Unpacking fitness differences between invaders in a multispecies context

Tomas Freire[1], Sten Madec[2], and Erida Gjini[1,*]

[1]Center for Computational and Stochastic Mathematics, Instituto Superior Tecnico, Lisbon, Portugal
[2]Institut Denis Poisson, University of Tours, Tours, France
[*]erida.gjini@tecnico.ulisboa.pt

November 7, 2024

Abstract

Ecosystems are constantly exposed to new strains of a given species. Which strain will be able to invade a resident multi-species community depends on the strain's relative fitness. Here we develop an approach to quantify fitness differences between two co-invading strains based on the replicator equation. By assuming that the resident species' frequencies remain constant during the initial phase of invasion, we are able to determine the invasion fitness differential between the two strains, which drives growth rate differences post-invasion. We then apply our approach to a critical current global problem: invasion of the gut microbiota by antibiotic resistant strains of the pathobiont *Escherichia coli*. Our results underscore the context-dependent nature of fitness and demonstrate how species frequencies in a host environment can explicitly modulate the selection coefficient between two strains. This framework can predict relative fitness in new environments and may contribute to design strategies to lower resistance levels in microbiomes.

1 Introduction

Understanding fitness differences between invaders within a multi-species context is a central question in evolutionary biology and ecology. The ability of different species or strains to invade and establish themselves in a new environment often depends on subtle differences in growth rates, which can be influenced by interactions with the resident species [1–3]. In this report, we present a theory-driven model framework [4] aimed at quantifying these fitness differences between two co-invading strains. We apply our framework to a series of co-invasion experimental data with *E.coli* strains in mice [5], quantifying fitness costs between drug-resistant and wild-type strains. This research aims to better understand how the composition of resident microbial communities influences initial selection dynamics, and whether it can be used to predict and steer competition in future environments, including for ecosystem control or biomedical applications.

2 Methods

2.1 Co-invasion in a multispecies context

The classical measurement of the selection coefficient for two strains, in a co-invasion scenario, can yield different values depending on the environment. Variations in resources, environmental conditions, and biotic or abiotic factors can all influence the outcome, as each of the two invaders may differ in relating to such gradient. A particular case of environmental variation is a multispecies context, studied for example by [5], where variation in co-invasion outcomes can be seen as a final consequence of host microbiota composition. Although this study reported different fitness costs for the same antibiotic-resistant strain relative to WT

strain in different mice, and proposed a resource-consumer model, it did not fully close the explanatory loop for linking specific microbiota variation to that in selection coefficients. Here, we aim to close this gap, using and highlighting a recently-derived replicator framework [4, 6].

2.2 Empirical invasion of two invaders

When two strains are mixed in equal proportions (1:1) and grow independently, their relative abundances initially follow the equation:

$$(r_A - r_B)t = \ln\left(\frac{n_A}{n_B}\right), \tag{1}$$

where r_A and r_B are their respective growth rates. From this linear relation with time, we can obtain the selection coefficient between the two strains, $s = r_A - r_B$.

2.3 A replicator equation model for multispecies dynamics

The frequency dynamics of N species (z_i) is modeled via a replicator equation framework [4, 6]:

$$\frac{dz_i}{d\tau} = \Theta z_i \left(\sum_{j \neq i} \lambda_i^j z_j - \sum_{1 \leq k \neq j \leq N} \sum \lambda_k^j z_j z_k \right). \tag{2}$$

In this equation, τ represents time, Θ a speed constant, and λ_i^j the pairwise initial growth rates between any species i and j, when i grows in an equilibrium set by j alone, also known as pairwise invasion fitness [7]. This replicator model has an equivalent Lotka-Volterra model for N species with equal growth rates, and interaction matrix (a_{ij}), where $\lambda_i^j = a_{ij} - a_{jj}$ for any two species.

2.4 Scenario of 2 strains invading a multi-species system

The initial growth rate of an invader in a system at state $z(\tau)$ depends on both invader and system traits and is given by:

$$r_{invader} = \sum_j \underbrace{\lambda_{inv}^j}_{\text{invader traits}} z_j - \underbrace{Q(z)}_{\text{system invasion resistance}} \tag{3}$$

where λ_{inv}^j, depends on invader traits, namely how this entity interacts with each existing species j in the system at time of invasion. Here, Q depends on the system as a whole and is equal for all invaders [6]. With this approach, and under the assumption of a linear selection coefficient, we attribute this linearity to the constant frequencies of the resident species in the system during the initial invasion period.

