bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Co-colonization interactions drive explicit frequency-dependent dynamics among N types

Sten Madec[a,c], Erida Gjini[b,d]

[a]*Institut Denis Poisson, University of Tours, Tours, France*
[b]*Instituto Gulbenkian de Ciência, Oeiras, Portugal*
[c]*sten.madec@univ-tours.fr*
[d]*egjini@igc.gulbenkian.pt*

Abstract

A fundamental question in ecology is the generation and maintenance of biodiversity. Classical approaches consider multi-species communities with Lotka-Volterra ODE models where inter- vs. intra-species interactions are key. Typically in high-dimensional systems, analysis is hard and model reduction is needed. Here, we describe and study a new system of multi-type interactions that arise in co-colonization, and develop a useful model reduction framework based on slow-fast dynamics under quasi-neutrality. We show that in a multi- type, Susceptible-Infected-Susceptible (SIS) system with co-colonization, neutral coexistence dynamics between N closely related strains occurs on a fast timescale, based on mean trait values, whereas the non-neutral selective forces act on a slow timescale, driven by the variance and asymmetries of the co-colonization interaction matrix. The explicit N equations for relative type frequencies on the slow manifold enable efficient computation of complex multi-strain dynamics, and provide analytical insight into high-dimensional community ecology, with potential applications to other multi-body systems.

Keywords: SIS model, slow-fast dynamics, weak selection, coexistence, multispecies community ecology, broken symmetries

Introduction

A fundamental question in ecology and evolutionary biology relates to how biodiversity is generated and maintained [1, 2, 3, 4]. Many factors have been shown to be important, including inter- vs. intra-species interactions [5], population size [6], number and functional links with resources [7], and movement in space [8]. Most research in this field has considered multi-species interactions with classical Lotka-Volterra models or with evolutionary game theoretic approaches.

In this Letter, we describe and study a new system of multi-type interactions that arise in the epidemiological dynamics of co-colonizing strains of the same microbial species. An example of such a system could be polymorphic *Streptococcus pneumonia* bacteria. We generalize a previously- introduced Susceptible-Infected-Susceptible (SIS) framework for 2 circulating strains with co-colonization [16, 9], to the context of N strains. We show that an explicit reduced system emerges from a time scale separation, expressing the total dynamics as a fast plus a slow component, related to broken symmetries in co-colonization interactions.

In our system, microbial strains can infect a host simultaneously, and here we concentrate only on the case of up to 2 strains co-colonizing a host, whereby single colonization by a resident strain alters the susceptibility to incoming strains, (increasing or decreasing it) by a factor K_{ij}, relative to uninfected hosts, without acquired immunity. The transmission and clearance rates of all strains are equal, except for the interactions in co-colonization given by an $N \times N$ matrix. In earlier theoretical studies such interaction has been studied in a slightly different context in models considering arbitrary infection multiplicity [10]. Later evolutionary frameworks have also considered vulnerability to co-infection [11]. In the experimental literature, for example in microbiota studies, the phenomenon of colonization resistance has been reported [12]. In pneumococcal bacteria, more relevant to our model, carriage of one serotype has been shown to alter the acquisition rate of a second serotype [13]. Some studies have shown how this trait at the host-pathogen interface impacts disease persistence [14], coexistence and vaccination effects [15, 16], and contributes to diversity in other traits, e.g. virulence [17] and antibiotic resistance [18].

Yet, robust mathematical frameworks to analyze the full spectrum of eco-epidemiological patterns emerging exclusively from co-colonization interactions among a high number of strains remain undeveloped. Here, we fill this gap. We model an arbitrary number of closely-related strains, and show that in the N-strain system with altered susceptibilities in co-colonization, neutral coexistence dynamics between types occurs on a fast timescale, whereas the non-neutral selective forces act on a slow timescale, driven explicitly by trait variation in the co-colonization interaction space. These dynamics can be very complex, and are nonlinearly modulated by global parameters such as overall transmission intensity R_0, and mean interaction coefficient between strains. We derive a closed analytic solution for relative type frequencies over long time-scales in a changing fitness landscape.

N-strain SIS model with co-colonization

We consider a multi-type infectious agent, transmitted via direct contact, following susceptible-infected-susceptible

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(SIS) epidemiological dynamics, with the possibility of co-colonization. The number of potentially co-circulating strains is N. With a set of ordinary differential equations, we describe the proportion of hosts in several compartments: susceptibles, S, hosts colonized by one type I_i, and co-colonized hosts I_{ij}, with two types of each combination, independent of the order of their acquisition. We have:

$$\begin{cases} \dot{S} = m(1-S) - S\sum_{j=1}^{N} F_j, \\ \dot{I}_i = F_i S - mI_i - I_i \sum_j K_{ij} F_j, \quad 1 \le i \le N \\ \dot{I}_{ij} = I_i K_{ij} F_j - mI_{ij}, \qquad 1 \le i, j \le N \end{cases} \tag{1}$$

where $F_i = \beta\left(I_i + \sum_{j=1}^{N} \frac{1}{2}(I_{ij} + I_{ji})\right)$ gives the force of infection of strain i. Remark that here we assume that hosts colonized with a mixture of two subtypes i and j transmit either with equal probability. For more information and interpretation of parameters see Table 1 (Supplementary materials 1). Notice that $S = 1 - \sum(I_i + I_{ij})$, thus the dimension of the system is indeed $N + N^2$. In practice, I_{ij} and I_{ji} cannot be distinguished so the dimension is effectively $N + N(N-1)/2$. For the special case of $N = 2$, this model has been described and analyzed elsewhere [16, 9]. We study pathogen diversity only in relation to how strains interact with each other upon co-colonization (K_{ij}), assuming equivalence in transmission β and clearance rate γ (although extension to further asymmetries in these traits is ongoing). The coefficients K_{ij} can be above or below 1, indicating facilitation or competition between any two types in co-colonization. In the above notation $m = \gamma + r$, encapsulating both clearance rate γ of colonization episodes and recruitment rate of susceptible hosts r, equal to the mortality rate from all compartments ($r = d$). For a model summary diagram see Fig.1a-b.

