ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Lakhani V, Tan L, Mukherjee S, Stewart WCL, Swords WE, Das J. 2018 Mutations in bacterial genes induce unanticipated changes in the relationship between bacterial pathogens in experimental otitis media. *R. Soc. open sci.* **5**: 180810.
http://dx.doi.org/10.1098/rsos.180810

Received: 1 June 2018
Accepted: 19 October 2018

Subject Category:
Biochemistry and biophysics

Subject Areas:
microbiology/biophysics/computational biology

Keywords:
polymicrobial infection, otitis media, maximum entropy estimation, Lotka–Volterra, Condorcet winner, Akkaike information criterion

Authors for correspondence:
W. Edward Swords
e-mail: wswords@uabmc.edu
Jayajit Das
e-mail: jayajit@gmail.com

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4285928.

THE ROYAL SOCIETY PUBLISHING

Mutations in bacterial genes induce unanticipated changes in the relationship between bacterial pathogens in experimental otitis media

Vinal Lakhani[1], Li Tan[2], Sayak Mukherjee[1], William C. L. Stewart[1], W. Edward Swords[2,3] and Jayajit Das[1,4,5,6]

[1]Battelle Center for Mathematical Medicine, The Research Institute at the Nationwide Children's Hospital, 700 Children's Drive, Columbus, OH 43205, USA
[2]Department of Microbiology and Immunology, Wake Forest School of Medicine, Winston-Salem, NC 27101, USA
[3]Division of Pulmonary, Allergy & Critical Care Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA
[4]Department of Pediatrics, [5]Department of Physics, and [6]Department of Biophysics Graduate Program, The Ohio State University, Columbus, OH 43210, USA

JD, 0000-0001-9649-4698

Otitis media (OM) is a common polymicrobial infection of the middle ear in children under the age of 15 years. A widely used experimental strategy to analyse roles of specific phenotypes of bacterial pathogens of OM is to study changes in co-infection kinetics of bacterial populations in animal models when a wild-type bacterial strain is replaced by a specific isogenic mutant strain in the co-inoculating mixtures. As relationships between the OM bacterial pathogens within the host are regulated by many interlinked processes, connecting the changes in the co-infection kinetics to a bacterial phenotype can be challenging. We investigated middle ear co-infections in adult chinchillas (*Chinchilla lanigera*) by two major OM pathogens: non-typeable *Haemophilus influenzae* (NTHi) and *Moraxella catarrhalis* (Mcat), as well as isogenic mutant strains in each bacterial species. We analysed the infection kinetic data using Lotka–Volterra population dynamics, maximum entropy inference and Akaike information criteria-(AIC)-based model selection. We found that changes in relationships between the bacterial pathogens that were not anticipated in the design of the co-infection experiments involving mutant strains are common and were strong

regulators of the co-infecting bacterial populations. The framework developed here allows for a systematic analysis of host–host variations of bacterial populations and small sizes of animal cohorts in co-infection experiments to quantify the role of specific mutant strains in changing the infection kinetics. Our combined approach can be used to analyse the functional footprint of mutant strains in regulating co-infection kinetics in models of experimental OM and other polymicrobial diseases.

1. Introduction

Otitis media (OM) is a common polymicrobial bacterial infection of the middle ear in children which is caused by three major bacterial pathogens: non-typeable *Haemophilus influenzae* (NTHi), *Moraxella catarrhalis* (Mcat) and *Streptococcus pneumoniae* (Sp) [1]. The relationships among these OM pathogens are both direct and indirect in nature. For example, quorum signals (autoinducer-2 or AI-2) secreted by NTHi help Mcat to form a biofilm and survive in the hostile middle ear environment [2]. This interaction represents a direct relationship (or an *active* interaction [3,4]) between NTHi and Mcat. In another case, NTHi stimulates the host immune response in the middle ear that suppresses the growth of Sp [5,6]; this interaction is an example of an indirect relationship (or a *passive* interaction) between NTHi and Sp. The qualitative (cooperative, competitive or neutral) and quantitative (interaction strength) nature of the *active* and *passive* interactions between the OM bacterial pathogens depend on phenotypes specific to bacterial strains and the host response [4,7]. Mechanistic understanding of how these interactions affect pathogenesis of polymicrobial diseases including OM has been a major research goal for developing vaccine candidates and other therapeutic strategies [8,9].

A common strategy to evaluate mechanistic roles of specific phenotypes of bacterial OM pathogens *in vivo* has been to co-infect animal models with bacterial pathogens obtained from clinical isolates and then assess changes in infection kinetics by replacing a wild-type bacterial strain with a mutant strain [2,6,10,11]. The mutant strains are designed to produce a loss or gain of specific bacterial phenotypes of interest. However, because the bacterial phenotypes probed by a mutant strain can be tightly intertwined with the phenotypes of the other bacterial pathogens, via *active* and *passive* interactions, this task could become challenging. For example, a mutant strain of Mcat lacking the ability to receive quorum signal from NTHi conceivably results in a decrease in the cooperative interaction from NTHi to Mcat. However, the same mutation could produce unanticipated changes in other relationships such as change in cooperation/competition of Mcat to NTHi. When these unanticipated changes are *strong* regulators of bacterial populations in co-infection experiments, we are required to revise our mechanistic understanding regarding the role of the specific mutation in influencing co-infection kinetics.

