Nonlinear Analysis: Real World Applications 34 (2017) 275–315

Contents lists available at ScienceDirect

Nonlinear Analysis: Real World Applications

www.elsevier.com/locate/nonrwa

An age-structured multi-strain epidemic model for antigenically diverse infectious diseases: A multi-locus framework

Alhaji Cherif[a,b,*], Janet Dyson[c], Philip K. Maini[b], Sunetra Gupta[d]

[a] *School of Mathematical and Statistical Sciences, Arizona State University, Tempe AZ 85281, USA*

[b] *Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom*

[c] *Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom*

[d] *Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom*

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form 20 September 2016
Accepted 21 September 2016

Keywords:
Age-structured epidemic model
Stability analysis
Antigenically diverse diseases
Strain community structures
Principle of competitive exclusion

ABSTRACT

The establishment of cross-protective responses and development of immunity within a host exert pressure on pathogens through cross-immunity mediated competition between antigenic forms. In this paper, we incorporate age-specificity in the multi-locus epidemic model used to study the pathogen-specific dynamic behaviours for infectious diseases with diverse co-circulating antigenic types. We establish results on the existence of a unique mild solution, and on the necessary conditions for local stability of the steady-states. In particular, we find that, when the reproductive number $\mathcal{R}_0 < 1$, all strains go to extinction. When $\mathcal{R}_0 > 1$, we show that there exist additional conditions which determine the stability of different types of endemic equilibria, namely weak and strong endemicity, where the weak endemic equilibria correspond to the existence of principle of competitive exclusions of pathogen-specific clusters, while strong endemicity represents the co-existence of all strains. Using numerical simulations, we also show that weak endemic equilibria yield dynamic features in which only one of the clusters containing discrete strain structures (e.g., of minimally, or non-overlapping antigenic types) persists while others go to extinction. For unique strong endemicity, we observe no strain structure, where antigenic types co-exist or exhibit cyclical strain structure with diverse dynamical behaviours (e.g., (quasi-)periodicity, intermittency, chaos). This demonstrates that pathogenic-specific dynamic features are ubiquitous and shows how cross-immunity between antigenic variants shape the maintenance and evolution of strain structures.

© 2016 Elsevier Ltd. All rights reserved.

* Corresponding author at: Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom.
E-mail address: cherif@asu.edu (A. Cherif).

http://dx.doi.org/10.1016/j.nonrwa.2016.09.006
1468-1218/© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In the past decades, we have seen emergence and re-emergence of infectious diseases from a range of extraordinarily diverse and rapidly evolving pathogens. These infectious diseases, which include influenza [1–6], malaria [7,8], dengue virus [9,10] and meningitis [11,12], are responsible for substantial annual morbidity and mortality, and are of utmost importance to public health interests. While mathematical models have successfully been useful in gaining insights into the dynamic features of most diseases, the hierarchical complexity with diverse evolutionary patterns generated by some of these diseases has posed a major challenge in examining the mechanisms governing their dynamics. One of the challenges is that many of these infectious diseases present themselves as multi-strain pathogens with diverse antigenic variants. Antigenically variable pathogens use antigenic variation as an immuno-evasion strategy, whereby the pathogens continuously alter their surface glycoproteins, as in the case of influenza, to prevent antibodies generated by a specific past strain from recognizing and destroying them. As a result, prior infection or exposure history with one strain of a pathogen may not always protect against another. If several infections are possible during an individual's lifetime, then a host's immunological and exposure history can constrain subsequent exposures, thereby substantially changing the dynamics of a disease. That is, during an outbreak, infected individuals may acquire immunity to the same circulating strain thereby restructuring the epidemiological landscape of the population which, in turn, constrains future epidemic outbreaks and drives the pathogen's antigenic evolution allowing spread to otherwise previously immunized individuals. Furthermore, a large number of immune hosts in the population can increase selective pressure exerted on the pathogens and influence the emergence of new antigenic variants. It has been observed that multiple exposures to antigenic variants of a pathogen and accumulation of immunity induce heterogeneity in both immune responses and population structures. In [13,14], the authors observe that both strain diversity and community structures are increased in the presence of contact structure. In [15, Ch. 7–8] and [16], it is shown that age-specificity can influence the interaction between epidemic dynamics and population-level immunological profile. Further, age-specificity has also been observed in immune response. In particular, Cobey and Pascual [17] showed that age may affect the specificity of the immune response, where children produce a monoclonal response while older individuals produce polyclonal responses to influenza infection. More recently, Hancock et al. [18] observe age-specificity in the strength of micro-neutralization of H1N1/2009 among subjects categorized by decade of birth. As a result, it is important to study age-specificity in order to determine the mechanisms responsible for the dynamics observed in these diseases, and the extent that these dynamic patterns reflect interaction of antigenic variation, strain community structure, population demographic patterns and immunological profiles.

Structured population models (e.g., spatial structure, temporal variation, multiple groups, differences in activity or risk factors) have been widely studied in the mathematical biology literature for many years. In particular, models written as evolutionary systems of partial differential equations (PDEs) for the density of individuals incorporating a specific structural variable, for example age and/or size for the age-structured and/or size-structured equations [19,20], phenotypic trait for the selection–mutation equations [21]; physiologically structured equations [22–24], have appeared. These structural variables are usually included in these models to enhance features not captured by standard unstructured population models (see [21,25]). In particular, structured models have been used to capture differential burdens due to heterogeneity across ages, classes, sizes, physiological traits, and so on, in the presence and/or absence of other prior predisposing risk factors. Furthermore, structural variables such as age and/or size have been incorporated into models in order to induce oscillatory dynamics in the cases where unstructured models do not exhibit such behaviours [26–29]. Therefore, studying epidemiological processes within a structural variable framework is critical to understanding factors that are important in disease transmission and evolution; and to designing effective organized public health measures. To understand the effect of

age-specificity on the interaction between epidemic dynamics, cross-immunity, population-level immunological mechanism and evolution and maintenance of strain structures, we consider a multi-locus framework and incorporate age-specificity into the model that has been used in a series of paper [12,30,6, 13,14,31,5,32–35,2,36–38] to investigate various infectious diseases with diverse antigenic forms.

In [30], Gupta and colleagues proposed a mathematical model of multiple strain system using a multi-locus framework to study the effects of cross-reactivity between antigenic variants on the evolution and maintenance of strain community structures. Similar multi-strain models have appeared in literature to study the ecology of pathogens [39–41,9,42–44,12,7,11,45,13,32–35,2,46,47]. Due to challenges in analysing higher dimensional nonlinear coupled dynamical systems and the combinatorial nature of multi-strain systems, some modellers have often relied on numerical investigations to study and characterize the dynamics of the models, while others have resorted to making simplifying assumptions on the strain space (e.g., linear, circular, square) with a limited number of strains. In [36], Chan and Yu employed a groupoid approach and centre-manifold reduction to study the existence and stability of partially synchronous steady states (or discrete strain structure), while Blyuss [37,48] used tools from equivariant dynamical system to analyse the steady state solutions for a four-strain system. Cherif [38] provided a framework for a general multi-locus system for any arbitrary strain size.

In this paper, we extend the multi-locus model presented in [30,38] to incorporate age-specificity, and perform a detailed analysis with a particular emphasis on the existence of a mild solution and on the stability of the equilibria. Extending the notions used in [38], along with the use of the quotient network concept employed by [36,49] and references therein, we provide general results. We describe the model in Section 2. In Section 3, we formulate the multi-locus–allele model as an abstract Cauchy semi-linear evolutionary problem on an infinite-dimensional Banach space and show the existence of a positive mild solution. The result on the existence of steady-state solutions is presented in Section 4. In addition, under some appropriate (sometimes restrictive) assumptions, we determine the stability of the equilibria, and numerical simulations are provided to illustrate the different dynamics observed in the system in Sections 5 and 6. respectively. We conclude with discussion and some remarks on further research directions in Section 7.

2. Model description: age-structured multi-locus–allele system

In [30], the authors use an overlapping compartmental epidemic model to introduce the multi-locus framework, where pathogenic strains are defined by multiple loci, each occupied by a varying number of alleles from a specified set, and individuals are characterized based on their immunological history. Each allelic combination defines a strain of the pathogen. Here, we extend the framework to incorporate age-specificity. Let $Z_i(t,a)$ denote the density of individuals of age a that have been exposed to and are now immune to strain type i at time t, $W_i(t,a)$ the density of individuals of age a who, at time t, have been exposed to any antigenic type sharing alleles with strain i, $Y_i(t,a)$ the density of individuals of age a, at time t, that are infectious with strain i, and $P(t,a)$ represent the total population of age a, at time t. Extending the model in [30,38,15], the model describing the age-specific dynamics of infectious diseases with diverse antigenic type takes the following form:

$$\begin{aligned}
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) Z_i(t,a) &= \left(P(t,a)-Z_i(t,a)\right)\Lambda_i(t,a)-\mu(a)Z_i(t,a),\\
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) W_i(t,a) &= \left(P(t,a)-W_i(t,a)\right)\sum_{j\sim i}\Lambda_j(t,a)-\mu(a)W_i(t,a),\\
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) Y_i(t,a) &= \left[\left(P(t,a)-W_i(t,a)\right)+\left(1-\gamma(a)\right)\left(W_i(t,a)-Z_i(t,a)\right)\right]\Lambda_i(t,a)\\
&\quad -\left(\mu(a)+\sigma(a)\right)Z_i(t,a),
\end{aligned} \tag{1}$$

with initial conditions: $Z_i(0,a) = Z_i^\circ(a) \in L^1_+(0,\infty)$, $W_i(0,a) = W_i^\circ(a) \in L^1_+(0,\infty)$ and $Y_i(0,a) = Y_i^\circ(a) \in L^1_+(0,\infty)$, and boundary conditions: $Z_i(t,0) = 0$, $W_i(t,0) = 0$, $Y_i(t,0) = 0$, for $i = 1,2,\ldots,\mathcal{N}$, and $j \sim i$ in Eq. (1) means j similar to i. The epidemiological parameters $\mu(a)$, $\sigma(a)$ and $\gamma(a)$ denote the age-specific per capita mortality and recovery rates and cross-immunity level, respectively. The parameter $\gamma(a)$ is a measure of immune selection and $0 \le \gamma(a) \le 1$, $a \in \mathbb{R}_+$. The force of infection, $\Lambda(t,a)$, is of the following form:

$$\Lambda_i(t,a) = \frac{\int_0^\infty \beta(a,s) Y_i(t,s) ds}{\int_0^\infty P(t,s) ds},$$

where $\beta(a,s)$ combines both age-specific transmission and contact rates between individuals of age a and s, $N(t) = \int_0^\infty P(t,s)ds$, and the total population $P(t,a)$ follows the Sinko–Streifer or McKendrick–von Forester equation

$$\begin{aligned} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) P(t,a) &= -\mu(a)P(t,a), \\ P(t,0) &= \int_0^\infty f(s)P(t,s)ds, \quad \text{and} \quad P(0,a) = P^\circ(a). \end{aligned} \tag{2}$$

It should be noted that, in demographic terms, there exists a maximum age $a_m < \infty$ such that the survival probability vanishes. That is, with μ locally integrable on $[0, a_m)$, the maximum age must satisfy $\int_0^{a_m} \mu(\tau)d\tau = +\infty$, which is necessary for the survival probability $\mathcal{S}(a) = e^{-\int_0^a \mu(\tau)d\tau}$ to vanish at the maximum age a_m.

We note that, since the age-specific mortality and fertility (including zero net-migration) rates are constant (i.e., time-invariant), we can use the stable age-distribution assumption (see, for example, discussions in [19,50,51] for details) to find that $P(t,a) \to P^\infty(a)$ as $t \to \infty$, where $P^\infty(a)$ is the stable population age-distribution. Using the notion of stable population, we make the following change of variables. Let $s_i = \frac{P - W_i}{P^\infty}$, $v_i = \frac{W_i - Z_i}{P^\infty}$, $z_i = \frac{Z_i}{P^\infty}$, and $y_i = \frac{Y_i}{P^\infty}$, where $P(t,a) = P^\infty(a) = \mathcal{S}(a)$ is the stable population. Then, the system can be rewritten as follows:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) s_i(t,a) = -s_i \sum_j \Omega_{ij}\lambda_j, \tag{3}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) v_i(t,a) = s_i \sum_{j \ne i} \Omega_{ij}\lambda_j - v_i\lambda_i, \tag{4}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) y_i(t,a) = \left[s_i + (1-\gamma(a))v_i\right]\lambda_i - \sigma(a)y_i, \tag{5}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) z_i(t,a) = (s_i + v_i)\lambda_i, \tag{6}$$

where we have rewritten $\sum_{j\sim i} \Lambda_j(t,a)$ as $\sum_j \Omega_{ij}\Lambda_j(t,a)$ to simplify the notation, where Ω_{ij} is the element of the similarity or relatedness matrix Ω, and $\Omega_{ij} = 1$ if strain i is similar to strain j or $i = j$, and $\Omega_{ij} = 0$ otherwise; in addition, the initial and boundary conditions are given as follows: $s_i(t,0) = 1$, $v_i(t,0) = z_i(t,0) = y_i(t,0) = 0$ and $s_i(0,a) = s_i^\circ(a) \in L^1_+(0,\infty)$, $v_i(0,a) = v_i^\circ(a) \in L^1_+(0,\infty)$, $z_i(0,a) = z_i^\circ(a) \in L^1_+(0,\infty)$, and $y_i(0,a) = y_i^\circ(a) \in L^1_+(0,\infty)$, $i = 1,2,\ldots,\mathcal{N}$ and $\mathcal{N}$ is the number of strains. The force of infection is given by:

$$\lambda_i(t,a) = \int_0^\infty \beta(a,s)\psi(s)y_i(t,s)ds, \quad \text{and} \quad \psi(a) = \frac{P^\infty(a)}{\int_0^\infty P^\infty(a)da},$$

where we have assumed a stable age distribution $P^\infty(a)$ and $\int_0^\infty \psi(s)ds = 1$ and ψ is the normalized stable age distribution. In addition, because the other equations are independent of the variable z_i, we can eliminate

z_i from the system. Hence, we will consider only Eqs. (3)–(5) to study existence and stability questions. First, we restate the technical and biologically reasonable assumptions:

Assumption 2.1. We assume the following properties on demographic patterns:

1. $\mathbf{H}_1$: $\mu(a) \geq 0$ *in* $[0,\infty)$ and there exists $a_m \in [0,\infty)$ such that $\mu(a)$ is locally integrable on $[0,a_m)$ and $\int_0^{a_m} \mu(\tau)d\tau = +\infty$.
2. $\mathbf{H}_2$: $\beta(a,s) \in L^\infty_+((0,\infty)\times(0,\infty))$ and $\beta(a,s) \geq 0$. In addition, $\beta(a,s) \leq \tilde{\beta}(a)$ with $\tilde{\beta}(a) \in L^1(0,\infty)$ and $\overline{\beta} = \text{ess sup } \beta(a,s)$.
3. $\mathbf{H}_3$: $\sigma(a) \in L^\infty_+(0,\infty)$ and $\sigma(a) \geq 0$ with $\overline{\sigma} = \text{ess sup } \sigma(a)$.
4. $\mathbf{H}_4$: $\gamma(a) \in L^\infty_+(0,\infty)$ and $0 \leq \gamma(a) \leq 1,\ a \in \mathbb{R}_+$ with $\overline{\gamma} = \text{ess sup } \gamma(a)$.
5. $\mathbf{H}_5$: $\psi(a) \in L^\infty_+(0,\infty)$ and $\psi(a) \geq 0$ with $\overline{\psi} = \text{ess sup } \psi(a)$.

Note that $P^\infty(a)$ is proportional to $\mathcal{S}(a)$, so is zero for $a \geq a_m$. Thus, as the equations depend only on μ through $P^\infty(a)$, condition $\mathbf{H}_1$ suffices and shows in particular that $\int_0^\infty P^\infty(a)da$ exists.

3. Existence of a positive solution

We now consider the well-posedness of the evolution problem described by Eqs. (3)–(5) on the Banach Space

$$\mathcal{X} := L^1\left((0,\infty);\mathbb{R}^{3\mathcal{N}}\right) = \left[L^1(0,\infty)\right]^{3\mathcal{N}},$$

endowed with the usual norm

$$\|\phi\| = \sum_{i=1}^{3}\sum_{j=1}^{\mathcal{N}} \|\phi_j^{(i)}\|,$$

for $\phi(a) = \left(\phi_1^{(1)}(a),\phi_2^{(1)}(a),\ldots,\phi_{\mathcal{N}}^{(1)}(a),\phi_1^{(2)}(a),\phi_2^{(2)}(a),\ldots,\phi_{\mathcal{N}}^{(2)}(a),\phi_1^{(3)}(a),\phi_2^{(3)}(a),\ldots,\phi_{\mathcal{N}}^{(3)}(a)\right)^T \in \mathcal{X}$, where $\|*\|$ is the norm of $L^1(0,\infty)$ and T denotes the transpose operation. In addition, $\phi_k = \left(\phi_1^{(k)},\phi_2^{(k)},\ldots,\phi_{\mathcal{N}}^{(k)}\right) \in \mathbb{R}^{\mathcal{N}}$, $k = 1,2,3$. The state space of the normalized age distribution for the system induced by the semi-flow Eqs. (3)–(5) is given as follows:

$$\Gamma := \{\phi := (\phi_1,\phi_2,\phi_3)^T \in \mathcal{X}_+ : 0_{\mathcal{N}\times 1} \leq \phi_1+\phi_2+\phi_3 \leq 1_{\mathcal{N}\times 1}\},$$

where $\mathcal{X}_+ = \left[L^1_+(0,\infty)\right]^{3\mathcal{N}}$ and $L^1_+(0,\infty)$ denotes the positive cone of $L^1(0,\infty)$ (i.e., $L^1_+ = \{f \in L^1 : f \geq 0 \text{ a.e.}\}$).

Let $\mathcal{A}$ be the linear operator on $\mathcal{X}$ defined by:

$$\mathcal{A}[\phi](a) := \begin{pmatrix} \mathcal{A}_1 & 0 & 0 \\ 0 & \mathcal{A}_2 & 0 \\ 0 & 0 & \mathcal{A}_3 \end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\\ \phi_3\end{pmatrix},$$

where $\mathcal{A}_1 := -\mathbb{I}_{\mathcal{N}\times\mathcal{N}}(\frac{\partial}{\partial a})$, $\mathcal{A}_2 := -\mathbb{I}_{\mathcal{N}\times\mathcal{N}}(\frac{\partial}{\partial a})$ and $\mathcal{A}_3 := -\mathbb{I}_{\mathcal{N}\times\mathcal{N}}(\frac{\partial}{\partial a} + \sigma(a))$. In addition, $\phi = (\phi_1,\phi_2,\phi_3)^T \in \mathcal{D}(\mathcal{A})$. The domain $\mathcal{D}(\mathcal{A})$ is defined as $\mathcal{D}(\mathcal{A}) := \{\phi \in \mathcal{X} : \phi_1,\phi_2,\phi_3 \text{ are absolutely continuous on } [0,\infty) \text{ and } \phi_1(0) = 1_{\mathcal{N}\times 1}, \phi_2(0) = 0_{\mathcal{N}\times 1}, \phi_3(0) = 0_{\mathcal{N}\times 1}\}$. We also let $\mathcal{F}$ be the nonlinear operator $\mathcal{F} : \mathcal{X} \to \mathcal{X}$ defined by:

$$\mathcal{F}[\phi](a) := \begin{pmatrix}\mathcal{F}_1[\phi](a)\\ \mathcal{F}_2[\phi](a)\\ \mathcal{F}_3[\phi](a)\end{pmatrix}, \quad \text{and} \quad \mathcal{F}_k[\phi](a) := \begin{pmatrix}\mathcal{F}_1^{(k)}[\phi](a)\\ \mathcal{F}_2^{(k)}[\phi](a)\\ \cdots \\ \mathcal{F}_{\mathcal{N}}^{(k)}[\phi](a)\end{pmatrix}, \quad k = 1,2,3,$$

where

$$\mathcal{F}_i^{(1)}[\phi](a) := -\phi_i^{(1)}(a)\sum_j \Omega_{ij}\lambda_j[\phi](a), \qquad \mathcal{F}_i^{(2)}[\phi](a) := \phi_i^{(1)}(a)\sum_{j\neq i}\Omega_{ij}\lambda_j[\phi](a) - \phi_i^{(2)}\lambda_i[\phi](a),$$
$$\mathcal{F}_i^{(3)}[\phi](a) := \left(\phi_i^{(1)}(a) + (1-\gamma(a))\,\phi_i^{(2)}(a)\right)\lambda_i[\phi](a), \quad i = 1,2,\ldots,\mathcal{N}.$$

The operator $\lambda_j[\phi](a) : \mathcal{X} \to L^1(0,\infty)$ is the linear operator:

$$\lambda_j[\phi](a) = \int_0^\infty \beta(a,s)\psi(s)\phi_j^{(3)}(s)ds.$$

Now let $u = (u_1,u_2,u_3)^T \in L^1((0,\infty),\mathbb{R}_+^{3\mathcal{N}})$, where $u_1 = (s_1,s_2,\ldots,s_{\mathcal{N}})$, $u_2 = (v_1,v_2,\ldots,v_{\mathcal{N}})$ and $u_3 = (y_1,y_2,\ldots,y_{\mathcal{N}})$, $u_k \in L^1((0,\infty),\mathbb{R}_+^{\mathcal{N}})$. Then, Eqs. (3)–(5) can be reformulated as the abstract semi-linear Cauchy evolution equation,

$$\frac{du}{dt} = \mathcal{A}u + \mathcal{F}(u), \tag{7}$$

where we observe that the linear operator $\mathcal{A}$ is the infinitesimal generator of a strongly continuous semigroup (C_0-semigroup) of bounded linear operators $\{T_{\mathcal{A}}(t)\}_{t\geq 0} := \{e^{t\mathcal{A}}\}_{t\geq 0}$ on $\mathcal{X}$. In addition, the linear operator $T_{\mathcal{A}}(t)$ is defined by

$$T_{\mathcal{A}}(t)[\phi] := \begin{pmatrix} T_1(t)[\phi] \\ T_2(t)[\phi] \\ T_3(t)[\phi] \end{pmatrix},$$

where $\{T_i(t)[\phi]\}_{t\geq 0}$, $i = 1,2$ is a nilpotent right translation C_0-semigroup of operators given by $T_1(t)[\phi] := \phi_1(a-t)\mathcal{H}(a-t) + \mathcal{H}(t-a)$, $T_2(t)[\phi] := \phi_2(a-t)\mathcal{H}(a-t)$, and $T_3[\phi]$ is a combination of the nilpotent right translation and bounded multiplication C_0-semigroup of operators given by $T_3(t)[\phi] := e^{-\int_{a-t}^a \sigma(\tau)d\tau}\phi_3(a-t)\mathcal{H}(a-t)$, where $\mathcal{H}(s)$ is a Heaviside function (see [52, pg. 4] and [53, Ch. I]). Moreover, it can be shown that the state space Γ is closed and convex, and it is invariant with respect to the semi-flow defined by $\{T_{\mathcal{A}}(t)\}_{t\geq 0}$ (i.e., $T_{\mathcal{A}}(t)[\Gamma] \subset \Gamma$). Let $\omega_1 = \max_{1\leq j\leq \mathcal{N}}\sum_{i=1}^{\mathcal{N}}|\Omega_{ij}|$ and $\omega_\infty = \max_{1\leq i\leq \mathcal{N}}\sum_{j=1}^{\mathcal{N}}|\Omega_{ij}|$, where Ω_{ij} is an element of Ω. Because Ω is symmetric, $\omega = \omega_1 = \omega_\infty$. Define $\Xi_i^{(2)}[\phi](a) := \sum_{j\neq i}\Omega_{ij}\lambda_j[\phi](a)$. Then, we state the following inequalities.

Inequality 3.1. *Suppose Assumption 2.1 holds. Let λ_i be defined as below. Then, $\lambda_i : \mathcal{X} \to L^1(0,\infty)$ and the following inequalities hold:*

$$\begin{aligned}
\lambda_i[\phi](a) &= \int_0^\infty \beta(a,s)\psi(s)\phi_i^{(3)}(s)ds \leq \overline{\beta}\int_0^\infty \psi(s)\phi_i^{(3)}(s)ds \leq \overline{\beta\psi}\|\phi\|,\\
\lambda_i[\phi](a) &= \int_0^\infty \beta(a,s)\psi(s)\phi_i^{(3)}(s)ds \leq \overline{\beta}\int_0^\infty \psi(s)ds = \overline{\beta}, \quad \phi\in\Gamma,\\
\|\lambda_i[\phi]\| &= \int_0^\infty \left|\int_0^\infty \beta(a,s)\psi(s)\phi_i^{(3)}(s)ds\right| da \leq \int_0^\infty \left(\int_0^\infty \beta(a,s)da\right)\psi(s)\left|\phi_i^{(3)}(s)\right| ds\\
&\leq \|\tilde{\beta}\|\int_0^\infty \psi(s)\left|\phi_i^{(3)}(s)\right| ds \leq \|\tilde{\beta}\|\overline{\psi}\|\phi\|.
\end{aligned}$$

From the above inequalities, it follows that

$$\begin{aligned}
\Xi_i^{(2)}[\phi](a) &= \sum_{j\neq i}\Omega_{ij}\lambda_j[\phi](a) \leq (\omega-1)\,\overline{\beta\psi}\|\phi\|,\\
\Xi_i^{(2)}[\phi](a) &= \sum_{j\neq i}\Omega_{ij}\lambda_j[\phi](a) \leq \sum_{j\neq i}\Omega_{ij}\overline{\beta} \leq (\omega-1)\,\overline{\beta}, \quad \phi\in\Gamma,\\
\|\Xi_i^{(2)}[\phi]\| &\leq (\omega-1)\,\|\tilde{\beta}\|\overline{\psi}\|\phi\|.
\end{aligned}$$

From Inequality 3.1, we can show that $\mathcal{F}$ satisfies the following properties:

Lemma 3.1. *The operator $\mathcal{F}$ is continuously Fréchet differentiable on $\mathcal{X}$.*

Proof. We calculate $\mathcal{F}[\phi+h]-\mathcal{F}[\phi]$:

$$\mathcal{F}[\phi+h]-\mathcal{F}[\phi]:=\begin{pmatrix}\mathcal{F}_1[\phi+h]-\mathcal{F}_1[\phi]\\ \mathcal{F}_2[\phi+h]-\mathcal{F}_2[\phi]\\ \mathcal{F}_3[\phi+h]-\mathcal{F}_3[\phi]\end{pmatrix}.$$

From the definition of norm in $\mathcal{X}$, it is sufficient to show that each component is continuously Fréchet differentiable. Let $\Xi_i^{(1)}[\phi](a):=\Xi_i^{(2)}[\phi](a)+\lambda_i[\phi](a)$. Then,

$$\begin{aligned}\mathcal{F}_i^{(1)}[\phi+h]-\mathcal{F}_i^{(1)}[\phi]&=-\left(\phi_i^{(1)}+h_i^{(1)}\right)\Xi_i^{(1)}[\phi+h]+\phi_i^{(1)}\Xi_i^{(1)}[\phi]\\&=\mathcal{L}_i^{(1)}[\phi,h]+\mathcal{R}_i^{(1)}[\phi,h],\end{aligned}\tag{8}$$

where $\mathcal{L}_i^{(1)}[\phi,h]:=-\phi_i^{(1)}\Xi_i^{(1)}[h]-h_i^{(1)}\Xi_i^{(1)}[\phi]$ and $\mathcal{R}_i^{(1)}[\phi,h]:=-h_i^{(1)}\Xi_i^{(1)}[h]$. From Inequality 3.1, we observe that

$$\|\mathcal{R}_i^{(1)}[\phi,h]\|=\|h_i^{(1)}\Xi_i^{(1)}[h]\|\le\omega\|\tilde{\beta}\|\overline{\psi}\|h\|\int_0^\infty\left|h_i^{(1)}(a)\right|da\le\omega\|\tilde{\beta}\|\overline{\psi}\|h\|^2,$$

so $\lim_{\|h\|\to0}\frac{\|\mathcal{R}_i^{(1)}[\phi,h]\|}{\|h\|}=0$, and hence, $\mathcal{F}_i^{(1)}$ is continuously Fréchet differentiable with Fréchet derivative $\mathcal{L}_i^{(1)}[\phi,h]$. Similarly, it can be shown that $\mathcal{F}_i^{(2)}[\phi]$ and $\mathcal{F}_i^{(3)}[\phi]$ are also continuously Fréchet differentiable. □

Theorem 3.2. *Suppose Assumption 2.1 holds. For each $u^0\in\Gamma$, there is a unique continuous mild solution $u(t,u^o)\in\Gamma$, $t\in[0,\infty)$ for Eq. (7) such that*

$$u(t)=T_{\mathcal{A}}(t)u^0+\int_0^t T_{\mathcal{A}}(t-s)\mathcal{F}(u(s))ds.\tag{9}$$

In addition, if $u^0\in\mathcal{D}(\mathcal{A})\cap\Gamma$, then the solution is a classical one.

