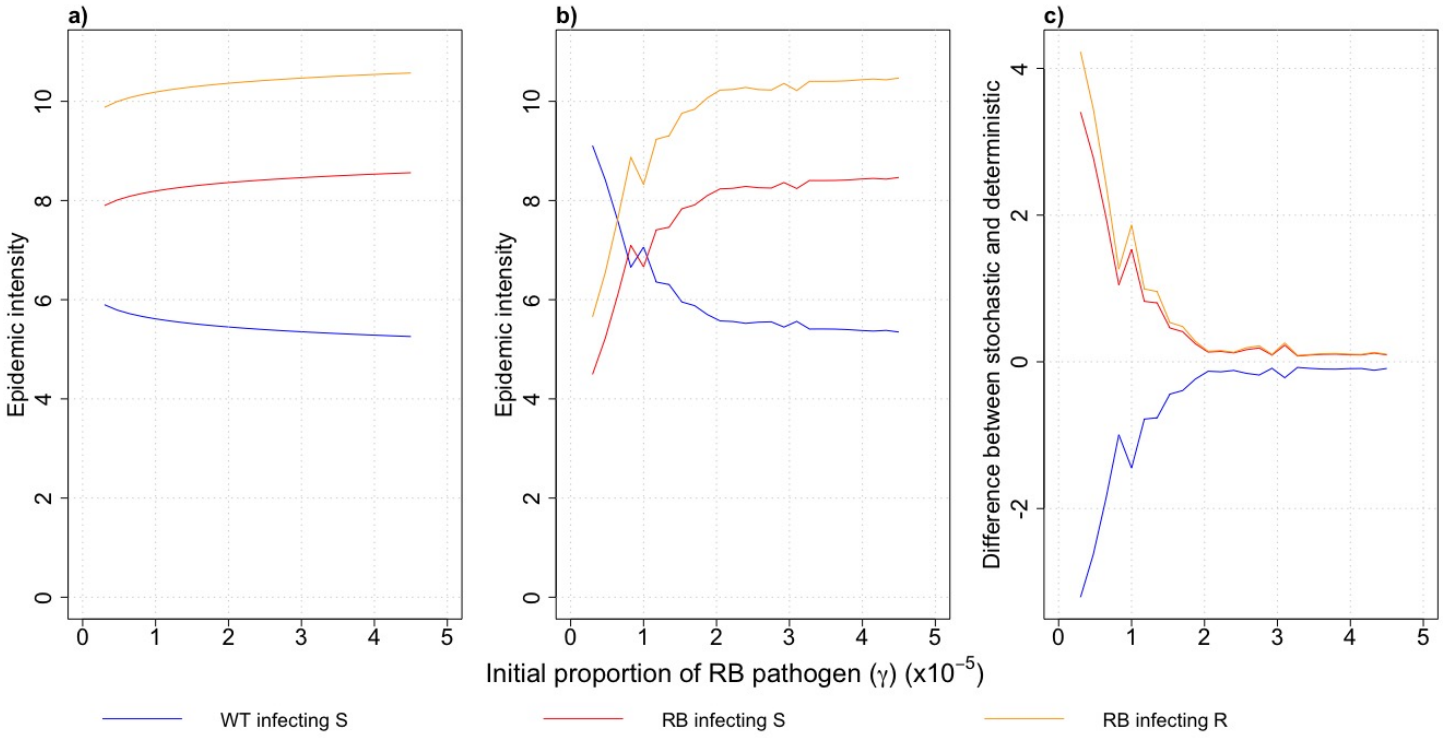

Figure 5: Comparison of EI across range of values of initial *rb* strain proportion (γ) for each group in the stochastic and deterministic models. (a) Epidemic intensity against γ for each infection class in a deterministic system. b) Mean epidemic intensity (of 100 runs) against γ for each infection class in a stochastic system. c) Absolute difference between deterministic and stochastic models per infection class. All other parameters are at their reference values.

Fitness cost of *rb* strain (δ)

In both the deterministic and the stochastic simulations, the EI decreases with increased δ (Figure 6a,b). This is probably because an increased fitness cost of the *rb* trait means that it's harder for the *rb* pathogen to invade as it's a worse competitor than the *wt* strain. This makes successful invasion by the *rb* strain less likely, especially as the initial plants infected with *rb* are in the S patch, in direct competition with the *wt* pathogen.

The proportional difference between the two model types shows an increase at low-to-intermediate δ values (Figure 6c). This is likely because as δ increases, the proportion of extinctions in the stochastic simulation increases. This causes a proportionally faster reduction in the EI than the deterministic, causing the difference between them to increase. However, both the proportional and absolute difference then rapidly drop again as both model types tend towards a *wt*-only epidemic