To this end, we address the above challenge by developing a framework that provides an answer to the following question: How is it possible to assess if unanticipated changes in the relationships induced by introducing mutant strains of OM pathogens in co-infection experiments are *strong* or *weak* regulators of the bacterial populations in the experiments? We define loss or gain of phenotype(s) in a specific mutant strain as a *weak* regulator when the interactions between bacterial species in co-infection experiments with the mutant strain are modified according to changes in the phenotype(s) as hypothesized for the mutant strain. The change in the phenotype(s) is defined as a *strong* regulator when additional unanticipated interactions are altered in co-infection experiments with the mutant strain. A more precise and formal definition of the *weak* and *strong* regulators is provided in the Material and methods section. The answer to the above question will provide a quantitative way to evaluate the mechanistic role of a bacterial gene in affecting the co-infection kinetics. Our framework combines (1) *in vivo* bacterial load measurements in an animal model, with (2) *in silico* approaches comprising Lotka–Volterra (LV) population dynamic models [12,13], maximum entropy (MaxEnt) inference [14–17] and Akaike information criterion (AIC)-based model selection [18]. The animal model we used is a *Chinchilla lanigera* experimental OM model [11], wherein the animals' middle ears are co-inoculated with NTHi (86-028NP) and Mcat. These strains may be wild-type or isogenic mutant strains.

Our study revealed three important findings. First, the unanticipated changes in the relationships between OM bacterial pathogens that substantially affect the co-infection kinetics are commonly present in co-infection with mutant bacterial strains in experimental OM. Second, several bacterial phenotypes are tightly correlated across co-infecting bacterial stains. Third, our combined framework provides a systematic way to deal with two common difficulties faced when analysing infection