The proof uses a Contraction Mapping theorem argument. First, it can be seen that for any $0<K\le\min\left(\frac{1}{\omega\overline{\beta}},\frac{1}{\overline{\beta}}\right)$, $(Id|_{\mathcal{X}}+K\mathcal{F})(\Gamma)\subset\Gamma$. Then, we recast Eq. (9) as

$$u(t)=e^{-K^{-1}t}T_{\mathcal{A}}(t)u^0+K^{-1}\int_0^t e^{-K^{-1}(t-s)}T_{\mathcal{A}}(t-s)\left[u+K\mathcal{F}(u)\right](s)ds.$$

But now, using the convexity of Γ, it can be seen that if

$$\begin{aligned}\mathcal{H}[u](t)&=e^{-K^{-1}t}T_{\mathcal{A}}(t)u^0+K^{-1}\int_0^t e^{-K^{-1}(t-s)}T_{\mathcal{A}}(t-s)\left[u+K\mathcal{F}(u)\right](s)ds\\&=e^{-K^{-1}t}T_{\mathcal{A}}(t)u^0+\left(1-e^{-K^{-1}t}\right)\frac{K^{-1}}{1-e^{-K^{-1}t}}\int_0^t e^{-K^{-1}(t-s)}T_{\mathcal{A}}(t-s)\left[u+K\mathcal{F}(u)\right](s)ds,\end{aligned}$$

then $\mathcal{H}:\Gamma\to\Gamma$. Finally, under Assumption 2.1 and Inequality 3.1, we can show that, for $\phi,\eta\in\Gamma$, $\|\mathcal{F}[\phi]-\mathcal{F}[\eta]\|\le L\|\phi-\eta\|$, where $L=2\mathcal{N}(1+\omega)\left(\overline{\beta}+\|\tilde{\beta}\|\overline{\psi}\right)$, so that $(Id|_{\mathcal{X}}+K\mathcal{F})$ is globally Lipschitz continuous on Γ. Hence, there exists a mild solution with continuous dependence on the initial condition. In addition, the mild solution is classical as a result of the continuous differentiability of $\mathcal{F}[\phi]$ (see Lemma 3.1) as in [52, Ch. 6].

4. Equilibria: steady-state analysis

From now on, we will assume the maximal age is finite and assume the following:

Assumption 4.1. Let a_m be the maximal age, which we assume to be finite. We further assume

1. $\mathcal{H}_7$: $\beta(a,s) \in L^\infty_+((0,a_m)\times(0,a_m))$ and $\lim_{h\to 0}\int_0^{a_m}|\beta(a+h,s)-\beta(a,s)|\,da = 0$ uniformly for $s \in \mathbb{R}_+$, where β is extended by $\beta(a,s)=0$ for almost all $(a,s)\in(-\infty,0)\bigcup(a_m,\infty)$.
2. $\mathcal{H}_8$: There exists a^* with $a_m > a^* > 0$ and $\varepsilon > 0$ such that $\beta(a,s) \geq \varepsilon$ for almost all $(a,s) \in (0,a_m)\times(a_m-a^*,a_m)$.

If $s_i^*(a)$, $v_i^*(a)$ and $y_i^*(a)$ is a steady state solution of Eqs. (3)–(5), then by integrating these equations in this case, we get:

$$s_i^*(a) = e^{-\int_0^a \overline{\Xi}_i^{(1)}[\lambda^*](\tau)d\tau}, \qquad v_i^*(a) = \int_0^a e^{-\int_0^b \overline{\Xi}_i^{(1)}[\lambda^*](\tau)d\tau}\overline{\Xi}_i^{(2)}[\lambda^*](b)e^{-\int_b^a \lambda_i^*(\tau)d\tau}db, \tag{10}$$

$$y_i^*(a) = \int_0^a e^{-\int_b^a \sigma(\tau)d\tau}\left[s_i^*(b)+(1-\gamma(b))v_i^*(b)\right]\lambda_i^*(b)db, \tag{11}$$

where $\overline{\Xi}_i^{(1)}[\lambda^*](a) := \sum_j \Omega_{ij}\lambda_j^*(a)$ and $\overline{\Xi}_i^{(2)}[\lambda^*](a) := \sum_{j\neq i}\Omega_{ij}\lambda_j^*(a)$. Let $\lambda^* := (\lambda_1^*,\lambda_2^*,\ldots,\lambda_{\mathcal{N}}^*)^T \in \mathcal{Z}_+$, where $\mathcal{Z} := [L^1(0,a_m)]^{\mathcal{N}}$ and

$$\lambda_i^*[y](a) = \int_0^{a_m}\beta(a,s)\psi(s)y_i^*(s)ds. \tag{12}$$

Then, substituting Eq. (11) into Eq. (12), we have,

$$\lambda_i^*[y^*](a) = \Phi_i[\lambda^*](a) = \Phi_i^{(1)}[\lambda^*](a) + \Phi_i^{(2)}[\lambda^*](a), \tag{13}$$

where

$$\Phi_i^{(1)}[\lambda^*](a) = \int_0^{a_m}\beta(a,s)\psi(s)\left[\int_0^s e^{-\int_b^s\sigma(\tau)d\tau}e^{-\int_0^b\overline{\Xi}_i^{(1)}[\lambda^*](\tau)d\tau}\lambda_i^*(b)db\right]ds, \tag{14}$$

$$\begin{aligned}\Phi_i^{(2)}[\lambda^*](a) = \int_0^{a_m}\beta(a,s)\psi(s)\Bigg(\int_0^s e^{-\int_b^s\sigma(\tau)d\tau}(1-\gamma(b))\Bigg[\int_0^b e^{-\int_0^{b^*}\overline{\Xi}_i^{(1)}[\lambda^*](\tau)d\tau}\\ \times\overline{\Xi}_i^{(2)}[\lambda^*](b^*)e^{-\int_{b^*}^b\lambda_i^*(\tau)d\tau}db^*\Bigg]\lambda_i^*(b)db\Bigg)ds.\end{aligned} \tag{15}$$

From Eq. (13), it is clear that one trivial solution is $\lambda_i^* = 0$ for all $i = 1,2,\ldots,\mathcal{N}$. This solution corresponds to the disease-free equilibrium, where no disease exists. To investigate the existence of non-trivial positive steady-state solutions for Eqs. (3)–(5), it suffices to find the fixed-points of the nonlinear operator $\Phi[\lambda^*](a)$ in Eq. (13), where $\lambda^* := (\lambda_1^*,\lambda_2^*,\ldots,\lambda_{\mathcal{N}}^*)^T$ and $\Phi[\lambda](a) := (\Phi_1[\lambda](a),\Phi_2[\lambda](a),\ldots,\Phi_{\mathcal{N}}[\lambda](a))^T$.

We now state the following properties for $\Phi[\lambda]$. We observe that, $\Phi : \mathcal{Z}_+ \to \mathcal{Z}_+$ is positive. In addition, it follows that

$$\begin{aligned}\|\Phi_i^{(1)}(\lambda)\| &= \int_0^{a_m}\left|\left[\int_0^{a_m}\beta(a,s)\psi(s)\left(\int_0^s e^{-\int_b^s\sigma(\tau)d\tau}\lambda_i(b)e^{-\int_0^b\overline{\Xi}_i^{(1)}[\lambda](\tau)d\tau}db\right)ds\right]\right|da\\ &\leq \int_0^{a_m}\left|\left[\int_0^{a_m}\beta(a,s)\psi(s)\left(\int_0^s\lambda_i(b)e^{-\int_0^b\overline{\Xi}_i^{(1)}[\lambda](\tau)d\tau}db\right)ds\right]\right|da\\ &\leq \int_0^{a_m}\left|\left[\int_0^{a_m}\beta(a,s)\psi(s)ds\right]\right|da \leq \int_0^{a_m}\left|\tilde{\beta}(a)\right|\left[\int_0^{a_m}\psi(s)ds\right]da \leq \|\tilde{\beta}\|.\end{aligned}$$

Similarly, it can be shown that $\|\Phi_i^{(2)}(\lambda)\| \leq \|\tilde{\beta}\|$ so that $\|\Phi(\lambda)\| \leq 2\|\tilde{\beta}\|\mathcal{N} := \mathcal{K}$. Hence $\Phi(\mathcal{Z}_+)$ is bounded. Moreover, it can be shown that $\Phi(\lambda)$ is Lipschitz continuous. We now observe that for $\Phi_i^{(1)}[\lambda](a)$, $\Phi_i^{(1)}[0+h](a) - \Phi_i^{(1)}[0](a) = \Phi_i^{(1)}[h](a) = \mathcal{T}_i[h](a) + \mathcal{R}_i^{(1)}[h](a)$, where $\mathcal{T}_i : \mathcal{Z} \to L^1(0, a_m)$ is given by

$$\mathcal{T}_i[\lambda](a) = \int_0^{a_m} \Psi(a,b)\lambda_i(b)db, \quad i = 1, 2, \ldots, \mathcal{N}, \tag{16}$$

$$\Psi(a,b) = \int_b^{a_m} \beta(a,s)\psi(s)\Theta(b,s)ds, \quad \text{and} \quad \Theta(b,s) = e^{-\int_b^s \sigma(\tau)d\tau}, \tag{17}$$

and $\mathcal{R}_i^{(1)}[h](a)$ is defined as follows

$$\mathcal{R}_i^{(1)}[h](a) = \int_0^{a_m} \beta(a,s)\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau} h_i^{(1)}(b)\left[e^{-\int_0^b \Xi_i^{(1)}[h](\tau)d\tau} - 1\right]db\right)ds.$$

We note that $\frac{\|\mathcal{R}_i^{(1)}[h](a)\|}{\|h\|} \leq \omega\|\tilde{\beta}\|\,\|h\|$, which goes to zero as $\|h\| \to 0$. Thus, $\mathcal{T}_i[h](a)$ is the Fréchet derivative of $\Phi_i^{(1)}[\lambda](a)$ at $\lambda = 0$. Similarly, the Fréchet derivative of $\Phi_i^{(2)}[\lambda](a)$ at $\lambda = 0$ is zero. Hence, the Fréchet derivative of $\Phi_i[\lambda](a)$ at $\lambda = 0$ is $\mathcal{T}_i[h](a)$. We conclude that the operator Φ is Fréchet differentiable at $\lambda = 0$ with derivative $\mathcal{T} : \mathcal{Z} \to \mathcal{Z}$ given by,

$$\mathcal{T}[\lambda](a) := \begin{pmatrix} \mathcal{T}_1[\lambda](a) \\ \mathcal{T}_2[\lambda](a) \\ \cdots \\ \mathcal{T}_{\mathcal{N}}[\lambda](a) \end{pmatrix}. \tag{18}$$

In addition, it can be seen that the operator $\mathcal{T}$ is a positive and bounded linear operator, and if $\lambda_i[\phi] \in L^1_+(0, a_m) \setminus \{0\}$, then $\mathcal{T}_i[\lambda](a) > 0$. Here $\mathcal{T} : \mathcal{Z} \to \mathcal{Z}$ defines the threshold or next-generation operator which transforms the distribution of infected population to the distribution of secondary cases in the initial phase of epidemiological invasion. As a result, the basic reproductive number $\mathcal{R}_0$ is given by the spectral radius of $\mathcal{T}$ (the spectral radius of $\mathcal{T}$ is the maximum eigenvalue of $\mathcal{T}$, see [54]), and is defined as the expected number of new infections caused by one infected individual during its entire infectious period. In addition, we note that $\mathcal{T}[\lambda]$ is symmetric across strains because we have assumed that all disease-related age-specific parameters are the same. We can therefore focus on just one of the next-generator operators.

Lemma 4.1. *Under Assumption 4.1, for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $|h| < \delta$,*

$$\int_0^{a_m} |\Psi(a+h,b) - \Psi(a,b)|\,da < \varepsilon, \quad \text{for all } b \in [0, a_m),$$

where Ψ is defined in Eq. (17).

Theorem 4.2. *The operator $\Phi : \mathcal{Z} \to \mathcal{Z}$ is completely continuous.*

Proof. Let $\mathcal{S}$ be a bounded subset of $L^1(0, a_m)$ and $F_i := \Phi_i(\mathcal{S}) \subset L^1(0, a_m)$ for each $i = 1, 2, \ldots, \mathcal{N}$. We want to prove that $\Phi(\mathcal{S})$ has a compact closure. We will use the Riesz–Fréchet–Kolmogorov Compactness Theorem (see [55, pg. 111, Theorem 4.26]). If $\lambda_i \in \mathcal{S}$, then we have $\|\lambda_i\| \leq c_0$ for some positive constant c_0. In addition, with the result of Inequality 3.1, we have

$$\begin{aligned}\|\Phi_i^{(1)}[\lambda]\| &= \int_0^{a_m}\left[\int_0^{a_m} |\beta(a,s)|\,\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau} e^{\int_0^b \left|\Xi_i^{(1)}[\lambda](\tau)\right|d\tau} |\lambda_i(b)|\,db\right)ds\right]da \\ &\leq e^{\omega c_o}\int_0^{a_m} \left|\tilde{\beta}(a)\right|\left[\int_0^{a_m} \psi(s)\left(\int_0^s |\lambda_i(b)|\,db\right)ds\right]da \leq e^{\omega c_o}\|\tilde{\beta}\|\,\|\lambda_i\| \leq e^{\omega c_o}\|\tilde{\beta}\|c_0.\end{aligned}$$

Using a similar argument, we observe that $\|\Phi_i^{(2)}[\lambda]\| \leq e^{\omega c_o}\|\tilde{\beta}\|(\omega-1)c_0^2$. Therefore, $\|\Phi_i[\lambda]\| \leq e^{\omega c_o}\|\tilde{\beta}\|c_0 + e^{\omega c_o}\|\tilde{\beta}\|(\omega-1)c_0^2$, and $\|\Phi[\lambda]\| \leq \mathcal{N}e^{\omega c_o}\|\tilde{\beta}\|c_0(1+(\omega-1)c_0)$. Hence, F_i are bounded. It now suffices to show that $\|\rho_h[\Phi[\lambda]](a) - \Phi[\lambda](a)\| \to 0$ uniformly, where $\rho_h[\Phi[\lambda]](a) = \Phi[\lambda](a+h)$. We observe that, using Assumption 4.1 (i.e. for every $\varepsilon > 0$, there exists δ such that $\int_0^{a_m} |\beta(a+h,s) - \beta(a,s)|\, da \leq \varepsilon$, for all $s \in \mathbb{R}_+$ and $|h| \leq \delta$), we have

$$\|\rho_h[\Phi_i[\lambda]](a) - \Phi_i[\lambda](a)\| \leq \|\rho_h[\Phi_i^{(1)}[\lambda]](a) - \Phi_i^{(1)}[\lambda](a)\| + \|\rho_h[\Phi_i^{(2)}[\lambda]](a) - \Phi_i^2[\lambda](a)\|,$$

where we calculate $\|\rho_h[\Phi_i^{(1)}[\lambda]](a) - \Phi_i^{(1)}[\lambda](a)\|$ as follows:

$$\begin{aligned}
\|\rho_h[\Phi_i^{(1)}[\lambda]](a) - \Phi_i^{(1)}[\lambda](a)\| &\leq \int_0^{a_m}\left[\int_0^{a_m} |\beta(a+h,s)-\beta(a,s)|\,\psi(s)\right.\\
&\quad\left.\times\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau} e^{\int_0^b \left|\overline{\Xi}_i^{(1)}[\lambda](\tau)\right|d\tau}\,|\lambda_i(b)|\,db\right)ds\right]da\\
&\leq e^{\omega c_o}\|\lambda\|\int_0^{a_m}\left[\int_0^{a_m}|\beta(a+h,s)-\beta(a,s)|\,\psi(s)ds\right]da\\
&\leq e^{\omega c_o}c_0\varepsilon.
\end{aligned}$$

Using a similar approach for $\Phi_i^{(2)}$, $\|\rho_h\left[\Phi_i^{(2)}[\lambda]\right](a) - \Phi_i^{(2)}[\lambda](a)\| \leq \varepsilon e^{\omega c_o}(\omega-1)c_0^2$. Hence, $\|\rho_h[\Phi[\lambda]](a) - \Phi[\lambda](a)\| \leq \mathcal{N}\varepsilon e^{\omega c_o}c_0(1+(\omega-1)c_0)$. Therefore, $F_i = \{\Phi_i(u) : u \in \mathcal{S}\}$ is precompact for each $i = 1,2,\ldots,\mathcal{N}$ in $L^1(0,a_m)$. Let $F = \{\Phi(u) : u \in \mathcal{S}\}$, then $F \subset F_1 \times F_2 \times \cdots \times F_{\mathcal{N}}$ implies that $\overline{F} \subset \overline{F_1 \times F_2 \times \cdots \times F_{\mathcal{N}}} = \overline{F}_1 \times \overline{F}_2 \times \cdots \times \overline{F}_{\mathcal{N}}$, where $\overline{F}$ denotes the closure of F. Because $\overline{F}_i$ is compact, $\overline{F}_1 \times \overline{F}_2 \times \cdots \times \overline{F}_{\mathcal{N}}$ is compact, so $\overline{F}$ is compact. Hence, Φ is completely continuous. □

To establish the properties of $\mathcal{T}[\lambda]$, we use the Perron–Frobenius theory for positive monotone operators on a partially ordered Banach space (see [56–58]). We make the following definitions:

Definition 4.1. 1. $\phi \in \mathcal{Y}_+$ is called a non-supporting (quasi-interior) point if $\langle F, \phi\rangle > 0$ for all $F \in \mathcal{Y}_+^* \setminus \{0\}$.
2. The operator $\mathcal{L} \in \mathcal{B}(\mathcal{Y})$ is called non-supporting if for every pair $v \in \mathcal{Y}_+ \setminus \{0\}$, $F \in \mathcal{Y}_+^* \setminus \{0\}$, there exists a positive $p = p(v,F) \in \mathbb{Z}_+$ such that $\langle F, \mathcal{L}^n v\rangle > 0$ for all $n \geq p$ (see [58]).

From [56,58], if the operator $\mathcal{L} \in \mathcal{B}(\mathcal{Y})$ is a compact, positive, bounded linear operator with spectral radius $r(\mathcal{L}) > 0$, then $r(\mathcal{L})$ is the largest eigenvalue of $\mathcal{L}$ with positive eigenfunction $v \in \mathcal{Y}_+ \setminus \{0\}$. In addition, if $\mathcal{L}$ is non-supporting, then $r(\mathcal{L})$ is the only positive eigenvalue with a positive eigenfunction and, furthermore, the eigenspace associated with the eigenvalue $r(\mathcal{L})$ is a one-dimensional subspace spanned by a quasi-interior point. Also, $r(\mathcal{L})$ is an eigenvalue of the adjoint operator $\mathcal{L}^* : \mathcal{Y}^* \to \mathcal{Y}^*$ with a strictly positive eigenfunction and the associated eigenspace is one-dimensional and is spanned by a strictly positive functional.

Note that $\mathcal{T} : \mathcal{Z} \to \mathcal{Z}$ is not nonsupporting. (From below the eigenspace of $\mathcal{T}$ associated with the eigenvalue $r(\mathcal{T})$ is not one dimensional. Furthermore, if we consider $F \in \mathcal{Z}^* \setminus \{0\}$ defined by $\langle F, \phi\rangle = \int_0^{a_m} \phi_1(a)da$, where $\phi = (\phi_1,\ldots,\phi_N)^T$, and then take $e_2 = (0,1,0,\ldots,0)^T \in \mathcal{Z}_+ \setminus \{0\}$, then for all $n \in \mathbb{Z}_+$, $\langle F, \mathcal{T}^n[e_2]\rangle = 0$.) However, if we define the associated operator $\mathcal{T}_0 : L^1(0,a_m) \to L^1(0,a_m)$ as below, then $\mathcal{T}_0$ is nonsupporting and we can deduce appropriate properties of $\mathcal{T}$ from those of $\mathcal{T}_0$.

Theorem 4.3. *Define* $\mathcal{T}_0 : L^1(0,a_m) \to L^1(0,a_m)$ *by*

$$\mathcal{T}_0[u](a) = \int_0^{a_m} \Psi(a,b)u(b)db,$$

so that, if $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{\mathcal{N}})$, *we have*

$$\mathcal{T}[\lambda](a) = (\mathcal{T}_0[\lambda_1](a), \mathcal{T}_0[\lambda_2](a), \cdots \mathcal{T}_0[\lambda_{\mathcal{N}}](a))^T.$$

Then, the following hold:

(i) $\mathcal{T}_0 : L^1(0, a_m) \to L^1(0, a_m)$ *and* $\mathcal{T} : \mathcal{Z} \to \mathcal{Z}$ *are both completely continuous.*
(ii) $\mathcal{T}_0 : L^1(0, a_m) \to L^1(0, a_m)$ *is non-supporting.*
(iii) $r(\mathcal{T}_0) = r(\mathcal{T})$, *and* $r(\mathcal{T})$ *is the only positive eigenvalue of* $\mathcal{T}$ *with a positive eigenfunction. The eigenspace of* $\mathcal{T}$ *associated with* $r(\mathcal{T})$ *is* $\mathcal{N}$*-dimensional and is of the form* ψV *for any* $V \in \mathbb{R}$, *where* $\psi \in L^1_+(0, a_m)$ *is the quasi-interior point which spans the eigenspace of* $\mathcal{T}_0$ *associated with the eigenvalue* $r(\mathcal{T}_0)$.
(iv) *The adjoint operator* $\mathcal{T}^* : \mathcal{Z}^* \to \mathcal{Z}^*$ *also has eigenvalue* $r(\mathcal{T})$ *with strictly positive eigenfunctions given by* $\langle G, u\rangle = \sum_{i=1}^{\mathcal{N}} \langle g, u_i\rangle \alpha_i$, *for* $\alpha_i > 0$, *where* $u = (u_1, \ldots, u_{\mathcal{N}})^T \in \mathcal{Z}$, *and* $g \in [L^1(0, a_m)]^* \setminus \{0\}$ *is the strictly positive functional which spans the eigenspace of* $\mathcal{T}_0^*$ *associated with the eigenvalue* $r(\mathcal{T}_0)$.

Proof. (i) The compactness of $\mathcal{T}_0$ is obtained by using Lemma 4.1 and similar arguments as in Theorem 4.2 via Riesz–Fréchet–Kolmogorov Compactness theorem. Since $\mathcal{T}_0$ and $\mathcal{T}$ are also bounded, they are completely continuous.

(ii) We follow the approach of Inaba, [59,60], and show that Definition 4.1 holds. We define $\tilde{s}(a) := \varepsilon$ if $a \in (a_m - a^*, a_m)$, $a_m \in (0, \infty)$, and $\tilde{s}(a) := 0$ otherwise. Then it follows from Assumption 4.1 that $\beta(a, s) \geq \tilde{s}(s)$ for all $(a, s) \in (0, a_m) \times (0, a_m)$. Let f be the linear functional on $L^1(0, a_m)$ defined as

$$\langle f, u\rangle = \int_0^{a_m} \left[\int_b^{a_m} \tilde{s}(\tau)\psi(\tau)\Theta(b, \tau)d\tau\right] u(b)db.$$

Then, the definition of $\tilde{s}$ implies that $f \in \left[L^1_+(0, a_m)\right]^*$. We observe that $\int_b^{a_m} \psi(s)\Theta(b, s)ds > 0$ for all $b \in (0, a_m)$. So, f is strictly positive. Hence, if $u \in L^1_+(0, a_m)$, then from $\beta(a, s,) \geq \tilde{s}(s)$, we have $\mathcal{T}_0[u](a) \geq \langle f, u\rangle e(a)$, where $e(a) = 1$ for all $a \in \mathbb{R}_+$. Therefore, for any $n \in \mathbb{Z}_+$, we have $\mathcal{T}_0^{n+1}u \geq \langle f, u\rangle\langle f, e\rangle^n e(a)$. Moreover, if the functional $F \in \left[L^1_+(0, a_m)\right]^* \setminus \{0\}$, then $\langle F, \mathcal{T}_0^n[u]\rangle \geq \langle f, u\rangle\langle f, e\rangle^{n-1}\langle F, e\rangle > 0$, for $n \geq 1$ for every any $u \in L^1_+(0, a_m) \setminus \{0\}$. Hence $\mathcal{T}_0$ is non-supporting.