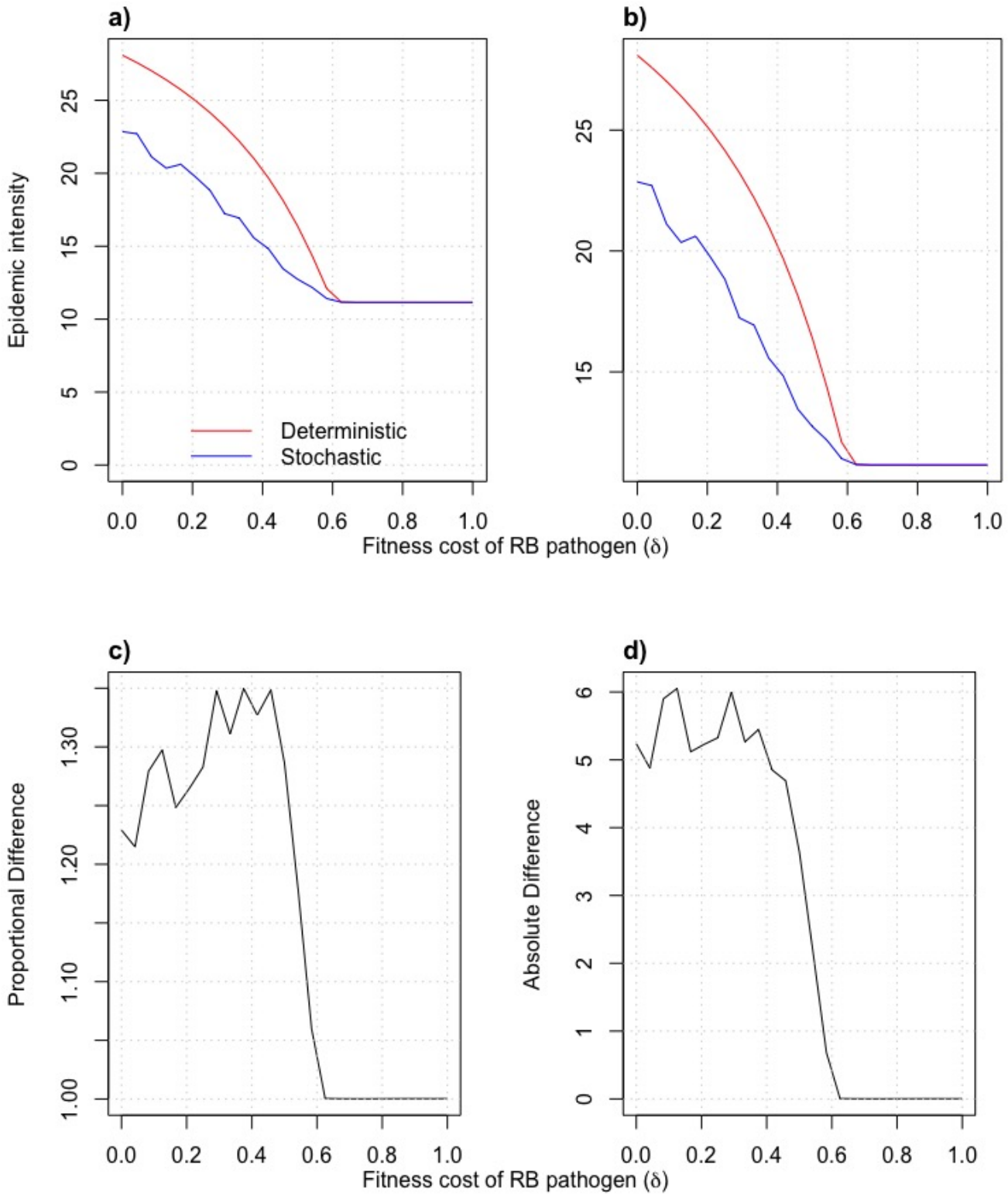

Figure 6: Comparison of EI for deterministic and stochastic models across values of fitness cost of *rb* strain (δ). (a) Epidemic intensity (A_{total}) in stochastic and deterministic systems over a range of values of δ. Stochastic line is the mean of 100 runs. (b) Enhanced x-axis of (a). (c-d) Proportional and absolute difference (respectively) between the mean stochastic and deterministic result. All other parameters are at their reference values.

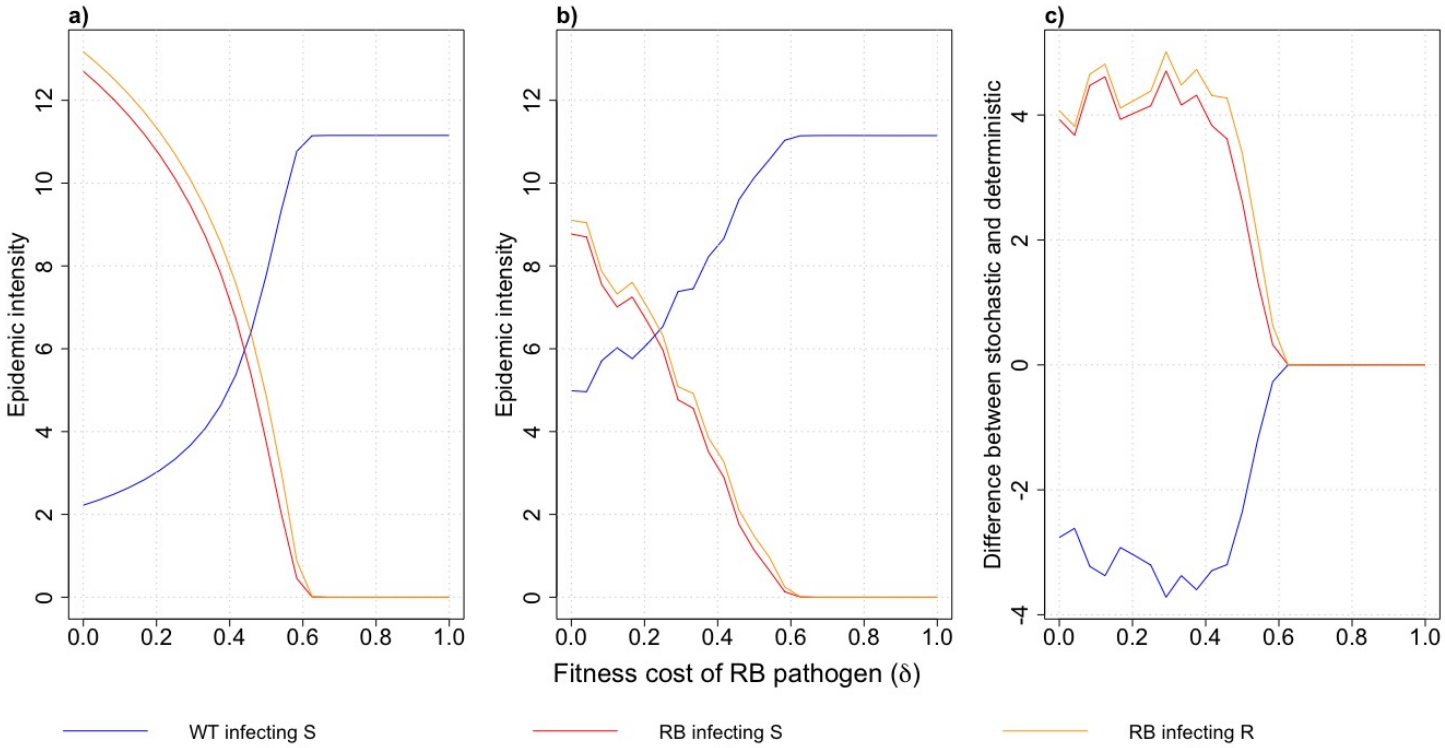

Figure 7: Comparison of EI across range of values of fitness cost of rb strain (δ) for each group in the stochastic and deterministic models. (a) Epidemic intensity against δ for each infection class in a deterministic system. b) Mean epidemic intensity (of 100 runs) against δ for each infection class in a stochastic system. c) Absolute difference between deterministic and stochastic models per infection class. All other parameters are at their reference values.

at wt equilibrium for these particular set of conditions. This is shown by Figure 7 as the EI of I_S^{rb} and I_S^{rb} decline to zero at high δ values, whereas the I_S^{wt} group increases and plateaus.