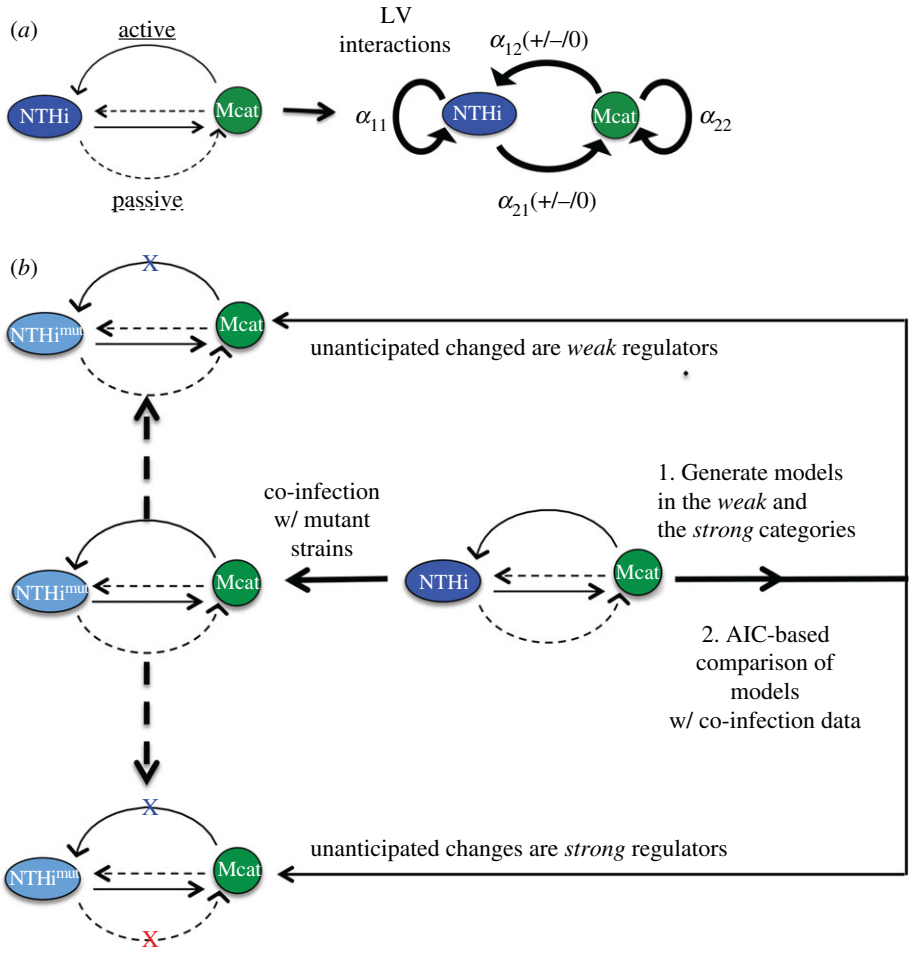


Figure 1. Schematic representation of our framework to determine roles of mutant bacterial strains in regulating co-infection kinetics. (*a*) Inter- and intra-species interactions between two bacterial species NTHi and Mcat residing within a host can be both *active* (solid lines) or *passive* (dashed lines) in nature. These interactions can be simplified and described by LV interaction parameters ($\{\alpha_{ij}\}$). α_{11} (greater than 0) and α_{22} (greater than 0) represent intra-species interactions for NTHi and Mcat, respectively. α_{12} and α_{21} represent the overall effect of Mcat on the growth of NTHi and NTHi on the growth of Mcat, respectively. α_{12} and α_{21} can be positive (competitive interaction), zero (neutral interaction) or negative (cooperative interaction). (*b*) Replacing a wild-type strain by a mutant strain in the co-infection experiments can change the LV interactions. These changes may be anticipated (blue 'X') or not anticipated (red 'X') based on the design of the experiment. Our framework uses data from co-infection experiments involving the wild-type strains to generate models that determine if these unanticipated changes are *weak* or *strong* regulators of bacterial kinetics. These models are compared to each other using AIC.

kinetic measurements in animal models: host–host variations of bacterial populations and small size of animal cohorts. Our framework can be used to design mutant strains to generate desired infection kinetics in experimental models of polymicrobial diseases such as OM and infections secondary to cystic fibrosis [3] with potential implications for therapeutic strategies [9].

2. Results

2.1. Development of a framework to assess effects of genetic mutation of the bacterial strains in co-infection kinetics

We developed a two species LV model to describe co-infection kinetics of populations of NTHi and Mcat strains within an individual chinchilla host (figure 1*a* and Material and methods section). For simplicity in the mathematical expressions of the probability distributions and interaction parameters, we will refer to NTHi strains (wild-type or mutant) as species 1 and Mcat strains (wild-type or mutant) as species 2 throughout the manuscript. The LV interactions (α_{11}, α_{12}, α_{21} and α_{22}) characterize the relationships

between the bacterial species that originate due to *active* and *passive* interactions (figure 1*a* and Material and methods section). Now we can pose the motivating challenge in terms of the LV interactions. How is it possible to assess if unanticipated changes in the LV interactions induced by mutant strains are *strong* or *weak* regulators of the bacterial populations *in vivo*? The unanticipated LV interactions in co-infection experiments with mutant strains are the ones that were not accounted for in the design of the experiment. For example, the *hag* mutant of Mcat does not adhere to the host's epithelial cell layer as well as the wild-type Mcat strain [19]. Hence, replacing the wild-type Mcat strain in the co-infection by NTHi (*wt*) + Mcat (*wt*) by the *hag* mutant strain should increase Mcat's self-inhibition; i.e. α_{22} should increase (see table 1 for details). Therefore, in the design of the co-infection with NTHi (*wt*) + Mcat (*hag*), one would anticipate a lower carrying capacity for Mcat or an increase of α_{22}. However, the changes in *passive* interactions induced by the *hag* mutant can also lead to changes in other LV interactions that were not anticipated, such as an increase in α_{21}. In addition, the host–host variations measured in the *NTHi* and *Mcat* populations were assumed to arise from the variations of the LV interactions ($\{\alpha_{ij}\}$) in our models. We developed a framework (figure 1*b*) to address the above question. The framework is executed in two main steps.

Step 1. The wild-type NTHi (or N_1) and the wild-type Mcat (or N_2) populations in individual chinchillas were measured in co-infection experiments carried out in a cohort of n number of chinchillas at a time T (7 or 14 days) post inoculation. These data were used to generate the reference data (denoted by the subscript r) ${}_nD_r$. This reference dataset was composed of the mean values ($\bar{N}_1, \bar{N}_2$), variances (σ_1^2, σ_2^2) and the covariance (ρ_{12}) calculated from the bacterial load measurements in the cohort; i.e. ${}_nD_r \equiv \{\bar{N}_1, \bar{N}_2, \sigma_1^2, \sigma_2^2, \rho_{12}\} \equiv \{{}_nD_r^j\}$ for $j = 1, \ldots, 5$. We estimated the probability distribution function $\hat{p}(N_1, N_2)$ of populations of wild-type NTHi (N_1) and wild-type Mcat (N_2) in the chinchilla cohort using MaxEnt (electronic supplementary material, figure S1), wherein the mean values, variances and the covariance as measured in the experiments (${}_nD_r$) were constrained in the calculation (electronic supplemental material, §S2). Using this $\hat{p}(N_1, N_2)$, we calculated the joint probability distribution of the interaction parameters $\hat{q}(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$ ($\equiv \hat{q}(\{\alpha_{ij}\})$ using a MaxEnt-based method (details in the Material and methods section and electronic supplementary material, figures S2–S3). The estimated joint probability distribution function $\hat{q}(\{\alpha_{ij}\})$was used to generate models that fall either in the *weak* or the *strong* category.