(iii) From (i), $r(\mathcal{T}_0)$ and $r(\mathcal{T})$ are maximal eigenvalues of $\mathcal{T}_0$ and $\mathcal{T}$, respectively. We observe that $\mathcal{T}_0\psi = r(\mathcal{T}_0)\psi$. If $e_1, e_2, \ldots, e_{\mathcal{N}}$ are the coordinate vectors, then $\mathcal{T}(\psi e_i) = r(\mathcal{T}_0)\psi e_i$. So, $r(\mathcal{T}_0)$ is also an eigenvalue of $\mathcal{T}$. Hence, $r(\mathcal{T}_0) \leq r(\mathcal{T})$. But if $\mathcal{T}u = r(\mathcal{T})u$, so

$$\mathcal{T}u = (\mathcal{T}_0u_1, \mathcal{T}_0u_2, \ldots, \mathcal{T}_0u_{\mathcal{N}})^T = r(\mathcal{T})(u_1, u_2, \ldots, u_{\mathcal{N}})^T,$$

then, $\mathcal{T}_0u_i = r(\mathcal{T})u_i$. But, there exists $u_i \neq 0$, so $\mathcal{T}_0u_i = r(\mathcal{T})u_i$. Therefore, $r(\mathcal{T}_0) \geq r(\mathcal{T})$. Hence, $r(\mathcal{T}) = r(\mathcal{T}_0)$. More generally, the same argument shows that λ is an eigenvalue of $\mathcal{T}$ if and only if λ is an eigenvalue of $\mathcal{T}_0$. Hence, $r(\mathcal{T})$ is the only positive eigenvalue of $\mathcal{T}$. Furthermore, the eigenspace of $\mathcal{T}$ is spanned by $\{\psi e_1, \psi e_2, \ldots, \psi e_{\mathcal{N}}\}$.

(iv) We note $\langle \mathcal{T}_0^*[g], u\rangle = \langle g, \mathcal{T}_0[u]\rangle$. Define $\langle G, u\rangle = \sum_{i=1}^{\mathcal{N}} \alpha_i\langle g, u_i\rangle$, where $u = (u_1, u_2, \ldots, u_{\mathcal{N}})^T$ and $\alpha_i > 0$. Then,

$$\begin{aligned}\langle \mathcal{T}^*[G], u\rangle &= \langle G, \mathcal{T}[u]\rangle \\ &= \left\langle G, \sum_{i=1}^{\mathcal{N}} \mathcal{T}_0[u_i]e_i \right\rangle = \sum_{i=1}^{\mathcal{N}} \alpha_i\langle g, \mathcal{T}_0[u_i]\rangle = \sum_{i=1}^{\mathcal{N}} \alpha_i\langle \mathcal{T}_0^*[g], u_i\rangle = \sum_{i=1}^{\mathcal{N}} \alpha_i r(\mathcal{T})\langle g, u_i\rangle \\ &= r(\mathcal{T})\langle G, u\rangle = r(\mathcal{T})\langle G, u\rangle.\end{aligned}$$

Therefore, $r(\mathcal{T})$ is an eigenvalue of $\mathcal{T}^*$ with eigenfunction G. In addition, G is strictly positive because if $u \neq 0$, then $u_i \neq 0$ for some i. Hence, $\langle g, u_i\rangle > 0$ and $\langle G, u\rangle > 0$. □

Using results from [56,58,57] and Krasnoselskii's fixed-point theorem (see [61]), we are now able to prove the following results.

Theorem 4.4 (*[61,60]*). *The following properties hold:*

1. *If $r(\mathcal{T}) < 1$, then $u = 0$ is the only solution of $u = \Phi(u)$ in $\mathcal{Z}_+$ and it is the disease-free equilibrium*
2. *If $r(\mathcal{T}) > 1$, then $u = \Phi(u)$ has at least one solution in $\mathcal{Z}_+ \setminus \{0\}$ and it is the endemic equilibrium.*

Proof. Following the approach of [60], suppose $r(\mathcal{T}) < 1$. We show that when $r(\mathcal{T}) < 1$, $u = 0$ is the only fixed point by contradiction. Suppose there exists a solution $\overline{u} \in \mathcal{Z}_+ \setminus \{0\}$ of $\overline{u} = \Phi(\overline{u})$, then $\overline{u} = \Phi(\overline{u}) \leq \mathcal{T}[\overline{u}]$. From Theorem 4.3, we let $\mathcal{F}_0^* \in [\mathcal{Z}_+]^* \setminus \{0\}$ denote a strictly positive eigenfunction of $\mathcal{T}^*$ corresponding to the eigenvalue $r(\mathcal{T})$. Then, by duality pairing, $\langle \mathcal{F}_0^*, \mathcal{T}[\overline{u}] - \overline{u} \rangle = (r(\mathcal{T}) - 1) \langle \mathcal{F}_0^*, \overline{u} \rangle$. Because $\mathcal{T}[\overline{u}] - \overline{u} \in \mathcal{Z}_+$ and $\mathcal{F}_0^*$ is strictly positive, we note that $\langle \mathcal{F}_0^*, \mathcal{T}[\overline{u}] - \overline{u} \rangle \geq 0$. In addition, we also observe that $\langle \mathcal{F}_0^*, \overline{u} \rangle > 0$ since, by assumption, $\overline{u} \in \mathcal{Z}_+ \setminus \{0\}$, and $\mathcal{F}_0^*$ is strictly positive. Then, we have $\langle \mathcal{F}_0^*, \mathcal{T}[\overline{u}] - \overline{u} \rangle = (r(\mathcal{T}) - 1) \langle \mathcal{F}_0^*, \overline{u} \rangle < 0$, because $r(\mathcal{T}) - 1 < 0$. This is a contradiction. Hence, $\overline{u} = 0$ must be the only solution of $\overline{u} = \Phi([\overline{u}]) \in \mathcal{Z}_+$.

Now suppose $r(\mathcal{T}) > 1$. To begin with, we show that the conditions of Krasnoselskii's fixed point theorem are satisfied (see [61]). We note that, from the previous discussion, $\Phi(\mathcal{Z}_+)$ is bounded with $\Phi(0) = 0$ and, from Theorem 4.2, Φ is a completely continuous operator in the Banach space $\mathcal{Z}$. In addition, Φ is Fréchet differentiable, with Fréchet derivative at zero given by the linear operator $\mathcal{T} = \Phi'(0)$. So, the first two conditions of Krasnoselskii's fixed point theorem are satisfied. From Theorem 4.3, there exists a positive eigenfunction of $\mathcal{T}$ corresponding to the eigenvalue $r(\mathcal{T})$ and by assumption, $r(\mathcal{T}) > 1$. Furthermore, from Theorem 4.3, $r(\mathcal{T})$ is the only positive eigenvalue with positive eigenfunction. So, $\mathcal{T}$ does not have a positive eigenfunction corresponding to the eigenvalue one [56,58]. Hence, the third condition of the Krasnoselskii's fixed point theorem is satisfied. Therefore, Φ has a non-zero positive fixed point. □

The proof above follows the approaches of [59,61], and establishes the properties of the reproductive number which we denote as $\mathcal{R}_0 = r(\mathcal{T})$. It provides a threshold-like property for the existence of steady states which are the solutions corresponding to the fixed point problem $\lambda = \Phi(\lambda)$. In the proof above, we have not discussed the case when $r(\mathcal{T}) = 1$ because this corresponds to the bifurcation point. Note that the Krasnoselskii Theorem [61] does not establish the uniqueness of the non-zero (non-trivial) positive (endemic) equilibrium. To establish uniqueness, we must provide additional properties. However, our goal in this section is to show the existence of a positive fixed point and the uniqueness problem is not necessary since we know, from [30,6,38,15] and numerical simulations presented in Section 6, that there are different concepts of endemicity corresponding to both the semi-trivial and non-trivial steady states.

5. Stability analysis of equilibria

5.1. Preliminary: quotient network and equilibrium clusters

In the previous section, we established a threshold condition defined by the spectral radius, $r(\mathcal{T})$, of the next-generation operator $\mathcal{T} : \mathcal{Z} \to \mathcal{Z}$. In this section, we consider $\mathcal{T}_0 : L^1(0, a_m) \to L^1(0, a_m)$, $\mathcal{T}_0 = \int_0^{a_m} \Psi(a, b)u(b)db$. From Theorem 4.3, $\mathcal{T}_0$ is nonsupporting and $r(\mathcal{T}_0) = r(\mathcal{T})$. In the case of the homogeneous systems, an explicit expression for the basic reproductive number can be obtained. However, in heterogeneous systems such as the model considered herein, it is not always straight-forward or may not be possible to obtain an explicit expression for $\mathcal{R}_0$. This is only possible in a few cases (e.g., proportionate mixing, intra-group mixing, see [54]). For ease of analysis, we will assume that there is a proportional mixing.

Assumption 5.1. The contact rate $\beta(a,s)$ is of the form $\beta(a,s) := \beta_1(a)\beta_2(s)$ where $\beta_1(a) \in L^\infty_+(0,a_m)$ and $\beta_2(s) \in L^\infty_+(0,a_m)$.

This assumption implies that the ages of susceptible and infected individuals are uncorrelated. It allows us to calculate the basic reproductive number explicitly, and to simplify the analysis in the case of the endemic equilibrium. Under proportionate mixing or separable mixing, we observe that the next-generation operator satisfies:

$$\mathcal{T}_0[u](a) = \int_0^{a_m} \left(\int_b^{a_m} \beta(a,s)\psi(s) e^{-\int_b^s \sigma(\tau)d\tau} ds \right) u(b)db, \tag{19}$$

$$= \beta_1(a) \int_0^{a_m} \left(\int_b^{a_m} \beta_2(s)\psi(s) e^{-\int_b^s \sigma(\tau)d\tau} ds \right) u(b)db. \tag{20}$$

Substituting β_1 for u and changing the order of integration yields

$$\mathcal{T}_0[\beta_1](a) = \beta_1(a) \int_0^{a_m} \left(\int_b^{a_m} \beta_2(s)\psi(s) e^{-\int_b^s \sigma(\tau)d\tau} ds \right) \beta_1(b)db$$

$$= \beta_1(a) \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^{s} \beta_1(b) e^{-\int_b^s \sigma(\tau)d\tau} db \right) ds. \tag{21}$$

So β_1 is an eigenfunction of $\mathcal{T}_0$ with positive eigenvalue. Therefore, under the proportionate mixing, the spectral radius of the next-generation operator $r(\mathcal{T}_0)$ is given by $\mathcal{R}_0$, where

$$\mathcal{R}_0 = \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^{s} \beta_1(b) e^{-\int_b^s \sigma(\tau)d\tau} db \right) ds. \tag{22}$$

Throughout this paper, we have assumed that strains have the same epidemiological parameters, where $\beta_i(a,s) = \beta(a,s)$, $\sigma_i(a) = \sigma(a)$ and $\psi_i(a) = \psi(a)$. In general, the epidemiological parameters for each strain can be different. That is, $\beta_i(a,s) \neq \beta_j(a,s)$, $\sigma_i(a) \neq \sigma_j(a)$ and $\psi_i(a) \neq \psi_j(a)$, $i \neq j$. In this case, we can define the strain-specific reproductive number $\mathcal{R}_0^{(i)}$ with proportionate mixing as

$$\mathcal{R}_0^{(i)} = \int_0^{a_m} \beta_{i,2}(s)\psi_i(s) \left(\int_0^{s} \beta_{i,1}(b) e^{-\int_b^s \sigma_i(\tau)d\tau} db \right) ds. \tag{23}$$

Then, the reproductive number $\mathcal{R}_0$ is defined as $\mathcal{R}_0 = \max_{1\leq i\leq \mathcal{N}}\{\mathcal{R}_0^{(i)}\}$. Previously, it was shown that, when $\mathcal{R}_0 < 1$, the dynamics are simple and there is only a trivial solution corresponding to the disease-free equilibrium (or disease extinction); and when $\mathcal{R}_0 > 1$, the system has at least one non-zero positive solution. When $\mathcal{R}_0 > 1$, numerical simulations [30,6,38,15, and refs. therein] suggest that the endemic equilibrium is not unique, and the dynamics are much more involved. In particular, the system can exhibit coexistence and the principle of competitive exclusion whereby in the presence of multiple strains only the strongest can survive depending on the strength of parameters modulating the competition.

Suppose we have a 2-strain system, then there are three possible endemic configurations. That is, if $\mathcal{R}_0 > 1$, then either strain 1 can persist while strain 2 goes to extinction, or strain 2 can persist while strain 1 goes to extinction, or both strains 1 and 2 can coexist. Similarly, in a 4-strain system, there are more possibilities. Fig. 1 illustrates all the possible endemic configurations. For instance, there are 15 possible persistent sets, namely $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{1,2\}$, $\{2,3\}$, $\{3,4\}$, $\{4,1\}$, $\{1,2,3\}$, $\{2,3,4\}$, $\{3,4,1\}$, $\{4,1,2\}$, $\{1,3\}$, $\{2,4\}$ and $\{1,2,3,4\}$. These are represented in Fig. 1, where persistent strains are marked by black circles. The 4-strain system illustrated in Fig. 1 is equivalent to a 2-locus–2-allele system resulting on a square lattice. Fig. 1 also illustrates different types of equilibrium clusters where black circles (•) denote strain pertaining to the persistent cluster while white circles (○) represent strains in extinct clusters. Using the idea summarized in Fig. 1, we introduce the following notion of positive solutions [51,38,15].

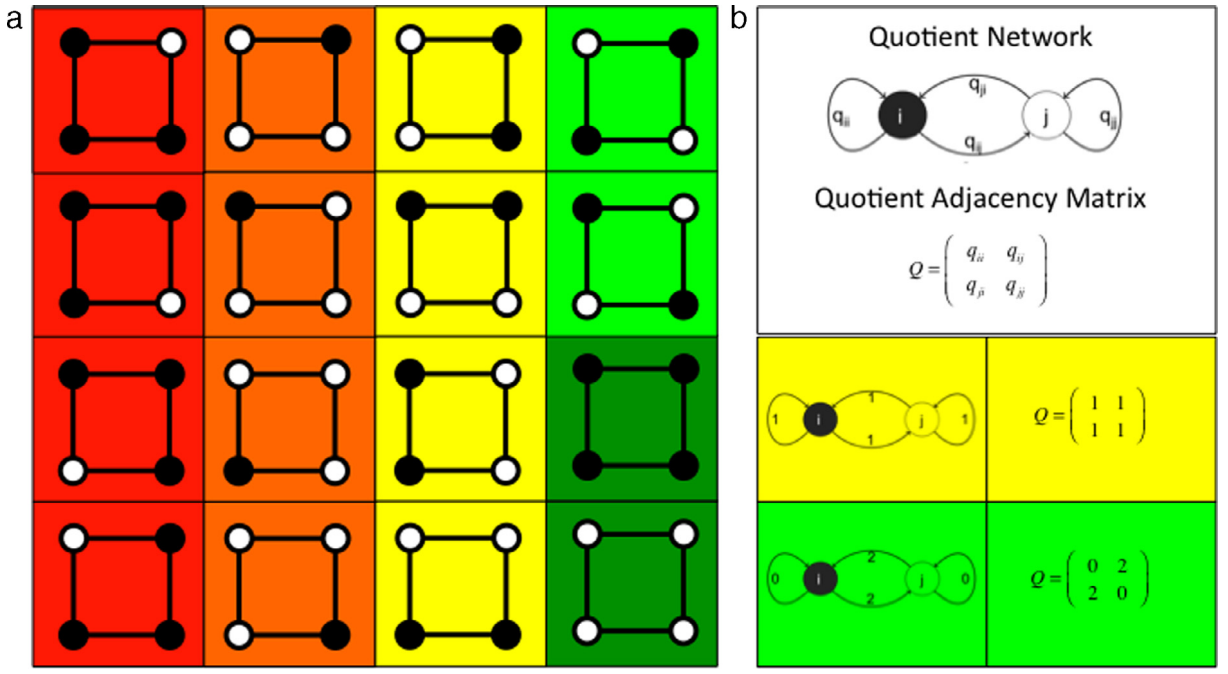

Fig. 1. Multi-locus system (Possible Equilibria for Four Strain System on a Square Lattice with $\mathbb{D}_4$-Symmetry) and quotient network representation. See text for full details. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Source: Diagram modified from [38].

Definition 5.1. 1. A steady state solution is called weakly endemic if $y_i(a) = y_i^*(a) \neq 0$ for some $i = 1, 2, \ldots, \mathcal{N}$.

2. A steady state solution is called strongly endemic if $y_i(a) = y_i^*(a) \neq 0$ for all $i = 1, 2, \ldots, \mathcal{N}$.

In the definition above, a weakly endemic equilibrium is equivalent to the principle of competitive exclusion (or partial coexistence) and strongly endemic equilibrium represents coexistence. In the proceeding sections, we will use the above notion with structural properties of the strain space to study the asymptotic behaviours of the model and the equilibria.

To study the general multi-locus system requires us to analyse $3\prod_{i=1}^{m} n_i$ systems (or $3n^m$ for uniform allelic combination), where n is the number of alleles and m is the number of loci. Instead of analysing this potentially high dimensional system, we use notions from graph theory [62–64]. In particular, we observe that the similarity matrix Ω can be rewritten as $\Omega = \mathbb{I} + \overline{\Omega}_G$, where $\overline{\Omega}_G$ is the adjacency matrix defining the similarity between strains i and j. Because the multi-locus framework induces a symmetric (undirected) adjacency matrix, we can use an automorphism partition to decompose the vertex set into disjoint structural equivalence classes called orbits. Let $\overline{\Omega}_G$ be the adjacency matrix of the graph $G = (V, E)$ with vertex V and edge E, and let $\mathcal{G} = Aut(G)$ be the automorphism group. Then, vertices in the same group orbit are structurally equivalent and are indistinguishable (i.e., vertices can be permuted without changing the adjacency structure). Hence, for every vertex $v \in V$, v belongs to the orbit $\Delta(v) = \{g \circ v \in V : g \in \mathcal{G}\}$. Using this notion, we can reduce the network into a coarser mesoscale network known as a quotient network, Q. Let $\Delta(v) = \{\Delta(v_1), \Delta(v_2), \ldots, \Delta(v_s)\}$ be the automorphism partition of the network G under the action of $\mathcal{G}$. Then, the adjacency matrix of Q is given by the constant q_{ij}, where q_{ij} is the number of edges starting from a vertex in $\Delta(v_i)$ and ending in vertices in $\Delta(v_j)$.

Using the quotient network framework and quotient of the adjacency matrix $\overline{\Omega}$, $\overline{\Omega}$ can be reduced to its quotient $Q_{\overline{\Omega}}$ with entry elements q_{ij} (see [62–64,49,65,66,36] for similar discussion). Then, the resulting reduced dynamic equations are given as follows:

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)s_i(t,a)=-s_i\left(\lambda_i+\sum_j q_{ij}\lambda_j\right), \tag{24}$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)v_i(t,a)=s_i\sum_j q_{ij}\lambda_j-v_i\lambda_i, \tag{25}$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)y_i(t,a)=\left[s_i+(1-\gamma(a))v_i\right]\lambda_i-\sigma(a)y_i. \tag{26}$$

To tie the above ideas to the existence of synchronous equilibrium clusters, we use the concept of minimal balanced colouring [49,65,66]. Here colouring is used in a manner different from the definition used in the graph-theoretical context, whereby the graph-theoretical notion of colouring refers to colouring of a graph as an assignment of colours to the vertices (i.e., one colour to each vertex so adjacent vertices have different colours). In our context, two adjacent vertices can have the same colour as long as their dynamics or functions are given by the same dynamic or functional equations up to a permutation of the relevant variables and/or provided the vertices are isomorphic. This is equivalent to vertices belonging to the same orbits. We provide the following definitions.

Definition 5.2. 1. A colouring of the vertices is balanced if each vertex of colour i gets the same number of inputs from the vertices of colour j, for all i, j.
2. A minimal balanced colouring is a balanced colouring with minimal number of colours.
3. A quotient network with k-colouring is balanced if the colouring is balanced.

The existence of synchronous equilibrium clusters yields a decomposition of the network into its support-disjoint subsets of the vertices as defined in the previous section. If this decomposition is invariant under the action of the vector-field, then the corresponding subspace is invariant, and the system can be reduced to synchronous equilibrium clusters. We note that if the parent network G has a trivial automorphism group (completely asymmetric), then the network is completely heterogeneous and all the vertices play a unique structural role. However, if the parent network admits non-trivial automorphism groups (i.e., it is symmetric), then the network can be reduced. In addition, it should be noted that quotient networks are heterogeneous because they contain only non-equivalent elements, and all vertices play different roles.

In Fig. 1, we show possible configurations of a persistent set for a 4-strain system on a square lattice with $\mathbb{D}_4$ symmetry, where either 1, 2, 3, or 4 strains can persist and the notion of a persistent set falls into one of the two notions of endemicity defined previously. It should be noted that it is impossible to study all balanced k-colouring for all possible multiple strain systems (even in the case of a multi-locus–allele system where the structure of the strain space is well defined and the symmetry properties are classical). Instead of studying all possible k-colouring, we focus on the case where we can interchange the notion of endemicity and persistent set. That is, if we use for example a 2-colouring framework where one colour presents a persistent synchronous equilibrium cluster and the other colour denotes extinct clusters, then the endemic equilibrium is weakly endemic and the persistent synchronous equilibrium cluster can still synchronize into sub-clusters commensurable with strain-specific reproductive fitness. In the strongly endemic case, all synchronous equilibrium clusters are persistent and the nature of synchronicity is more complex. In the case where all strains approach the equilibrium point, the synchronous clustering behaviours are simple, in that all strains synchronize into sub-clusters commensurable to their reproductive successes. However, in the case where the dynamic behaviours are cyclical, the synchronous behaviours of the synchronous equilibrium clusters are far more complex.

In the case of a multi-locus system on hypergraphs, the quotient network is bi-nodal (see an example for the 2-locus–2-allele system in Fig. 1). While using the definition of a balanced colour, the quotient network of the hypercube (Mendelian graph) generated by a multi-locus–allele strain space is balanced if $q_{11}+q_{12}=q_{21}+q_{22}$. Using Fig. 1, we note that all the representations in red panels are unbalanced, and only those in yellow and green are balanced, see Fig. 1(b) where the q-values are shown. Using the notion of a 2-colouring and the bi-nodal structure of the quotient network for the strain space of interest, the dynamic equations further reduce to,

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)s_i(t,a)=-s_i\left[(1+q_{ii})\lambda_i+q_{ij}\lambda_j\right], \tag{27}$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)v_i(t,a)=s_i\left[q_{ii}\lambda_i+q_{ij}\lambda_j\right]-v_i\lambda_i, \tag{28}$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)y_i(t,a)=\left[s_i+(1-\gamma(a))v_i\right]\lambda_i-\sigma(a)y_i, \tag{29}$$

for $i\neq j$, and $i,j=1,2$. The notion of strong endemicity is reduced to $y_i(a)=y_i^*(a)=y^*(a)$ while we have $y_i(a)=y_i^*(a)=y^*(a)$ and $y_j(a)=y_j^*(a)=0$ for all $i\neq j$ for weak endemicity. Using this reduction of the model to a bi-nodal quotient network, we can now proceed with the stability analysis of the equilibria under a 2-colouring framework where we denote the potential persistent strain set with a black vertex (•) while the extinct strain set is represented by a white vertex (○), in particular, synchronous equilibrium clusters in Fig. 1.