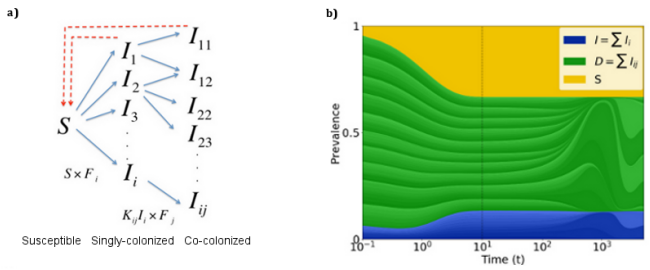

Figure 1: **Model summary diagram. a.** Epidemiological model structure. The force of infection for each strain i is F_i, and governs the transition from susceptible to singly-colonized state, and from singly-colonized to co-colonized state. Clearance happens at equal rate for single and co-colonization, back to the susceptbile state. Co-colonization rate by strain j of singly-colonized hosts with i is altered by a factor K_{ij} relative to uncolonized hosts. **b)** Considering $K_{ij} = k + \varepsilon\alpha_{ij}$, the global epidemiological dynamics can be decomposed in a fast and slow component. On the fast time-scale ($o(1/\varepsilon)$), strains follow neutral dynamics, driven by mean parameters, where S, I and D are conserved. On a slow time-scale, εt, complex non-neutral dynamics between strains takes place, depicted here by the constituent variations within the blue and green.

Results

Re-writing the system in aggregated form

In order to derive a reduced model for such a general N strain system, displaying complicated dynamics (Fig. 1b), we use the following notation:

$$J_i = I_i + \sum_{j=1}^{N} \frac{1}{2}(I_{ij} + I_{ji}) \text{ and } T = \sum_i I_i + \sum_i I_{ii} + \sum_{i \neq j} I_{ij}. \tag{2}$$

for the prevalence of strain i in the population, and the total prevalence of all strains, respectively. Total prevalence satisfies $T = \sum_i J_i$ and the forces of infection are: $F_i = \beta J_i$. With these notations, the system (1) reads :

$$\begin{cases} \dot{S} = m(1-S) - \beta S T, \\ \dot{T} = \beta S T - mT, \\ \dot{I}_i = \beta J_i S - mI_i - \beta I_i \sum_{j=1}^{N} K_{ij} J_j, \quad 1 \le i \le N \\ \dot{J}_i = (\beta S - m) J_i + \frac{\beta}{2} \sum_{j=1}^{N} \left(I_i K_{ij} J_j - I_j K_{ji} J_i\right), \quad 1 \le i \le N \\ \dot{I}_{ij} = I_i K_{ij} \beta_j J_j - mI_{ij}, \qquad 1 \le i, j \le N \end{cases} \tag{3}$$

This system of $2 + N + N^2$ equations admit a triangular structure: i) First we describe block of 2 equations (S, T). ii) Next, we study the block of $2N$ equations (I_i, J_i), which is the most complicated. iii) Lastly, we deal with the block of the N^2 equations of I_{ij}, which is simple once the dynamics of I_i and J_i are known. In the following we describe briefly these three blocks.

The 1st part (S, T).

$$\begin{cases} \dot{S} = m(1-S) - \beta S T, \\ \dot{T} = \beta S T - mT \end{cases}.$$

Clearly, if the basic reproduction number $R_0 = \frac{\beta}{m} > 1$, then $(S, T) \to \left(\frac{1}{R_0}, 1 - \frac{1}{R_0}\right)$. Hence, without loss of generality, we consider that $(S, T) = \left(\frac{1}{R_0}, 1 - \frac{1}{R_0}\right)$.

The 2nd block of equations (I_i, J_i). If we reduce the system to the invariant manifold $(S, T) = (\mathbf{S}, \mathbf{T}) = \left(\frac{1}{R_0}, 1 - \frac{1}{R_0}\right)$, we obtain the following sub-system of $2N$ differential equations:

$$\begin{cases} \dot{I}_i = m(J_i - I_i) - \beta I_i \sum_{j=1}^{N} K_{ij} J_j \\ \dot{J}_i = \frac{\beta}{2} \sum_{j=1}^{N} \left(I_j K_{ji} J_i - I_i K_{ij} J_j\right) \end{cases}, \quad 1 \le i \le N \tag{4}$$

together with: $\sum_{i=1}^{N} J_i = 1 - \frac{1}{R_0} = \frac{\beta - m}{m}$. The dynamics of (4) are complex and we use a slow-fast approach to analyze it.

The co-colonization block I_{ij}. This is the simple system

$$\dot{I}_{ij} = \beta I_i K_{ij} J_j - mI_{ij}, \qquad 1 \le i, j \le N \tag{5}$$

Once it is known that $(I_i, J_j) \to (I_i^*, J_j^*)$, then (5) implies $I_{ij} \to \frac{\beta}{m} I_i^* K_{ij} J_j^*$. Thus, once the dynamics (4)-(5) are explicit, so are the dynamics of (1).

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Slow-fast representation of (4)

Since we model N closely-related strains, we can write each co-colonization coefficient as: $K_{ij} = k+\varepsilon\alpha_{ij}$, where $0 \leq \varepsilon << 1$. Replacing these in (4), and re-arranging, we obtain:

$$\begin{cases} \dot{I} = m(\mathbf{T} - I) - \beta k T^* I - \varepsilon\beta \sum_{i=1}^{N}\sum_{j=1}^{N} I_i \alpha_{ij} J_j \\ \dot{I}_i = m(J_i - I_i) - \beta k \mathbf{T} I_i - \varepsilon\beta I_i \sum_{j=1}^{N} \alpha_{ij} J_j \\ \dot{J}_i = \frac{\beta k}{2}(I J_i - I_i \mathbf{T}) + \frac{\varepsilon\beta}{2}\sum_{j=1}^{N}\left(I_j \alpha_{ji} J_i - I_i \alpha_{ij} J_j\right) \end{cases}, \quad (6)$$

where $1 \leq i \leq N$, and $I = \sum I_j$.