Step 2. We generated a test dataset using data from co-infection experiments where *at least* one of the wild-type bacterial strains was replaced by a mutant strain. The test dataset ${}_{n'}D_x$ contains the population means, variances and covariance for the cohort which contained n' number of chinchillas for the same time T (7 or 14 days) post inoculation. The subscript x in ${}_{n'}D_x$, to be determined in our analysis, quantifies the role of the mutation in the co-infection: $x = weak$ or $x = strong$. As described above, this distinction indicates whether the unanticipated changes in LV interactions induced by the mutant strain(s) were *weak* or *strong* regulators of the bacterial populations. To safeguard against small sizes of n' (approx. 10 or less), we performed bootstrapping [23] on the data, wherein we sampled n' data points with replacement from the original. In this way, we generated t sets and determined x in each of those t samples of ${}_{n'}D_x$. These t samples of ${}_{n'}D_x$, e.g. ${}^{(1)}_{n'}D_x, \ldots, {}^{(t)}_{n'}D_x$, were generated by evaluating ${}_{n'}D_x$ in t independent groups containing n' number of animals each. Next we evaluated which model(s) generated in *step 1* for a particular co-infection experiment best described the t samples of the test dataset $({}^{(1)}_{n'}D_x, \ldots, {}^{(t)}_{n'}D_x)$; subsequently, we assigned the category (*weak* or *strong*) of the best model to x. The best model was found (whenever possible) by comparing the Akaike information criterion (AIC) values for each model in a head-to-head pairwise manner for each of the t samples. The 'Condorcet Winner' was the model which was preferred over all others in head-to-head comparisons [24]. We chose the Condorcet winner as the best model. Further details are provided in the Material and methods section and in the electronic supplementary material.

2.2. Application of the framework on synthetic data

To test the efficacy of our method, we generated synthetic data and applied the framework developed in the previous section. We had the following three goals in mind: (i) validate the framework, (ii) determine how the strengths of the mutations and/or the host immune response affect interspecies interactions and (iii) determine the dependence of the model selection on the sample size or the number (n') of animals in the cohort. We generated the synthetic data by numerical solution of coupled ODEs that described LV-type population kinetics involving two interacting bacterial species and a host immune response (see Material and methods section). The parameters describing the inter- and intra-species bacterial interactions as well as the host immune response were drawn from uniform distribution within

Table 1. Hypothesized changes in the interactions for co-infections with mutant strains of NTHi and Mcat. For each set of co-infected strains, we indicate the hypothesized changes in the four interaction parameters as suggested by the published literature. An open, dashed circle (◌) indicates no change. A filled circle (●) indicates an increase. A small, open circle (∘) indicates a decrease, and a bull's-eye (◉) indicates the parameter could increase or decrease. As shown in figure 1*a*, the LV interaction parameters α_{11} and α_{22} represent the strength of NTHi's and Mcat's self-inhibition, respectively. The α_{12} parameter represents the interaction of Mcat on NTHi, and α_{21} represents the inverse. For the latter two, a decreasing (or more negative) value indicates cooperation, while an increasing (or more positive) value indicates competition.

co-infecting strains	hypothesized changes in the interactions α_{11} α_{12} α_{21} α_{22}	justification
NTHi (*wt*) + Mcat (*hag*)	◌ ◌ ◌ ●	The *hag* mutation decreases adhesion of Mcat to the substrate and results in poor biofilm formation [20,21]. This mutant could lead to lower biomass or decreased carrying capacity (or higher α_{22}) for Mcat.
NTHi (*wt*) + Mcat (*mcaB*)	◌ ◌ ● ◌	The *mcaB* mutant is unable to sense the AI-2 molecules secreted by NTHi (Li *et al*., unpublished data). Thus, this Mcat mutant will experience less cooperation from NTHi, or increased values of α_{21}.
NTHi (*wt*) + Mcat (*aaa*)	◌ ◌ ◌ ∘	The *aaa* mutant has the Hag programme constitutively on (Li *et al*., unpublished data). Therefore, it makes dense biofilms without any help from NTHi and has a larger carrying capacity or decreased value of α_{22}.
NTHi (*luxS*) + Mcat (*wt*)	◉ ◌ ◉ ◌	The *luxS* mutant is unable to produce AI-2, produces less dense biofilms and as a result causes more acute disease with higher levels of early inflammation. Thus the mutant can produce lower cooperation to Mcat (higher α_{21}) due to the lack of quorum sensing, and lower values of carrying capacity (higher α_{11}) due to poor biofilm quality [22]. However, the increased inflammation could provide nutrients for self-growth (lower α_{11}) or Mcat growth (lower α_{21}).
NTHi (*wt*) + Mcat (*mclR*)	◌ ● ● ◌	The *mclR* mutant does not respond to NTHi quorum sensing and forms poor biofilms (Li *et al*., unpublished data). Thus, Mcat will receive less cooperation from NTHi (higher α_{21}) and give less cooperation to NTHi (higher α_{12}).
NTHi (*wt*) + Mcat (*dtgt*)	◌ ◌ ● ●	The *dtgt* mutant is unable to respond to quorum signals from NTHi and also has a mutation in the *hag* promoter. The latter effect disrupts Mcat production of this adhesion protein and thus adhesion to the substrate (presumably host epithelia but also potentially aggregation with other bacteria) (Li *et al*., unpublished data). Thus, the mutant receives less cooperation from NTHi (higher α_{21}) and has a lower carrying capacity (higher α_{22}).
NTHi (*luxS*) + Mcat (*mcaB*)	◉ ◌ ● ◌	As described above for the NTHi (*luxS*) mutant, α_{11} and α_{21} may either increase or decrease, whereas for the Mcat (*mcaB*) mutant, α_{21} increases. Because of the latter mutation, we hypothesize an increase in α_{21}.
NTHi (*luxS*) + Mcat (*hag*)	◉ ◌ ◉ ●	As described above for the *luxS* mutant, α_{11} and α_{21} can either increase or decrease depending on the existing conditions. α_{22} will increase based on the *hag* mutant as mentioned above.