5.2. Linearized equation

We now consider the stability of both the disease-free and the endemic equilibria. To study the linear stability, we consider small perturbations from the equilibrium solutions and let $s_i(t,a)=s_i^*(a)+\overline{s}_i(t,a)$, $v_i(t,a)=v_i^*(a)+\overline{v}_i(t,a)$, and $y_i(t,a)=y_i^*(a)+\overline{y}_i(t,a)$. Then Eqs. (27)–(29) become

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)\overline{s}_i(t,a)=-\left(\overline{s}_i+s_i^*\right)\left[(1+q_{ii})\overline{\lambda}_i+q_{ij}\overline{\lambda}_j\right]-\overline{s}_i\left[(1+q_{ii})\lambda_i^*+q_{ij}\lambda_j^*\right],$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)\overline{v}_i(t,a)=\left(\overline{s}_i+s_i^*\right)\left[q_{ii}\overline{\lambda}_i+q_{ij}\overline{\lambda}_j\right]-\left(\overline{v}_i+v_i^*\right)\overline{\lambda}_i+\overline{s}_i\left[q_{ii}\lambda_i^*+q_{ij}\lambda_j^*\right]-\overline{v}_i\lambda_i^*,$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)\overline{y}_i(t,a)=\left[\left(\overline{s}_i+s_i^*\right)+(1-\gamma(a))\left(\overline{v}_i+v_i^*\right)\right]\overline{\lambda}_i-\sigma(a)\overline{y}_i+\left[\overline{s}_i+(1-\gamma(a))\overline{v}_i\right]\lambda_i^*,$$

where $\overline{s}_i(t,0)=0$, $\overline{v}_i(t,0)=0$, $\overline{y}_i(t,0)=0$, and

$$\overline{\lambda}_i(t,a)=\int_0^{a_m}\beta(a,s)\psi(s)\overline{y}_i(t,s)ds,\quad\text{and}\quad\lambda_i^*(a)=\int_0^{a_m}\beta(a,s)\psi(s)y_i^*(s)ds.$$

Now, let the linear operator $\mathcal{A}$ on $\mathcal{X}_6:=[L^1(0,a_m)]^6$ be defined as before (with bi-nodal configuration, i.e., $\mathcal{N}=2$), and $u=(u_1,u_2,u_3)^T\in\mathcal{D}(\mathcal{A})$, where $u_k=\left(u_1^{(k)},u_2^{(k)}\right)$, $k=1,2,3$, $u_j^{(1)}=\overline{s}_j$, $u_j^{(2)}=\overline{v}_j$, $u_j^{(3)}=\overline{y}_j$, $j=1,2$. Then, we can reformulate the above dynamic equations into the abstract semilinear Cauchy evolution equation,

$$\frac{du}{dt}=\mathcal{A}u+\mathcal{G}(u), \tag{30}$$

where $\mathcal{G}$ is a nonlinear operator, $\mathcal{G}:\mathcal{X}_6\to\mathcal{X}_6$ and is defined by:

$$\mathcal{G}[u](a):=\begin{pmatrix}\mathcal{G}_1[u](a)\\ \mathcal{G}_2[u](a)\\ \mathcal{G}_3[u](a)\end{pmatrix},\quad\text{and}\quad\mathcal{G}_k[u](a):=\begin{pmatrix}\mathcal{G}_1^{(k)}[u](a)\\ \mathcal{G}_2^{(k)}[u](a)\end{pmatrix},\quad k=1,2,3,$$

where

$$\begin{aligned}
\mathcal{G}_i^{(1)}[u](a) &:= -\left(u_i^{(1)}(a)+s_i^*(a)\right)\left[(1+q_{ii})\overline{\lambda}_i[u](a)+q_{ij}\overline{\lambda}_j[u](a)\right]-u_i^{(1)}(a)\left[(1+q_{ii})\lambda_i^*(a)+q_{ij}\lambda_j^*(a)\right],\\
\mathcal{G}_i^{(2)}[u](a) &:= \left(u_i^{(1)}(a)+s_i^*(a)\right)\left[q_{ii}\overline{\lambda}_i[u](a)+q_{ij}\overline{\lambda}_j[u](a)\right]-\left(u_i^{(2)}(a)+v_i^*(a)\right)\overline{\lambda}_i[u](a)-u_i^{(2)}(a)\lambda_i^*(a)\\
&\quad+u_i^{(1)}(a)\left[q_{ii}\lambda_i^*(a)+q_{ij}\lambda_j^*(a)\right],\\
\mathcal{G}_i^{(3)}[u](a) &:= \left[\left(u_i^{(1)}(a)+s_i^*(a)\right)+(1-\gamma(a))\left(u_i^{(2)}(a)+v_i^*(a)\right)\right]\overline{\lambda}_i[u](a)\\
&\quad+\left[u_i^{(1)}(a)+(1-\gamma(a))u_i^{(2)}(a)\right]\lambda_i^*(a),\quad i\neq j\ i=1,2
\end{aligned}$$

are components of the column vectors $\mathcal{G}_k,\ k=1,2,3$. The operator $\overline{\lambda}_j[\phi](a):\mathcal{X}_6\to L^1(0,a_m)$ is a linear operator:

$$\overline{\lambda}_j[u](a)=\int_0^{a_m}\beta(a,s)\psi(s)u_j^{(3)}(s)ds,\quad j=1,2,$$

and $\overline{\lambda}[u]=\left(\overline{\lambda}_1[u],\overline{\lambda}_2[u]\right)^T$. In addition, we note that, using the Riesz–Fréchet–Kolmogorov Compactness Theorem (see [55, pg. 111, Theorem 4.26]) and Assumption 4.1, we can show that the linear operator $\overline{\lambda}[u]$ is bounded and compact. Also, since $\mathcal{F}[\phi](a)$ in Eq. (7) is differentiable, the nonlinear operator $\mathcal{G}[u]$ has a Fréchet derivative at 0 given by $\mathcal{B}$, where

$$\mathcal{B}[u](a):=\begin{pmatrix}\mathcal{B}_1[u](a)\\ \mathcal{B}_2[u](a)\\ \mathcal{B}_3[u](a)\end{pmatrix},\quad\text{and}\quad \mathcal{B}_k[u](a):=\begin{pmatrix}\mathcal{B}_1^{(k)}[u](a)\\ \mathcal{B}_2^{(k)}[u](a)\end{pmatrix},\quad k=1,2,3$$

and

$$\begin{aligned}
\mathcal{B}_i^{(1)}[u](a) &:= -s_i^*(a)\left[(1+q_{ii})\overline{\lambda}_i[u](a)+q_{ij}\overline{\lambda}_j[u](a)\right]-u_i^{(1)}(a)\left[(1+q_{ii})\lambda_i^*(a)+q_{ij}\lambda_j^*(a)\right],\\
\mathcal{B}_i^{(2)}[u](a) &:= s_i^*(a)\left[q_{ii}\overline{\lambda}_i[u](a)+q_{ij}\overline{\lambda}_j[u](a)\right]-v_i^*(a)\overline{\lambda}_i[u](a)+u_i^{(1)}(a)\left[q_{ii}\lambda_i^*(a)+q_{ij}\lambda_j^*(a)\right]-u_i^{(2)}(a)\lambda_i^*(a),\\
\mathcal{B}_i^{(3)}[u](a) &:= \left[s_i^*(a)+(1-\gamma(a))v_i^*(a)\right]\overline{\lambda}_i[u](a)+\left[u_i^{(1)}(a)+(1-\gamma(a))u_i^{(2)}(a)\right]\lambda_i^*(a),
\end{aligned}$$

for $i=1,2,\ i\neq j$ and if $i=1$, then $j=2$ and vice versa. The linearized equation is

$$\frac{du}{dt}=\mathcal{A}u+\mathcal{B}u. \tag{31}$$

In the remainder of the section, we follow the standard approach of [60,50,19]. But first, we make the following definitions [53].

Definition 5.3. Let $\mathcal{L}:D(\mathcal{L})\subset\mathcal{Y}\to\mathcal{Y}$ be a closed linear operator. Then,

1. Spectrum of $\mathcal{L}$: $\Sigma(\mathcal{L}):=\{\zeta\in\mathbb{C}:\zeta-\mathcal{L}\text{ is not bijective}\}$,
2. Resolvent set of $\mathcal{L}$: $\rho(\mathcal{L}):=\mathbb{C}\setminus\Sigma(\mathcal{L})$,
3. Resolvent of $\mathcal{L}$: $R(\zeta,\mathcal{L}):=(\zeta-\mathcal{L})^{-1}$ at $\zeta\in\rho(\mathcal{L})$,
4. Point spectrum of $\mathcal{L}$: $\Sigma_p(\mathcal{L}):=\{\zeta\in\mathbb{C}:\zeta-\mathcal{L}\text{ is not injective}\}$,
5. Spectral radius of $\mathcal{L}$: $r(\mathcal{L}):=\sup\{|\zeta|:\zeta\in\Sigma(\mathcal{L})\}$,
6. Spectral bound of $\mathcal{L}$: $b(\mathcal{L}):=\sup\{\mathbb{R}e\zeta:\zeta\in\Sigma(\mathcal{L})\}$.

5.3. Stability of disease-free equilibrium: disease extinction

To study the stability of the disease-free equilibrium, we set $y_i^*(a) = 0$ and $\lambda_i^*(a) = 0$ for all $i = 1, 2$ and for all $a \in [0, a_m)$. If $\lambda_i^*(a) = 0$, then $v_i^*(a) = 0$ and $s_i^*(a) = 1$ for all i and for all $a \in [0, a_m)$. Then, $\mathcal{B}$ in Eq. (31) above reduces to

$$\begin{aligned}
\mathcal{B}_i^{(1)}[u](a) &:= \left[(1+q_{ii})\overline{\lambda}_i[u](a) + q_{ij}\overline{\lambda}_j[u](a)\right], \\
\mathcal{B}_i^{(2)}[u](a) &:= \left[q_{ii}\overline{\lambda}_i[u](a) + q_{ij}\overline{\lambda}_j[u](a)\right], \\
\mathcal{B}_i^{(3)}[u](a) &:= \overline{\lambda}_i[u](a).
\end{aligned}$$

Let $\omega_0(\mathcal{A}+\mathcal{B})$ denote the growth bound of a semigroup $T_{\mathcal{A}+\mathcal{B}}(t)$, which is defined as $\omega_0 = \inf_{t>0} \frac{1}{t}\log\|T_{\mathcal{A}+\mathcal{B}}(t)\|$. We note that $b(\mathcal{A}+\mathcal{B}) \le \omega_0(\mathcal{A}+\mathcal{B})$, where $b(\mathcal{A}+\mathcal{B})$ is defined in Definition 5.3(6). We now state the following lemma:

Lemma 5.1. *Let $\{T_{\mathcal{A}+\mathcal{B}}(t)\}_{t\ge 0}$ be the C_0-semigroup generated by the perturbed operator $\mathcal{A}+\mathcal{B}$. Then $T_{\mathcal{A}+\mathcal{B}}(t)$ is eventually norm continuous; hence, the growth bound $\omega_0(\mathcal{A}+\mathcal{B})$ of the semigroup $\{T_{\mathcal{A}+\mathcal{B}}(t)\}_{t\ge 0}$ and the spectral bound $b(\mathcal{A}+\mathcal{B})$ of the generator $\mathcal{A}+\mathcal{B}$ are related by*

$$\omega_0(\mathcal{A}+\mathcal{B}) = b(\mathcal{A}+\mathcal{B}). \tag{32}$$

Proof. From the previous section, we note that $\mathcal{A}$ generates a nilpotent translation C_0-semigroup $\{T_{\mathcal{A}}(t)\}_{t\ge 0}$. Hence, it is eventually norm continuous (see [53, pg. 112]). That is, there exists a $t_0 \ge 0$ such that the map $t \to T_{\mathcal{A}}(t)$ is continuous from (t_0, ∞) to the space of bounded linear operators on Banach space $\mathcal{X}_6$. In addition, we know that the bounded linear operator $\lambda[u](a)$ is compact, so that $\mathcal{B}[u]$ is compact. Hence, $T_{\mathcal{A}+\mathcal{B}}(t)$ is eventually norm continuous (see [53, pg. 166, Proposition 3.1.14]). Then, the spectral mapping theorem of eventual norm continuity implies that the operator $\mathcal{A}+\mathcal{B}$ satisfies the spectral determined growth condition $\omega_0(\mathcal{A}+\mathcal{B}) = b(\mathcal{A}+\mathcal{B})$ (see [53, Theorem 4.3.10 and Corollary 4.3.11–4.3.12]). □

To determine the stability of the equilibrium solutions, we consider the resolvent operator $R(\zeta, \mathcal{A}+\mathcal{B}) = (\zeta\mathbb{I} - [\mathcal{A}+\mathcal{B}])^{-1}$ of $\mathcal{A}+\mathcal{B}$. That is, let $(\zeta\mathbb{I} - [\mathcal{A}+\mathcal{B}])\, w = x,\ w \in \mathcal{D}(\mathcal{A}),\ x \in \mathcal{X}_{6,+},\ \zeta \in \mathbb{C}$. Then, we have

$$\begin{aligned}
\frac{\partial w_i^{(1)}(a)}{\partial a} &= x_i^{(1)}(a) - \zeta w_i^{(1)}(a) - \left[(1+q_{ii})\overline{\lambda}_i[w](a) + q_{ij}\overline{\lambda}_j[w](a)\right], \\
\frac{\partial w_i^{(2)}(a)}{\partial a} &= x_i^{(2)}(a) - \zeta w_i^{(2)}(a) + \left[q_{ii}\overline{\lambda}_i[w](a) + q_{ij}\overline{\lambda}_j[w](a)\right], \\
\frac{\partial w_i^{(3)}(a)}{\partial a} &= x_i^{(3)}(a) - \zeta w_i^{(3)}(a) + \overline{\lambda}_i[w](a) - \sigma(a) w_i^{(3)}(a).
\end{aligned}$$

Integrating the above equations and making the relevant substitutions (i.e., $s_i^*(a) = 1$ for all $i = 1, 2$ and for all $a \in [0, a_m)$), we obtain

$$\begin{aligned}
w_i^{(1)}(a) &= \int_0^a e^{-\int_b^a \zeta d\tau} \left(x_i^{(1)}(b) - \left[(1+q_{ii})\overline{\lambda}_i[w](b) + q_{ij}\overline{\lambda}_j[w](b)\right]\right) db \\
w_i^{(2)}(a) &= \int_0^a e^{-\int_b^a \zeta d\tau} \left(x_i^{(2)}(b) + \left[q_{ii}\overline{\lambda}_i[w](b) + q_{ij}\overline{\lambda}_j[w](b)\right]\right) db \\
w_i^{(3)}(a) &= \int_0^a e^{-\int_b^a \zeta + \sigma(\tau) d\tau} \left(x_i^{(3)}(b) + \overline{\lambda}_i[w](b)\right) db.
\end{aligned}$$

Substituting $w_i^{(3)}$ into $\overline{\lambda}_i[w](a)$, we have

$$\begin{aligned}\overline{\lambda}_i[w](a) &= \int_0^{a_m} \beta(a,s)\psi(s)w_i^{(3)}(s)ds \\ &= \int_0^{a_m} \beta(a,s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\left(x_i^{(3)}(b)+\overline{\lambda}_i[w](b)\right)db\right)ds \\ &= \mathcal{T}_\zeta[x_i^{(3)}](a)+\mathcal{T}_\zeta[\overline{\lambda}_i](a),\end{aligned}$$

where the linear operator $\mathcal{T}_\zeta$ on $L^1_+(0,a_m)$ is defined as:

$$\mathcal{T}_\zeta[\phi_i](a)=\int_0^{a_m}\left(\int_s^{a_m}\beta(a,h)\psi(h)\Theta(s,h)e^{-\zeta(h-s)}dh\right)\phi_i(s)ds, \quad \text{and} \quad \Theta(s,h)=e^{-\int_s^h\sigma(\tau)d\tau},$$

so $\mathcal{T}_0[\phi]=\mathcal{T}_\zeta[\phi]|_{\zeta=0}$. Then,

$$\overline{\lambda}_i[w](a)=(\mathbb{I}-\mathcal{T}_\zeta)^{-1}\mathcal{T}_\zeta[x_i^{(3)}](a), \tag{33}$$

provided that the inverse exists. In addition, we note that substituting $\overline{\lambda}_i[w](b)$ into $w_i^{(2)}(a)$ and $w_i^{(3)}(a)$, the equations can be expressed in terms of only $x_i^{(k)}(a),\ k=1,2,3$. Let

$$\mathcal{V}_\sigma^\zeta[\phi_i](a)=\int_0^a e^{-\int_b^a\zeta+\sigma(\tau)d\tau}\phi_i(b)db, \quad \text{and} \quad \mathcal{J}_\sigma^\zeta[\phi_i](a)=\int_0^a e^{-\int_b^a\zeta+\sigma(\tau)d\tau}(\mathbb{I}-\mathcal{T}_\zeta)^{-1}\mathcal{T}_\zeta[\phi_i](b)db.$$

Therefore, $w_i^{(1)}(a)$, $w_i^{(2)}(a)$ and $w_i^{(3)}(a)$ can be rewritten as

$$w_i^{(1)}(a)=\mathcal{V}_0^\zeta[x_i^{(1)}](a)-\left[(1+q_{ii})\mathcal{J}_0^\zeta[x_i^{(3)}](a)+q_{ij}\mathcal{J}_0^\zeta[x_j^{(3)}](a)\right], \tag{34}$$

$$w_i^{(2)}(a)=\mathcal{V}_0^\zeta[x_i^{(2)}](a)+\left[q_{ii}\mathcal{J}_0^\zeta[x_i^{(3)}](a)+q_{ij}\mathcal{J}_0^\zeta[x_j^{(3)}](a)\right], \tag{35}$$

$$w_i^{(3)}(a)=\mathcal{V}_\sigma^\zeta[x_i^{(3)}](a)+\mathcal{J}_\sigma^\zeta[x_i^{(3)}](a). \tag{36}$$

Lemma 5.2. *The linear operator $\mathcal{T}_\zeta$ is bounded for all ζ, positive and non-supporting for real ζ, and compact for all ζ, while the operators $\mathcal{V}_\sigma^\zeta$ and $\mathcal{J}_\sigma^\zeta$ (for $\zeta\in\mathbb{C}\setminus\mathbb{E}$) are compact, where $\mathbb{E}:=\{\zeta\in\mathbb{C},\ 1\in\Sigma(\mathcal{T}_\zeta)\}$.*

Proof. From $\mathcal{H}_7$ and $\mathcal{H}_8$ in Assumption 4.1, the proof of boundedness, positivity, non-supporting and compactness of $\mathcal{T}_\zeta$ follows directly from similar arguments in Theorems 4.3 and 4.2 via the Riesz–Fréchet–Kolmogorov Compactness Theorem (see [55, pg. 111, Theorem 4.26]). Similarly, we note that $\mathcal{V}_\sigma^\zeta$ is a Volterra operator with complex-valued (or real-valued) continuous exponential kernel, so it is compact. It follows that if $\zeta\in\mathbb{C}\setminus\mathbb{E}$, then $(\mathbb{I}-\mathcal{T}_\zeta)^{-1}$ exists and is bounded. So $(\mathbb{I}-\mathcal{T}_\zeta)^{-1}\mathcal{T}_\zeta$ is compact. We also observe that $\mathcal{J}_\sigma^\zeta$ is a composition of a compact operator $\mathcal{V}_\sigma^\zeta$ with a bounded operator $(\mathbb{I}-\mathcal{T}_\zeta)^{-1}\mathcal{T}_\zeta$. Therefore, $\mathcal{J}_\sigma^\zeta$ is compact. □

Lemma 5.3. *For $\zeta\in\rho(\mathcal{A}+\mathcal{B})=\mathbb{C}\setminus\Sigma(\mathcal{A}+\mathcal{B})$, the resolvent $R(\zeta,\mathcal{A}+\mathcal{B}):\mathcal{X}_6\to\mathcal{X}_6$ is of the form*

$$R(\zeta,\mathcal{A}+\mathcal{B})x=\begin{pmatrix}\mathbb{I}_{2\times2}\mathcal{V}_0^\zeta & 0_{2\times2} & -(\mathbb{I}_{2\times2}+Q)\mathcal{J}_0^\zeta \\ 0_{2\times2} & \mathbb{I}_{2\times2}\mathcal{V}_0^\zeta & Q\mathcal{J}_0^\zeta \\ 0_{2\times2} & 0_{2\times2} & \mathbb{I}_{2\times2}(\mathcal{V}_\sigma^\zeta+\mathcal{J}_\sigma^\zeta)\end{pmatrix}x,$$

where $Q=\begin{pmatrix}q_{11} & q_{12}\\ q_{21} & q_{22}\end{pmatrix}$, $\mathbb{I}_{2\times2}$ is a 2×2 identity matrix, and $0_{2\times2}$ is a 2×2 zero matrix. Also,

$$\mathcal{V}_\sigma^\zeta[\phi_i]+\mathcal{J}_\sigma^\zeta[\phi_i]=\int_0^a e^{-\int_b^a\zeta+\sigma(\tau)d\tau}(\mathbb{I}-\mathcal{T}_\zeta)^{-1}\phi_i(b)db. \tag{37}$$

In addition, the resolvent is compact, and

$$\Sigma(\mathcal{A}+\mathcal{B}) = \Sigma_p(\mathcal{A}+\mathcal{B}) = \{\zeta \in \mathbb{C} : 1 \in \Sigma_p(\mathcal{T}_\zeta)\} = \mathbb{E}, \tag{38}$$

where $\Sigma(\mathcal{A}+\mathcal{B})$ and $\Sigma_p(\mathcal{A}+\mathcal{B})$ are spectrum and point spectrum of $\mathcal{A}+\mathcal{B}$, respectively.

Proof. Let $(\zeta - [\mathcal{A}+\mathcal{B}])\, w = x,\ w \in \mathcal{D}(\mathcal{A}),\ x \in \mathcal{X}_6,\ \zeta \in \mathbb{C}$. From our previous calculations, we obtain the expressions for $w_i^{(k)}(a)$ in Eqs. (34)–(36), which are uniquely determined if $1 \in \rho(\mathcal{T}_\zeta)$. We note that the expression for $w_i^{(3)}$ can be simplified. That is,

$$\begin{aligned} w_i^{(3)}(a) &= \int_0^a e^{-\int_b^a \zeta+\sigma(\tau)d\tau} x_i^{(3)}(b)db + \int_0^a e^{-\int_b^a \zeta+\sigma(\tau)d\tau} (\mathbb{I}-\mathcal{T}_\zeta)^{-1}\mathcal{T}_\zeta[x_i^{(3)}](b)db \\ &= \int_0^a e^{-\int_b^a \zeta+\sigma(\tau)d\tau} (\mathbb{I}-\mathcal{T}_\zeta)^{-1}(\mathbb{I}-\mathcal{T}_\zeta)\, x_i^{(3)}(b) + (\mathbb{I}-\mathcal{T}_\zeta)^{-1}\mathcal{T}_\zeta[x_i^{(3)}](b)db \\ &= \int_0^a e^{-\int_b^a \zeta+\sigma(\tau)d\tau} (\mathbb{I}-\mathcal{T}_\zeta)^{-1} x_i^{(3)}(b)db = \mathcal{V}_\sigma^\zeta[x_i^{(3)}](a) + \mathcal{J}_\sigma^\zeta[x_i^{(3)}](a). \end{aligned}$$

Hence, putting the above expressions into matrix notation, we obtain the expression for the resolvent $R(\zeta, \mathcal{A}+\mathcal{B})x$. Since $\mathcal{J}_\sigma^\zeta$ and $\mathcal{V}_\sigma^\zeta$ are compact from Lemma 5.2 and $R(\zeta, \mathcal{A}+\mathcal{B})x$ consists of components with linear combinations of an integral operator with continuous kernel and a compact operator, then $R(\zeta, \mathcal{A}+\mathcal{B})x$ is compact on $\mathcal{X}_6$. Hence, $\mathcal{A}+\mathcal{B}$ has a compact resolvent. Because the spectra of operators with compact resolvent consist of isolated eigenvalues with finite multiplicities, $\Sigma(\mathcal{A}+\mathcal{B}) = \Sigma_p(\mathcal{A}+\mathcal{B})$, and it follows that $\mathbb{C}\setminus\mathbb{E} \subset \rho(\mathcal{A}+\mathcal{B})$ (see Ch. 7–8 of [67]; Theorem 6.29 of [68, pg. 187]). That is, $\Sigma(\mathcal{A}+\mathcal{B}) = \Sigma_p(\mathcal{A}+\mathcal{B}) \subset \mathbb{E}$ (see [68, Theorem 6.29]). Furthermore, from the compactness of $\mathcal{T}_\zeta$, we know that $\Sigma(\mathcal{T}_\zeta)\setminus\{0\} = \Sigma_p(\mathcal{T}_\zeta)\setminus\{0\}$, and for $\zeta \in \mathbb{E}$, there exists an eigenfunction η_ζ such that $\mathcal{T}_\zeta\eta_\zeta = \eta_\zeta$. Then it follows that,

$$\begin{aligned} \phi_i^{(1)}(a) &= -\int_0^a e^{-\int_b^a \zeta d\tau} \left[(1+q_{ii})\eta_{i,\zeta} + q_{ij}\eta_{j,\zeta}\right](b)db \\ \phi_i^{(2)}(a) &= \int_0^a e^{-\int_b^a \zeta d\tau} \left[q_{ii}\eta_{i,\zeta} + q_{ij}\eta_{j,\zeta}\right](b)db \\ \phi_i^{(3)}(a) &= \int_0^a e^{-\int_b^a \zeta+\sigma(\tau)d\tau} \eta_{i,\zeta}(b)db, \quad i,j = 1,2,\ i \neq j, \end{aligned}$$

are the components of an eigenvector of $\mathcal{A}+\mathcal{B}$ corresponding to the eigenvalue ζ. Then, $\mathbb{E} \subset \Sigma(\mathcal{A}+\mathcal{B}) = \Sigma_p(\mathcal{A}+\mathcal{B})$. Hence, $\Sigma(\mathcal{A}+\mathcal{B}) = \Sigma_p(\mathcal{A}+\mathcal{B}) = \mathbb{E}$. □

We now consider the asymptotic properties of the spectral radius $r(\mathcal{T}_\zeta)$.

Lemma 5.4. *The following limiting properties hold:*

1. $\lim_{\zeta\to-\infty} r(\mathcal{T}_\zeta) = \infty$,
2. $\lim_{\zeta\to\infty} r(\mathcal{T}_\zeta) = 0$.