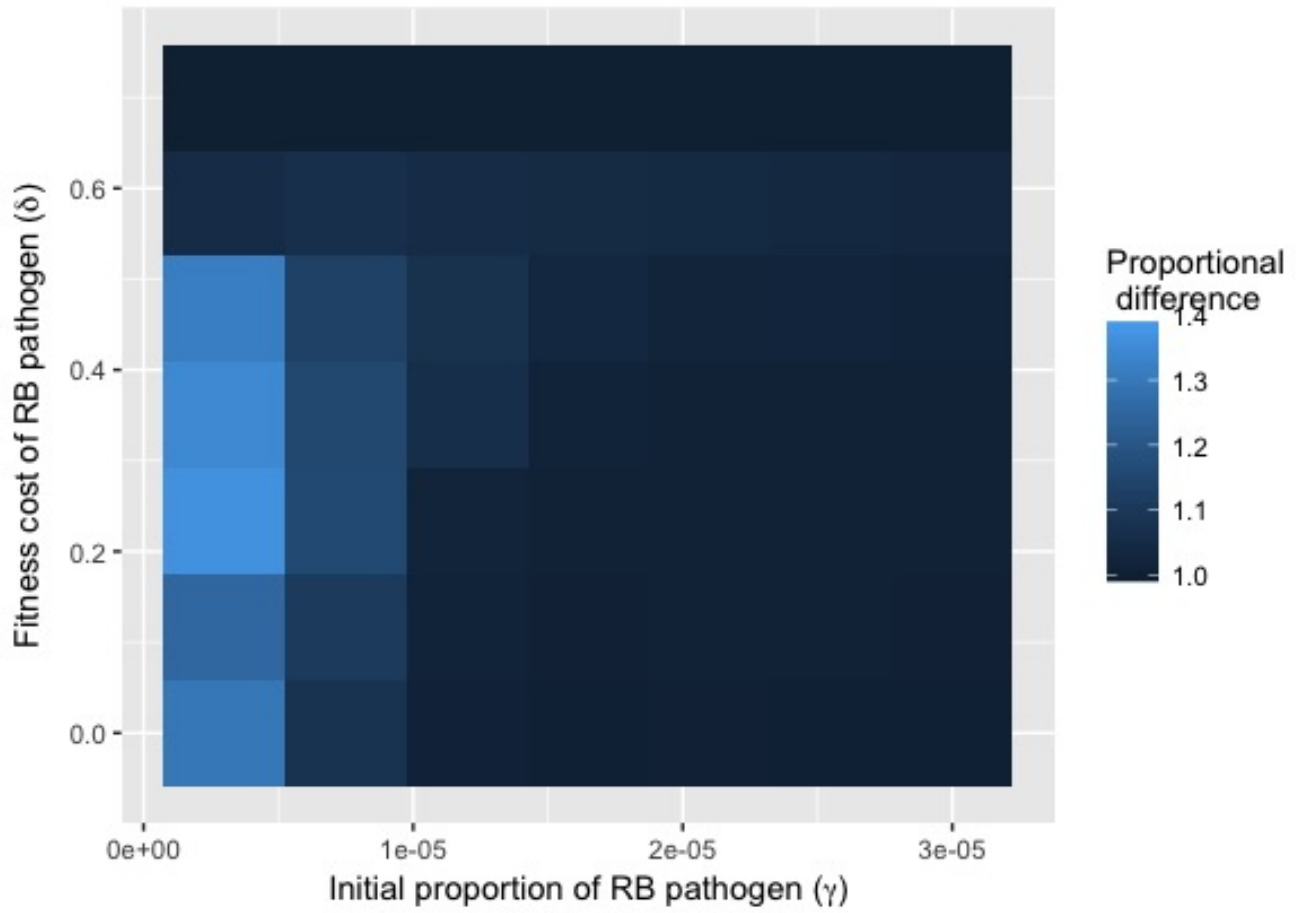

Figure 8: Heat map of proportional difference between stochastic and deterministic EIs (fill colour) across range of values of initial *rb* strain proportion (γ) and fitness cost of *rb* trait (δ). All other parameters are at their reference values.

Relationship between robustness (γ) and fitness cost (δ)

Figure 8 shows that γ has a much more substantial effect on the difference between the stochastic and deterministic models than δ does. Increasing the initial levels of *rb* pathogen (ie. robustness) creates a sharp decline in difference between the two model types (and hence proportion of extinctions in the stochastic model) regardless of the value of the reproductive fitness cost (δ) of the pathogen. However, increasing δ does slightly increase the proportional difference in EI between the stochastic and deterministic simulations at higher γ levels.

The difference is only substantial at low initial levels of the *rb* strain, and when the fitness cost is less than ˜0.6. These conditions are where extinction in the stochastic model is most likely.

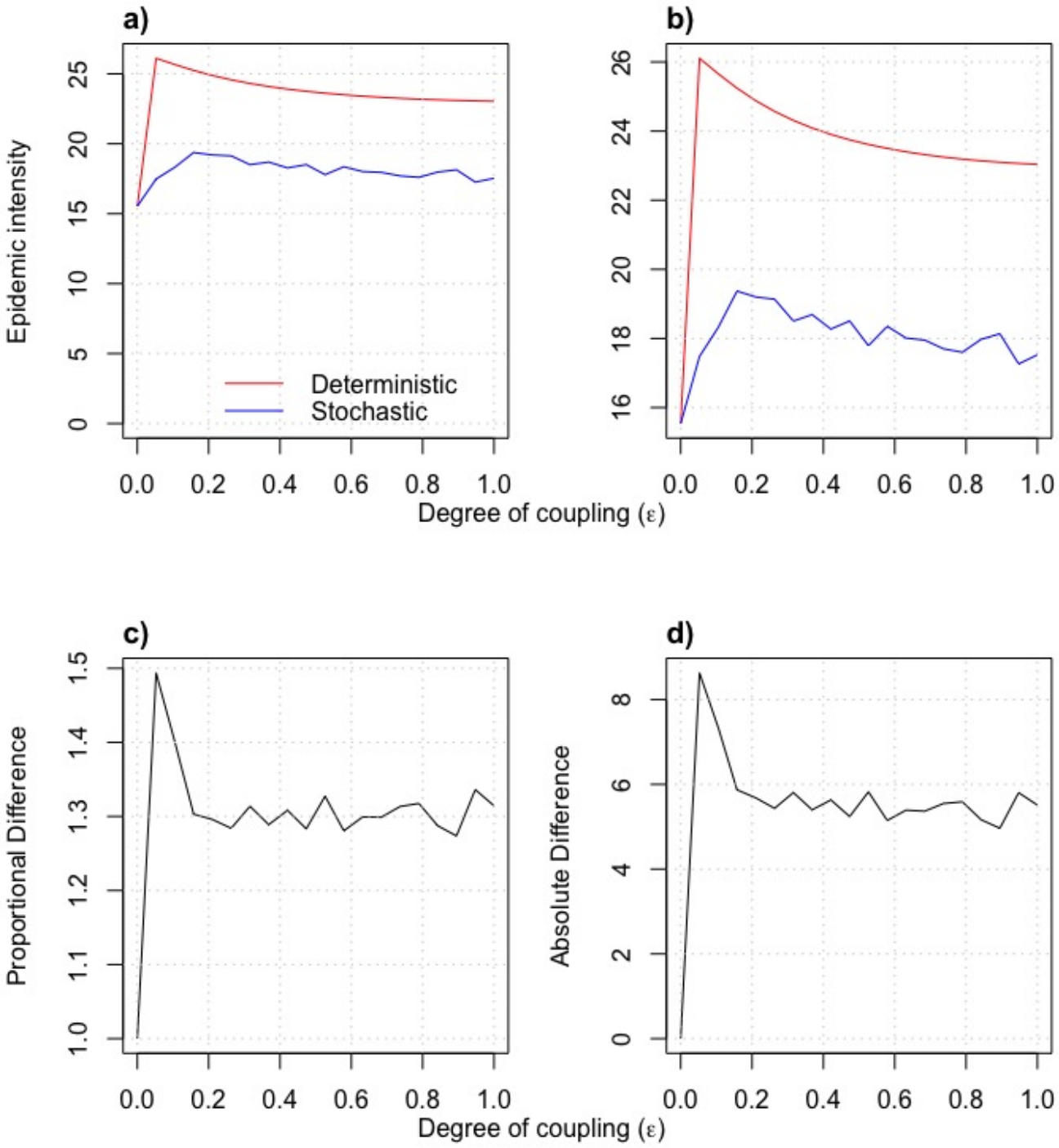

Figure 9: Comparison of EI for deterministic and stochastic models across values of coupling coefficient (ϵ). (a) Epidemic intensity (A_{total}) in stochastic and deterministic systems over a range of values of ϵ. Stochastic line is the mean of 100 runs. (b) Enhanced x-axis of (a). (c-d) Proportional and absolute difference (respectively) between the mean stochastic and deterministic result. All other parameters are at their reference values.