(*Continued*.)

Table 1. (*Continued.*)

co-infecting strains	hypothesized changes in the interactions α_{11} α_{12} α_{21} α_{22}	justification
NTHi (*luxS*) + Mcat (*dtgt*)	⬤ ◌ ⬤ ●	As described above for the *luxS* and *dtgt* mutants, α_{11} and α_{21} can either increase or decrease depending on the existing conditions, and α_{22} will increase.
NTHi (*luxS*) + Mcat (*aaa*)	⬤ ◌ ⬤ ∘	As described above for the *luxS* mutant, α_{11} and α_{21} can either increase or decrease depending on the existing conditions. α_{22} will decrease based on the *aaa* mutant as mentioned above.
NTHi (*luxS*) + Mcat (*mclR*)	⬤ ◌ ● ◌	As described above for the NTHi (*luxS*) mutant, α_{11} and α_{21} can either increase or decrease, whereas for Mcat (*mclR*) mutant, α_{21} increases. Because of the latter mutation, we hypothesize an increase in α_{21}.

specific ranges to generate host–host variations of the co-infection kinetics (see Material and methods). We chose the parameter range for the wild-type strains such that it produced steady-state population values similar to those observed *in vivo* (electronic supplementary material, figures S6–S8).

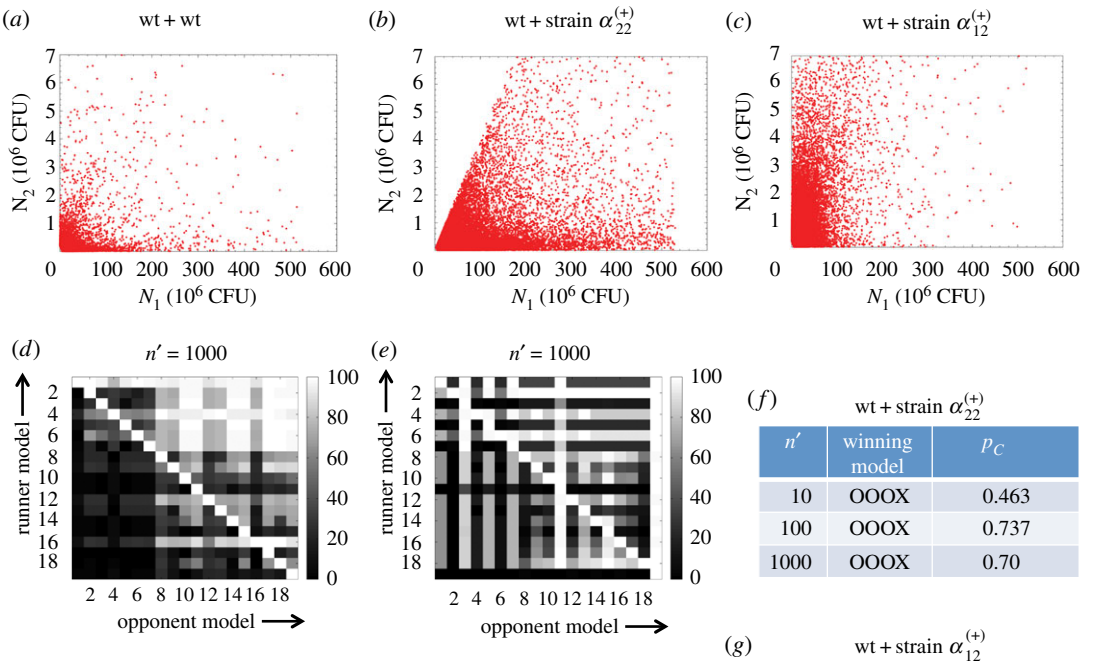
The average statistical variables ${}_nD_r \equiv \{\bar{N}_1, \bar{N}_2, \sigma_1^2, \sigma_2^2\rho_{12}\}$, calculated from the numerical solutions at $T =$ day 7, produced the reference dataset. We generated test datasets $\{{}_{n'}D_x\}$, wherein one of the wild-type strains was replaced by a mutant strain. Specifically, we considered the following two mutant strains for species 1. The $(\alpha_{11}^{(+)})$ mutant, which is an increase in the α_{11} parameter, possesses increased self-competition for species 1 compared to wild-type. The $(\alpha_{21}^{(+)})$ mutant, which is an increase in the α_{21} parameter, possesses increased competition of species 1 towards species 2 as compared to wild-type. We also considered two mutant strains for species 2. The $(\alpha_{22}^{(+)})$ mutant, which is an increase in the α_{22} parameter, possesses increased self-competition for species 2 compared to wild-type. The $(\alpha_{12}^{(+)})$ mutant, which is an increase in the α_{12} parameter, possesses increased competition of species 2 towards species 1 as compared to wild-type. The mutants were generated by changing the ranges of the associated parameters from that of the wild-type strains (see Material and methods). For example, the range of α_{11} used to generate the mutant strain $\alpha_{11}^{(+)}$ spanned a smaller range $[a', b]$ compared that of the wild-type strain, $[a, b]$, where, $a < a'$. Each of these mutations was performed with low, moderate and large strengths based on the relative change in the magnitude of range for the parameters. We also solved the co-infection kinetics in the presence of no, weak and strong host immune response. Thus, in total we considered 9×4 different mutation experiments *in silico* (electronic supplementary material, figures S9–S12). The test datasets $\{{}_{n'}D_x\}$ were obtained from the co-infection kinetics involving the above mutant strains.