Neutral system

If $\varepsilon = 0$, then from above, we obtain the *Neutral model*

$$\begin{cases} \dot{I} = m\mathbf{T} - (m + \beta k \mathbf{T}) I \\ \dot{I}_i = m J_i - (m + \beta k \mathbf{T}) I_i \\ \dot{J}_i = \frac{\beta k}{2}(I J_i - I_i \mathbf{T}) \end{cases}, \quad 1 \leq i \leq N \quad (7)$$

The first equation gives $I(t) = \mathbf{I} + e^{-t(m+\beta k \mathbf{T})}(I(0) - \mathbf{I}) \to \mathbf{I} := \frac{m\mathbf{T}}{m+\beta k \mathbf{T}}$. Co-colonization prevalence is simply derived as: $\mathbf{D} = \mathbf{T} - \mathbf{I}$. Fixing $I = \mathbf{I}$, yields the (degenerate) linear system:

$$\begin{pmatrix} \dot{I}_i \\ \dot{J}_i \end{pmatrix} = \begin{pmatrix} -(m+\beta k \mathbf{T}) & m \\ -\frac{\beta k \mathbf{T}}{2} & \frac{\beta k \mathbf{I}}{2} \end{pmatrix} \begin{pmatrix} I_i \\ J_i \end{pmatrix} = A_0 \begin{pmatrix} I_i \\ J_i \end{pmatrix}. \quad (8)$$

Matrix A_0 has the two eigenvalues 0 and $-\xi = tr(A_0) < 0$, and eigenvectors $V_0 = \begin{pmatrix} \mathbf{I} \\ \mathbf{T} \end{pmatrix}$, $V_\xi = \begin{pmatrix} 2\mathbf{T} \\ \mathbf{D} \end{pmatrix}$. We define H_i and z_i as:

$$\begin{pmatrix} H_i \\ z_i \end{pmatrix} = \begin{pmatrix} 2\mathbf{T} & \mathbf{I} \\ \mathbf{D} & \mathbf{T} \end{pmatrix}^{-1} \begin{pmatrix} I_i \\ J_i \end{pmatrix}. \quad (9)$$

We have

$$\begin{cases} \dot{H}_i = -\xi H_i \\ \dot{z}_i = 0 \end{cases}.$$

Thus

$$H_i = \frac{\mathbf{IT}}{2(\mathbf{T})^2 - \mathbf{DI}}\left[\frac{I_i}{\mathbf{I}} - \frac{J_i}{\mathbf{T}}\right]$$

tends to zero, and

$$z_i = \left(\frac{I_i}{\mathbf{I}}\right) + \frac{2(\mathbf{T})^2}{2(\mathbf{T})^2 - \mathbf{DI}}\left(\frac{J_i}{\mathbf{T}} - \frac{I_i}{\mathbf{I}}\right),$$

remains constant in a neutrally stable manner. H_i measures the difference between the part of strain i on the single infected versus the part of the strain i on the total number of infected. Thus, $H_i \to 0$ means that the proportion of strain i in single colonization, tends to be the same as the proportion of strain i in the total mass of all colonization. This implies it also tends to be equal to the proportion of strain i in co-colonization. Because on the slow manifold, we have $H_i = 0$, we can infer that during fast dynamics z_i tends to:

$$z_i = \frac{J_i}{\mathbf{T}} = \frac{I_i}{\mathbf{I}}, \quad (10)$$

whereby on the slow manifold z_i describes exactly the frequency of strain i among all hosts that are colonized, with $\sum z_i = 1$.

The slow-fast dynamics

In (6) I plays a role in the order 0 term of the last two equations. Thus, we use the ansatz: $I(t) = \mathbf{I} + \varepsilon X(t) + O(\varepsilon)$, and rewrite (6) (for details see Supplementary Materials 2). After some algebraic manipulations, using newly defined variables we obtain the explicit slow-fast system:

$$\begin{cases} \dot{X} = -(m + \beta k \mathbf{T}) X - \beta Q(\mathbf{H}, \mathbf{z}) + O(\varepsilon) \\ \dot{H}_i = -\xi H_i + O(\varepsilon) \\ \dot{z}_i = \varepsilon\delta\left(g_{i,2}(\mathbf{H}, \mathbf{z}) + \mathbf{T}(\mathbf{D} H_i + \mathbf{T} z_i) X + O(\varepsilon)\right) \end{cases}, \quad (11)$$

for $1 \leq i \leq N$, where $\delta = \frac{\beta}{2(\mathbf{T})^2 - \mathbf{ID}} > 0$ is a constant.
If $\varepsilon \to 0$ then z_i remains constant and $(X(t), \mathbf{H}(t)) \to (\phi(\mathbf{z}), 0)$ (shown in detail in S2). However, on the slow time scale $\tau = \varepsilon t$ in (11) , z_i is not constant, and letting $\varepsilon \to 0$, one obtains the slow dynamics on z_i, given by the N-dimensional system:

$$\dot{z}_i = \delta\left(g_{i,2}(0, \mathbf{z}) + \mathbf{T}^2 z_i \phi(\mathbf{z})\right), \quad 1 \leq i \leq N$$

which reads explicitly

$$\dot{z}_i = \Theta z_i \left(\sum_{\substack{j=1 \\ j \neq i}}^{N} \left[\mu(\alpha_{ji} - \alpha_{ij}) + \alpha_{ji}\right] z_j + \alpha_{ii} z_i - q(z)\right), \quad (12)$$

for $1 \leq i \leq N$, where the constants $\Theta, \mu > 0$ are:

$$\Theta = \frac{\beta \mathbf{TID}}{2\mathbf{T}^2 - \mathbf{ID}}\ ;\ \mu = \frac{\mathbf{I}}{\mathbf{D}} = \frac{1}{k(R_0 - 1)},$$

and the term $q(z)$ is given by:

$$q(z) = \sum_{1 \leq k, j \leq N} \alpha_{kj} z_k z_j.$$

The constant Θ is the rate that sets the tempo of multi-strain "motion" on the slow manifold towards an equilibrium. Notice that Θ depends specifically on the absolute transmission rate of the pathogen β, but also on mean traits R_0 and k, via the conserved aggregated quantities $\mathbf{T}$, $\mathbf{I}$, $\mathbf{D}$. The quantity μ on the other hand represents the ratio between single colonization and co-colonization prevalence from the neutral system, a factor that amplifies the importance of type asymmetry.

The function $q(z)$ represents the time-changing impact of all the strains on their environment, which in turn modifies the fitness landscape for each in an equal manner. A more explicit way to see $q(z)$ is in terms of relative change in 'effective' mean interaction coefficient between all extant types in the system, which if negative, indicates a global trend toward more competition in co-colonization, and if positive, indicates a global trend toward more facilitation. At this stage, we can apply quasi-stationarity methods (see [19, 20, 21, 22]) to show that the solution of (11) tends to the solution of (12) as $\varepsilon \to 0$ locally uniformly in time on $[t_0, +\infty[$ for some $t_0 > 0$.