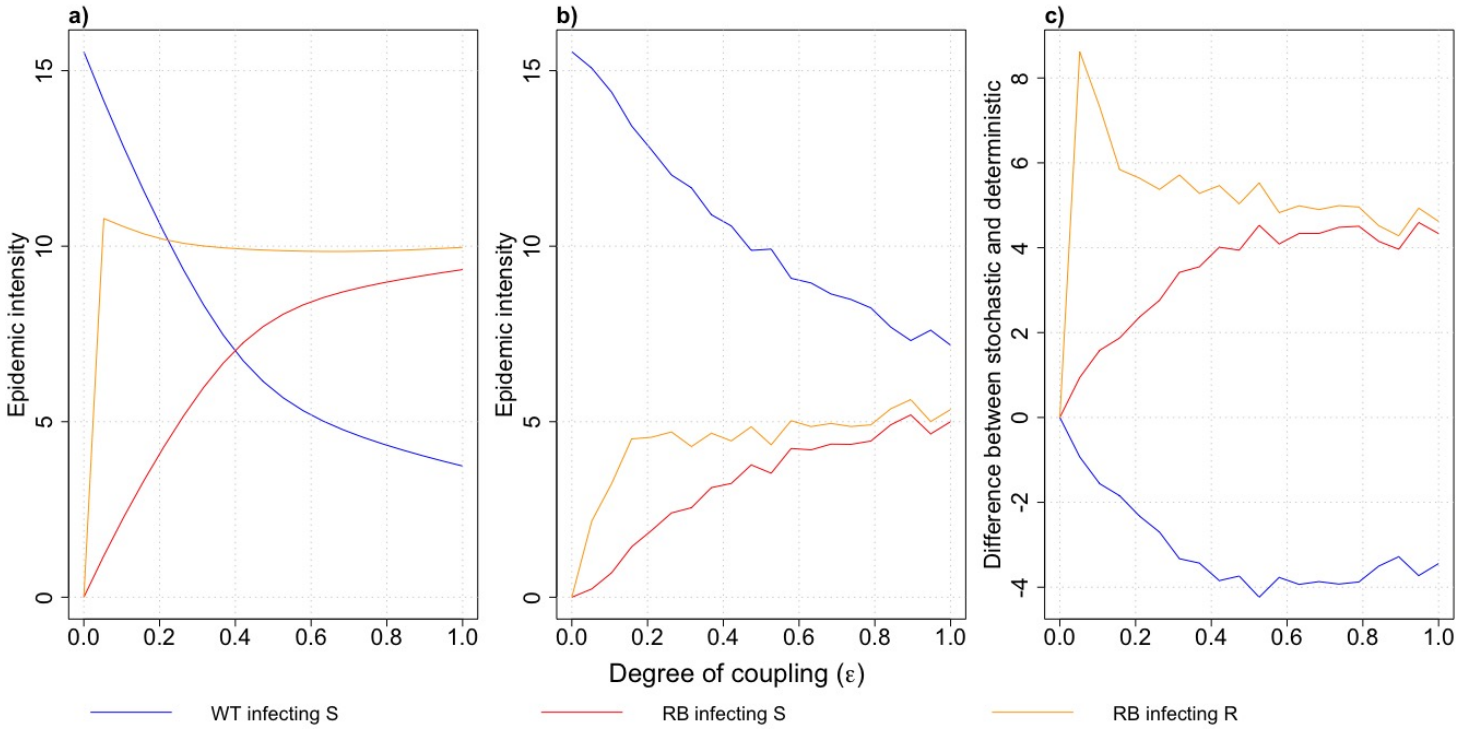

Figure 10: Comparison of EI across range of values of coupling coefficient (ϵ) for each group in the stochastic and deterministic models. (a) Epidemic intensity against ϵ for each infection class in a deterministic system. b) Mean epidemic intensity (of 100 runs) against ϵ for each infection class in a stochastic system. c) Absolute difference between deterministic and stochastic models per infection class. All other parameters are at their reference values.

3.3 Effect of landscape characteristics

Level of landscape connectivity (ϵ)

The stochastic model has a considerably lower EI than the deterministic at all levels of landscape connectivity (except $\epsilon = 0$) when the parameters are at their reference values (Figure 9). At $\epsilon = 0$, the S and R patches are completely separate. There can therefore be no infection of the R patch and the rb strain is out-competed on the S patch, meaning both model types are essentially at wt-only equilibrium. The difference - both proportional and absolute - is highest when coupling is very low (˜0.05), where the deterministic model has a 50% higher EI than the stochastic (Figure 9c,d). With increasing ϵ, the difference between the two model types remains fairly constant, with both models showing declining EI.

The reason for these differences can be seen in Figure 10. The biggest difference between the two models is seen in the I_R^{rb} group. In the deterministic model, the I_R^{rb} EI is consistently high and changes little across the different ϵ values. This is due to the fact that as the wt pathogen cannot infect the R patch, the rb strain has no competition. By contrast, the I_S^{wt} EI declines as the degree of mixing increases, and I_S^{rb} increases. This is because the rb strain is a worse competitor on the S patch due to its reproductive fitness cost (δ), so at low coupling it can't compete with wt strain. However, increasing ϵ causes increasing dilution effects, whereby the force of infection from the wt strain is wasted due to its landing on an R plant that it cannot infect. Therefore as ϵ increases, the wt strain becomes a worse competitor so the rb strain can out-compete it on the S patch.

However, in the stochastic model, the same trend in I_R^{rb} is not seen (Figure 10b). The rb strain is much less successful at all levels of coupling, but the difference is highest at low ϵ. This is likely due to the fact that the initial rb infection is only on the S patch. Therefore at low levels of mixing, the likelihood that the rb pathogen goes extinct before infecting the R patch is increased, both due to the lower chance that an rb pathogen will land on an R crop and due to the wt pathogen being a much better competitor of the S patch.

Relationship between landscape connectivity (ϵ) and robustness of R gene (γ)

At anything other than a very low initial proportion of rb pathogen (high robustness of R gene), there is very little difference between the stochastic and deterministic epidemic intensities at any level of spatial heterogeneity (Figure 11). Therefore, as was the case for δ (Figure 8), γ has a much more substantial effect on the difference between the stochastic and deterministic models than ϵ does. As was seen in Figure 9, the degree of coupling has little effect on the proportional difference between the deterministic and stochastic models.