When 2 strains, A and B (at relative ratio 1:1) co-invade a multi-species host environment, the initial difference in their growth rates is:

$$r_A - r_B = \sum_{j=1}^{N} (\lambda_A^j - \lambda_B^j) z_j = \mathbf{z}\Delta_\Lambda, \tag{4}$$

where the common quadratic term Q has cancelled out, and where $\mathbf{z}$ is the row vector of species frequencies in the resident, and Δ_Λ is the column vector invasion fitness differences between the two strains relative to resident species. Assuming low initial frequencies ($z_A = z_B \ll \sum_j z_j \approx 1$), compared to those of the residents, the initial influence of the two invaders can be neglected on the system. Under many realizations of such co-invasion in different resident systems, we have: $\mathbf{s} = F\Delta_\Lambda$ where $\mathbf{s} = (r_A - r_B)_{i=1..M}$ are observed growth rate differences between A and B in M different hosts, and F the $M \times N$ matrix of their multi-species frequency compositions.

Table 1: **Invasion fitness differential between two invading strains of E.coli relative to host microbiota.** Solution of a linear system for Data 1, and regression estimates (plus 95%CI) for Data 2 ($R^2 = 0.999$), Data 3 ($R^2 = 0.999$), and 1+2 jointly fitted ($R^2 = 0.976$). A fitness cost ($\Delta_\Lambda^j < 0$) for each antibiotic-resistant strain is confirmed relative to WT, with respect to any phyla ($j = B, P$ or V), except for the rifampicin-resistant strain relative to species $j = F$, displaying a positive signal in Data 1 and 2, separately and jointly, but without reaching individual significance.

Host microbiota Species i	**Data 1** Coefficients (Δ_Λ^j)	**Data 2** Coefficients (Δ_Λ^j)	**Data 3** Coefficients (Δ_Λ^j)	**Data 1+2** Coefficients (Δ_Λ^j)
Bacteroidetes (B)	-0.83	-1.78 $[-10.17, 6.61]$	-1.26^{***} $[-1.5, -1.02]$	-1.16^{***} $[-1.80, -0.52]$
Proteobacteria (P)	-2.17	-9.17 $[-73.97, 55.63]$	-1.02^{**} $[-1.61, -0.44]$	-2.92^{**} $[-5.08, -0.75]$
Verrucomicrobia (V)	-2.77	-2.00 $[-8.05, 4.04]$	$-$	-2.26^{*} $[-3.99, -0.52]$
Firmicutes (F)	0.22	8.98 $[-15.92, 33.89]$	-3.32^{*} $[-6.45, -0.19]$	1.07 $[-1.23, 3.36]$

3 Results

3.1 Linking growth rate differences to microbiota compositions

We assume the frequencies of N resident species are measured just prior to invasion, in M replicates (hosts) of the same co-invasion with 2 strains, collected in data matrix F. The empirical initial growth rate differential between strains A and B, is denoted by a vector $\mathbf{s}$, which links to F via:

$$F\Delta_\Lambda = \mathbf{s} \tag{5}$$

$$\begin{pmatrix} z_1^{(1)} & \cdots & z_N^{(1)} \\ z_1^{(2)} & \cdots & z_N^{(2)} \\ \vdots & & \vdots \\ z_1^{(M)} & \cdots & z_N^{(M)} \end{pmatrix} \begin{pmatrix} \lambda_A^1 - \lambda_B^1 \\ \lambda_A^2 - \lambda_B^2 \\ \vdots \\ \lambda_A^N - \lambda_B^N \end{pmatrix} = \begin{pmatrix} \hat{r_A}^{(1)} - \hat{r_B}^{(1)} \\ \hat{r_A}^{(2)} - \hat{r_B}^{(2)} \\ \vdots \\ \hat{r_A}^{(N)} - \hat{r_B}^{(N)} \end{pmatrix}.$$

Thus, each row of F is the frequency vector of multispecies composition in each host (assumed constant), Δ_Λ the invasion fitness difference vector between the two strains, and $\mathbf{s}$ the observed selection coefficient in each host. The goal is to estimate the vector Δ_Λ. When $N = M$, we can obtain Δ_Λ as an exact solution of the linear system: $\Delta_\Lambda = F^{-1}\mathbf{s}$ (requires F to be invertible). When $N < M$, the number of species is less than or equal to the number of hosts, we can use least squares regression: $\Delta_\Lambda = (F^T F)^{-1} F^T \mathbf{s}$. This allows us to leverage many classical statistical results regarding the uncertainty of estimation and prediction. When $N > M$, the number of species exceeds that of hosts, and one can proceed with some aggregation of species to reduce the number of variables, and apply the same method.