Proof. Here, we follow the approach of Inaba [60]. Let $f_\zeta \in \left[L_+^1(0,a_m)\right]^*$ be the strictly positive linear functional on $L^1(0,a_m)$ defined as

$$\langle f_\zeta, u\rangle = \int_0^{a_m} \left(\int_b^{a_m} \tilde{s}(\tau)\psi(\tau)\Theta(b,\tau)e^{-\zeta(\tau-b)}d\tau\right) u(b)db, \tag{39}$$

where $\int_b^{a_m} \psi(s)\Theta(b,s)ds > 0$ for all $b \in (0,a_m)$, $\tilde{s}(a) := \varepsilon$ if $a \in (a_m - a^*, a_m)$, $a_m \in (0,\infty)$ and $\tilde{s}(a) := 0$ otherwise, and $\beta(a,s) \geq \tilde{s}(s)$ for all $(a,s) \in (0,a_m)\times(0,a_m)$ (see $\mathcal{H}_8$ in Assumption 4.1). In addition, we

observe that $\Theta(b,\tau) = e^{-\int_b^\tau \sigma(s)ds} \geq e^{-\tilde{\sigma}a_m} := c$, for $a_m \geq \tau \geq b \geq 0$. We also observe that, when $\zeta = 0$, $f_0 = f$ in Theorem 4.3. Because $\beta \geq \tilde{s}$, we note that, for $u \in L^1_+(0,a_m)$

$$\mathcal{T}_\zeta[u](a) \geq \langle f_\zeta, u\rangle e(a), \quad e(a) = 1 \text{ for all } a \in \mathbb{R}_+.$$

Since $\mathcal{T}_\zeta$ is compact (see Lemma 5.2) and non-supporting (see Lemma 5.2), $r(\mathcal{T}_\zeta)$ is an eigenvalue of $\mathcal{T}_\zeta^*$ with strictly positive eigenfunctional $F_\zeta \in \left[L^1_+(0,a_m)\right]^* \setminus \{0\}$ satisfying

$$\langle F_\zeta, \mathcal{T}_\zeta[u]\rangle \geq \langle f_\zeta, u\rangle\langle F_\zeta, e\rangle > 0.$$

Hence, $r(\mathcal{T}_\zeta)\langle F_\zeta, u\rangle \geq \langle f_\zeta, u\rangle\langle F_\zeta, e\rangle$. Suppose $u(a) = e(a) = 1$ for all $a \in [0, a_m)$, then $r(\mathcal{T}_\zeta) \geq \langle f_\zeta, e\rangle$. Therefore,

$$\begin{aligned}
\lim_{\zeta\to-\infty} r(\mathcal{T}_\zeta) &\geq \lim_{\zeta\to-\infty}\langle f_\zeta, e\rangle = \lim_{\zeta\to-\infty}\int_0^{a_m}\left(\int_b^{a_m}\tilde{s}(\tau)\psi(\tau)\Theta(b,\tau)e^{-\zeta(\tau-b)}d\tau\right)db \\
&\geq \varepsilon c\lim_{\zeta\to-\infty}\int_{a_m-a^*}^{a_m}\psi(\tau)\left(\int_0^{\tau}e^{-\zeta(\tau-b)}db\right)d\tau \ \geq \varepsilon c\lim_{\zeta\to-\infty}\int_{a_m-a^*}^{a_m}\psi(\tau)\left(\frac{1-e^{-\zeta\tau}}{\zeta}\right)d\tau \\
&\geq \varepsilon c\lim_{\zeta\to-\infty}\left(\frac{1-e^{-\zeta(a_m-a^*)}}{\zeta}\right)\int_{a_m-a^*}^{a_m}\psi(\tau)d\tau = \infty.
\end{aligned}$$

For the second part of the proof, let

$$\langle g_\zeta, u\rangle = \overline{\beta}\int_0^{a_m}\left(\int_b^{a_m}\psi(\tau)e^{-\zeta(\tau-b)}d\tau\right)u(b)db. \tag{40}$$

We note that

$$\begin{aligned}
\mathcal{T}_\zeta[u](a) &= \int_0^{a_m}\left(\int_b^{a_m}\beta(a,s)\psi(s)\Theta(b,s)e^{-\zeta(s-b)}ds\right)u(b)db \\
&\leq \int_0^{a_m}\left(\int_b^{a_m}\beta(a,s)\psi(s)e^{-\zeta(s-b)}ds\right)u(b)db \ \leq \overline{\beta}\int_0^{a_m}\left(\int_b^{a_m}\psi(s)e^{-\zeta(s-b)}ds\right)u(b)db.
\end{aligned}$$

Hence, $\mathcal{T}_\zeta[u](a) \leq \langle g_\zeta, u\rangle e(a)$, $e(a) = 1$, $a \in \mathbb{R}_+$, $\zeta \in \mathbb{R}$. So from the above calculation, we have

$$\begin{aligned}
\lim_{\zeta\to\infty} r(\mathcal{T}_\zeta) &\leq \lim_{\zeta\to\infty}\langle g_\zeta, e\rangle = \overline{\beta}\lim_{\zeta\to\infty}\int_0^{a_m}\left(\int_b^{a_m}\psi(s)e^{-\zeta(s-b)}ds\right)db \\
&= \overline{\beta}\lim_{\zeta\to\infty}\int_0^{a_m}\psi(s)\left(\int_0^{s}e^{-\zeta(s-b)}db\right)ds \\
&= \lim_{\zeta\to\infty}\overline{\beta}\int_0^{a_m}\psi(s)\left(\frac{1-e^{-\zeta s}}{\zeta}\right)ds \leq \lim_{\zeta\to\infty}\frac{\overline{\beta}}{\zeta}\int_0^{a_m}\psi(s)ds = 0. \quad \square
\end{aligned}$$

Lemma 5.5. *The spectral radius of the next-generation operator $r(\mathcal{T}_\zeta)$ is continuous and strictly decreasing as a function of $\zeta \in \mathbb{R}$. In addition, there exists $\zeta_0 \in \mathbb{R}$ such that $r(\mathcal{T}_{\zeta_0}) = 1$, and $\zeta_0 \in \mathbb{E} = \{\zeta \in \mathbb{C} : 1 \in \Sigma_p(\mathcal{T}_\zeta)\}$.*

Proof. It is a simple matter to show that $\mathcal{T}_\zeta$ is continuous. Because $r(\mathcal{T}_\zeta)$ is in the point spectrum of $\mathcal{T}_\zeta$ and not an accumulation point of $\Sigma(\mathcal{T}_\zeta)$, $r(\mathcal{T}_\zeta)$ is also continuous (see [68, Section 4.3.5]). From Lemma 5.2, we know that $\mathcal{T}_\zeta$ is nonsupporting and compact for $\zeta \in \mathbb{R}$. If $\zeta < \zeta'$, then we have $\mathcal{T}_\zeta \geq \mathcal{T}_{\zeta'}$, and $\mathcal{T}_\zeta \neq \mathcal{T}_{\zeta'}$ and $r(\mathcal{T}_\zeta) > 0$. Hence, from [57], $r(\mathcal{T}_\zeta) > r(\mathcal{T}_{\zeta'})$. Therefore, $r(\mathcal{T}_\zeta)$ is strictly decreasing for $\zeta \in \mathbb{R}$. Moreover, from Lemma 5.4 and applying the intermediate value theorem, we see that $r(\mathcal{T}_\zeta) = 1$ has a root $\zeta_0 \in \mathbb{R}$. From the compactness and nonsupporting properties of $\mathcal{T}_\zeta$, we can conclude that $1 = r(\mathcal{T}_{\zeta_0}) \in \Sigma_p(\mathcal{T}_{\zeta_0})$. Hence, $\zeta_0 \in \{\zeta \in \mathbb{C} : 1 \in \Sigma_p(\mathcal{T}_\zeta)\} = \mathbb{E}$. $\square$

Corollary 5.6. *There exists a unique $\zeta_0 \in \mathbb{R}\bigcap\mathbb{E}$ such that $r(\mathcal{T}_{\zeta_0}) = 1$ and the following implications hold:*

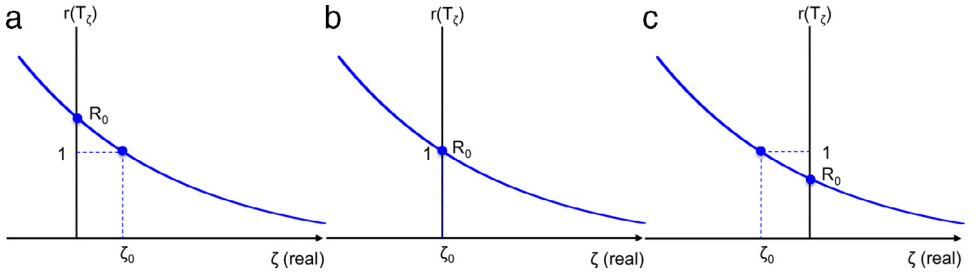

Fig. 2. Sketch for the Proof of Corollary 5.6. The threshold property of $r(\mathcal{T}_\zeta)$ as a function of ζ. For (a) $\mathcal{R}_0 > 1$, $\zeta_0 > 0$, (b) $\mathcal{R}_0 = 1$, $\zeta_0 = 0$, (c) $\mathcal{R}_0 < 1$, $\zeta_0 < 0$.

1. $\zeta_0 > 0$, *if and only if* $r(\mathcal{T}_0) = r(\mathcal{T}) > 1$,
2. $\zeta_0 = 0$, *if and only if* $r(\mathcal{T}_0) = r(\mathcal{T}) = 1$,
3. $\zeta_0 < 0$, *if and only if* $r(\mathcal{T}_0) = r(\mathcal{T}) < 1$.

Proof. The proof follows directly from Lemma 5.5 and monotonicity of $r(\mathcal{T}_\zeta)$ (see the illustration in Fig. 2). We note that when $\zeta = 0$, $\mathcal{R}_0 = r(\mathcal{T}_0)$. □

Lemma 5.7. *If there exists a* $\zeta \in \mathbb{E}$ *and* $\zeta_0 \neq \zeta$, *then* $\mathbb{R}e\zeta < \zeta_0$.

Proof. Suppose that $\zeta_0 \in \mathbb{E}$, $1 \in \Sigma_p(\mathcal{T}_\zeta)$, and $\mathcal{T}_\zeta[\phi](a) = \phi(a)$ for some $\phi \in L^1(0, a_m)$. Let $|\phi|(a) = |\phi(a)|$, then

$$|\phi|(a) = |\mathcal{T}_\zeta[\phi](a)| \leq \int_0^{a_m} \left(\int_b^{a_m} \beta(a,s)\psi(s)\Theta(b,s)|e^{-\zeta(s-b)}|ds \right) \phi(b)db$$

$$\leq \int_0^{a_m} \left(\int_b^{a_m} \beta(a,s)\psi(s)\Theta(b,s)e^{-\mathbb{R}e\zeta(s-b)}ds \right) \phi(b)db = \mathcal{T}_{\mathbb{R}e\zeta}[\phi](a), \quad \text{for all } a \in [0, a_m). \tag{41}$$

That is, $|\phi|(a) = |\mathcal{T}_\zeta\phi|(a) \leq \mathcal{T}_{\mathbb{R}e\zeta}|\phi|(a)$. Let $F_{\mathbb{R}e\zeta} \in [L^1(0, a_m)]^*$ be a strictly positive eigenfunctional of $\mathcal{T}^*_{\mathbb{R}e\zeta}$. By duality pairing, we have $r(\mathcal{T}_{\mathbb{R}e\zeta})\langle F_{\mathbb{R}e\zeta}, |\phi|\rangle \geq \langle F_{\mathbb{R}e\zeta}, |\phi|\rangle$. Since $\langle F_{\mathbb{R}e\zeta}, |\phi|\rangle > 0$, we have $r(\mathcal{T}_{\mathbb{R}e\zeta}) \geq 1$. From Lemma 5.5, $\mathbb{R}e\zeta \leq \zeta_0$. Suppose $\mathbb{R}e\zeta = \zeta_0$, then $|\phi| \leq \mathcal{T}_{\zeta_0}|\phi|$. If we assume that $|\phi| < \mathcal{T}_{\zeta_0}|\phi|$, then taking duality pairing with the eigenfunctional F_0 of $\mathcal{T}_{\zeta_0}$ corresponding to $r(\mathcal{T}_{\zeta_0}) = 1$ yields

$$\begin{aligned} \langle F_{\zeta_0}, |\phi|\rangle &< \langle F_{\zeta_0}, \mathcal{T}_{\zeta_0}|\phi|\rangle = \langle \mathcal{T}^*_{\zeta_0} F_{\zeta_0}, |\phi|\rangle \\ &= r(\mathcal{T}_{\zeta_0})\langle F_{\zeta_0}, |\phi|\rangle = \langle F_{\zeta_0}, |\phi|\rangle, \end{aligned} \tag{42}$$

which is a contradiction. Hence, $\mathcal{T}_{\zeta_0}|\phi|(a) = |\phi|(a)$. Now let $\phi_0 \in L^1_+(0, a_m)$ be a nonsupporting eigenvector of $\mathcal{T}_{\zeta_0}$ corresponding to the eigenvalue $r(\mathcal{T}_{\zeta_0}) = 1$, then $|\phi| = c\phi_0$ for some constant $c > 0$. So $\phi(a) = c\phi_0(a)e^{i\kappa(a)}$ for some real-valued function $\kappa : [0, a_m) \to \mathbb{R}$. We observe that $\mathcal{T}_{\zeta_0}\phi_0 = |\mathcal{T}_\zeta\phi|$; and substituting $\phi(a) = c\phi_0(a)e^{i\kappa(a)}$ into $\mathcal{T}_\zeta|\phi| = |\mathcal{T}_\zeta\phi|$, we have $|\mathcal{T}_\zeta\phi(a)| = |\phi|(a) = c\phi_0(a) = c\mathcal{T}_{\zeta_0}\phi_0$. That is, without loss of generality, $c = 1$,

$$\begin{aligned} \mathcal{T}_{\zeta_0}\phi_0(a) &= \int_0^{a_m} \left(\int_b^{a_m} \beta(a,s)\psi(s)\Theta(b,s)e^{-\zeta_0(s-b)}ds \right) \phi_0(b)db \\ &= \left| \int_0^{a_m} \left(\int_b^{a_m} \beta(a,s)\psi(s)\Theta(b,s)e^{-(\zeta_0 + i\mathbb{I}m\zeta)(s-b)}ds \right) \phi_0(b)e^{i\kappa(b)}db \right|. \end{aligned}$$

It follows that $f(s,b) = \beta(a,s)\psi(s)\Theta(b,s)e^{-(\zeta_0+i\mathbb{I}m\zeta)(s-b)}\phi_0(b)e^{i\kappa(b)} = f_0(s,b)e^{i\overline{\kappa}}$, for some constant real number $\overline{\kappa} = \kappa(b) - \mathbb{I}m\zeta(s-b)$ and eigenfunction f_0. Then,

$$\begin{aligned}\mathcal{T}_\zeta\phi(a) &= \int_0^{a_m}\left(\int_b^{a_m}\beta(a,s)\psi(s)\Theta(b,s)e^{-\zeta(s-b)}ds\right)\phi_0(b)e^{i\kappa(b)}db\\ &= \int_0^{a_m}\left(\int_b^{a_m}\beta(a,s)\psi(s)\Theta(b,s)e^{-\mathbb{R}e\zeta(s-b)+i\overline{\kappa}}ds\right)\phi_0(b)db = e^{i\overline{\kappa}}\mathcal{T}_{\mathbb{R}e\zeta}\phi_0(a) = e^{i\overline{\kappa}}\phi_0(a). \quad (43)\end{aligned}$$

Hence, $e^{i\kappa}\phi_0(a) = \mathcal{T}_\zeta\phi(a) = e^{i\overline{\kappa}}\phi_0(a)$. Therefore, from $\mathcal{T}_\zeta\phi = \phi$, $e^{\mathbb{I}m\zeta(s-b)} = 1$ for $0 \le b \le s \le a_m$. So $\overline{\kappa} = \kappa(a)$, which implies that $\mathbb{I}m\zeta = 0$. □

Theorem 5.8. *Let $b(\mathcal{A}+\mathcal{B})$ be the spectral bound of $\mathcal{A}+\mathcal{B}$, then $\zeta_0 = b(\mathcal{A}+\mathcal{B})$.*

Proof. From Lemma 5.3, $b(\mathcal{A}+\mathcal{B}) = \sup\{\mathbb{R}e\zeta : \zeta \in \mathbb{E}\}$, and from Lemma 5.7, we have $\zeta_0 = b(\mathcal{A}+\mathcal{B})$. Hence, the assertion holds. □

Here we define exponential stability of $\mathcal{A}+\mathcal{B}$ as follows: for any $\omega > \omega_0$, there exists M_ω such that for all $t \ge 0$, $\|T_{\mathcal{A}+\mathcal{B}}\| \le M_\omega e^{\omega t}$. So the semigroup is exponentially stable if $\omega_0 < 0$. Throughout this paper, we will use locally asymptotically stable to mean that the linearized system is stable. Using the linear and global stability notions of [19,60], we state the following stability condition.

Theorem 5.9. *If $\mathcal{R}_0 < 1$, then the disease-free equilibrium is locally asymptotically stable; and it is unstable when $\mathcal{R}_0 > 1$.*

Proof. We note that, from Lemma 5.1 and Theorem 5.8, $\zeta_0 = b(\mathcal{A}+\mathcal{B}) = \omega_0(\mathcal{A}+\mathcal{B})$. Then, by Corollary 5.6, $\mathcal{R}_0 = r(\mathcal{T}_0) < 1$ implies that $\zeta_0 < 0$. So $\omega_0(\mathcal{A}+\mathcal{B}) < 0$ and $\|T_{\mathcal{A}+\mathcal{B}}(t)\| \to 0$ as $t \to \infty$. Hence, the disease-free equilibrium is locally asymptotically stable. However, if $r(\mathcal{T}_0) > 1$, then $\zeta_0 > 0$ and $\omega_0(\mathcal{A}+\mathcal{B}) > 0$, so the disease-free equilibrium is unstable. □

Corollary 5.10. *If $\mathcal{R}_0 < 1$, then the disease-free equilibrium for initial data in Γ is globally asymptotically stable.*

Proof. We note that the original nonlinear Cauchy evolution equation is,

$$\frac{du}{dt} = \mathcal{A}u + \mathcal{F}(u). \quad (44)$$

We focus on the y component of u (i.e., $u_3 = (u_1^{(3)}, u_2^{(3)})^T$), and observe that, for $u \in \Gamma$, $u_i^{(1)} + (1-\gamma)u_i^{(2)} \le u_i^{(1)} + u_i^{(2)} \le 1$. Hence, using the same notation as in Section 3, for $u \in \Gamma$ and $i = 1,2$, $\mathcal{F}_i^3[u](a) = (u_i^{(1)} + (1-\gamma)u_i^{(2)})\lambda_i[u] \le \lambda_i[u]$ and $\mathcal{B}_i^{(3)}[u](a) = s_i^*(a)\lambda_i[u] = \lambda_i[u]$, as $s_i^*(a) = 1$. Now,

$$\frac{du_3}{dt} = \mathcal{A}_3u + \mathcal{F}_3(u) \quad (45)$$

can be integrated to obtain

$$\begin{aligned}u_3(t) &= T_3(t)u_0 + \int_0^t T_3(t-s)\mathcal{F}_3(u(s))ds\\ &\le T_3(t)u_0 + \int_0^t T_3(t-s)\mathcal{B}_3(u(s))ds = T_{\mathcal{A}_3+\mathcal{B}_3}u_o, \quad u_0 \in \Gamma. \quad (46)\end{aligned}$$

Hence, $0 \le u_3(t) \le T_{\mathcal{A}_3+\mathcal{B}_3}(t)u_0$. From Theorem 5.9, the semigroup $T_{\mathcal{A}+\mathcal{B}}$ is exponentially stable. So $T_{\mathcal{A}+\mathcal{B}}(t) \to 0$ as $t \to \infty$. Therefore, $T_{\mathcal{A}_3+\mathcal{B}_3}(t) \to 0$ as $t \to \infty$ and $u_3(t) \le T_{\mathcal{A}_3+\mathcal{B}_3}(t)u_0 \to 0$, where

$T_{\mathcal{A}+\mathcal{B}}(t) = (T_{\mathcal{A}_1+\mathcal{B}_1}(t), T_{\mathcal{A}_2+\mathcal{B}_2}(t),\ T_{\mathcal{A}_3+\mathcal{B}_3}(t))^T \to 0$. So $y_i(t) := u_i^{(3)}(t) \to 0$ as $t \to \infty$, and $\lambda_i \to 0$, $v_i(t) := u_i^{(2)}(t) \to 0$ and $u_i^{(1)} = s_i \to 1$. □

5.4. Endemic equilibria

So far, we have established that, when $\mathcal{R}_0 < 1$, the trivial (disease-free) equilibrium is the only biological relevant steady state and is globally asymptotically stable. Moreover, when $\mathcal{R}_0 > 1$, the disease-free equilibrium is unstable, and there is at least one non-trivial solution. Here we establish the stability of the non-trivial solution(s) using the different notions of endemicity. From Eqs. (27)–(29), the endemic steady state solution is given by:

$$s_1^*(a) = e^{-\int_0^a (1+q_{11})\lambda_1^*(\tau)+q_{12}\lambda_2^*(\tau)d\tau}, \qquad s_2^*(a) = e^{-\int_0^a (1+q_{22})\lambda_2^*(\tau)+q_{21}\lambda_1^*(\tau)d\tau}, \tag{47}$$

$$v_1^*(a) = e^{-\int_0^a \lambda_1^*(\tau)d\tau}\left(1 - e^{-\int_0^a q_{11}\lambda_1^*(\tau)+q_{12}\lambda_2^*(\tau)d\tau}\right), \tag{48}$$

$$v_2^*(a) = e^{-\int_0^a \lambda_2^*(\tau)d\tau}\left(1 - e^{-\int_0^a q_{22}\lambda_2^*(\tau)+q_{21}\lambda_1^*(\tau)d\tau}\right), \tag{49}$$

$$y_1^*(a) = \int_0^a e^{-\int_b^a \sigma(\tau)d\tau}\lambda_1^*(b)\{s_1^*(b) + (1-\gamma(b))v_1^*(b)\}db, \tag{50}$$

$$y_2^*(a) = \int_0^a e^{-\int_b^a \sigma(\tau)d\tau}\lambda_2^*(b)\{s_2^*(b) + (1-\gamma(b))v_2^*(b)\}db. \tag{51}$$

Let $c_i = \int_0^{a_m} \beta_2(s)\psi(s)y_i^*(s)ds,\ i = 1, 2$, then $\lambda_i^*(a) = c_i\beta_1(a)$ and let $B(a) = \int_0^a \beta_1(\tau)d\tau$. Substituting y_i^* into c_i, we have

$$c_1 = \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau}c_1\beta_1(b)q_1(b)db\right)ds = c_1H_1(c_1, c_2, \mathcal{R}_0),$$
$$c_2 = \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau}c_2\beta_1(b)q_2(b)db\right)ds = c_2H_2(c_1, c_2, \mathcal{R}_0),$$

where

$$q_1(b) = \gamma(b)e^{-B(b)[(1+q_{11})c_1+q_{12}c_2]} + (1-\gamma(b))e^{-B(b)c_1}$$
$$q_2(b) = \gamma(b)e^{-B(b)[(1+q_{22})c_2+q_{21}c_1]} + (1-\gamma(b))e^{-B(b)c_2}.$$

We note that $c_i = 0$ for $i = 1, 2$ corresponds to the disease-free equilibrium, and $c_i \neq 0$ for some i corresponds to a weakly endemic equilibrium while $c_i \neq 0$ for all i corresponds to a strongly endemic equilibrium. Moreover, we observe that $H_1(c_1, c_2, \mathcal{R}_0)$ and $H_2(c_1, c_2, \mathcal{R}_0)$ are analytic in $c = (c_1, c_2)^T$ (see [69,70]). In addition, we note that $H_1(0, 0, \mathcal{R}_0) = H_2(0, 0, \mathcal{R}_0) = \mathcal{R}_0$.

We also observe that,

$$\frac{\partial H_1(0,0,\mathcal{R}_0)}{\partial c_1} = -\int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau}\beta_1(b)B(b)\left(1+q_{11}\gamma(b)\right)db\right)ds$$
$$\frac{\partial H_1(0,0,\mathcal{R}_0)}{\partial c_2} = -\int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau}\beta_1(b)B(b)q_{12}\gamma(b)db\right)ds$$
$$\frac{\partial H_2(0,0,\mathcal{R}_0)}{\partial c_1} = -\int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau}\beta_1(b)B(b)q_{21}\gamma(b)db\right)ds$$
$$\frac{\partial H_2(0,0,\mathcal{R}_0)}{\partial c_2} = -\int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau}\beta_1(b)B(b)\left(1+q_{22}\gamma(b)\right)db\right)ds.$$

Hence, $\frac{\partial H_i(0,0,\mathcal{R}_0)}{\partial c_j} \leq 0,\ i,j=1,2$. Define the following,

$$m_0(\mathcal{R}_0) = \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau}\beta_1(b)B(b)db\right)ds, \tag{52}$$

$$m_1(\mathcal{R}_0) = \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \sigma(\tau)d\tau}\beta_1(b)B(b)\gamma(b)db\right)ds, \tag{53}$$

and let $\mathcal{M}(c_1,c_2,\mathcal{R}_0)$ be the Jacobian matrix of $\overline{H} = (H_1,H_2)^T$. Then,

$$\mathcal{M}(0,0,\mathcal{R}_0) := -\begin{pmatrix} m_0(\mathcal{R}_0)+q_{11}m_1(\mathcal{R}_0) & q_{12}m_1(\mathcal{R}_0) \\ q_{21}m_1(\mathcal{R}_0) & m_0(\mathcal{R}_0)+q_{22}m_1(\mathcal{R}_0)\end{pmatrix}. \tag{54}$$

Suppose that γ is chosen such that the determinant $\mathcal{DM}(0,0,1) = m_0(1)^2 + m_0(1)\ m_1(1)(q_{11}+q_{22}) + m_1(1)^2(q_{11}q_{22}-q_{12}q_{21})$ is nonzero. So $\mathcal{M}^{-1}(0,0,1)$ exists and is bounded. We can now apply the Implicit Function Theorem of Hildebrandt & Graves (see [71] and [72, Theorem 4.B]) to show that there exists a branching solution when $\mathcal{R}_0 > 1$ for $\mathcal{R}_0$ small or equivalently we can find a solution $c(\mathcal{R}_0) = (c_1(\mathcal{R}_0), c_2(\mathcal{R}_0))^T$ bifurcating from the trivial solution at $\mathcal{R}_0 = 1$. We have,

Theorem 5.11. *If $\overline{H}(c_1,c_2,\mathcal{R}_0) : \mathbb{R}^2_+ \times \mathbb{R}_+ \to \mathbb{R}^2_+$ satisfies*

1. $\overline{H}(0,0,1) = (1,1)^T$ *and*
2. $\mathcal{DM}(0,0,1)$ *is nonzero (i.e., $\mathcal{M}(0,0,1)$ is nonsingular),*

then there is a solution family $c(\mathcal{R}_0)$ such that $\overline{H}(c_1(\mathcal{R}_0),c_2(\mathcal{R}_0),\mathcal{R}_0) = (1,1)^T$ for all $\mathcal{R}_0 > 1$ near $\mathcal{R}_0 = 1$.

Remark 5.12. To further deduce that c_1 and c_2 are continuous functions of $\mathcal{R}_0$, requires the continuity of $H_i(c_1,c_2,\mathcal{R}_0)$ for $i=1,2$ with respect to all its arguments $(c_1,c_2,\mathcal{R}_0)$. It is an easy matter to show that $H_i(c_1,c_2,\mathcal{R}_0)$ for $i=1,2$ is continuous with respect to (c_1,c_2). However, we are not able to determine the continuity of $H_i(c_1,c_2,\mathcal{R}_0)$ with respect to $\mathcal{R}_0$ because $H_i(c_1,c_2,\mathcal{R}_0)$ is not explicitly expressed in terms of $\mathcal{R}_0$ even though we know that $H_i(0,0,\mathcal{R}_0) = \mathcal{R}_0$.