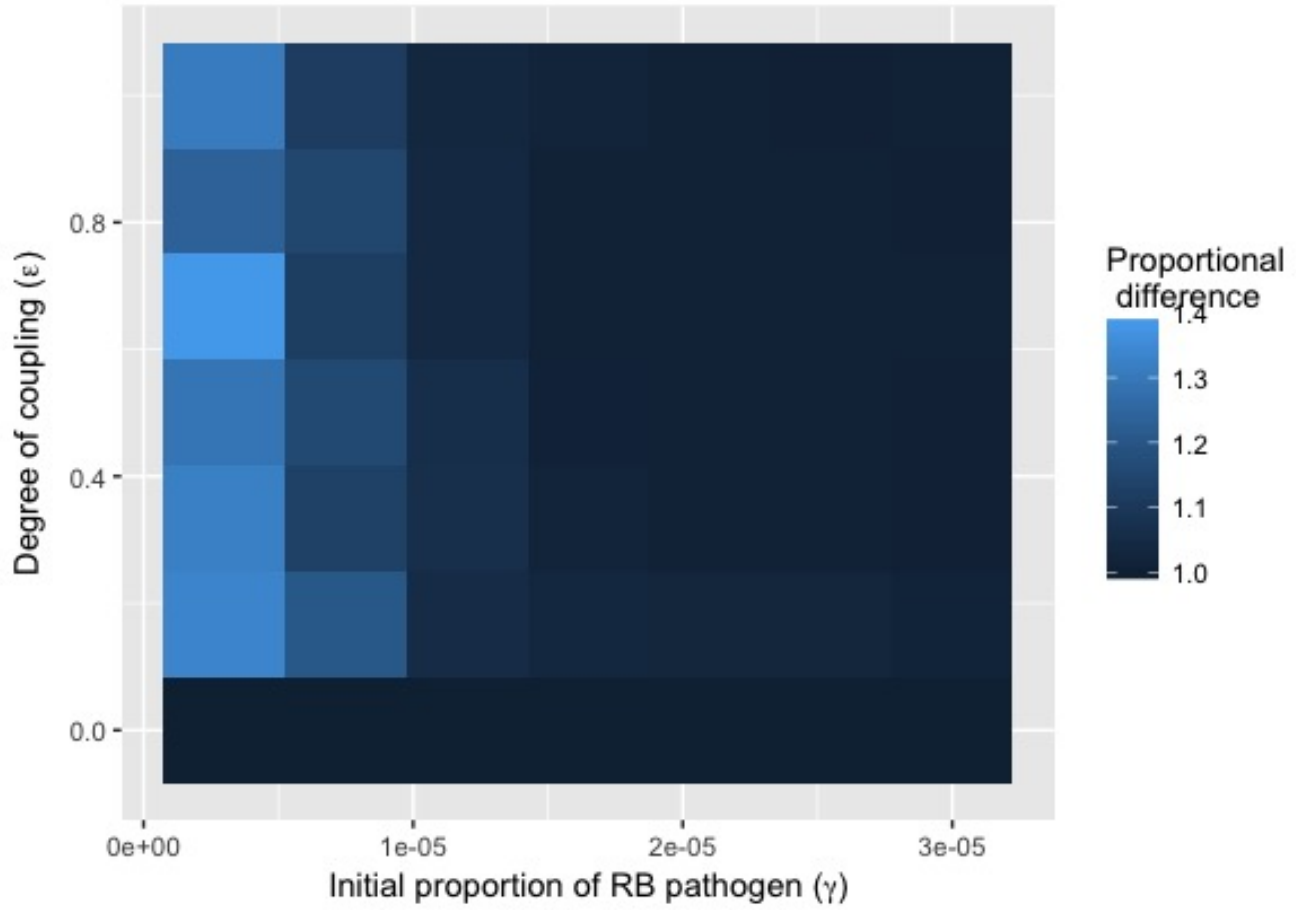

Figure 11: Heat map of proportional difference between stochastic and deterministic EIs (fill colour) across range of values of initial *rb* strain proportion (γ) and coupling coefficient (ϵ). All other parameters are at their reference values. Note that x and y axes don't cover full range of parameter constraints in Table 1 as there was no difference at these values.

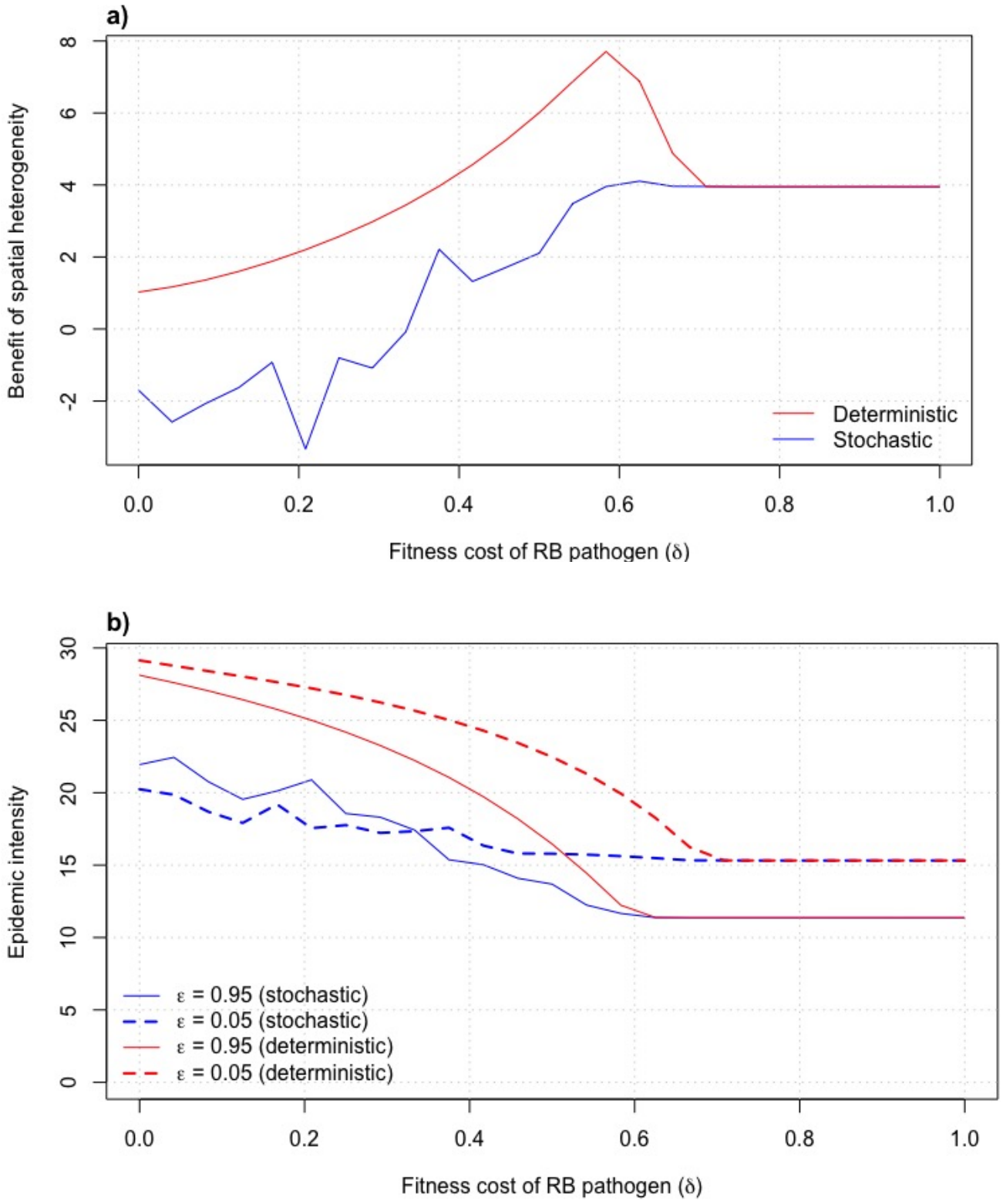

Figure 12: Relationship between the fitness cost of the *rb* pathogen and the degree to which increased coupling reduces the difference between deterministic and stochastic EI. (a) Difference between EI at high coupling and low coupling for deterministic and stochastic models across values of *rb* trait fitness cost (δ). (b) EI of high and low coupling for deterministic and stochastic models across values of δ. All other parameters are at their reference values.