We show results for two mutant strains $\alpha_{12}^{(+)}$ and $\alpha_{22}^{(+)}$with moderate strength mutations in the absence of any host immune response (figure 2*b*,*c*). The rest of the mutants are described in the electronic supplementary material. We followed the steps described in electronic supplementary material §S1 to determine the nature of the mutation (or x) in a test dataset ${}_{n'}D_x$. First, we used ${}_nD_r$ to generate models that belonged to the *weak* or the *strong* category corresponding to the co-infection strain#1(wt)+strain#2($\alpha_{22}^{(+)}$) or strain#1(wt)+strain#2($\alpha_{12}^{(+)}$). Next, we compared the models with the samples ($t > 100$) of the test dataset and evaluated the Condorcet winner model. We found that for the co-infection with strain#1(wt) + strain#2($\alpha_{22}^{(+)}$), a model in the $x = weak$ category was the Condorcet winner (figure 2*d* and *f*). By contrast, for the co-infection with strain#1(wt) + strain#2($\alpha_{12}^{(+)}$), a model in the $x = strong$ category was the Condorcet winners (figure 2*e*,*g*). These results can be explained by correlations among the LV interaction parameters (electronic supplementary material, figure S6) pertaining to the reference dataset ${}_nD_r$. The correlations of α_{22} with the other LV parameters for the above co-infection (strain#1(*wt*)+strain#2(*wt*)) are substantially small (less than 0.06); therefore, increasing α_{22} alone, as in the $\alpha_{22}^{(+)}$ strain, will minimally affect the other parameters. Therefore, the changes in other LV parameters (unanticipated changes) induced by the increase in α_{22} for the $\alpha_{22}^{(+)}$ strain will be small. These unanticipated changes would play a *weak* role in regulating the bacterial populations. By contrast, α_{12} is correlated strongly with several other interaction parameters (e.g. α_{21}, Corr ≈ -0.5) (electronic supplementary material, figure S6). Thus increasing α_{12} even by a moderate amount, as in the $\alpha_{12}^{(+)}$ strain, will produce large changes in the other LV parameters. Those unanticipated changes will generate a *strong* effect on the bacterial populations. These expectations about the effects of the correlations among the parameters were consistent with the results obtained from our framework. Therefore, these results validate our framework. The roles of the above mutant strains in affecting co-infection kinetics for different mutation strengths change depending on the mutation strength and/or the presence of the host immune response (electronic supplementary material, figures S9–S12). Therefore, strengths of the mutations and the host immune response are important in determining the influence of the mutations in the co-infection kinetics.

We checked the dependence of the Condorcet winner on the sample size n' in the test dataset. We found that even for small sample sizes ($n' = 10$), the framework picked the correct Condorcet winner; however, the margin of victory increased with larger n' (figure 2*f*,*g*).