3.2 From model to data: *E.coli* strains invade different microbiota

We extracted and prepared data from [5], and applied the model to 3 co-invasion experiments: 1) Rif. vs. WT in single-caged mice, 2) Rif. vs. WT in co-housed mice, and 3) Strep. vs. WT in co-housed mice (see Supplementary Dataset and Fig.1), with microbiota resolution at the phylum level ($N = 4$ species). We also combined data 1 and 2 under a joint regression, as they shared the same co-invaders. This led to a larger sample size ($M = 9$), hence more statistical power. Results are in Table 1, revealing the ecological interaction basis of fitness cost in these antibiotic-resistant strains.

3.3 Quality of estimation and prediction

From the regression framework, we obtain explicitly the uncertainty around each coefficient as: $var(\hat{\Delta}_\Lambda^j) = SSR/(M-N)(F^T F)^{-1}_{jj}$, where SSR is the sum of squared residuals. In a new microbiota composition c, we can predict the selection coefficient between the two strains as: $\hat{\theta} = \mathbf{c}\hat{\Delta}_\Lambda$, where in mean the standard error is $s_{\hat{\theta}} = \hat{\sigma}\sqrt{c(F^T F)^{-1}c^T}$, with $\hat{\sigma} = \sqrt{\frac{SSR}{M-N}}$, while for a particular case, $s_d = \hat{\sigma}\sqrt{1 + c(F^T F)^{-1}c^T}$, depending both on past (F) and current (c) data (See SI Text 2 for details).