Relationship between landscape connectivity (ϵ) and fitness cost of R gene (δ)

Figure 12a shows the degree to which increased spatial heterogeneity (increased landscape connectivity of R and S crops) decreases the EI at different levels of reproductive fitness cost of rb. There is a clear difference between the deterministic and stochastic models: the benefit of increased coupling of S and R cultivar is lower when stochasticity is considered. Both models show increased benefit of higher coupling at higher levels of δ, although the stochastic model doesn't have a peak like the deterministic does. However, unlike in the deterministic model, in the case of the stochastic model the benefit starts out negative, meaning that at low values of δ increased coupling increases rather than decreases the size of the epidemic (Figure 12b). The stochastic simulations are quite variable, but as the EI at low coupling is consistently lower across $0 \leq \delta \leq 0.35$, this is likely to be a true result and not just an artefact of the inherent randomness of the stochastic model.

In the deterministic simulation, the reason that the EI is lower at higher levels of spatial heterogeneity is because mixing of the two crop patches causes dilution effects, as discussed earlier. However, in the stochastic model where extinctions are possible, this dynamic changes at lower levels of rb fitness cost. Reduced fitness cost means that dilution effects on the rb strain are reduced, and so the rb pathogen becomes a better competitor on the S patch. This means that when the patches are very mixed, the wt pathogen is more likely to go extinct than the rb (Figure 13a), as it still experiences dilution effects because it cannot infect the R cultivar. This may allow a higher total EI as the rb strain has no competition. In addition, high levels of coupling may also reduce the extent of rb strain extinction, in two ways. Firstly, extinction of rb at the start of the epidemic before it can pass from the S to the R crop is less likely. Secondly, if extinction does occur on one patch it's much more likely that it will be reinfected with the rb pathogen from the other patch. These effects are present at all levels of δ, but although it may mean there's a lower proportion of rb extinctions at high ϵ than low ϵ (Figure 13a), at higher δ values there are more extinction and reinfection events at high ϵ, suggesting that the rb pathogen is maintained at very low levels.

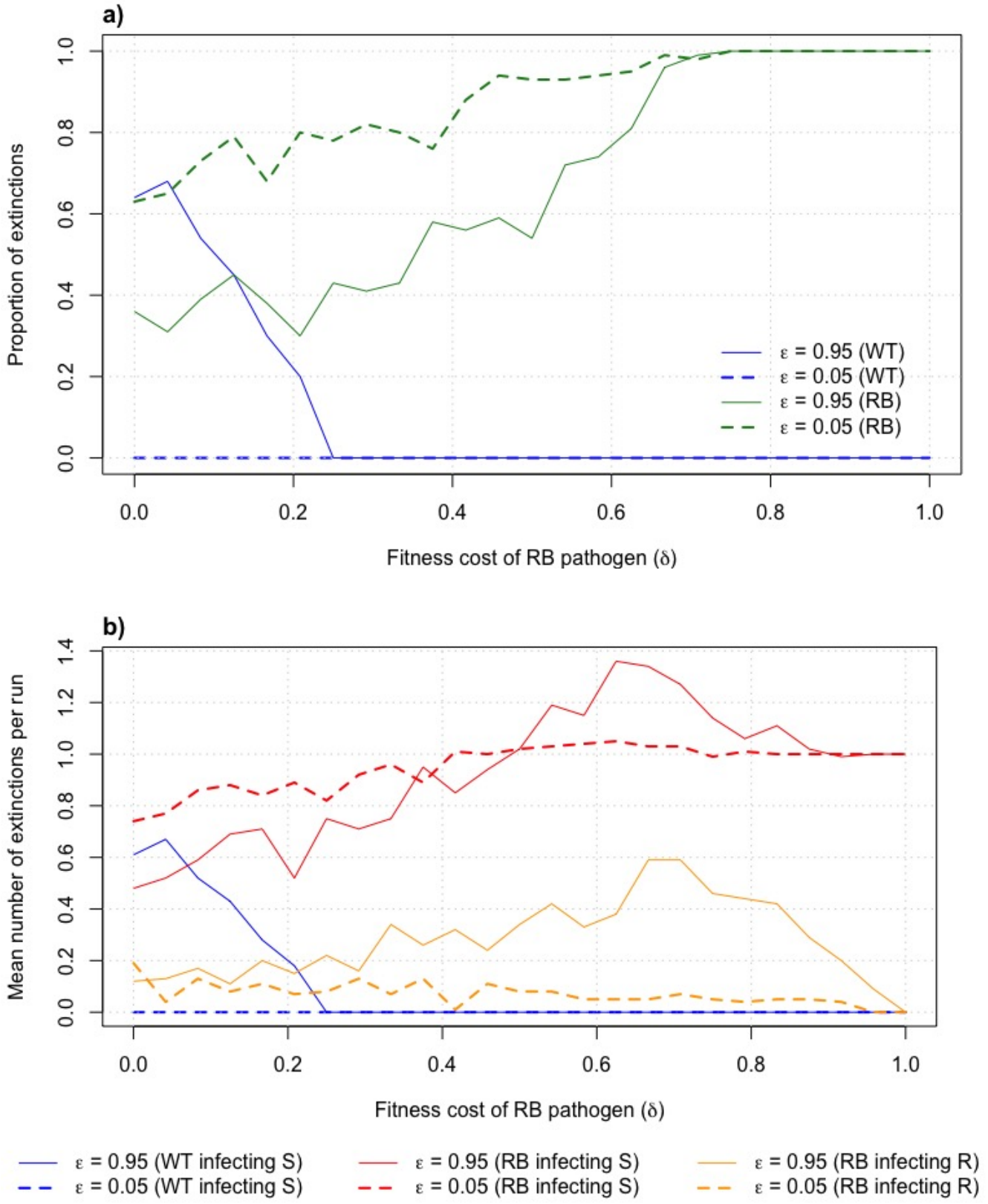

Figure 13: Extinction trends of stochastic model at high and low coupling for *rb* and *wt* strain across range of values of *rb* fitness cost (δ). (a) Proportion of runs resulting in extinction for *rb* and *wt* strains across values of δ at high and low coupling. (a) Local extinction events (ie. mean number of extinction and reinfection events) for each infection class in stochastic model, for high and low coupling and across range of δ values. All other parameters are at their reference values.

4 Discussion

My results show that there are certainly cases where stochasticity reveals dynamics the deterministic model doesn't. This was also found by Iacono et al.[12] who found that adding stochasticity to a model resulted in the resistance durability becoming dependent on the cropping ratio where it previously wasn't[3]. I will discuss under which conditions incorporating stochasticity is most appropriate, ie. where it produced different results to using a deterministic model. This will be discussed in terms of the three parameters analysed: the fitness cost of the *rb* trait, robustness of R gene, and the epidemiological process of the degree of landscape connectivity. In general, stochasticity is required when there's a high risk of extinction of the *rb* (and in some cases the *wt*) pathogen, so scenarios where this is the case will be highlighted. The results of this model consider the invasion of an *rb* pathogen, but could likely be extrapolated to any scenario where there are very low initial frequencies of pathogen, such as immigration events.