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Link with pairwise invasion fitnesses λ_i^j

Next, we provide another equivalent very useful representation of the model by using the notion of pairwise invasion fitness [23]. Let λ_i^j be the exponential growth rate of the strain i evaluated when introduced at the trivial endemic equilibrium of the strain j alone. If the fitness $\lambda_i^j > 0$, strain i will invade, and viceversa. By considering the rate of growth of strain i in an equilibrium set by j (dz_i/dt (12) in the special case where all $z_k = 0$ for $k \neq j$), we find the exact formulation of pairwise invasion fitness is given by:

$$\lambda_i^j = \mu(\alpha_{ji} - \alpha_{ij}) + \alpha_{ji} - \alpha_{jj}. \tag{13}$$

We can thus recast the system (12) using only these fitness notations (for details see Supplementary Material 3) :

$$\frac{d}{d\tau}\mathbf{z}_i = \Theta\mathbf{z}_i \cdot \left(\sum_{j\neq i} \lambda_i^j z_j - \sum_{1\leq k\neq j\leq N}\sum \lambda_j^k z_j z_k \right) \tag{14}$$

The above can be made more compact by denoting $\Lambda = (\lambda_i^j)_{i,j}$ the pairwise invasion fitness matrix, and using vector notation:

$$\frac{d}{d\tau}\mathbf{z} = \Theta\mathbf{z} \cdot (\Lambda\mathbf{z} - \mathbf{z}^t\Lambda\mathbf{z}). \tag{15}$$

It is this matrix Λ that defines all 'edges' of the rescaled interaction network between N strains. For $N = 2$, similar to the classical Lotka-Volterra model, in our model, as already shown [9], there are only four possible outcomes between 2 strains (edge linking 1 and 2): **i)** $\lambda_1^2 > 0$, $\lambda_2^1 > 0$: stable coexistence of 1 and 2; **ii)** $\lambda_1^2 < 0$, $\lambda_2^1 < 0$: bistability of 1-only and 2-only; **iii)** $\lambda_1^2 > 0$, $\lambda_2^1 < 0$: 1-only competitive exclusion; **iv)** $\lambda_1^2 < 0$, $\lambda_2^1 > 0$: 2-only competitive exclusion. Knowing all pairwise invasion fitnesses between each couple of strains, via (15) we can reconstitute the ultimate dynamics of the full system with N types and co-colonization.

Going back to the original $N + N(N-1)/2$ variables

Recall the original system with N strains is given by the SIS model with co-colonization interactions (1). While the decomposition $K_{ij} = k + \varepsilon\alpha_{ij}$ is mathematically non-unique, and can be applied with respect to any reference k, provided that the resulting ε is small, a possible convenient choice is to define k as the average of the original interaction matrix entries K_{ij}:

$$k = \frac{\sum_{i,j} K_{ij}}{N^2}, \tag{16}$$

and to define deviation from neutrality, ε, as the root mean square distance of the K_{ij} from the mean interaction coefficient:

$$\varepsilon = \sqrt{\frac{\sum_{i,j}(K_{ij} - k)^2}{N^2}},$$

thus representing the standard deviation of the K_{ij} traits. The direction of deviation from neutrality (bias) for the interaction between strain i and j is then obtained as:

$$\alpha_{ij} = \frac{K_{ij} - k}{\varepsilon}.$$

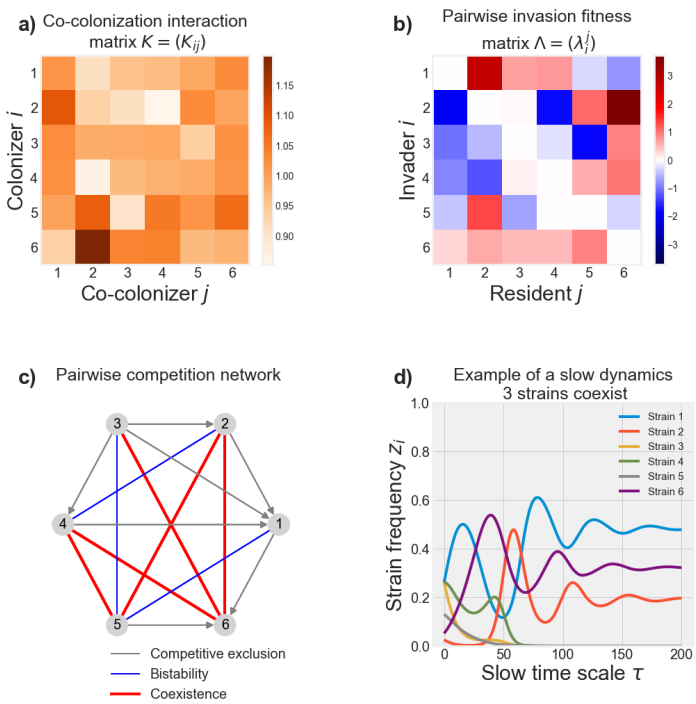

Figure 2: **Example dynamics of our model for $N = 6$. a)** The matrix of interaction coefficients in co-colonization, generated randomly, with mean $k = 1$, and standard deviation $\varepsilon = 0.1$. **b)** The corresponding pairwise invasion fitness matrix for assumed $R_0 = 2$. **c)** The network where each edge displays the outcome of pairwise competition between any couple of strains. **d)** Slow epidemiological dynamics resulting from these interactions. A dynamic display of the trajectory is shown in Supplementary Movie 1.

Thus, the matrix $A = \left(\alpha_{ij}\right)_{1\leq i,j\leq N}$ is the *standardized interaction matrix*, with $\|A\|_2 = \sqrt{\sum_{1\leq i,j\leq N}\alpha_{ij}^2} = N$. This matrix A, and the ratio μ, determine the pairwise invasion fitness matrix (13), which contains nearly all the qualitative information about the non-neutral dynamics (15). So far, we have shown that provided the deviation from neutrality, ε, is small, the behavior of (12) describes very well the long term dynamics of (1). To recover the original variables, after solving for strain frequencies z_i on the slow manifold, as the dynamics of (1) are well approached by those of (12), we can use the relations:

$$S(t) = \mathbf{S} := \frac{m}{\beta} = \frac{1}{R_0}, \quad T(t) = \mathbf{T} := 1 - \mathbf{S} = 1 - \frac{1}{R_0},$$

$$\mathbf{I}(t) := \frac{m\mathbf{T}}{m + \beta k\mathbf{T}} = \frac{\mathbf{T}}{1 + R_0 k\mathbf{T}},$$

to obtain the total prevalence of susceptibles S, total prevalence of colonized individuals T, and prevalence of single colonization I. The prevalence of co-colonization is simply $\mathbf{D} = \mathbf{T} - \mathbf{I}$. Further, to recover strain-specific single colonization, and co-colonization prevalences, we apply:

$$I_i(t) := \mathbf{I}z_i(\varepsilon t), \qquad I_{ij}(t) = k\frac{\mathbf{IT}}{\mathbf{S}}z_i(\varepsilon t)z_j(\varepsilon t).$$

where the slow time scale is $\tau = \varepsilon t$, and the slow variables $\mathbf{z}(\tau) = (z_i(\varepsilon t))_i$ verify $\sum_i z_i(\varepsilon t) = 1$ and follow dynamics (15).