We note that near $c_i = 0,\ i=1,2$ and $\mathcal{R}_0 = 1$, the level set of $\overline{H}(c_1,c_2,\mathcal{R}_0)$ at $(c_1,c_2,\mathcal{R}_0) = (0,0,1)$ given by $\{(c_1,c_2,\mathcal{R}_0) \in \mathbb{R}^2_+ \times \mathbb{R}_+ : \overline{H}(c_1,c_2,\mathcal{R}_0) = (1,1)^T\}$ is one-dimensional. Moreover, we observe that from the notion of endemicity discussed previously, there are two possible cases (for balanced quotient):

1. Weakly endemic: $c_1 > 0$ and $c_2 = 0$ or $c_1 = 0$ and $c_2 > 0$.
2. Strongly endemic: $c_1 > 0$ and $c_2 > 0$.

We note that, for the first case, whenever one solution type exists the other also exists. However, which equilibrium is selected will depend on the initial condition, and if one of the weakly endemic equilibria is (un)stable, the other weakly endemic state will be (un)stable due to the inherent symmetry. We now state the following:

Lemma 5.13. *Let $H_i(c_1,c_2,\mathcal{R}_0),\ i=1,2$ be defined as above. If $\mathcal{R}_0 > 1$, then the following holds:*

1. *In the case of balanced quotient (i.e., $q_{11}+q_{12} = q_{22}+q_{21} = \alpha_0$), there exists a unique strongly endemic equilibrium, $c_1 = c_2 = c > 0$ such that $H(c,c,\mathcal{R}_0) = H_1(c,c,\mathcal{R}_0) = H_2(c,c,\mathcal{R}_0) = 1$.*
2. *There exist weakly endemic equilibria, $c_1 > 0$ and $c_2 = 0$, and $c_1 = 0$ and $c_2 > 0$, such $H_1(c_1,0,\mathcal{R}_0) = 1$ and $H_2(0,c_2,\mathcal{R}_0) = 1$, respectively.*

However, if $\mathcal{R}_0 < 1$, then there is no non-negative non-trivial steady state.

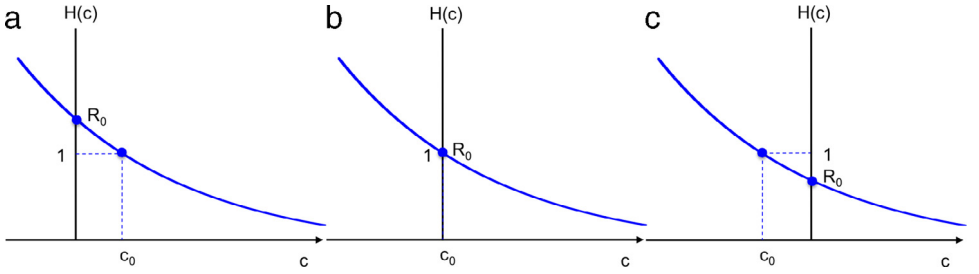

Fig. 3. Sketch for the Proof of Lemma 5.13. The threshold property of $\mathcal{R}_0$ and $H(c, c, \mathcal{R}_0)$. For (a) $\mathcal{R}_0 > 1$, $c_0 > 0$, (b) $\mathcal{R}_0 = 1$, $c_0 = 0$, (c) $\mathcal{R}_0 < 1$, $c_0 < 0$.

Proof. For the first result, we note that $c_i = c_j = c,\ i \neq j$ for balanced quotient since $c_i,\ i = 1, 2$ is commensurable to the reproductive fitness and all strains have the same epidemiological parameters. To find c such that $H(c, c, \mathcal{R}_0) = H_1(c, c, \mathcal{R}_0) = H_2(c, c, \mathcal{R}_0) = 1$, we observe that $H(c, c, \mathcal{R}_0)$ is differentiable with respect to c and $\frac{\partial H(c,c,\mathcal{R}_0)}{\partial c} < 0$ for all $c > 0$. Hence, $H(c, c, \mathcal{R}_0)$ is strictly monotonically decreasing, and it also satisfies the following limiting behaviours,

1. $\lim_{c\to+\infty} H(c, c, \mathcal{R}_0) = 0$,
2. $\lim_{c\to-\infty} H(c, c, \mathcal{R}_0) = +\infty$.

Using similar arguments as in Corollary 5.6 and the Intermediate Value Theorem, the result follows (see Fig. 3). Hence, when $\mathcal{R}_0 > 1$, there is a unique solution for $H(c, c, \mathcal{R}_0) = 1$, and c is positive. When $\mathcal{R}_0 < 1$, there is no solution on the right-hand half plane. The characteristic equation $H(c, c, \mathcal{R}_0) = 1$ has a negative solution which is biologically irrelevant. For the second part, we observe that $\frac{\partial H_1(c_1,0,\mathcal{R}_0)}{\partial c_1} < 0$ and $\frac{\partial H_2(0,c_2,\mathcal{R}_0)}{\partial c_2} < 0$ hold for all c_1 and c_2, respectively, and $H_1(c_1, 0, \mathcal{R}_0)$ and $H_2(0, c_2, \mathcal{R}_0)$ are both monotonically decreasing functions. Hence, the results follow from the Intermediate Value theorem. □

In the next sections, we investigate the asymptotic behaviours of these equilibria. We will assume $\mathcal{R}_0 > 1$ but sufficiently close to 1 in the remainder of this section.

5.4.1. Stability of weakly endemic equilibria: Principle of competitive exclusion

To study the stability of an endemic equilibrium, we start with the linearized equation as in the previous section, and consider the resolvent of $\mathcal{A} + \mathcal{B}$ (i.e. consider the equation $(\zeta\mathbb{I} - (\mathcal{A} + \mathcal{B}))\, w = x$). Then, we have for $i, j = 1, 2,\ i \neq j$,

$$\begin{aligned}
\frac{\partial w_i^{(1)}(a)}{\partial a} &= x_i^{(1)}(a) - \zeta w_i^{(1)}(a) + \mathcal{B}_i^{(1)}[w](a),\\
\frac{\partial w_i^{(2)}(a)}{\partial a} &= x_i^{(2)}(a) - \zeta w_i^{(2)}(a) + \mathcal{B}_i^{(2)}[w](a),\\
\frac{\partial w_i^{(3)}(a)}{\partial a} &= x_i^{(3)}(a) - \zeta w_i^{(3)}(a) + \mathcal{B}_i^{(3)}[w](a) - \sigma(a) w_i^{(3)}(a), \quad i = 1, 2
\end{aligned}$$

where $\mathcal{B}_i^{(k)}[w](a) = \overline{\mathcal{B}}_i^{(k)}[w](a) + \tilde{\mathcal{B}}_i^{(k)}[w](a),\ k = 1, 2, 3$, and

$$\begin{aligned}
\overline{\mathcal{B}}_i^{(1)}[w](a) &:= -s_i^*(a)\left[(1 + q_{ii})\overline{\lambda}_i[w](a) + q_{ij}\overline{\lambda}_j[w](a)\right],\\
\tilde{\mathcal{B}}_i^{(1)}[w](a) &:= -w_i^{(1)}(a)\left[(1 + q_{ii})\lambda_i^* + q_{ij}\lambda_j^*\right](a),\\
\overline{\mathcal{B}}_i^{(2)}[w](a) &:= s_i^*(a)\left[q_{ii}\overline{\lambda}_i[w](a) + q_{ij}\overline{\lambda}_j[w](a)\right] - v_i^*(a)\overline{\lambda}_i[w](a),
\end{aligned}$$

$$\tilde{\mathcal{B}}_i^{(2)}[w](a) := w_i^{(1)}\left[q_{ii}\lambda_i^* + q_{ij}\lambda_j^*\right](a) - w_i^{(2)}(a)\lambda_i^*(a),$$
$$\overline{\mathcal{B}}_i^{(3)}[w](a) := \left[s_i^*(a) + (1-\gamma(a))v_i^*(a)\right]\overline{\lambda}_i[w](a),$$
$$\tilde{\mathcal{B}}_i^{(3)}[w](a) := \left[w_i^{(1)}(a) + (1-\gamma(a))w_i^{(2)}(a)\right]\lambda_i^*(a),$$

for $i,j=1,2,\ i\neq j$ and $w_i^{(k)}(0)=0,\ k=1,2,3$. Here $w=(w_1,w_2,w_3)^T$, $w_k=\left(w_1^{(k)},w_2^{(k)}\right)$, for $k=1,2,3$, and $w_j^{(1)}=\overline{s}_j,\ w_j^{(2)}=\overline{v}_j$ and $w_j^{(3)}=\overline{y}_j$. Let $c_i=\int_0^{a_m}\beta_2(s)\psi(s)y_i^*(s)ds$, $\overline{c}_i=\int_0^{a_m}\beta_2(s)\psi(s)w_i^{(3)}(s)ds$ and $B(a)=\int_0^a\beta_1(\tau)d\tau$. Then $\lambda_i^*(a)=c_i\beta_1(a)$ and $\overline{\lambda}_i[w](a)=\overline{c}_i\beta_1(a)$. To study the stability of the weakly endemic equilibrium, we use the definition of weak endemicity and either set (i) $c_1\neq 0$ and $c_2=0$ or (ii) $c_1=0$ and $c_2\neq 0$. Here we focus on (i), for the stability of (ii) follows directly from (i) and the analyses of the two cases are identical. If $c_1\neq 0$ and $c_2=0$, substituting the relevant terms in the above expression, we obtain the following:

$$\frac{\partial w_1^{(1)}(a)}{\partial a} = x_1^{(1)}(a) - \zeta w_1^{(1)}(a) - s_1^*(a)\left[(1+q_{11})\overline{\lambda}_1[w](a) + q_{12}\overline{\lambda}_2[w](a)\right] - w_1^{(1)}(a)(1+q_{11})\lambda_1^*(a),$$
$$\frac{\partial w_2^{(1)}(a)}{\partial a} = x_2^{(1)}(a) - \zeta w_2^{(1)}(a) - s_2^*(a)\left[(1+q_{22})\overline{\lambda}_2[w](a) + q_{21}\overline{\lambda}_1[w](a)\right] - w_2^{(1)}(a)q_{21}\lambda_1^*(a),$$
$$\begin{aligned}\frac{\partial w_1^{(2)}(a)}{\partial a} &= x_1^{(2)}(a) - \zeta w_1^{(2)}(a) + s_1^*(a)\left[q_{11}\overline{\lambda}_1[w](a) + q_{12}\overline{\lambda}_2[w](a)\right] - v_1^*(a)\overline{\lambda}_1[w](a)\\ &\quad + w_1^{(1)}(a)\left[q_{11}\lambda_1^*\right](a) - w_1^{(2)}(a)\lambda_1^*(a),\end{aligned}$$
$$\frac{\partial w_2^{(2)}(a)}{\partial a} = x_2^{(2)}(a) - \zeta w_2^{(2)}(a) + s_2^*(a)\left[q_{22}\overline{\lambda}_2[w](a) + q_{21}\overline{\lambda}_1[w](a)\right] - v_2^*(a)\overline{\lambda}_2[w](a) + w_2^{(1)}(a)q_{21}\lambda_1^*(a),$$
$$\begin{aligned}\frac{\partial w_1^{(3)}(a)}{\partial a} &= x_1^{(3)}(a) - \zeta w_1^{(3)}(a) + \left[s_1^*(a) + (1-\gamma(a))v_1^*(a)\right]\overline{\lambda}_1[w](a)\\ &\quad + \left[w_1^{(1)}(a) + (1-\gamma(a))w_1^{(2)}(a)\right]\lambda_1^*(a) - \sigma(a)w_1^{(3)}(a),\end{aligned}$$
$$\frac{\partial w_2^{(3)}(a)}{\partial a} = x_2^{(3)}(a) - \zeta w_2^{(3)}(a) + \left[s_2^*(a) + (1-\gamma(a))v_2^*(a)\right]\overline{\lambda}_2[w](a) - \sigma(a)w_2^{(3)}(a).$$

Integrating the above equations, we obtain

$$w_1^{(1)}(a) = \int_0^a e^{-\int_b^a \zeta + (1+q_{11})\lambda_1^*(\tau)d\tau}\left(x_1^{(1)}(b) - s_1^*(b)\left[(1+q_{11})\overline{\lambda}_1[w](b) + q_{12}\overline{\lambda}_2[w](b)\right]\right)db, \tag{55}$$
$$w_2^{(1)}(a) = \int_0^a e^{-\int_b^a \zeta + q_{21}\lambda_1^*(\tau)d\tau}\left(x_2^{(1)}(b) - s_2^*(b)\left[(1+q_{22})\overline{\lambda}_2[w](b) + q_{21}\overline{\lambda}_1[w](b)\right]\right)db, \tag{56}$$
$$\begin{aligned}w_1^{(2)}(a) = \int_0^a e^{-\int_b^a \zeta + \lambda_1^*(\tau)d\tau}&\left(x_1^{(2)}(b) + s_1^*(b)\left[q_{11}\overline{\lambda}_1[w](b) + q_{12}\overline{\lambda}_2[w](b)\right]\right.\\ &\left. - v_1^*(b)\overline{\lambda}_1[w](b) + w_1^{(1)}(b)q_{11}\lambda_1^*(b)\right)db,\end{aligned} \tag{57}$$
$$\begin{aligned}w_2^{(2)}(a) = \int_0^a e^{-\int_b^a \zeta d\tau}&\left(x_2^{(2)}(b) + s_2^*(b)\left[q_{22}\overline{\lambda}_2[w](b) + q_{21}\overline{\lambda}_1[w](b)\right]\right.\\ &\left. - v_2^*(b)\overline{\lambda}_2[w](b) + w_2^{(1)}(b)q_{21}\lambda_1^*(b)\right)db,\end{aligned} \tag{58}$$
$$\begin{aligned}w_1^{(3)}(a) = \int_0^a e^{-\int_b^a \zeta + \sigma(\tau)d\tau}&\left(x_1^{(3)}(b) + \left[s_1^*(b) + (1-\gamma(a))v_1^*(b)\right]\overline{\lambda}_1[w](b)\right.\\ &\left. + \left[w_1^{(1)}(b) + (1-\gamma(a))w_1^{(2)}(b)\right]\lambda_1^*(b)\right)db,\end{aligned} \tag{59}$$
$$w_2^{(3)}(a) = \int_0^a e^{-\int_b^a \zeta + \sigma(\tau)d\tau}\left(x_2^{(3)}(b) + \left[s_2^*(b) + (1-\gamma(a))v_2^*(b)\right]\overline{\lambda}_2[w](b)\right)db. \tag{60}$$

Recall that $\overline{\lambda}_i[w](a) = \overline{c}_i\beta_1(a)$, so substituting the above expressions into

$$\overline{c}_1 = \int_0^{a_m} \beta_2(s)\psi(s)w_1^{(3)}(s)ds, \quad \text{and} \quad \overline{c}_2 = \int_0^{a_m} \beta_2(s)\psi(s)w_2^{(3)}(s)ds,$$

we obtain a two-dimensional matrix, $(\mathcal{X}_1, \mathcal{X}_2)^T = (\mathbb{I} - \mathcal{H}(\zeta, c_1(\mathcal{R}_0), 0))\,(\overline{c}_1, \overline{c}_2)^T$, where $\mathcal{H}(\zeta, c_1(\mathcal{R}_0), 0)$ is a 2×2 matrix with element $\mathcal{H}_{ij}(\zeta, c_1(\mathcal{R}_0), 0)_{1\le i,j\le 2}$. The elements of $\mathcal{H}_{ij}$ of $\mathcal{H}$ and $\mathcal{X}_i$ $(1 \le i, j \le 2)$ are defined as follows:

$$\begin{aligned}
\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) = {} & \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} Q_1^*(b)\beta_1(b)db \right) ds - \int_0^{a_m} \beta_2(s) \\
& \times \psi(s) \left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} \lambda_1^*(b) \left[\int_0^b \beta_1(b')(1+q_{11})s_1^*(b') \right.\right. \\
& \left.\left. \times e^{-\int_{b'}^b \zeta+(1+q_{11})\lambda_1^*(\tau)d\tau} db' \right] db \right) ds + q_{11} \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s \lambda_1^*(b) \right. \\
& \left. \times e^{-\int_b^s \zeta+\sigma(\tau)d\tau} (1-\gamma(b)) \left[\int_0^b s_1^*(b')\beta_1(b')e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau} db' \right] db \right) ds \\
& - \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s \lambda_1^*(b)e^{-\int_b^s \zeta+\sigma(\tau)d\tau} (1-\gamma(b)) \left[\int_0^b v_1^*(b') \right.\right. \\
& \left.\left. \times \beta_1(b')e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau} db' \right] db \right) ds - \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s \lambda_1^*(b) \right. \\
& \left. \times e^{-\int_b^s \zeta+\sigma(\tau)d\tau} (1-\gamma(b)) \left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau} q_{11}\lambda_1^*(b') \right.\right. \\
& \left.\left. \times \left(\int_0^{b'} s_1^*(r)(1+q_{11})\beta_1(r)e^{-\int_r^{b'} \zeta+(1+q_{11})\lambda_1^*(\tau)d\tau} dr \right) db' \right] db \right) ds, \\
\mathcal{H}_{12}(\zeta, c_1(\mathcal{R}_0), 0) = {} & \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} \lambda_1^*(b) \left[\int_0^b s_1^*(b')q_{12}\beta_1(b') \right.\right. \\
& \left.\left. \times e^{-\int_{b'}^b \zeta+(1+q_{11})\lambda_1^*(\tau)d\tau} db' \right] db \right) ds + q_{12} \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s \lambda_1^*(b) \right. \\
& \left. \times e^{-\int_b^s \zeta+\sigma(\tau)d\tau} (1-\gamma(b)) \left[\int_0^b s_1^*(b')\beta_1(b')e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau} db' \right] db \right) ds \\
& + q_{11}q_{12} \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} \lambda_1^*(b)(1-\gamma(b)) \left[\int_0^b \lambda_1^*(b') \right.\right. \\
& \left.\left. \times e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau} \left(\int_0^{b'} e^{-\int_r^{b'} \zeta+(1+q_{11})\lambda_1^*(\tau)d\tau} s_1^*(r)\beta_1(r)dr \right) db' \right] db \right) ds \\
\mathcal{H}_{21}(\zeta, c_1(\mathcal{R}_0), 0) = {} & 0 \\
\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) = {} & \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} Q_2^*(b)\beta_1(b)db \right) ds \\
\mathcal{X}_1(\zeta, c_1(\mathcal{R}_0), 0) = {} & \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} x_1^{(3)}(b)db \right) ds + \int_0^{a_m} \beta_2(s)\psi(s) \\
& \times \left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} \lambda_1^*(b) \left[\int_0^b e^{-\int_{b'}^b \zeta+(1+q_{11})\lambda_1^*(\tau)d\tau} x_1^{(1)}(b')db' \right] db \right) ds
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^{a_m}\beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda_1^*(b)(1-\gamma(b))\left[\int_0^b x_1^{(2)}(b')\right.\right.\\
&\left.\left.\times e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau}db'\right]db\right)ds+\int_0^{a_m}\beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\right.\\
&\times\lambda_1^*(b)(1-\gamma(b))\left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau}q_{11}\lambda_1^*(b')\left(\int_0^{b'} x_1^{(1)}(r)\right.\right.\\
&\left.\left.\left.\times e^{-\int_r^{b'} \zeta+(1+q_{11})\lambda_1^*(\tau)d\tau}dr\right)db'\right]db\right)ds\\
\mathcal{X}_2(\zeta,c_1(\mathcal{R}_0),0) &= \int_0^{a_m}\beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}x_2^{(3)}(b)db\right)ds\\
Q_i^*(b) &= s_i^*(b)+(1-\gamma(b))\,v_i^*(b),\quad i=1,2\\
s_1^*(b) &= e^{-(1+q_{11})c_1B(b)},\qquad s_2^*(b)=e^{-q_{21}c_1B(b)},\\
v_1^*(b) &= e^{-c_1B(b)}-e^{-(1+q_{11})c_1B(b)},\quad \text{and}\quad v_2^*(b)=1-e^{-q_{21}c_1B(b)}.
\end{aligned}
$$

Let $\mathcal{M}_1(\zeta,\mathcal{R}_0)=(\mathbb{I}-\mathcal{H}(\zeta,c(\mathcal{R}_0)))$ and the determinant of $\mathcal{M}_1(\zeta,c(\mathcal{R}_0))$ be given by $\mathcal{D}_1\mathcal{M}(\zeta,\mathcal{R}_0)$, where for a weakly endemic equilibrium $c(\mathcal{R}_0)=(c_1(\mathcal{R}_0),0)^T$ (or $c(\mathcal{R}_0)=(0,c_2(\mathcal{R}_0))$). From the structure (i.e. upper block matrix) of $\mathcal{H}(\zeta,c(\mathcal{R}_0))$ (and hence $\mathcal{M}_1(\zeta,\mathcal{R}_0)$), we can calculate the determinant as $\mathcal{D}_1\mathcal{M}(\zeta,\mathcal{R}_0)=(1-\mathcal{H}_{11}(\zeta,c(\mathcal{R}_0)))(1-\mathcal{H}_{22}(\zeta,c(\mathcal{R}_0)))$. We note that when $\mathcal{R}_0=1$ and $c_i=0,\ i=1,2$, the determinant $\mathcal{D}_1\mathcal{M}(\zeta,c(\mathcal{R}_0))=0$ determines the eigenvalues of the linearized equations for the disease-free equilibrium. Moreover, we observe that at $\zeta=0$ and $\mathcal{R}_0=1$, $\mathcal{H}_{11}(\zeta=0,c_1(\mathcal{R}_0)=0,0)=\mathcal{H}_{22}(\zeta=0,c_1(\mathcal{R}_0)=0,0)=\mathcal{R}_0$ and $\mathcal{D}_1\mathcal{M}(\zeta,\mathcal{R}_0)=0$. From the previous discussion, the resolvent equation is only solvable when $\mathcal{M}_1(\zeta,\mathcal{R}_0)$ is invertible. That is, if the determinant $\mathcal{D}_1\mathcal{M}(\zeta,\mathcal{R}_0)$ of $\mathcal{M}_1(\zeta,\mathcal{R}_0)$ is nonzero. Therefore, to determine the stability of the weakly endemic equilibrium, we first give the following results:

Lemma 5.14. *Let $\Sigma^{wee}(\mathcal{A}+\mathcal{B})$ and $\Sigma_p^{wee}(\mathcal{A}+\mathcal{B})$ be the spectrum and point spectrum of $\mathcal{A}+\mathcal{B}$, respectively, for the weakly endemic system. Then, $\Sigma^{wee}(\mathcal{A}+\mathcal{B})=\Sigma_p^{wee}(\mathcal{A}+\mathcal{B})=\{\zeta\in\mathbb{C}:\mathcal{D}_1\mathcal{M}(\zeta,\mathcal{R}_0)=0\}$.*

Proof. We observe that the resolvent equation is solvable if and only if $\mathcal{M}_1(\zeta,\mathcal{R}_0)$ is invertible, i.e., $\mathcal{D}_1\mathcal{M}(\zeta,\mathcal{R}_0)\neq 0$. So $\Sigma^{wee}(\mathcal{A}+\mathcal{B})=\{\zeta\in\mathbb{C}:\mathcal{D}_1\mathcal{M}(\zeta,\mathcal{R}_0)=0\}$. Hence $\rho(\mathcal{A}+\mathcal{B})=\{\zeta\in\mathbb{C}:\mathcal{D}_1\mathcal{M}(\zeta,\mathcal{R}_0)\neq 0\}$. Because $\mathcal{M}_1(\zeta,\mathcal{R}_0)$ is invertible for $\zeta\in\rho(\mathcal{A}+\mathcal{B})$, it can be shown that the resolvent $R^{wee}(\zeta,\mathcal{A}+\mathcal{B})$ is compact for any $\zeta\in\rho(\mathcal{A}+\mathcal{B})$ using similar arguments as in Lemma 5.3. That is, we first obtain w in terms of $\overline{c}_i,\ i=1,2$ and x from Eqs. (55)–(60) by substituting for $\overline{\lambda}_i(a)=\beta_1(a)\overline{c}_i$. Then, we find $\overline{c}_i$ in terms of $\mathcal{X}_i,\ i=1,2$ from $(\overline{c}_1,\overline{c}_2)^T=\mathcal{M}_1^{-1}(\zeta,\mathcal{R}_0)\,(\mathcal{X}_1,\mathcal{X}_2)^T$ and $\mathcal{X}_i$ in terms of x. We note that $\mathcal{M}_1^{-1}(\zeta,\mathcal{R}_0)$ exists and is bounded. Thus, we observe that, after the relevant substitutions, the right-hand sides of w in Eqs. (55)–(60) are compositions of compact and bounded operators. Hence, the resolvent is compact. Therefore, $\Sigma^{wee}(\mathcal{A}+\mathcal{B})=\Sigma_p^{wee}(\mathcal{A}+\mathcal{B})$ (see [68, pg. 187], [19, Proposition 4.11]). □

Corollary 5.15. *$\mathcal{H}_{11}(\zeta,c_1(\mathcal{R}_0),0)=1$ or $\mathcal{H}_{22}(\zeta,c_1(\mathcal{R}_0),0)=1$, if and only if $\zeta\in\Sigma_p^{wee}(\mathcal{A}+\mathcal{B})$.*

Theorem 5.16. *Suppose $\mathbb{R}e\zeta<0$ for all $\zeta\in\Sigma_p^{wee}(\mathcal{A}+\mathcal{B})$. Then the weakly endemic equilibrium point is locally asymptotically stable.*

Proof. We first note that the operator $\mathcal{A}$ generates a nilpotent translation C_0-semigroup $\{T_{\mathcal{A}}(t)\}_{t>0}$, and that the operator $\mathcal{B}$ can be broken up into two operators $\mathcal{B}=\overline{\mathcal{B}}+\tilde{\mathcal{B}}$. Then, we observe that $T_{\mathcal{A}+\tilde{\mathcal{B}}}$ is a semigroup of a stable population with a finite age interval, and hence it is eventually compact (see

[59, Proposition 2.5]). From Assumption 4.1, $\overline{\mathcal{B}}$ is compact. Therefore, the C_0-semigroup $\{T_{\mathcal{A}+\mathcal{B}}(t)\}_{t>0}$ is eventually norm continuous (see [53, pg. 166, Proposition 3.1.14]). Hence, $\omega_0(\mathcal{A}+\mathcal{B}) = b(\mathcal{A}+\mathcal{B}) := \sup\{\mathbb{R}e\zeta : \zeta \in \Sigma^{wee}(\mathcal{A}+\mathcal{B})\}$ (see [53, pg. 281, Corollaries 4.3.11–4.3.12]). So, from Lemma 5.14, we have $b(\mathcal{A}+\mathcal{B}) = \sup\{\mathbb{R}e\zeta : \zeta \in \Sigma_p^{wee}(\mathcal{A}+\mathcal{B})\} = \zeta_0$, say. In addition, by Theorem 2.4.18 in [53, pg. 113], the set $\mathcal{S} = \{\zeta \in \Sigma^{wee}(\mathcal{A}+\mathcal{B}) : \mathbb{R}e\zeta \geq -1\}$ is bounded and consists of isolated eigenvalues with finite multiplicities. Therefore, under the given assumption, it follows that the set $\mathcal{S}$ is finite and $\zeta_0 < 0$. Thus, $\lim_{t\to\infty} \|T_{\mathcal{A}+\mathcal{B}}(t)\| \to 0$. Hence, the weakly endemic equilibrium is locally asymptotically stable. □

We can now consider the concrete conditions under which the weakly endemic equilibrium is stable, i.e., there is no $\zeta \in \Sigma_p^{wee}(\mathcal{A}+\mathcal{B})$ with non-negative real part. We note that the condition $\mathcal{D}_1\mathcal{M}(\zeta, \mathcal{R}_0) = 0$ is equivalent to the following: $\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) = 1$ or $\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) = 1$. In the proceeding, we characterize $\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) = 1$ and $\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) = 1$. We state the following properties:

Lemma 5.17. *The function $\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0)$ is continuous in ζ and strictly decreasing with respect to $\zeta \in \mathbb{R}$, and satisfies the following limiting behaviours:*

1. $\lim_{\zeta\to\infty} \mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) = 0$,
2. $\lim_{\zeta\to-\infty} \mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) = +\infty$.