4.1 When is stochasticity appropriate?

Robustness of R gene (γ)

The analysis has shown that the robustness of the R gene is the most important factor to consider when deciding whether to incorporate stochasticity into the model. My findings showed that both the stochastic and deterministic models saw an increase in resistance durability (ie. a decrease in epidemic intensity) with increasing robustness of the R gene. This agrees with the literature[3], as studies have shown an observed increase in resistance durability with the number of mutations needed for resistance breakdown, which can be used as a proxy for the robustness of the R gene[1,9,18]. However, the increase in durability was only slight in the deterministic model. It was far more pronounced in the stochastic model, meaning that the biggest difference between the two models was seen when the R gene robustness was highest. Therefore this is when use of a stochastic model would be most beneficial to more accurately represent the true dynamics of the system. In practice, due to the falling costs of high-throughput sequencing techniques[19], it is now possible to predict what the background frequency of a pathogen will be and whether it's at

equilibrium[20,21]. This could also be used to predict how long the *rb* pathogen population takes to reach this equilibrium[1], as it's assumed to be instantaneous in my model.

The robustness of the R gene appears to be the most significant factor of the ones analysed, as when compared with both the fitness cost of the *rb* trait and the degree of landscape connectivity, it had far more effect than either of them on the extent of the difference in resistance durability between the stochastic and deterministic models. Considering that the determining factor differentiating the two model types is the possibility of extinction in the stochastic simulations, this result is expected to some extent as smaller initial frequencies of pathogens (ie. a higher R gene robustness) makes the chance of random extinction much more likely. In addition, as the reference value for the initial *rb* pathogen population was the equivalent of only 1 individual being initially infected with the *rb* strain ($\gamma = 3 \times 10^{-6}$), this likely amplified the effects of other results in the study. However, there is evidence that such a low baseline frequency of *rb* pathogen is realistic if the *rb* trait required two mutations on the R gene. Stam and McDonald[22] calculated that if the R gene only required a single mutation to overcome then 10×10^9 *rb* pathogens could be produced per hectare per day. However, if two mutations were required this number drops drastically to only 10 *rb* pathogens per hectare per day. Even if such a high level of R gene robustness is not always realistic, it does represent an ideal R gene as it produces a high level of resistance durability.

A high robustness means that producing a mutation(s) that creates a viable resistance-breaking trait is unlikely to occur, so may require a very specific mutation or even multiple mutations in order to do so, as Stam and McDonald[22] suggest. Although there are studies that assume that qualitative resistance (ie. resistance conveyed by a single R gene with distinct resistance phenotypes, as seen in this study) doesn't create very robust or durable resistance in the same way quantitative resistance (resistance conveyed across multiple quantitative trait loci) can[23–25], the evidence for this is contentious[26]. There is evidence that shows that qualitative R genes can convey durable resistance in the field[27–29], such as the recessive gene mlo for powdery mildew resistance in barley, Rpg1 for resistance to stem rust in barley, and the dominant I gene for resistance to bean

common mosaic virus in common bean[23,24,30–32]. This suggests that a scenario of a robust qualitative resistance is viable, and in these scenarios stochastic modelling would be most appropriate.

Alternatively, as the 'robustness' parameter (γ) is the initial proportion of the rb strain, it could also translate to a scenario where the rb pathogen has just migrated from elsewhere. As stochasticity is most appropriate when the initial rb proportion is low, in this context this would mean that the migration was likely a long distance dispersal (LDD) event, where very few spores travel huge distances. For example, Stem rust outbreaks on wheat in the UK are caused by airborne spores originating in southwest Europe and northwest Africa[14]. However, immigration of the rb pathogen is less biologically accurate for my system than emergence is, as if the rb pathogen arrived via an LDD event, it would not necessarily exclusively be on the S cultivar at $t = 0$.

Reproductive fitness cost of rb trait (δ)

The results showed that the reproductive fitness cost of the rb trait did have an effect on the difference in resistance durability between the two model types to an extent, as from low to medium fitness costs the difference between the stochastic and deterministic models increased due to an increasing number of extinction events. The epidemic intensity in both models decreased with increasing fitness costs, which is supported in the literature[18,20,33,34], but the stochastic declined at a greater rate, due to the occurrence of extinction events. However, as this was seen when γ was equivalent to 1 individual infected with rb at $t = 0$, this effect is likely to decline with increasing initial rb levels.

Landscape connectivity (ϵ)

On its own, ϵ doesn't appear to have much effect on the difference in resistance durability between the deterministic and stochastic models. Although there was a large difference between the two model types, it was fairly constant across nearly all values of ϵ, suggesting that this could just be a symptom of the very low γ level (ie. high robustness of R gene) used as a reference value that was the true cause for the extinctions.

However, the addition of stochasticity did reveal dynamics that the deterministic model did not in the case of the relationship between ϵ and δ. From the stochastic model, it can be seen that there is a trade-off between the beneficial longer-term effects of higher spatial heterogeneity and the detrimental shorter-term effects. In the long-term, when the disease is established, a higher degree of mixing is beneficial as the dilution effects it creates[15] increases extinctions of the *rb* pathogen and so increases resistance durability by maintaining the epidemic at lower levels. However, at emergence of the *rb* pathogen when it's at very low frequencies, a higher spatial heterogeneity increases the likelihood that the *rb* strain will pass from the S to the R patch, where it doesn't have to compete with the *wt* strain. Colonisation of the R patch appears to be essential to a successful invasion. These effects are amplified at low fitness costs as it proportionally benefits the pathogen in a high coupling environment more than a low coupling one. This is because it relaxes the dilution effects felt by the *rb* pathogen at high coupling but doesn't benefit the *wt* pathogen in the same way, as it still cannot infect the R patch due to the resistance being complete. This trade-off effect doesn't seem to be reflected in the literature; both theoretical and experimental studies show that the mixing of cultivar types can reduce the rate of disease spread[4,15,35]. This may be due to the fact that this effect only occurs at lower levels of reproductive fitness costs, and that results of different systems are often highly varied, partly due to the specific conditions of the experiment/model, and partly due to the inherent variability exemplified by the wide range of plant pathogens and hosts[12]. In addition, as the resistance is complete, this may convey a more exaggerated result, ie. *wt* extinctions can occur which would eliminate competition and allow a much higher *rb* strain epidemic.

In addition, the stochastic model captured another interesting dynamic that the deterministic couldn't. The results showed that high levels of spatial heterogeneity also increased the degree to which extinction of the *rb* strain in one patch could be 'saved' by re-infection from the other. This effect has been seen in animal populations, and is known as the 'rescue effect'[36,37].

4.2 Future directions

A limitation of my model is that although one of the main areas of interest was the initial levels of the *rb* pathogen, I assumed that the *rb* trait was present from the start and didn't consider its emergence. Future research could incorporate emergence frequencies of the pathogen to make the system even more stochastic. In addition, I only considered demographic stochasticity, but environmental stochasticity, defined as variability arising from randomness of environmental factors[13] would also be an interesting area in which to compare stochastic and deterministic models. Finally, I assumed complete resistance of the R patch, which was useful in order to determine key drivers of differences resulting from stochasticity, but was also an assumption that could be relaxed in future studies.