2.3. Analysis of the *in vivo* data

We analysed co-infection kinetics in *Chinchilla lanigera* co-inoculated with a mixture of wild-type NTHi and Mcat strains or mixtures of NTHi and Mcat strains wherein at least one of the bacterial strains was a mutant strain. The chinchillas were inoculated by injecting 10^3 CFU of NTHi and 10^4 CFU of Mcat directly into the middle ears of the animals. The co-infection experiments investigated the *hag*, *mcaB*, *aaa*, *mclR* and *dtgt* mutant strains of Mcat and the *luxS* mutant strain of NTHi (table 1; electronic supplementary material,



(f) wt + strain $\alpha_{22}^{(+)}$

n'	winning model	p_C
10	OOOX	0.463
100	OOOX	0.737
1000	OOOX	0.70

(g) wt + strain $\alpha_{12}^{(+)}$

n'	winning model	p_C
10	OOOX	5.3×10^{-6}
100	OOOX	0.0019
1000	OOOX	0.0516

Figure 2. Application of the scheme on synthetic data. (*a*) Values of N_1, N_2 pairs (10^4 pairs) obtained from steady-state solutions of the ODEs corresponding to the LV model where $\{\alpha_{ij}\}$ were drawn from uniform distributions in the following ranges: $2.74\times 10^{-3} \leq \alpha_{11} \leq 0.2$, $-200 \leq \alpha_{12} \leq 5$, $-5 \leq \alpha_{21} \leq 0.1$ and $1.9 \leq \alpha_{22} \leq 140$. The solutions where either N_1 or N_2 went to zero values or became very large ($N_1 > 530 \times 10^6$ or $N_2 > 7 \times 10^6$) were not included in the synthetic dataset. (*b*) Synthetic data (10^5 data points) for a co-infection with the mixture wt + $\alpha_{22}^{(+)}$ strain. The $\alpha_{22}^{(+)}$ strain was generated by increasing the lower range of α_{22} to 120. (*c*) Synthetic data (10^5 data points) for the co-infection for the mixture wt + $\alpha_{12}^{(+)}$ strain. The $\alpha_{12}^{(+)}$ strain was generated by increasing the lower range of α_{12} to -2. (*d*) The percentage of the time a runner model won against an opponent model in head-to-head comparison of AICs for the models describing the synthetic data in (*b*). The $m = 19$ different models are indexed by integers. The percentages shown were obtained for $t = 100$ trials, each with a sample size of $n' = 1000$. A bright row indicates the winning model. (*e*) Results in head-to-head comparisons between the models presented similar to the data in (*c*). (*f*,*g*) The probability p_C for the Condorcet winner to win all the pairwise encounters in the t samples is shown for increasing sample size n'. p_C for the Condorcet winner model (#*i*) is calculated using $p_C = \prod_{j(\neq i)} f_{ij}$, where f_{ij} (greater than 1/2) denotes the fraction of the t samples where the Condorcet winner model #*i* was preferred over model #*j*. The product is calculated for all the $m - 1$ pairwise combinations where m number of models were considered. p_C increased with the sample size (n'). The winning models are denoted in the second column by the changes in α_{11}, α_{12}, α_{21} and α_{22} for the wild-type+wild-type co-infection. O indicates no change, X indicates an increase and o indicates a decrease.

table S1). The properties of the mutant strains and their hypothesized effects on the LV interaction parameters are described in table 1. The populations of the NTHi and Mcat strains were measured at 7 days (day 7) and 14 days (day 14) post inoculation (figure 3; electronic supplementary material, table S1).

The data were collected from both ears of the chinchillas in cohorts containing more than five animals. The bacterial counts showed large to moderate host–host variations (electronic supplementary material, figure S1). These variations could arise from the host–host differences in the physiology, anatomy and immune responses in the upper respiratory tract of the outbred population of chinchillas. The mean NTHi population was substantially larger (greater than 100 fold) than that of Mcat for the co-inoculation with the wild-type NTHi and Mcat strains (figure 3). The mean populations of the wild-type NTHi strain at day 7 were lower if co-inoculated with any mutant Mcat strain (except Mcat (*mcaB*)) rather than if co-inoculated with wild-type Mcat (electronic supplementary material, table S1). By contrast, the mean populations of the Mcat strains for the same co-infections