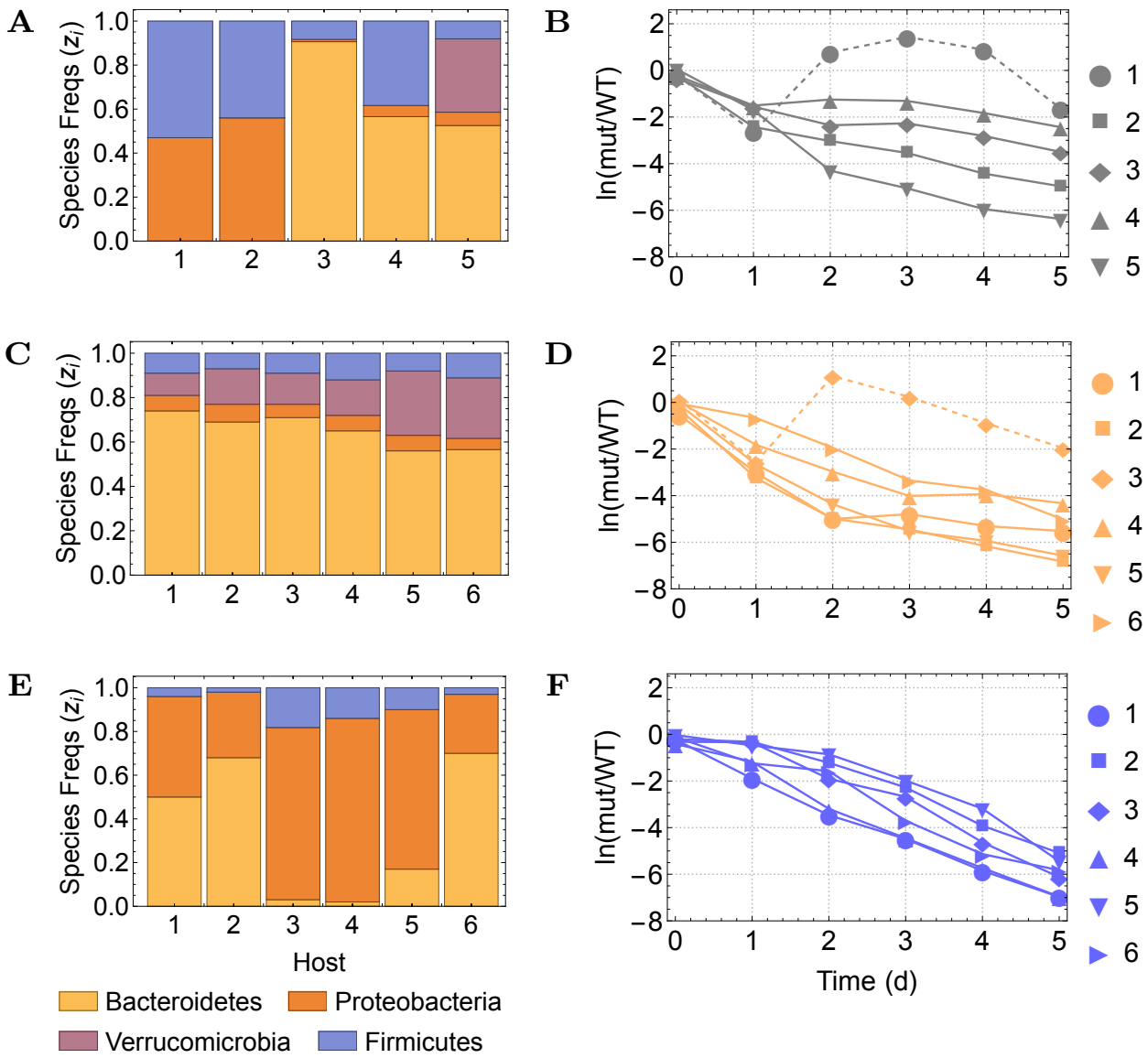

Figure 1: **Microbiota compositions and initial relative growth between two invaders. Data publicly available and extracted from experiments with E.coli strains in mice [8]. A-B** Data 1 ($N = 5$, $M = 4$): Mice were single-housed and injected with WT + a rifampicin-resistant mutant; **C-D** Data 2 ($N = 6$, $M = 4$): Mice were co-housed and injected with WT + a rifampicin-resistant mutant; **E-F** Data 3 ($N = 6$, $M = 3$): Mice were co-housed and injected with WT + a streptomycin-resistant mutant. Left panels show host microbiota compositions just prior to invasion. Right panels show plots of $ln(Res/Wt)$. The slopes $r_A - r_B$ were estimated from the linear fit with time, and their significance, assessed by an F-test. Only hosts with a significant non-zero slope were retained for analysis (mouse 1 in Data 1 and mouse 3 in Data 2 were removed, depicted in dashed lines in B,D) . For details see SI dataset and SI Text 1.

3.4 Validity test and applications

Having quantified Δ_Λ, including its uncertainty, we could predict $r_A - r_B$ when the same strains A and B invade other hosts, with any microbiota compositions. To illustrate this predictive power, we validated our method on the same dataset, but utilizing a *leave-one-out cross-validation* approach (Fig.2, SI Text 3), which confirmed very good regression performance. Yet, even in this small dataset, sensitivity of linear regression to outliers or multicollinearity may lead to inaccurate predictions, warranting for care and preliminary tests.