4.3 Conclusions

In conclusion, the current study has shown that there are scenarios where using a deterministic model to track the effects of genetic and epidemiological processes on the durability of disease resistance is an oversimplification that doesn't capture the true dynamics of the system. This highlights the importance of understanding the dynamics of the system to be modelled in order to produce the most accurate representation of reality.

5 Acknowledgements

I would like to thank Dr Nik Cunniffe for his support and guidance throughout the project, and Ben Watkinson-Powell for helpful discussions and suggesting directions for the practical aspects.

References

1. Fabre, F., Rousseau, E., Mailleret, L. & Moury, B. Durable strategies to deploy plant resistance in agricultural landscapes. New Phytologist 193, 1064–1075 (2012).

2. Johnson, R. The Concept of Durable Resistance. Phytopathology 69, 198 (1979).

3. van den Bosch, F. & Gilligan, C. A. Measures of Durability of Resistance. Phytopathology 93, 616–625 (2003).

4. Burdon, J. J., Zhan, J., Barrett, L. G., Papaix, J. & Thrall, P. H. Addressing the Challenges of Pathogen Evolution on the World's Arable Crops. Phytopathology 106, 1117–1127 (2016).

5. Flor, H. H. Current Status of the Gene-For-Gene Concept. Annual Review of Phytopathology 9, 275–296 (1971).

6. Cui, H., Tsuda, K. & Parker, J. E. Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. Annual Review of Plant Biology 66, 487–511 (2015).

7. Watkinson-Powell, B., Gilligan, C. A. & Cunniffe, N. J. When does spatial diversification usefully maximise the durability of crop disease resistance? bioRxiv 540013 (2019). doi:10.1101/540013

8. Bousset, L., Sprague, S. J., Thrall, P. H. & Barrett, L. G. Spatio-temporal connectivity and host resistance influence evolutionary and epidemiological dynamics of the canola pathogen Leptosphaeria maculans. Evolutionary Applications 11, 1354–1370 (2018).

9. Harrison, B. D. Virus variation in relation to resistance-breaking in plants. Euphytica 124, 181–192 (2002).

10. Brown, J. K. M. Recombination and selection in populations of plant pathogens. Plant Pathology 44, 279–293 (1995).

11. Lannou, C. & Mundt, C. C. Evolution of a pathogen population in host mixtures: simple

race–complex race competition. Plant Pathology 45, 440–453 (1996).

12. Iacono, G. L., Bosch, F. van den & Gilligan, C. A. Durable Resistance to Crop Pathogens: An Epidemiological Framework to Predict Risk under Uncertainty. PLOS Computational Biology 9, e1002870 (2013).

13. van den Bosch, F. & Gilligan, C. A. Models of Fungicide Resistance Dynamics. Annual Review of Phytopathology 46, 123–147 (2008).

14. Encyclopaedia of cereal diseases — AHDB. Available at: https://ahdb.org.uk/knowledge-library/encyclopaedia-of-cereal-diseases. (Accessed: 22nd March 2019)

15. Mundt, C. C., Cowger, C. & Garrett, K. A. Relevance of integrated disease management to resistance durability. Euphytica 124, 245–252 (2002).

16. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361 (1977).

17. Keeling, M. J. & Rohani, P. Modeling Infectious Diseases in Humans and Animals. (Princeton University Press, 2011).

18. Lecoq, H., Moury, B., Desbiez, C., Palloix, A. & Pitrat, M. Durable virus resistance in plants through conventional approaches: a challenge. Virus Research 100, 31–39 (2004).

19. Brockhurst, M. A., Colegrave, N. & Rozen, D. E. Next-generation sequencing as a tool to study microbial evolution. Molecular Ecology 20, 972–980 (2011).

20. Janzac, B., Montarry, J., Palloix, A., Navaud, O. & Moury, B. A Point Mutation in the

Polymerase of Potato virus Y Confers Virulence Toward the Pvr4 Resistance of Pepper and a High Competitiveness Cost in Susceptible Cultivar. MPMI 23, 823–830 (2010).

21. Torres-Barceló, C., Daròs, J.-A. & Elena, S. F. Compensatory Molecular Evolution of HC-Pro, an RNA-Silencing Suppressor from a Plant RNA Virus. Mol Biol Evol 27, 543–551 (2010).

22. Stam, R. & McDonald, B. A. When resistance gene pyramids are not durable—the role of pathogen diversity. Molecular Plant Pathology 19, 521–524 (2018).

23. Kelly, J. D. & Vallejo, V. QTL Analysis of Multigenic Disease Resistance in Plant Breeding. in Multigenic and Induced Systemic Resistance in Plants (eds. Tuzun, S. & Bent, E.) 21–48 (Springer US, 2006). doi:10.1007/0-387-23266-4_3

24. Lindhout, P. The perspectives of polygenic resistance in breeding for durable disease resistance. Euphytica 124, 217–226 (2002).

25. Parlevliet, J. E. Durability of resistance against fungal, bacterial and viral pathogens; present situation. Euphytica 124, 147–156 (2002).

26. St.Clair, D. A. Quantitative Disease Resistance and Quantitative Resistance Loci in Breeding. Annual Review of Phytopathology 48, 247–268 (2010).

27. Johnson, R. A Critical Analysis of Durable Resistance. Annu. Rev. Phytopathol. 22, 309–330 (1984).

28. Johnson, R. Durable resistance: definition of, genetic control, and attainment in plant breeding. Phytopathology 71, 567–568 (1981).

29. Stuthman, D. D., Leonard, K. J. & Miller-Garvin, J. Breeding Crops for Durable Resistance to Disease. in Advances in Agronomy 95, 319–367 (Academic Press, 2007).

30. Ayliffe, M., Singh, R. & Lagudah, E. Durable resistance to wheat stem rust needed. Current Opinion in Plant Biology 11, 187–192 (2008).

31. Johnson, R. CLASSICAL PLANT BREEDING FOR DURABLE RESISTANCE TO DISEASES. Journal of Plant Pathology 82, 3–7 (2000).

32. Keller, B., Bieri, S., Bossolini, E. & Yahiaoui, N. Cloning Genes and QTLs for Disease Resistance in Cereals. in Genomics-Assisted Crop Improvement: Vol 2: Genomics Applications in Crops (eds. Varshney, R. K. & Tuberosa, R.) 103–127 (Springer Netherlands, 2007). doi:10.1007/978-1-4020-6297-1_5

33. Fraile, A., Pagan, I., Anastasio, G., Saez, E. & Garcia-Arenal, F. Rapid Genetic Diversification and High Fitness Penalties Associated with Pathogenicity Evolution in a Plant Virus. Mol Biol Evol 28, 1425–1437 (2011).

34. Janzac, B., Fabre, F., Palloix, A. & Moury, B. Constraints on evolution of virus avirulence factors predict the durability of corresponding plant resistances. Molecular Plant Pathology 10, 599–610 (2009).

35. Zhan, J. & McDonald, B. A. Field-based experimental evolution of three cereal pathogens using a mark–release–recapture strategy. Plant Pathology 62, 106–114 (2013).

36. Blasius, B., Huppert, A. & Stone, L. Complex dynamics and phase synchronization in spatially extended ecological systems. Nature 399, 354–359 (1999).

37. Earn, D. J. D., Levin, S. A. & Rohani, P. Coherence and Conservation. Science 290, 1360–1364 (2000).