Thus, there exists a unique solution $\zeta_{2,0} \in \mathbb{R}$ of $\mathcal{H}_{22}(\zeta_{2,0}, c_1(\mathcal{R}_0), 0) = 1$ and $\zeta_{2,0}$ is a simple root. Moreover, each of the following implications holds:

1. $\zeta_{2,0} < 0$, *if* $\mathcal{H}_{22}(0, c_1(\mathcal{R}_0), 0) < 1$,
2. $\zeta_{2,0} = 0$, *if* $\mathcal{H}_{22}(0, c_1(\mathcal{R}_0), 0) = 1$,
3. $\zeta_{2,0} > 0$, *if* $\mathcal{H}_{22}(0, c_1(\mathcal{R}_0), 0) > 1$.

Furthermore, any solution $\zeta \neq \zeta_{2,0}$ of $\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) = 1$ is such that $\mathbb{R}e\zeta < \zeta_{2,0}$.

Proof. We observe that $\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0)$ is continuous and strictly monotonically decreasing in ζ. Furthermore, using Assumption 2.1, we note that

$$\begin{aligned}\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) &= \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} Q_2^*(b)\beta_1(b)db\right)ds \\ &\leq \overline{\beta}\int_0^{a_m}\left(\int_0^s \psi(s)e^{-\int_b^s \zeta d\tau}db\right)ds \leq \frac{\overline{\beta}}{\zeta} \to 0 \quad \text{as } \zeta\to\infty\end{aligned}$$

as in Lemma 5.4(2). For the second part, we observe that $\Theta(b,\tau) = e^{-\int_b^\tau \sigma(s)ds} \geq e^{-\tilde{\sigma}a_m} := c$ for $a_m \geq s \geq b \geq 0$. Let $\overline{Q}_2 = \sup_{b\in[0,a_m]} Q_2^*(b)$. Then, from Assumption 4.1(2),

$$\begin{aligned}\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) &= \int_0^{a_m}\left(\int_b^{a_m} \beta_2(s)\psi(s)e^{-\int_b^s \zeta+\sigma(\tau)d\tau} Q_2^*(b)\beta_1(b)ds\right)db \\ &\geq \varepsilon c\overline{Q}_2\int_{a_m-a^*}^{a_m} \psi(s)\left(\int_0^s e^{-\int_b^s \zeta d\tau}db\right)ds \to \infty \quad \text{as } \zeta\to-\infty\end{aligned}$$

as in Lemma 5.4(1).

It follows that there is a unique real value $\zeta_{2,0} \in \mathbb{R}$ such that $\mathcal{H}_{22}(\zeta_{2,0}, c_1(\mathcal{R}_0), 0) = 1$. Since, $\mathcal{H}_{22}(\zeta_{2,0}, c_1(\mathcal{R}_0), 0)$ is strictly decreasing, it is a simple root. Therefore, the assertions follow.

Moreover, let ζ be a solution different from $\zeta_{2,0}$, then

$$\begin{aligned}\mathcal{H}_{22}(\zeta_{2,0}, c_1(\mathcal{R}_0), 0) = 1 &= \mathbb{R}e\left(\int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} Q_2^*(b)\beta_1(b)db\right)ds\right)\\ &= \left(\int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \mathbb{R}e\zeta+\sigma(\tau)d\tau}\cos(\mathbb{I}m\zeta(s-b))Q_2^*(b)\beta_1(b)db\right)ds\right)\\ &< \left(\int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \mathbb{R}e\zeta+\sigma(\tau)d\tau} Q_2^*(b)\beta_1(b)db\right)ds\right).\end{aligned}$$

Thus, since $\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0)$ is strictly decreasing function, it follows that $\mathbb{R}e\zeta < \zeta_{2,0}$. □

We observe that $\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) > 0$ and $\mathcal{H}_{22}(0,0,0) = \mathcal{R}_0$. Similarly, $\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) > 0$ and $\mathcal{H}_{11}(0,0,0) = \mathcal{R}_0$. In addition, $\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0)$ can be expressed as a difference of two functions, namely

$$\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) = \mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0) - \mathcal{H}_{11}^{(2)}(\zeta, c_1(\mathcal{R}_0), 0), \tag{61}$$

where

$$\begin{aligned}\mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0) &= \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} Q_1^*(b)\beta_1(b)db\right)ds\\ &\quad + \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda_1^*(b)(1-\gamma(b))\right.\\ &\quad \left.\times\left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau} s_1^*(b')q_{11}\beta_1(b')db'\right]db\right)ds,\end{aligned} \tag{62}$$

$$\begin{aligned}\mathcal{H}_{11}^{(2)}(\zeta, c_1(\mathcal{R}_0), 0) &= \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda_1^*(b)\left[\int_0^b s_1^*(b')(1+q_{11})\beta_1(b')\right.\right.\\ &\quad \left.\left.\times e^{-\int_{b'}^b \zeta+(1+q_{11})\lambda_1^*(\tau)d\tau}db'\right]db\right)ds + \int_0^{a_m} \beta_2(s)\psi(s)\\ &\quad \times\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda_1^*(b)(1-\gamma(b))\left[\int_0^b v_1^*(b')\beta_1(b')\right.\right.\\ &\quad \left.\left.\times e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau}db'\right]db\right)ds + \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s \lambda_1^*(b)\right.\\ &\quad \times(1-\gamma(b))e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau} q_{11}\lambda_1^*(b')\left(\int_0^{b'} s_1^*(r)\right.\right.\\ &\quad \left.\left.\left.\times(1+q_{11})\beta_1(r)e^{-\int_r^{b'} \zeta+(1+q_{11})\lambda_1^*(\tau)d\tau}dr\right)db'\right]db\right)ds.\end{aligned} \tag{63}$$

Lemma 5.18. *Let $\mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0)$ be defined as above. Then $\mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0)$ is continuous in ζ and monotonically decreasing with respect to ζ and satisfies,*

1. $\lim_{\zeta\to\infty} \mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0) = 0$,
2. $\lim_{\zeta\to-\infty} \mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0) = \infty$.

Proof. Clearly, $\mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0)$ is continuous with respect to ζ, and is strictly monotonically decreasing. In addition, let $\overline{B}_1 = \int_0^{a_m} \beta_1(\tau)d\tau$. Then,

$$\begin{aligned}\mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0) &= \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} Q_1^*(b)\beta_1(b)db\right)ds \\ &+ \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda_1^*(b)(1-\gamma(b))\right. \\ &\times \left.\left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau} s_1^*(b')q_{11}\beta_1(b')db'\right]db\right)ds \\ &\leq \overline{\beta}\left(1+q_{11}c_1\overline{B}_1\right)\int_0^{a_m}\left(\int_0^s \psi(s)e^{-\int_b^s \zeta d\tau}db\right)ds \to 0 \quad \text{as } \zeta \to \infty\end{aligned}$$

as in Lemma 5.17(1). Moreover, we observe that

$$\begin{aligned}\mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0) &= \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} Q_1^*(b)\beta_1(b)db\right)ds \\ &+ \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda_1^*(b)(1-\gamma(b))\right. \\ &\times \left.\left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda_1^*(\tau)d\tau} s_1^*(b')q_{11}\beta_1(b')db'\right]db\right)ds, \\ &\geq \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} Q_1^*(b)\beta_1(b)db\right)ds \to \infty\end{aligned}$$

as $\zeta \to -\infty$ as in Lemma 5.4(1) and Lemma 5.17(2). □

Corollary 5.19. *$\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0)$ is continuous with respect to ζ, and for ζ real, satisfies the following*

1. $\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) \leq \mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0)$, *and*
2. $\lim_{\zeta\to\infty} \mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) = 0$.

Lemma 5.20. *There exists a unique solution $\zeta_1 \in \mathbb{R}$ such that $\mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0) = 1$. In addition,*

1. *If $\mathcal{H}_{11}^{(1)}(0, c_1(\mathcal{R}_0), 0) < 1$, then $\zeta_1 < 0$,*
2. *If $\mathcal{H}_{11}^{(1)}(0, c_1(\mathcal{R}_0), 0) = 1$, then $\zeta_1 = 0$,*
3. *If $\mathcal{H}_{11}^{(1)}(0, c_1(\mathcal{R}_0), 0) > 1$, then $\zeta_1 > 0$.*

Moreover, if $\zeta \in \{\zeta \in \mathbb{C} : \mathcal{H}_{11}^{(1)}(\zeta, c_1(\mathcal{R}_0), 0) = 1\}$ then if $\zeta \neq \zeta_1$, $\mathbb{R}e\zeta < \zeta_1$.

Proof. The proof is similar to Lemma 5.17. □

Lemma 5.21. *Let $\mathcal{R}_0 > 1$ but sufficiently close to 1. If $\mathcal{H}_{11}(0, c_1(\mathcal{R}_0), 0) > 1$, then the characteristic equation admits a positive root.*

Proof. We note that, from the properties stated in Lemma 5.18 and Corollary 5.19, when $\mathcal{H}_{11}(0, c_1(\mathcal{R}_0), 0) > 1$, there is a solution $\zeta_0 > 0$ such that $\mathcal{H}_{11}(\zeta_0, c_1(\mathcal{R}_0), 0) = 1$. That is, let $\Delta(\zeta) = \mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) - 1 = 0$. We observe that $\Delta(0) > 0$ and $\Delta(\infty) < 0$. Then, by the Intermediate Value Theorem, there exists a point $\zeta_0 \in (0, \infty)$ such that $\Delta(\zeta_0) = 0$ or $\mathcal{H}_{11}(\zeta_0, c_1(\mathcal{R}_0), 0) = 1$. This corresponds to instability. □

We now give conditions for $\mathcal{H}_{11}(0, c_1(\mathcal{R}_0), 0) = 1$ to have only roots with negative real part.

Lemma 5.22. *Take $\mathcal{R}_0 > 1$. Suppose $\mathcal{H}_{11}^{(1)}(0, c_1, 0) < 1$, so that, for c_1 sufficiently small and positive, $\mathcal{H}_{11}^{(1)}(\zeta, c_1, 0) = 1$ has a unique real negative solution $\zeta_1(c_1)$. Then for c_1 sufficiently small and positive $\mathcal{H}_{11}(\zeta, c_1, 0) = 1$ has no roots with $\mathbb{R}e\zeta \geq \frac{\zeta_1(0)}{2}$. Hence, for c_1 sufficiently small and positive, all the roots of $\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) = 1$ have negative real part.*

Proof. Assume $\mathcal{H}_{11}^{(1)}(0, c_1, 0) < 1$. From Lemma 5.20, $\mathcal{H}_{11}^{(1)}(\zeta, c_1, 0) = 1$ has a unique real negative solution $\zeta_1(c_1)$. Let $\mathcal{D} = \{\zeta : \mathbb{R}e\zeta \geq \frac{\zeta_1(0)}{2}\}$.

Now let $1 - H_{11}(\zeta, c_1, 0) = \Delta_1(\zeta, c_1) + \Delta_2(\zeta, c_1)$, where

$$\Delta_1(\zeta, c_1) = 1 - \mathcal{H}_{11}^{(1)}(\zeta, c_1, 0), \quad \text{and} \quad \Delta_2(\zeta, c_1) = \mathcal{H}_{11}^{(2)}(\zeta, c_1, 0).$$

We show that for $\zeta \in \mathcal{D}$ and c_1 small enough, $|\Delta_1(\zeta, c_1)| > |\Delta_2(\zeta, c_1)|$ and hence $\mathcal{H}_{11}(\zeta, c_1, 0) = 1$ has no roots in $\mathcal{D}$.

Set $\eta = \mathcal{H}_{11}^{(1)}\left(\frac{\zeta_1(0)}{2}, 0, 0\right)$ and note that $\eta < 1$, because $\mathcal{H}_{11}^{(1)}(\zeta_1(0), 0, 0) = 1$ and $\mathcal{H}_{11}^{(1)}(\zeta, c_1, 0)$ is decreasing in ζ. In addition, $\mathcal{H}_{11}^{(1)}(\zeta, c_1, 0)$ is continuous in c_1. Therefore, for $\zeta \in \mathcal{D}$ and c_1 small enough, we have

$$\left|\mathcal{H}_{11}^{(1)}(\zeta, c_1, 0)\right| \leq \mathcal{H}_{11}^{(1)}\left(\frac{\zeta_1(0)}{2}, c_1, 0\right) < \frac{1+\eta}{2}.$$

Hence,

$$\left|1 - \mathcal{H}_{11}^{(1)}(\zeta, c_1, 0)\right| > \frac{1-\eta}{2}.$$

Moreover, we observe that $\Delta_2(\zeta, 0) = \mathcal{H}_{11}^{(2)}(\zeta, 0, 0) = 0$ and $\mathcal{H}_{11}^{(2)}(\zeta, c_1, 0)$ is continuous in c_1. For $\zeta \in \mathcal{D}$ and c_1 small enough, we have

$$|\Delta_2(\zeta, c_1)| = |\mathcal{H}_{11}^{(2)}(\zeta, c_1, 0)| \leq \mathcal{H}_{11}^{(2)}\left(\frac{\zeta_1(0)}{2}, c_1, 0\right) < \frac{1-\eta}{2}.$$

Hence,

$$|\mathcal{H}_{11}(\zeta, c_1, 0) - 1| \geq |\Delta_1(\zeta, c_1)| - |\Delta_2(\zeta, c_1)| > 0.$$

Thus, $\mathcal{H}_{11}(\zeta, c_1, 0) = 1$ has no roots in $\mathcal{D}$. □

In the above discussion, we have shown the existence and non-existence of roots (positive real and complex with positive real parts) depending on the conditions imposed on $\mathcal{H}_{11}(0, c_1(\mathcal{R}_0), 0)$ and $\mathcal{H}_{11}^{(1)}(0, c_1(\mathcal{R}_0), 0)$. In particular, we note that the existence and exclusion of positive real roots depends only on $\mathcal{H}_{11}(0, c_1(\mathcal{R}_0), 0)$, while exclusion of complex roots (with positive real part) depends on $\mathcal{H}_{11}(0, c_1(\mathcal{R}_0), 0)$ and $\mathcal{H}_{11}^{(1)}(0, c_1(\mathcal{R}_0), 0)$. Given the above results in Lemmas 5.21 and 5.22, we establish the condition for the stability of the weakly endemic steady state.

Theorem 5.23. *Take $\mathcal{R}_0 > 1$ and $\mathcal{R}_0 - 1$ is sufficiently small.*

1. *If $\mathcal{H}_{11}^{(1)}(0, c_1(\mathcal{R}_0), 0) < 1$ and $\mathcal{H}_{22}(0, c_1(\mathcal{R}_0), 0) < 1$, then the weakly endemic equilibrium (i.e., discrete strain structure) is locally asymptotically stable.*
2. *If $\mathcal{H}_{11}(0, c_1(\mathcal{R}_0), 0) > 1$ or $\mathcal{H}_{22}(0, c_1(\mathcal{R}_0), 0) > 1$, then the weakly endemic equilibrium is unstable.*

Proof. On the one hand, we note that, from Lemma 5.22, if $\mathcal{H}_{11}^{(1)}(0, c_1(\mathcal{R}_0), 0) < 1$, then the characteristic equation $\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0) = 1$ does not admit (real or complex) non-negative characteristic roots. In

addition, from Lemma 5.17, we observe that the dominant root of $\mathcal{H}_{22}(\zeta, c_1(\mathcal{R}_0), 0) = 1$ is negative when $\mathcal{H}_{22}(0, c_1(\mathcal{R}_0), 0) < 1$. Thus, from Corollary 5.15 and Theorem 5.16, the weakly endemic equilibrium point is locally asymptotically stable. On the other hand, if one of the conditions is violated, then there is a non-negative eigenvalue $\zeta_0 > 0$, so $\omega_0(\mathcal{A} + \mathcal{B}) > 0$. □

5.4.2. Stability of a strongly endemic equilibrium: coexistence

To study the stability of a strongly endemic equilibrium, we consider the linearized resolvent equation as done in the previous section. We note that, from the definition of strong endemicity, $y_i^*(a) \neq 0$ for all i. In addition, because the epidemiological parameters are the same for all strains, $\lambda_i = \lambda_j$ and $y_i = y_j$ for all $i,\ j$. Therefore, it is sufficient to look at the asymptotic behaviours of the system as a single strain system provided we take into consideration the balanced condition for the quotient network. That is, $q_{ii} + q_{ij} = q_{jj} + q_{ji}$. Now let $\alpha_0 = q_{ii} + q_{ij} = q_{jj} + q_{ji}$ and $\alpha = \alpha_0 + 1$. Then the system above reduces to

$$\begin{aligned}
\frac{\partial w^{(1)}(a)}{\partial a} &= x^{(1)}(a) - \zeta w^{(1)}(a) - \alpha s^*(a)\overline{\lambda}[w](a) - \alpha w^{(1)}(a)\lambda^*(a),\\
\frac{\partial w^{(2)}(a)}{\partial a} &= x^{(2)}(a) - \zeta w^{(2)}(a) + (\alpha-1)\, s^*(a)\overline{\lambda}[w](a) - v^*(a)\overline{\lambda}[w](a) + w^{(1)}(a)\,(\alpha-1)\,\lambda^*(a) - w^{(2)}(a)\lambda^*(a),\\
\frac{\partial w^{(3)}(a)}{\partial a} &= x^{(3)}(a) - \zeta w^{(3)}(a) + [s^*(a) + (1-\gamma(a))v^*(a)]\,\overline{\lambda}[w](a)\\
&\quad + \left[w^{(1)}(a) + (1-\gamma(a))w^{(2)}(a)\right]\lambda^*(a) - \sigma(a)w^{(3)}(a),
\end{aligned}$$

where the steady state solutions $s^*(a)$, $v^*(a)$ and $y^*(a)$ reduce to

$$\begin{aligned}
&s^*(a) = e^{-\int_0^a \alpha\lambda^*(\tau)d\tau}, \qquad v^*(a) = e^{-\int_0^a \lambda^*(\tau)d\tau} - e^{-\int_0^a \alpha\lambda^*(\tau)d\tau}, \qquad Q^*(a) = s^*(a) + (1-\gamma(a))v^*(a),\\
&y^*(a) = \int_0^a e^{-\int_b^a \sigma(\tau)d\tau}\,(1-\gamma(b))\,\lambda^*(b)\left[e^{-\int_0^b \lambda^*(\tau)d\tau} - e^{-\int_0^b \alpha\lambda^*(\tau)d\tau}\right]db\\
&\qquad + \int_0^a e^{-\int_b^a \sigma(\tau)d\tau}\lambda^*(b)e^{-\int_0^b \lambda^*(\tau)d\tau}db.
\end{aligned}$$

Integrating the resolvent equation above, we have

$$\begin{aligned}
w^{(1)}(a) &= \int_0^a e^{-\int_b^a \zeta+\alpha\lambda^*(\tau)d\tau}\left(x^{(1)} - \alpha s^*(b)\overline{\lambda}[w](b)\right)db,\\
w^{(2)}(a) &= \int_0^a e^{-\int_b^a \zeta+\lambda^*(\tau)d\tau}\left(x^{(2)} + (\alpha-1)s^*(b)\overline{\lambda}[w](b) - v^*(b)\overline{\lambda}[w](b) + (\alpha-1)w^{(1)}(b)\lambda^*(b)\right)db,\\
w^{(3)}(a) &= \int_0^a e^{-\int_b^a \sigma(\tau)+\zeta d\tau}\left(x^{(3)}(b) + q^*(b)\overline{\lambda}[w](b) + \left[w^{(1)}(b) + (1-\gamma(b))w^{(2)}(b)\right]\lambda^*(b)\right)db.
\end{aligned}$$

Let $c = \int_0^{a_m} \beta_2(s)\psi(s)y^*(s)ds$, $\overline{c} = \int_0^{a_m} \beta_2(s)\psi(s)w^{(3)}(s)ds$, and $B(a) = \int_0^a \beta_1(\tau)d\tau$. Then $\lambda^*(a) = c\beta_1(a)$ and $\overline{\lambda}[w](a) = \overline{c}\beta_1(a)$.

The strongly endemic steady state only exists when $\mathcal{R}_0 > 1$ (see Lemma 5.13). To study the stability of the strongly endemic equilibrium, we substitute the above expression for $w_i^{(k)}$, $i = 1, 2$ into $\overline{c}$ using strong endemicity. We obtain an equation of the form $\mathcal{X} = (1 - \mathcal{H}(\zeta, c(\mathcal{R}_0)))\,\overline{c}$, where

$$\begin{aligned}
\mathcal{H}(\zeta, c(\mathcal{R}_0)) &= \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}Q^*(b)\beta_1(b)db\right)ds\\
&\quad - \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda^*(b)\left[\int_0^b e^{-\int_{b'}^b \zeta+\alpha\lambda^*(\tau)d\tau}\alpha s_1^*(b')\right.\right.\\
&\quad \left.\left.\times\,\beta_1(b')db'\right]db\right)ds + \int_0^{a_m} \beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda^*(b)\right.
\end{aligned}$$

$$\times (1-\gamma(b))\left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda^*(\tau)d\tau}(\alpha-1)s^*(b')\beta_1(b')db'\right]db\Bigg)ds$$
$$-\int_0^{a_m}\beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda^*(b)(1-\gamma(b))\right.$$
$$\times\left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda^*(\tau)d\tau}v^*(b')\beta_1(b')db'\right]db\Bigg)ds-\int_0^{a_m}\beta_2(s)\psi(s)$$
$$\times\left[\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda^*(b)(1-\gamma(b))\left(\int_0^b e^{-\int_{b'}^b \zeta+\lambda^*(\tau)d\tau}(\alpha-1)\right.\right.$$
$$\left.\left.\times\lambda^*(b')\left[\int_0^{b'} e^{-\int_\eta^{b'} \zeta+\alpha\lambda^*(\tau)d\tau}\alpha s^*(\eta)\beta_1(\eta)d\eta\right]db'\right)db\right]ds,$$

$$\mathcal{X}=\int_0^{a_m}\beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}x^{(3)}(b)db\right)ds+\int_0^{a_m}\beta_2(s)\psi(s)$$
$$\times\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda^*(b)\left[\int_0^b e^{-\int_{b'}^b \zeta+\alpha\lambda^*(\tau)d\tau}x^{(1)}(b')db'\right]db\right)ds$$
$$+\int_0^{a_m}\beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda^*(b)(1-\gamma(b))\left[\int_0^b x^{(2)}(b')\right.\right.$$
$$\left.\left.\times e^{-\int_{b'}^b \zeta+\lambda^*(\tau)d\tau}db'\right]db\right)ds+\int_0^{a_m}\beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\right.$$
$$\times(1-\gamma(b))\lambda^*(b)\left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda^*(\tau)d\tau}(\alpha-1)\lambda^*(b')\left(\int_0^{b'} x^{(1)}(\eta)\right.\right.$$
$$\left.\left.\left.\times e^{-\int_\eta^{b'} \zeta+\alpha\lambda^*(\tau)d\tau}d\eta\right)db'\right]db\right)ds.$$

We state the following conditions:

Lemma 5.24. *Let $\Sigma^{see}(\mathcal{A}+\mathcal{B})$ and $\Sigma_p^{see}(\mathcal{A}+\mathcal{B})$ be the spectrum and point spectrum of $\mathcal{A}+\mathcal{B}$, respectively, for the strongly endemic case, then $\Sigma^{see}(\mathcal{A}+\mathcal{B})=\Sigma_p^{see}(\mathcal{A}+\mathcal{B})=\{\zeta\in\mathbb{C}:\mathcal{H}(\zeta,c(\mathcal{R}_0))=1\}$.*

Proof. The resolvent equation is solvable if and only if $1-\mathcal{H}(\zeta,c(\mathcal{R}_0))\neq 0$. So $\Sigma^{see}(\mathcal{A}+\mathcal{B})=\{\zeta\in\mathbb{C}:\mathcal{H}(\zeta,c(\mathcal{R}_0))=1\}$. Hence, $\rho(\mathcal{A}+\mathcal{B})=\{\zeta\in\mathbb{C}:\mathcal{H}(\zeta,c(\mathcal{R}_0))\neq 1\}$. Using similar arguments to those in Lemma 5.14, if $\zeta\in\rho(\mathcal{A}+\mathcal{B})$, then $\mathcal{A}+\mathcal{B}$ has a compact resolvent (see [68, pg. 187] and [19, Proposition 4.11]). Hence, the assertion $\Sigma^{see}(\mathcal{A}+\mathcal{B})=\Sigma_p^{see}(\mathcal{A}+\mathcal{B})$ holds. □

Corollary 5.25. *$\mathcal{H}(\zeta,c(\mathcal{R}_0))=1$, if and only if $\zeta\in\Sigma_p^{see}(\mathcal{A}+\mathcal{B})$.*

Let $\mathcal{H}(\zeta,c(\mathcal{R}_0))=\mathcal{H}^{(1)}(\zeta,c(\mathcal{R}_0))-\mathcal{H}^{(2)}(\zeta,c(\mathcal{R}_0))$, where

$$\mathcal{H}^{(1)}(\zeta,c(\mathcal{R}_0))=\int_0^{a_m}\beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}Q^*(b)\beta_1(b)db\right)ds$$
$$+\int_0^{a_m}\beta_2(s)\psi(s)\left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau}\lambda^*(b)(1-\gamma(b))\right.$$
$$\left.\times\left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda^*(\tau)d\tau}(\alpha-1)s^*(b')\beta_1(b')db'\right]db\right)ds,$$

$$\mathcal{H}^{(2)}(\zeta, c(\mathcal{R}_0)) = \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} \lambda^*(b) \left[\int_0^b e^{-\int_{b'}^b \zeta+\alpha\lambda^*(\tau)d\tau} \alpha s_1^*(b') \times \beta_1(b')db' \right] db \right) ds + \int_0^{a_m} \beta_2(s)\psi(s) \left(\int_0^s \lambda^*(b)(1-\gamma(b)) \times e^{-\int_b^s \zeta+\sigma(\tau)d\tau} \left[\int_0^b e^{-\int_{b'}^b \zeta+\lambda^*(\tau)d\tau} v^*(b')\beta_1(b')db' \right] db \right) ds + (\alpha-1) \times \int_0^{a_m} \beta_2(s)\psi(s) \left[\int_0^s e^{-\int_b^s \zeta+\sigma(\tau)d\tau} \lambda^*(b)(1-\gamma(b)) \left(\int_0^b \lambda^*(b') \times e^{-\int_{b'}^b \zeta+\lambda^*(\tau)d\tau} \left[\int_0^{b'} e^{-\int_\eta^{b'} \zeta+\alpha\lambda^*(\tau)d\tau} \alpha s^*(\eta)\beta_1(\eta)d\eta \right] db' \right) db \right] ds.$$

We observe that the functional form of $\mathcal{H}(\zeta, c(\mathcal{R}_0))$ is similar to the functional form of $\mathcal{H}_{11}(\zeta, c_1(\mathcal{R}_0), 0)$. In fact, $\mathcal{H}(\zeta, c(\mathcal{R}_0)) = \mathcal{H}_{11}(\zeta, c(\mathcal{R}_0), 0)$, when $\lambda^*(a) = \lambda_1^*(a)$, and q_{11} is replaced by $\alpha - 1$. Thus, the results and proofs of Lemma 5.18, Corollary 5.19, Lemmas 5.20–5.22, all carry over to this case and we have:

Theorem 5.26. *If $\mathcal{H}^{(1)}(0, c(\mathcal{R}_0)) < 1$, then the strongly endemic equilibrium (i.e., co-existence, fully synchronized steady-state, or no strain structure) is locally asymptotically stable. If $\mathcal{H}(0, c(\mathcal{R}_0)) > 1$, then the strongly endemic equilibrium is unstable.*

6. Numerical simulations and discussion: the effect of cross-immunity on the dynamics

In this section, we investigate the effect of the serotypic architectures and the various dynamic features of the model. In particular, we look at the dynamic behaviours of the model under both uniform and differential allelic combinations as we vary the cross-immunity level γ (i,.e., $\gamma(a) = \gamma$ for all $a \in [0, a_m]$) as an exploratory parameter. Although it is possible to use the age-specific parameters $\beta(a, s) = qc(a, s)$, $\mu(a)$, $\sigma(a)$ where q is the disease-specific transmission probability and $c(a, s)$ is the contact rate, we assume, for simplicity, that the age-specific parameters $\beta(a, s) = \beta$, $\mu(a) = \mu$ and $\sigma(a) = \sigma$, because the dynamic behaviours are preserved except that the abundance level of each strain for individual age–cohorts is commensurable with the age-specific reproductive number (see [30,6,38,15]). The method of lines was used with the following 15 discretized age–cohorts [0–4], [5–9], [10–14], [15–19], [20–24], [24–29], [30–34], [35–39], [40, 44], [45, 49], [50, 54], [55, 59], [60, 64], [65, 69] and [70–100] years.