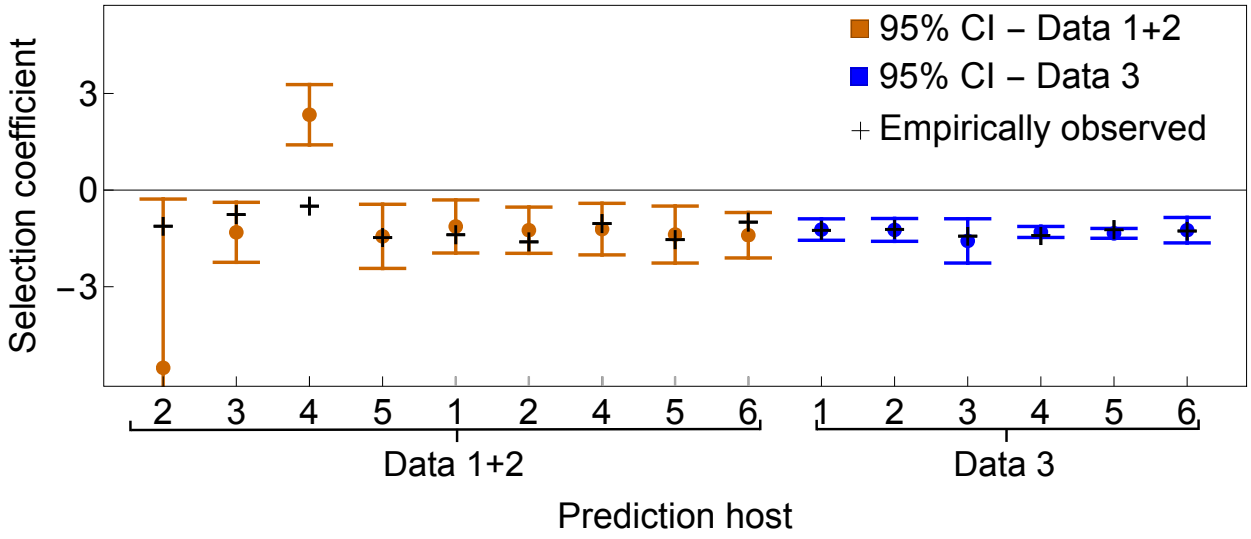

Figure 2: **Model predictions vs. observed selection coefficients.** We re-estimated invasion fitness parameters via a *leave-one-out cross-validation* procedure. For each host not used in estimation, we predicted the selection coefficient (and 95%CI), based on its microbiota composition, c. The $1-\alpha$ confidence interval for predicted $\hat{\theta} = \hat{\Delta}_\Lambda c$ is: $(\hat{\theta} \pm t_{\alpha/2(M-N)} s_{\hat{\theta}})$ in mean, or $(\hat{\theta} \pm t_{\alpha/2(M-N)} s_d)$ for a particular case (shown here for tests with Data 1+2 and Data 3). See also SI Text 3.

On the application front, we also explored how this method could inform the design of optimal microbiomes to steer selection coefficients towards desired regimes (Fig.3, SI Text 4). In general, determining the right microbiota compositions and diversity levels, that keep selection coefficient s below or above a threshold, can be framed and solved as a constrained optimization problem.

4 Discussion

To model invader dynamics in a multi-species system, we propose the replicator equation framework [6,9]. It effectively and minimally describes frequency-dependent dynamics, and lends itself to the contextual nature of fitness cost between 2 strains upon co-invasion. Assuming that the frequencies of resident species remain constant in the initial invasion phase, the problem simplifies to a linear system, where depending on data dimensions, either an exact solution, or optimal regression-based estimate can be found. We illustrated this *proof-of-concept* on *E.coli* invasion data in mice with different microbiota [5], yet the framework is applicable to other multispecies composition and co-invasion scenarios. The next challenge lies in optimizing diversity resolution for robust inference, and integrating short- and long-term invasion outcomes [10], along key gradients [11], and across systems.

Acknowledgements

We thank Ermanda Dekaj for initial extraction of the data, and Isabel Gordo and Susana Vinga for helpful discussions. The work is supported by the Portuguese Foundation for Science and Technology via grant number 2022.03060.PTDC.

Supporting Information

SI dataset. All microbiota composition data and selection coefficients between two invaders, extracted from [5, 8].