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

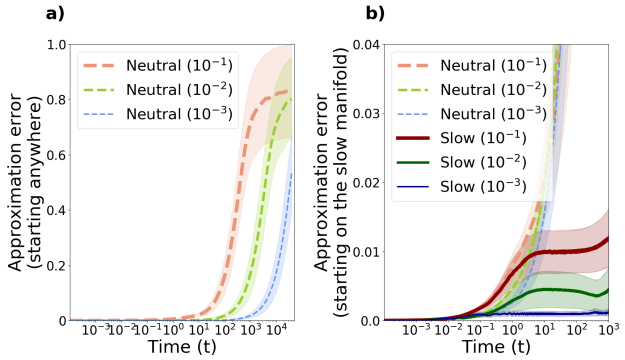

Figure 3: **Accuracy of our approximation over time a)** Error between the neutral model and the full system, starting from random initial conditions. We summarize simulations for different ε, where for each ε, 20 co-colonization matrices of different size ($N \in [2, 15]$) were randomly generated, together with initial conditions. On a fast time-scale ($o(1/\varepsilon)$) the neutral model is a very good approximation to the real system. **b)** Error of the neutral and the slow dynamics approximation, starting from the slow manifold (10). The error (mean over all model variables) between the original system trajectories and the neutral vs. slow dynamics. We generated random co-colonization matrices ($N \in [2, 15]$) for 3 values of ε (10^{-1},10^{-2},10^{-3}) while keeping mean $k = 1$. The K_{ij} were drawn from the normal distribution $N(k, \varepsilon^2)$. On the slow manifold, the slow-dynamics approximate well the original system, unlike the neutral model.

In Figure 2 and in Supplementary Movie 1, we provide illustrations of the modeling framework and some dynamics that are possible for an arbitrary number of strains. Some further illustrations for other values of parameters are given in Supplementary Figure 1 and Supplementary Movie 2.

Accuracy of the approximation and simulation speed

The slow-fast model being a good approximation of the original system, more precisely, means that there exists $t_0 > 0$ such that for any $T > t_0$ and any $\varepsilon > 0$ small enough, there is $M > 0$ such that for any $\tau \in (t_0, T)$, the error between the two is small: $\|S\left(\frac{\tau}{\varepsilon}\right) - \mathbf{S}\| + \sum_{i=1}^{N} \|I_i\left(\frac{\tau}{\varepsilon}\right) - \mathbf{I}z_i(\tau)\| + \sum_{1\le i,j\le N} \|I_{ij}\left(\frac{\tau}{\varepsilon}\right) - \frac{\mathbf{IT}}{\mathbf{S}} z_i(\tau)z_j(\tau)\| \le M\varepsilon$. We confirm this with numerical simulations in Figure 3a)-b), where the neutral model is shown to be a good approximation of the original system in a fast time-scale, and the slow-dynamics reduction a good approximation on the longer time-scale. Furthermore, we verify that the slow-fast approximation enables exploration of dynamics in a much more computationally efficient manner over a very large number of strains N (Fig.4), which is difficult with the original system.

Discussion

In this study, we investigated how co-colonization interaction networks drive coexistence in a system with N closely related types. We derived a slow-fast decomposition for the dynamics, which links explicitly variation in the interaction matrix with the slow time scale under which multi-type selective dynamics unfold. Model reduction for small deviations from neutrality allowed us to express strain frequencies on the slow manifold via only N equations (instead of $N + N(N-1)/2$ equations in the full system). These can be easily re-mapped to the more complex epidemiological variables of the original system. The slow variables z_i ($1 \le i \le N$), describing relative strain frequencies in the host population, necessarily equal in single and co-colonization, could be used as a practical test for quasi-neutrality, when strain prevalence data are available.

Our focus here was on presenting the timescale decomposition method for this system. The wider and more complete ecological picture, as well as the analysis of diversity-stability relationships in coexistence will be the focus of another paper [24]. We believe this work makes a crucial step towards the full characterization of conservative multi-type SIS dynamics.

The slow-fast framework allows to reduce the complexity of multi-strain systems, and understand transient and long-term coexistence on conserved manifolds near neutrality. We expect our results to stimulate further progress for the investigation of coexistence and evolution in multi-strain epidemiological systems such as the one of pneumococcal bacteria [25], where explicit mathematical results linking neutral and niche mechanisms, for N large, are thusfar missing. Although motivated by such a pathogen with altered susceptibilities in co-colonization, the global transmission dynamics captured here have broader implications. As an example, information on the resilience of a group of strains may be derived from the sign of $q(z)$ in our system (12): if $q(z) > 0$ then each strain is less competitive within the group but the overall community is more resistant to invasion by a new strain. Our model's generality and closed analytical form for relative type dynamics provide a new bridge between epidemiology and community ecology, invite comparison with the classical Lotka-Volterra model, and offer extension to other multi-body problems in physics and genetics.

We acknowledge support by the University of Tours and EURAXESS for an invited researcher stay of Erida Gjini at IDP, in Tours during 2019.

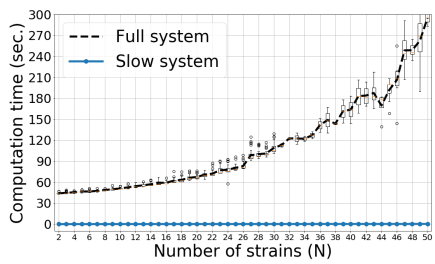

Figure 4: **Computation speed of the approximation.** Computation time increases as a function of number of strains N in the full system (black dashed line) but not in the slow-fast dynamic aproximation (blue solid line) for $T = 5000$ time units. The line shows the mean over 20 random matrices for each N, with $\varepsilon = 0.1$, $k = 1$ and $R_0 = 2$. For details see Supplementary Figure 2.

References

[1] G. F. Gause, *The struggle for existence* (Williams and Wilkins, Baltimore, MD, 1934).