We note that, for both uniform and variable age-specific parameters, the dynamic features of the strains naturally tend towards the dynamic patterns observed in the unstructured models discussed in [30,38,15, and references therein]. In particular, the system (1) exhibits three distinct dynamic features with very rich temporal and clustering properties, and self-organization/synchrony that depends on the relative degrees of cross-immunity (see Fig. 4). In Fig. 4(a), we observe the coexistence of all strain for low cross-immunity, where all strains tend to a stable equilibrium which are usually commensurable to their respective reproductive fitnesses. However, because we have assumed that all the strains have the same epidemiological parameters and hence reproductive successes, the strains approach the same abundance level regardless of the allelic architectures. That is, all strains approach the same steady states whether we have uniform (e.g., 2-locus–2-allele, 3-locus–2-allele systems) or differential allelic variations (e.g., 2-locus system with 2 alleles at one locus and 3 alleles at the other locus). These dynamic patterns correspond to fully synchronized no structure structures (NSS) with a maximum diversity and low discordance.

At the intermediary cross-immunity level, the model exhibits cyclic (both periodic, intermittency chaotic) temporal dynamics with exotic cluster-specific sequential domination (see Fig. 4(b)–(d)). However, the cyclical dynamical behaviours and their cluster-specific structures at the intermediary cross-immunity level

are subtle, as the model exhibits various complex forms of synchronization and dynamic patterns. In particular, we observe that dominant types are cyclically replaced, where we can detect sharp epidemics dominated by, among a number of other behaviours: (i) anti-phase periodic solution with alternating dominance of discordant pairs (including large amplitude periodic cycles), (ii) a single strain with erratic multi-year abundance similar to the dynamic feature of influenza epidemics with irregular emergence and re-emergence of certain pathogenic forms (seen in y_i, not shown), (iii) ordered alternating appearance of a single antigenic type in periodic or quasi-periodic form similar to discrete periodic travelling waves in strain/antigenic space with spatio-temporal symmetry, (iv) chaotic solutions with erratic appearance and disappearance of synchrony between discordant antigenic types or discrete sets of minimally overlapping or non-overlapping structures (periodic solutions with intermittent chaotic episodes, chaotic short-period cycles with decoherence between discordant sets, large-amplitude chaotic cycles with spatio-temporal symmetry), and (v) phase-synchronization with (periodic or chaotic) uncorrelated amplitudes. These dynamic behaviours represent cyclical strain structures (CSS).

For high cross-immunity, the system exhibits the principle of competitive exclusion where certain clusters containing discordant serotypes or containing sets of serotypes (i.e., sub-population of strains) with minimal (see Fig. 4(e)) or non-overlapping structures persist while others go to extinction (see Fig. 4(f)). In particular, for uniform allelic variations, only discrete strain structures with a single discordant pair with non-overlapping structures are possible (see Fig. 4(f)), whereas in the case of differential allelic combinations we see discrete strain structures with both non-overlapping (see Fig. 4(f), where persistent strains are discordant to each other and share no alleles in common) and minimally overlapping serotypic structures (see Fig. 4(e) where the dominant persistent strain is discordant to the other two strains that shares alleles in common). In general, the dominant persistent sets share the least number of alleles while the other persistent strains share the most in the case of discrete strain structures with minimally overlapping serotypic structures. These dynamic patterns correspond to partially synchronized discrete strain structures (DSS). Some of these dynamic patterns described above have been discussed previously in [30,6,5,38,15]. The only difference between these behaviours and those discussed previously is that each age cohort follows similar dynamics, and there might be age-specific synchronizations, which may not hold in the case of a non-uniform age-specific scenario. In addition, for a non-uniform age-specific scenario, we can observe localized epidemics within certain age–cohort in the region where the system exhibits the highest single strain dominance (or lowest strain diversity) and the combination of age-specific transmissibility/reproductive number and the contact structures, and the cross-immunity influence the organization and dynamical behaviours of strain structures.

7. Conclusion

In this article, we have examined a multi-locus–allele epidemic model describing the dynamics of antigenically variable infectious diseases. We reformulated the model into an abstract semilinear Cauchy evolution equation and investigated the existence of positive solutions. In addition, we employed a fixed point approach to study the steady-states, and the spectral theory of operators to prove the existence of a threshold condition on the next-generation operator. In particular, we showed that if the spectral radius of the next-generation operator $\mathcal{T}$, $r(\mathcal{T})$, is less than 1, then the disease-free equilibrium is the only equilibrium and is globally asymptotically stable. If $r(\mathcal{T}) > 1$, then the disease-free equilibrium is unstable. Moreover, we showed that there is at least one nontrivial equilibrium. However, using the spectral radius $r(\mathcal{T})$ of the next generation operator $\mathcal{T}$, the condition for existence of endemic equilibrium is necessary but not sufficient to establish the stability of the endemic equilibrium. To establish the stability of endemic equilibria, we rely on insights gained from the mesoscale property captured by the quotient network of the strain space and from the analysis presented in [38]. We defined two notions of endemicity (e.g., weakly and strongly endemic

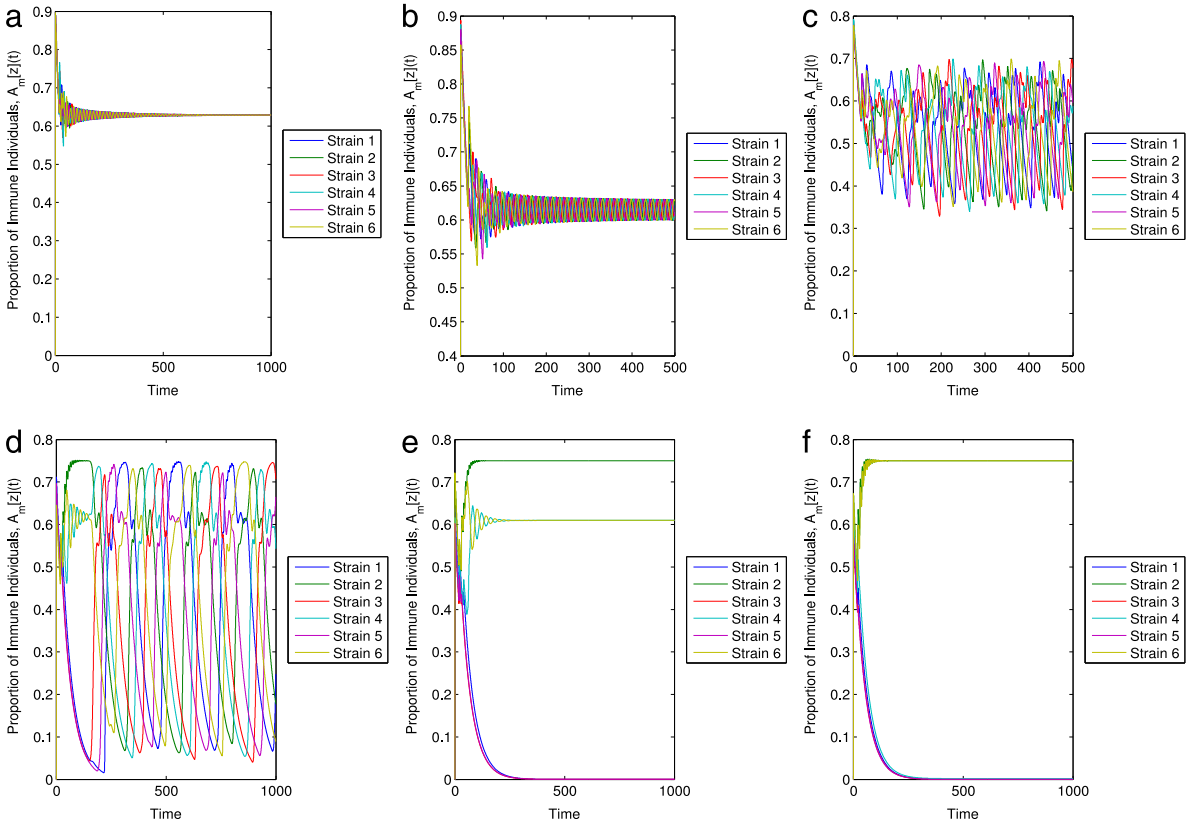

Fig. 4. Average temporal evolution 2-locus–(2,3)-allele system for Eq. (1). Temporal dynamics of a 2-locus system with 2 alleles on one locus and 3 alleles on the other locus show (a) the coexistence of strains approaching the same endemic steady-state, $\gamma = 0.5$, (b) discrete periodic travelling wave solutions and spatio-temporal symmetry across all age–cohorts, $\gamma = 0.54$, (c) chaotic dynamics ($\gamma = 0.75$) and (d) large-amplitude chaotic solutions ($\gamma = 0.90$), (e)–(f) discrete strain structure where one discordant pair persists while the other pairs go to extinction (principle of competitive exclusion) for $\gamma = 0.95$ and $\gamma = 1.0$, respectively. Here, the parameters are as follows: $\beta(a,s) = \beta = 40,\ \mu(a) = \mu = 0.02,\ \sigma(a) = \sigma = 9.98$; and $\mathcal{A}_m[\phi](t) = \frac{1}{|\Omega_m|}\int_{\Omega_m}\phi(t,a)da,\ \Omega_m = [0, a_m]$.

equilibria) corresponding to the principle of competitive exclusion and co-existence, and used these ideas along with the symmetric structure of the strain space and the uniformity of parameters across strains to simplify our analysis and to provide additional conditions necessary for the stability of the system. We provide numerical simulations corresponding to the analytical results. Furthermore, we observe the existence of discrete antigenic forms among pathogens can either fully or partially self-organize, where (i) strains exhibit no strain structures and coexist or (ii) antigenic variants sort into non-overlapping or minimally overlapping clusters that either undergo the principle of competitive exclusion exhibiting discrete strain structures, or co-exist cyclically. These observations are consistent with numerical simulations for unstructured models based on strain theory in [12,30,39,35,73,8,33,34,6,36,37,48,38,15, and references therein].

The analysis presented herein is based on some assumptions that might be more restrictive for epidemic models. On a fundamental level, we assume a proportionate mixing pattern for the contact rate, a demographically stable population and the uniformity of parameters across strains. The structure of the strain interaction network and its 2-coloured balanced quotient is also restrictive. In particular, we have focused only on a general strain space structure with Mendelian structure or hyperoctahedral symmetry, where the 2-coloured quotient networks are bi-nodal, and the dimension of the system can be reduced. For a more general strain space with no obvious well known symmetry property, we suspect that our analysis can be carried out in a similar manner as long as all the quotient networks can be obtained using tools from

algebraic graph theory and all the different combination of synchronous equilibrium clusters can be found. Although, the numerical results of the model described in [38,15, and the relevant references therein] and in Section 6 show the existence of cyclical behaviours, such dynamics and their orbital asymptotical stability were not investigated herein. This is left as an open problem. The age-specific model described herein can be extended to incorporate different level of cross-immunity including temporary immunity and vaccine-acquired immunity, to investigate the impact of different functional forms of $\gamma(a)$ with other functional forms of $\beta(a, s)$, $\mu(a)$ and $\sigma(a)$ as compared to its age-independent analogue on the dynamic behaviours of the model described in [38,15], to name but a few.

Acknowledgement

This work was supported in part by National Science Foundation under the auspice of the Graduate Research Fellowship Program DGE 1047766 (to A.C.).

References

[1] D.J.D. Earn, J. Dushoff, S.A. Levin, Ecology and evolution of the flu, Trends Ecol. Evol. 17 (7) (2002) 334–340.
[2] N.M. Ferguson, A.P. Galvani, R.M. Bush, Ecological and immunological determinants of influenza evolution, Nature 422 (2003) 428–433.
[3] D.J. Smith, et al., Variable efficacy of repeated annual influenza vaccination, Proc. Natl. Acad. Sci. USA 96 (1999) 14001–14006.
[4] D.J. Smith, et al., Mapping the antigenic and genetic evolution of influenza virus, Science 305 (5682) (2004) 371–376.
[5] C.O. Buckee, et al., Role of stochastic processes in maintaining discrete strain structure in antigenically diverse pathogen populations, Proc. Natl. Acad. Sci. 108 (2011) 15504–15509.
[6] M. Recker, et al., The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types, Proc. Natl. Acad. Sci. 104 (2007) 7711–7716.
[7] S. Gupta, K. Trenholme, R.M. Anderson, K.P. Day, Antigenic diversity and the transmission dynamics of Plasmodium falciparum, Science 263 (1994) 961–963.
[8] M. Recker, et al., Transient cross-reactive immune responses can orchestrate antigenic variation in malaria, Nature 429 (2004) 555–558.
[9] J.R. Gog, B.T. Grenfell, Dynamics and selection of many-strain pathogens, Proc. Natl. Acad. Sci. 99 (2002) 17209–17214.
[10] M. Recker, et al., Immunological serotype interactions and their effect on the epidemiological pattern of dengue, Proc Roy. Soc. London B 276 (2009) 2541–2548.
[11] S. Gupta, R.M. Anderson, Population structure of pathogens: the role of immune selection, Parasitol. Today 15 (1999) 497–501.
[12] S. Gupta, M.C.J. Maiden, I.M. Feavers, S. Nee, R.M. May, R.M. Anderson, The maintenance of strain structure in populations of recombining infectious agents, Nat. Med. 2 (1996) 437–442.
[13] C.O. Buckee, et al., The effects of host contact structure on pathogen diversity and strain structure, Proc. Natl. Acad. Sci. 101 (29) (2004) 10839–10844.
[14] C.O. Buckee, et al., Host community structure and the maintenance of pathogen diversity, Proc. R. Soc. B: Biol. Sci. 274 (1619) (2007) 1715–1721.
[15] A. Cherif, Mathematical evolutionary epidemiology: Limited epitopes, evolution of strain structures and age-specificity (D. Phil Thesis), University of Oxford, 2015.
[16] A.J. Kucharski, J.R. Gog, Age profile of immunity to influenza: Effect of original antigenic sin, J. Theor. Popul. Biol. 81 (2012) 102–112.
[17] S. Cobey, M. Pascual, Consequences of host heterogeneity, epitope immunodominance, and immune breadth for strain competition, J. Theoret. Biol. 270 (1) (2001) 80–87.
[18] K. Hancock, V. Veguilla, X. Lu, W. Zhong, E.N. Butler, H. Sun, F. Liu, L. Dong, J.R. DeVos, P.M. Gargiullo, T.L. Brammer, N.J. Cox, T.M. Tumpey, J.M. Katz, Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus, N. Engl. J. Med. 361 (2009) 1945–1952.
[19] G.F. Webb, Theory of Nonlinear Age-Dependent Population Dynamics, Marcel Dekker, New York, 1985.
[20] G.F. Webb, Structured population dynamics, Pol. Acad. Sci. Banach Center Publ. 63 (2004) 177–186.
[21] R. Burger, The Mathematical Theory of Selection, Recombination, and Mutation, John Wiley & Sons Ltd., Chichester, 2000.
[22] O. Diekmann, P. Getto, Boundedness, global existence and continuous dependence for nonlinear dynamical systems describing physiologically structured populations, J. Differential Equations 215 (2) (2005) 268–319.
[23] O. Diekmann, M. Gyllenberg, H. Huang, M. Kirkilionis, J.A.J. Metz, H.R. Thieme, On the formulation and analysis of general deterministic structured population models. II. Nonlinear theory, J. Math. Biol. 43 (2) (2001) 157–189.
[24] O. Diekmann, J.A.J. Metz (Eds.), The Dynamics of Physiologically Structured Populations, in: Lecture Notes in Biomathematics, vol. 68, Springer-Verlag, Berlin, 1986.

[25] B. Perthame, Transport Equations in Biology, Birkhauser Verlag, Basel, 2007.
[26] C. Castillo-Chavez, et al., Epidemiological models with age structure, proportionate mixing and cross-immunity, J. Math. Biol. 27 (1989) 233–258.
[27] P. Magal, S. Ruan, Center manifolds for semilinear equations with non-dense domain and applications on Hopf bifurcation in age structured models, Mem. Amer. Math. Soc. 202 (951) (2009).
[28] P. Magal, S. Ruan, On semilinear Cauchy problems with non-densely defined domain, Adv. Differential Equations 14 (2009) 1041–1084.
[29] P. Magal, S. Ruan, Sustained oscillations in an evolutionary epidemiological model of influenza A drift, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 466 (2010) 965–992.
[30] S. Gupta, N.M. Ferguson, R.M. Anderson, Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents, Science 280 (1998) 912–915.
[31] C.O. Buckee, et al., Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis, Proc. Natl. Acad. Sci. 105 (2008) 15082–15087.
[32] C.O. Buckee, S. Gupta, A network approach to understanding pathogen population structure, in: V. Sintchenko (Ed.), Infectious Disease Informatics, Springer Verlag, New York, 2010.
[33] M. Recker, S. Gupta, A model for pathogen population structure with cross-protection depending on the extent of overlap in antigenic variant repertoires, J. Theoret. Biol. 232 (2005) 363–373.
[34] V. Calvez, et al., Cluster formation for multi-strain infections with cross-immunity, J. Theoret. Biol. 233 (2005) 75–83.
[35] N.M. Ferguson, V. Andreasen, The influence of different forms of cross-protective immunity on the population dynamics of antigenically diverse pathogens, in: S. Blower, C. Castillo-Chavez, K.L. Cooke, D. Kirschner, P. van der Driessche (Eds.), Mathematical Approaches for Emerging and Re-emerging Infections, in: IMA Volumes in Mathematics and its Applications, Springer, New York, 2001.
[36] B.S. Chan, P. Yu, Bifurcation, stability, and cluster formation of multi-strain infection models, J. Math. Biol. 67 (6–7) (2013) 1507–1532.
[37] K.B. Blyuss, The effects of symmetry on the dynamics of antigenic variation, J. Math. Biol. 66 (2013) 115–137.
[38] A. Cherif, Mathematical analysis of a multiple strain, multi-locus-allele system for antigenically variable infectious diseases revisited, Math. Biosci. 267 (2015) 24–40.
[39] V. Andreasen, et al., The dynamics of cocirculating influenza strains conferring partial cross-immunity, J. Math. Biol. 35 (1997) 825–842.
[40] J. Lin, V. Andreasen, S.A. Levin, Dynamics of influenza A drift: the linear three-strain model, Math. Biosci. 162 (1999) 33–51.
[41] J.B. Plotkin, J. Dushoff, S.A. Levin, Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus, Proc. Natl. Acad. Sci. 99 (2002) 6263–6268.
[42] J.R. Gog, J. Swinton, A status-based approach to multiple strain dynamics, J. Math. Biol. 44 (2002) 169–184.
[43] M.G.M. Gomes, G. Medley, D.J. Nokes, On the determinants of population structure in antigenically diverse pathogens, Proc. R. Soc. Lond. Ser. B 269 (2002) 227–233.
[44] A.D. Cliff, P. Haggett, J.K. Ord, Spatial Aspects of Influenza Epidemics, Pion, London, 1986.
[45] S. Gupta, A. Galvani, The effects of host heterogeneity on pathogen population structure, Proc Roy. Soc. B 354 (1999) 711–719.
[46] A. Sasaki, Y. Haraguchi, Antigenic drift of viruses within a host: a finite site model with demographic stochasticity, J. Mol. Evol. 51 (2000) 245–255.
[47] M. Nuno, M. Martcheva, C. Castillo-Chavez, Immune level approach For multiple strain pathogens, J. Biol. Syst. 17 (4) (2009) 713–737.
[48] K.B. Blyuss, Analysis of symmetries in models of multi-strain infections, J. Math. Biol. 69 (2014) 1431–1459.
[49] M. Golubitsky, I. Stewart, A. Torok, Patterns of synchrony in coupled cell networks with multiple arrows, SIAM J. Appl. Dynam. Sys. 4 (1) (2005) 78–100.
[50] M. Iannelli, Mathematical Theory of Age-Structured Population Dynamics, in: Applied Mathematics Monograph C.N.R., vol. 7, Giardini editorie stampatori, Pisa, 1995.
[51] H.R. Thieme, Mathematics in Population Biology, Princeton University Press, New Jersey, 2003.
[52] A. Pazy, Semigroups of Linear Operators and Applications to Partial Differential Equations, Springer, Berlin, 1983.
[53] K.-J. Engel, R. Nagel, One-Parameter Semigroups for Linear Evolution Equations, Springer, Berlin, 2000.
[54] O. Diekmann, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, Wiley, NY, 2000.
[55] H. Brezis, Functional Analysis, Sobolev Spaces, and Partial Differential Equations, in: Universitext, Springer, New York, 2010.
[56] M.G. Krein, M.A. Rutman, Linear operators leaving invariant a cone in a Banach space, Uspekhi Mat. Nauk 3 (1948) 3–95 (in Russian); English translation: Amer. Math. Soc. Transl. 1 (10):199–325, 1950.
[57] I. Marek, Frobenius theory of positive operators: comparison theorems and applications, SIAM J. Appl. Math. 19 (1970) 607–628.
[58] I. Sawashima, On spectral properties of some positive operators, Natur. Sci. Rep. Ochanomizu Univ. 15 (1964) (1964) 53–64.
[59] H. Inaba, A semigroup approach to the strong ergodic theorem of the multi-state stable population process, Math. Popul. Stud. 1 (1) (1988) 49–77.
[60] H. Inaba, Threshold and stability results for an age-structured epidemic model, J. Math. Biol. 28 (1990) 411–434.
[61] M.A. Krasnoselskii, Positive Solutions of Operator Equations, Noordhoff, Groningen, 1964.
[62] N. Biggs, Algebraic Graph Theory, Cambridge University Press, New York, 1974.
[63] C. Godsil, G. Royle, Algebraic Graph Theory, in: Graduate Texts in Mathematics, vol. 207, Springer, New York, 2001.

[64] J. Lauri, R. Scapellato, Topics in Graph Automorphisms and Reconstruction, Cambridge University Press, New York, 2003.
[65] Y. Wang, M. Golubitsky, Two-color patterns of synchrony in lattice dynamical systems, Nonlinearity 18 (2005) 631–657.
[66] M. Golubitsky, I. Stewart, Nonlinear dynamics of networks: the groupoid formalism, Bull. Amer. Math. Soc. 43 (2006) 305–364.
[67] E. Kreyszig, Introductory Functional Analysis, John Wiley & Sons, New York, 1978.
[68] T. Kato, Perturbation Theory for Linear Operators, second ed., Springer, Berlin, 1984.
[69] W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1966.
[70] S. Steinberg, Meromorphic families of compact operators, Arch. Ration. Mech. Anal. 31 (1968) 372–379.
[71] T. Hildebrandt, L. Graves, Implicit functions and their differentials in general analysis, Trans. Amer. Math. Soc. 29 (1927) 127–153.
[72] E. Zeidler, Nonlinear Functional Analysis and its Applications I, Fixed Point Theorems, Springer, New York, 1986.
[73] J.H.P. Dawes, J.R. Gog, The onset of oscillatory dynamics in models of multiple disease strains, J. Math. Biol. 45 (2002) 471–510.