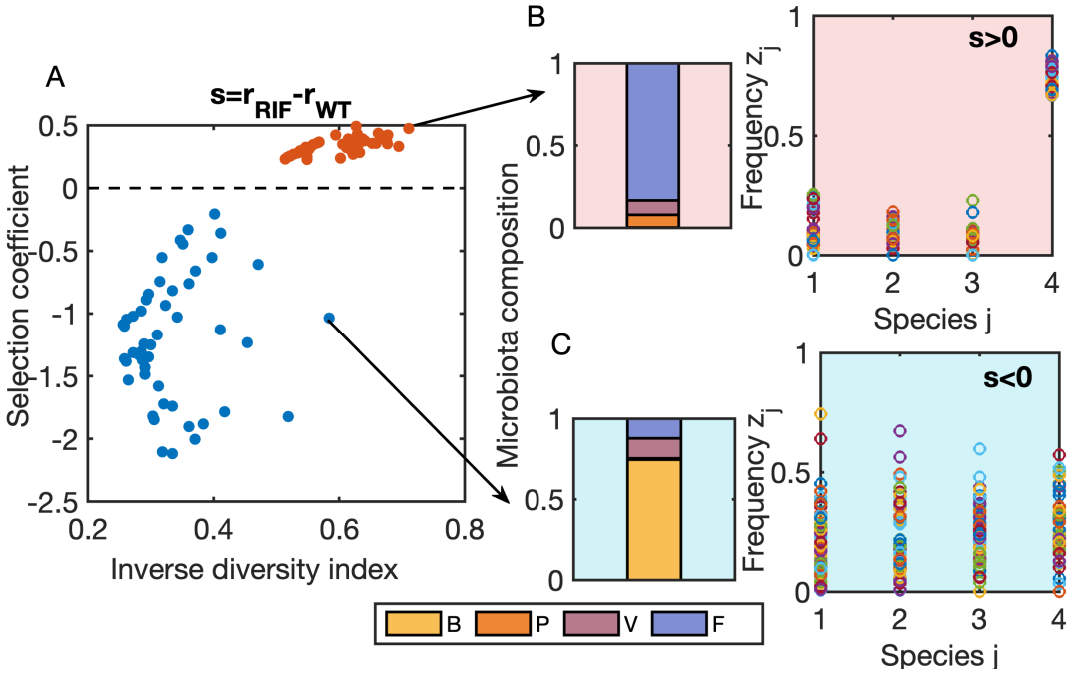

Figure 3: **Searching microbiota compositions for target selection coefficients between invaders.** We used the estimated invasion fitness differential for the Rif-WT pair of strains [8], and sought microbiota compositions yielding selection coefficients s in 2 target regimes: positive $s > 0.2$, and negative $s < -0.2$ (red and blue circles in **A**). We efficiently explored the space of possible solutions via minimizing $\sum_i z_i^2$ so as to favor more diversity, related to Simpson's index. In this search, the species frequencies z_i potentially converge to different admissible values (**B-C**), that satisfy inequality constraints $\Delta_\Lambda \mathbf{z} < s_{crit}$, but vary in diversity.

SI Text. Extended methods and references.

References

[1] Helen M. Kurkjian, M. Javad Akbari, and Babak Momeni. The impact of interactions on invasion and colonization resistance in microbial communities. *PLOS Computational Biology*, 17(1):1–18, 01 2021.

[2] Yael Litvak and Andreas J. Bäumler. The founder hypothesis: A basis for microbiota resistance, diversity in taxa carriage, and colonization resistance against pathogens. *PLOS Pathogens*, 15(2):1–6, 02 2019.

[3] Xuesong He, Jeffrey S McLean, Lihong Guo, Renate Lux, and Wenyuan Shi. The social structure of microbial community involved in colonization resistance. *The ISME Journal*, 8(3):564–574, 10 2013.

[4] S. Madec and E. Gjini. Predicting n-strain coexistence from co-colonization interactions: Epidemiology meets ecology and the replicator equation. *Bulletin of Mathematical Biology*, 82:142, 2020.

[5] Luis Leonidas Cardoso, Paulo Durão, Massimo Amicone, and Isabel Gordo. Dysbiosis individualizes the fitness effect of antibiotic resistance in the mammalian gut. *Nature Ecology & Evolution*, 4(9):1268–1278, 2020.

[6] Erida Gjini and Sten Madec. Towards a mathematical understanding of invasion resistance in multi-species communities. *Royal Society Open Science*, 10(11):231034, 2023.

[7] S. Geritz, E. Kisdi, G. MeszeŃA, et al. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology*, 12:35–57, 1998.

[8] Data (https://doi.org/10.5061/dryad.j0zpc869r), 2020. Accessed: 2023-09-30.

[9] Josef Hofbauer and Karl Sigmund. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 40(4):479–519, 2003.

[10] Jean-François Arnoldi, Matthieu Barbier, Ruth Kelly, György Barabás, and Andrew L Jackson. Invasions of ecological communities: Hints of impacts in the invader's growth rate. *Methods in Ecology and Evolution*, 13(1):167–182, 2022.

[11] Erida Gjini and Sten Madec. The ratio of single to co-colonization is key to complexity in interacting systems with multiple strains. *Ecology and Evolution*, 11(13):8456–8474, 2021.