[2] R. H. MacArthur, *The theory of island biogeography*, vol. 1 (Princeton University Press, 1967).

[3] D. Tilman, Proceedings of the National Academy of Sciences USA **101**, 10854 (2004).

[4] S. P. Hubbell, *The unified neutral theory of biodiversity and biogeography (MPB-32)*, vol. 32 (Princeton University Press, 2001).

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[5] D. Tilman, The American Naturalist **129**, 769 (1987).
[6] C. Taylor, D. Fudenberg, A. Sasaki, and M. A. Nowak, Bulletin of mathematical biology **66**, 1621 (2004).
[7] R. A. Armstrong and R. McGehee, Theoretical population biology **9**, 317 (1976).
[8] M. A. Nowak and R. M. May, Nature **359**, 826 (1992).
[9] E. Gjini and S. Madec, Theoretical Ecology **10**, 129 (2017).
[10] F. R. Adler and R. C. Brunet, Theoretical population biology **40**, 369 (1991).
[11] S. Alizon, Interface Focus (2013).
[12] R. A. Britton and V. B. Young, Trends in microbiology **20**, 313 (2012).
[13] M. Lipsitch, O. Abdullahi, A. D'Amour, et al., Epidemiology **23**, 510 (2012).
[14] M. Gaivão, F. Dionisio, and E. Gjini, Bulletin of mathematical biology **79**, 2068 (2017).
[15] M. Lipsitch, PNAS **94**, 6571 (1997).
[16] E. Gjini, C. Valente, R. Sá-Leão, and M. G. M. Gomes, Journal of Theoretical Biology **388**, 50 (2016).
[17] S. Alizon, J. C. de Roode, and Y. Michalakis, Ecology letters **16**, 556 (2013).
[18] N. G. Davies, S. Flasche, M. Jit, and K. E. Atkins, Nature ecology & evolution **3**, 440 (2019).
[19] A. N. Tikhonov, Matematicheskii sbornik **73**, 575 (1952).
[20] C. Lobry and T. Sari, Electronic Journal of Differential Equations **19**, 1 (1998).
[21] C. Lobry and T. Sari, Contrôle non linéaire et Applications **64**, 155 (2005).
[22] F. Hoppensteadt, Transactions of the American Mathematical Society **123**, 521 (1966).
[23] G. Meszéna, M. Gyllenberg, F. J. Jacobs, and J. A. Metz, Physical review letters **95**, 078105 (2005).
[24] E. Gjini and S. Madec, in prep. (2019).
[25] S. Cobey and M. Lipsitch, Science (New York, NY) **335**, 1376 (2012).
[26] K. Dietz, Statistical methods in medical research **2**, 23 (1993).

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPLEMENTARY MATERIAL

Co-colonization interactions drive explicit frequency-dependent dynamics among N-types

by *Sten Madec* and *Erida Gjini*

1. Table of illustrative model parameters for a biological system (e.g. pneumococcus bacteria transmission)

Table S1: Model parameters and their biological interpretation.

Parameter	Interpretation	Features	Illustrative value	Reference
β	Transmission rate of pathogen	$\beta > 0$	2.3 (month)$^{-1}$	[16]
K_{ij}	Interaction coefficient between strains i and j in co-colonization	Alteration of susceptibility to j when colonized by i $\begin{cases} K_{ij} \in [0,1] \text{ competition} \\ K_{ij} > 1 \text{ facilitation} \end{cases}$	0.1	[13]
r	Susceptible recruitment rate	$r > 0$	0.02 (month)$^{-1}$	[16]
d	Natural host mortality rate	$d = r$	0.02 (month)$^{-1}$	
γ	Clearance rate of colonization	$\gamma > 0$	0.8 (month)$^{-1}$	[16]
m	Net outflow rate from the colonization state	Inverse duration of any colonization episode	$m = r + \gamma$	
N	Number of strains in the system	Arbitrary, $N > 1$	30-60	[13]
R_0	Basic reproduction number of the pathogen, $R_0 = \beta/m$, here assumed equal among strains	$\begin{cases} R_0 < 1 \text{ Disease extinction} \\ R_0 > 1 \text{ Persistence} \end{cases}$	2-3	[26],[16]

2. Derivation details for the slow-fast dynamics

Using the ansatz: $I(t) = \mathbf{I} + \varepsilon X(t) + O(\varepsilon)$, system (6) reads

$$\begin{cases} \dot{X} = -(m + \beta k \mathbf{T})X - \beta \sum_{i=1}^{n} \sum_{j=1}^{n} I_i \alpha_{ij} J_j + O(\varepsilon) \\ \dot{\begin{pmatrix} I_i \\ J_i \end{pmatrix}} = A \begin{pmatrix} I_i \\ J_i \end{pmatrix} + \frac{\varepsilon \beta k}{2} X J_i \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \varepsilon \beta \mathbf{W}_i(\mathbf{I}, \mathbf{J}) + O(\varepsilon) \end{cases}$$

wherein we have denoted:

$$\mathbf{W}_i(\mathbf{I}, \mathbf{J}) = \begin{pmatrix} -\sum_j I_i \alpha_{ij} J_j \\ \frac{1}{2} \sum_j I_j \alpha_{ji} J_i - I_i \alpha_{ij} J_j \end{pmatrix}$$

The final slow fast form is given by using the new variables H_i and z_i defined in (9). Hence, define

$$\begin{pmatrix} h_{i,1}(\mathbf{I}, \mathbf{J}) \\ h_{i,2}(\mathbf{I}, \mathbf{J}) \end{pmatrix} = \begin{pmatrix} 2\mathbf{T} & \mathbf{I} \\ \mathbf{D} & \mathbf{T} \end{pmatrix}^{-1} \begin{pmatrix} -2 \sum_j \alpha_{ij} J_j \\ \sum_j I_j \alpha_{ji} J_i - I_i \alpha_{ij} J_j \end{pmatrix}$$

and let us introduce the $\mathbf{G}_i = (g_{i,1}, g_{i,2})^T$ as [1]

$$g_{i,s}(\mathbf{H}, \mathbf{z}) = (2(\mathbf{T})^2 - \mathbf{ID}) h_{i,s}(2\mathbf{TH} + \mathbf{Iz}, \mathbf{DH} + \mathbf{Tz}), \qquad s = 1, 2.$$

and

$$Q(\mathbf{H}, \mathbf{z}) = \sum_{i,j} \alpha_{i,j} (2\mathbf{T}H_i + \mathbf{I}z_i) \left(\mathbf{D}H_j + \mathbf{T}z_j \right).$$

we have

$$\dot{X} = -(m + \beta k \mathbf{T})X - \beta Q(\mathbf{H}, \mathbf{z}) + O(\varepsilon)$$

[1] Denoting $\mathbf{G} = (\mathbf{G}_i)_{1 \le i \le N}$ and $\mathbf{W} = (\mathbf{w}_i)_{1 \le i \le N}$, this simply means that $\mathbf{G} = \phi^{-1} \circ \mathbf{W} \circ \phi$ where $\phi(\mathbf{H}, \mathbf{z}) = (\mathbf{I}, \mathbf{J})$ is the linear change of variables defined in (9).

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Finally, using these new unknows, one obtains the expression of the slow fast system (11):

$$\begin{cases} \dot{X} = -(m+\beta k\mathbf{T})X - \beta Q(\mathbf{H},\mathbf{z}) + O(\varepsilon) \\ \dot{H}_i = -\mu H_i + O(\varepsilon) \\ \dot{z}_i = \varepsilon\delta\left(g_{i,2}(\mathbf{H},\mathbf{z}) + \mathbf{T}(\mathbf{D}H_i + \mathbf{T}z_i)X + O(\varepsilon)\right) \end{cases}, \quad 1 \le i \le N \tag{17}$$

where $\delta = \frac{\beta}{2(\mathbf{T})^2 - \mathbf{ID}}$ is a positive constant. If $\varepsilon \to 0$ then z_i remains constant and $(X(t), \mathbf{H}(t)) \to (\phi(\mathbf{z}), 0)$ where:

$$\phi(\mathbf{z}) = -\frac{\beta Q(0,\mathbf{z})}{m+\beta k\mathbf{T}} = -\frac{\beta\mathbf{IT}}{m+\beta k\mathbf{T}}\sum_{i,j}\alpha_{ij}z_iz_j. \tag{18}$$

On the slow time scale $\tau = \varepsilon t$ in (11) , and letting $\varepsilon \to 0$, one obtains the slow dynamics

$$\begin{cases} 0 = -(m+\beta k\mathbf{T})X - \beta Q(\mathbf{H},\mathbf{z}) \\ 0 = -\mu H_i \\ \dot{z}_i = \delta\left(g_{i,2}(\mathbf{H},\mathbf{z}) + k(\mathbf{T})^2 z_i X\right) \end{cases}, \quad 1 \le i \le N \tag{19}$$

Taking $(X, \mathbf{H}) = (\phi(\mathbf{z}), 0)$ yields the slow dynamics reducted to the slow manifold

$$\dot{z}_i = \delta\left(g_{i,2}(0,\mathbf{z}) + \beta k\mathbf{T}\phi(\mathbf{z})\right), \quad 1 \le i \le N$$

where

$$\begin{aligned} g_{i,2}(0,\mathbf{z}) &= \mathbf{IT}\sum_{j=1}^{N}\left[\mathbf{D}z_i\alpha_{i,j}z_j + \mathbf{T}(z_j\alpha_{j,i}z_i - z_i\alpha_{i,j}z_j)\right] \\ &= \mathbf{ITD}\sum_{j=1}^{N}\left(\frac{\mathbf{T}}{\mathbf{D}}\alpha_{j,i} - \frac{\mathbf{I}}{\mathbf{D}}\alpha_{i,j}\right)z_iz_j \end{aligned} \tag{20}$$

$\phi(\mathbf{z})$ is given by (18) and verifies

$$k\mathbf{T}^2\phi(\mathbf{z}) = -\mathbf{IT}\frac{k\mathbf{T}^2}{m+\beta k\mathbf{T}}\sum_{i,j}\alpha_{ij}z_iz_j = -\mathbf{ITD}\sum_{i,j}\alpha_{ij}z_iz_j$$

Denoting $\mu = \mathbf{I}/\mathbf{D}$ and $\Theta = \dfrac{\beta\mathbf{TID}}{2(\mathbf{T})^2 - \mathbf{ID}}$, the explicit slow dynamics is then given by system (12) which may be rewritten as

$$\frac{d}{d\tau}\mathbf{z} = \Theta\mathbf{z}\cdot(\mu(A^t - A)\mathbf{z} + A^t\mathbf{z} - \mathbf{z}^tA\mathbf{z}). \tag{21}$$

Remark that we have:

$$\frac{d}{d\tau}\sum_j z_j = \Theta\sum_{i,j}\alpha_{i,j}z_iz_j\left(1 - \sum_j z_j\right). \tag{22}$$

which implies that the simplexe $\mathcal{P} = \left\{\mathbf{z} \in \mathbb{R}_+^N,\ \sum z_j = 1\right\}$ is invariant under (12).

3. Characterization of the model via invasion fitnesses

We focus on the trivial solutions of (12), which are the extreme competitive exclusion equilibria $E_i = (\delta_{ik})_{1\le k\le n}$ where δ_{ik} is the Kronecker delta: $\delta_{ik} = \begin{cases} 1 \text{ if } i = k, \\ 0 \text{ if } i \ne k. \end{cases}$

In particular, we note λ_i^j the invasion fitness of the strain i with respect to the trivial solution E_j. More precisely, λ_i^j is the per-capita growth rate of the strain i evaluated at E_j. Since the i-th line of the system reads $\frac{d}{d\tau}z_i = z_ih_i(z_1, \cdots, z_n)$, we have the explicit formula:

$$\lambda_i^j = h_i(E_j)$$

that is

$$\lambda_i^j = \mu(\alpha_{ji} - \alpha_{ij}) + \alpha_{ji} - \alpha_{jj}.$$

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We have that the line i of (21) is

$$\dot{z}_i = z_i \left(\sum_{\substack{j=1 \\ j\neq i}}^{N} \Big(\mu(\alpha_{ji} - \alpha_{ij}) + \alpha_{ji} \Big) z_j + \alpha_{ii} z_i - \sum_{j=1}^{N} \sum_{k=1}^{N} \alpha_{kj} z_k z_j \right).$$

With the notations introduced above, this can be rewritten as:

$$\dot{z}_i = z_i \left(\sum_{\substack{j=1 \\ j\neq i}}^{N} \lambda_i^j z_j + \sum_{j=1}^{N} \alpha_{jj} z_j - \sum_{j=1}^{N} \sum_{k=1}^{N} \alpha_{kj} z_k z_j \right). \tag{23}$$

From $\sum_{j=1}^{N} z_j = 1$, we infer

$$\begin{aligned} \sum_{j=1}^{n} \alpha_{jj} z_j - \sum_{k,j} \alpha_{kj} z_k z_j &= \sum_{j=1}^{N} z_j (\alpha_{jj} - \sum_{k=1}^{N} \alpha_{kj} z_k) \\ &= \sum_{j=1}^{N} z_j (\alpha_{jj} \sum_{k=1}^{N} z_k - \sum_{k=1}^{N} \alpha_{kj} z_k) \\ &= \sum_{j=1}^{N} \sum_{k=1}^{N} z_j (\alpha_{jj} - \alpha_{kj}) z_k \end{aligned}$$

finally, the fact that $\sum_{j=1}^{N} \sum_{k=1}^{N} (\alpha_{jk} - \alpha_{kj}) z_j z_k = 0$ yields

$$\sum_{j=1}^{n} \alpha_{jj} z_j - \sum_{k,j} \alpha_{kj} z_k z_j = - \sum \sum_{1 \leq k \neq j \leq N} \lambda_j^k z_j z_k.$$

Plugging this into (23), we obtain a new version of system (21), enterely parametrized by the fitnesses λ_i^j :

$$\begin{cases} \frac{d}{d\tau} z_i = \Theta z_i \cdot \left(\sum_{j\neq i} \lambda_i^j z_j - \sum \sum_{1 \leq k \neq j \leq N} \lambda_j^k z_j z_k \right), \quad i = 1, \cdots, N \\ z_1 + \cdots + z_N = 1. \end{cases} \tag{24}$$

or more shortly by noting $\lambda_i^i = 0$ and the *pairwise invasion fitness matrix* $\Lambda = (\lambda_i^j)_{1 \leq i,j \leq N}$:

$$\begin{cases} \frac{d}{d\tau} \mathbf{z} = \Theta \mathbf{z} \cdot (\Lambda \mathbf{z} - \mathbf{z}^t \Lambda \mathbf{z}) \\ z_1 + \cdots + z_N = 1. \end{cases} \tag{25}$$

3. Supplementary Figures and Movies

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Movie 1. Here we illustrate the dynamic trajectory of the system (Figure 2 of the paper). A. Network of pairwise invasion fitnesses between strains, where edges can be of 4 types: coexistence (red), bistability (blue), exclusion of i, exclusion of j (directed gray arrows toward the winner). The system trajectory z_i is shown as a movement in a 6-dimensional space. B. Trajectory of the strain frequencies over time.

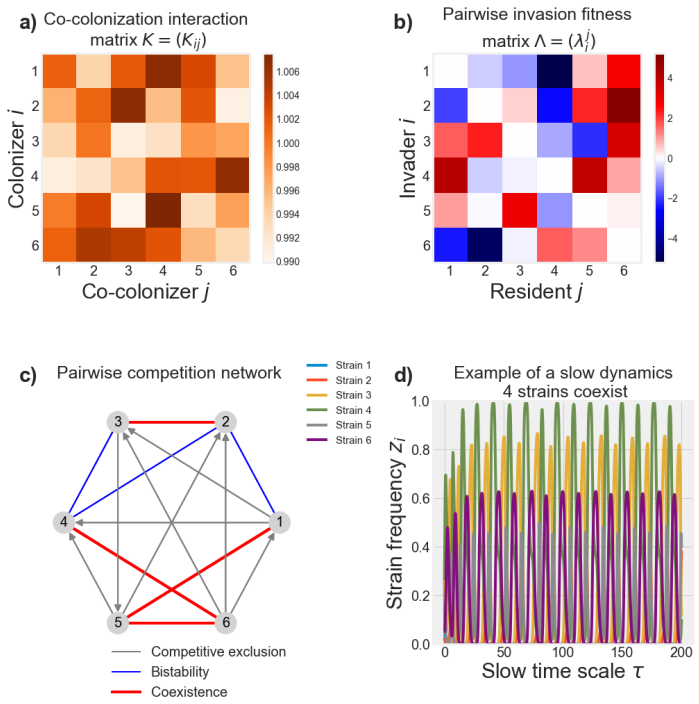

Figure S1: **Illustration of more complex dynamics**. Here we simulate the model in its reduced form for another set of parameters, leading to limit cycle coexistence between 5 strains in an $N = 6$ system. See associated movie S2.

bioRxiv preprint doi: https://doi.org/10.1101/722587; this version posted August 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Movie 2. Here we illustrate the dynamic trajectory of the system (Figure 1 of the Supplements). A. Network of pairwise invasion fitnesses between strains, where edges can be of 4 types: coexistence (red), bistability (blue), exclusion of i, exclusion of j (directed gray arrows toward the winner). The system trajectory z_i is shown as a movement in a 6-dimensional space. B. Trajectory of the strain frequencies over time.

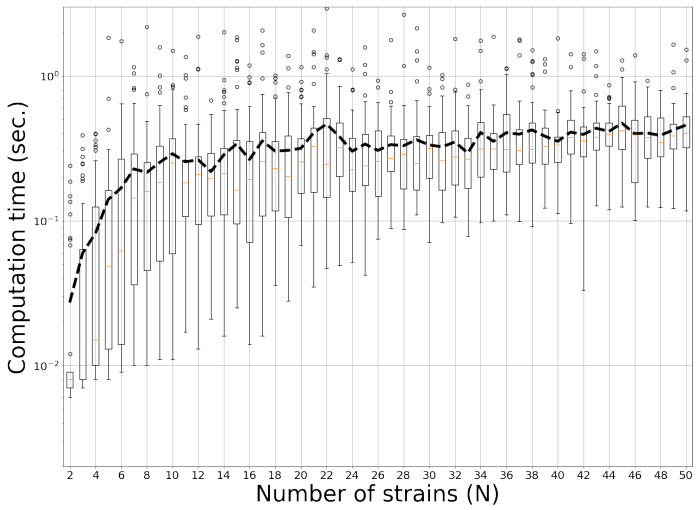

Figure S2: **Supplementary Figure 2. Computation speed of the slow-dynamics approximation in detail.** Here we plot the computation time as a function of number of strains N, in terms of number of seconds on a PC to simulate the slow-fast dynamic aproximation (boxplots) for $T = 5000$ time units starting from the slow manifold. The line shows the mean over 20 random matrices for each N, where $\varepsilon = 0.1$ and mean co-colonization coefficient $k = 1$ and $R_0 = 2$.