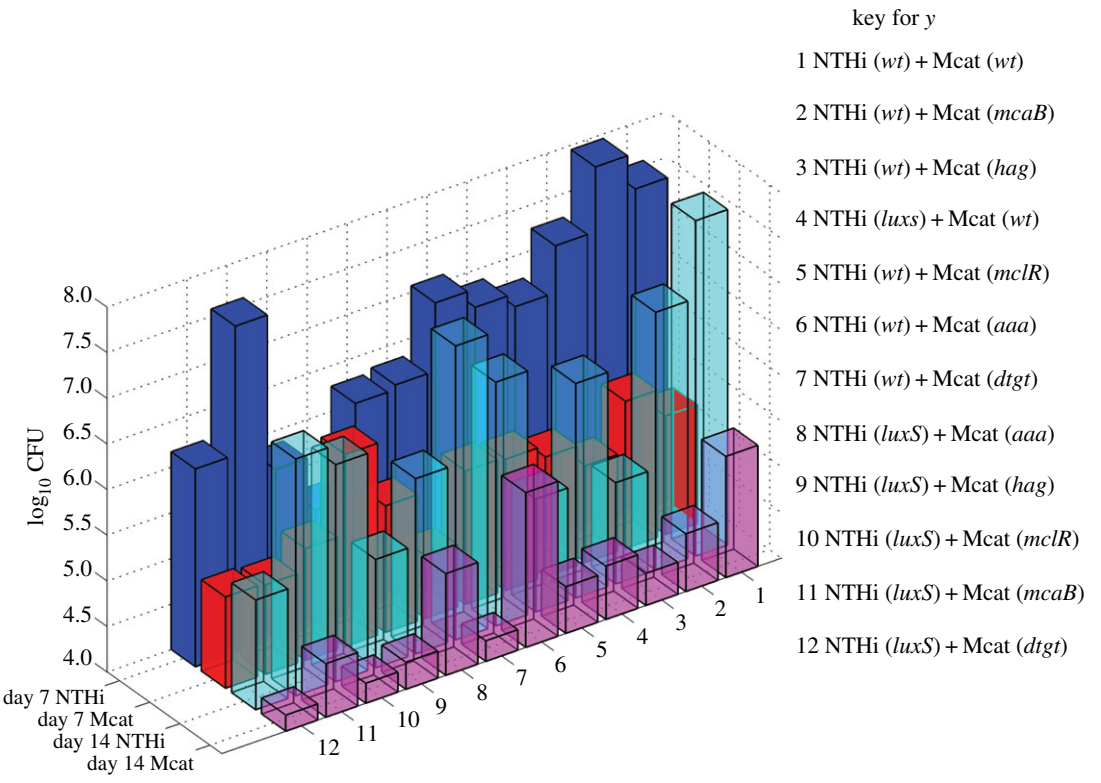


Figure 3. Mean populations of NTHi and Mcat strains in co-infection experiments. Shows the mean populations of NTHi and Mcat strains at day 7 and day 14 post inoculation calculated from counts in the bullae measured in greater than 5 chinchillas for each of the 12 different cases of co-inoculation. The combinations of the bacterial strains used for co-inoculating the chinchillas are listed.

showed small changes (increase or decrease) (figure 3; electronic supplementary material, table S1). Co-inoculation with NTHi (*luxS*) + Mcat (*wt*) resulted in a negligible change in the mean Mcat population but a large decrease in the mean NTHi population at day 7 as compared to the NTHi (*wt*) + Mcat (*wt*) experiment. When both the NTHi (*wt*) and the Mcat (*wt*) strains were replaced by their mutant strains in the co-inoculating mixtures, the mean populations of both the strains decreased at day 7. At day 14, the mean populations of the NTHi strains decreased, compared to the NTHi (*wt*) + Mcat (*wt*) experiment, in all the experiments with any mutant strains. The covariances between the populations of the NTHi and the Mcat strains were negative for the majority of the cases investigated here (electronic supplementary material, table S1). In a few cases, such as NTHi (*wt*) + Mcat (*hag*) NTHi or NTHi (*wt*) + Mcat (*mcaB*) the covariances were positive (electronic supplementary material, table S1). Overall, the data showed a complex pattern as further explained below.

Changes in the bacterial counts due to mutations pointed towards the presence of unanticipated changes in the bacterial relationships in regulating bacterial populations. For example, compared to co-infecting with NTHi (*wt*) + Mcat (*wt*), co-infecting with NTHi (*wt*) + Mcat (*hag*) substantially decreases the NTHi population (almost by half), whereas the Mcat population only decreases a small amount. As the *hag* mutant has lower adherence and poor biofilm formation capability compared to its wild-type counterpart, we would expect the Mcat population to decrease while having minimal consequence to the NTHi population. Because the NTHi population was substantially reduced here, it suggests a potential change in the interaction from Mcat towards NTHi. We used our framework to quantify the roles of specific NTHi and Mcat mutations in regulating the bacterial populations in co-infection experiments. As described in the previous section, the data from the co-infection experiments with the wild-type strains generated our reference dataset ${}_nD_r$. The models in the *weak* or the *strong* category were generated using ${}_nD_r$ and were compared against the test datasets $\{{}_{n'}D_x\}$. The test datasets were obtained from the co-infection experiments that involved *at least* one mutant strain. Multiple samples of a test dataset were obtained by using bootstrapping [23]. Our analysis showed that for the majority of the cases, models with additional interactions (*strong* models) better described the data (at both day 7 and day 14) compared to models with no additional interactions (*weak* models) (figure 4). We found that the *weak* model described the data obtained at day 7 for the co-

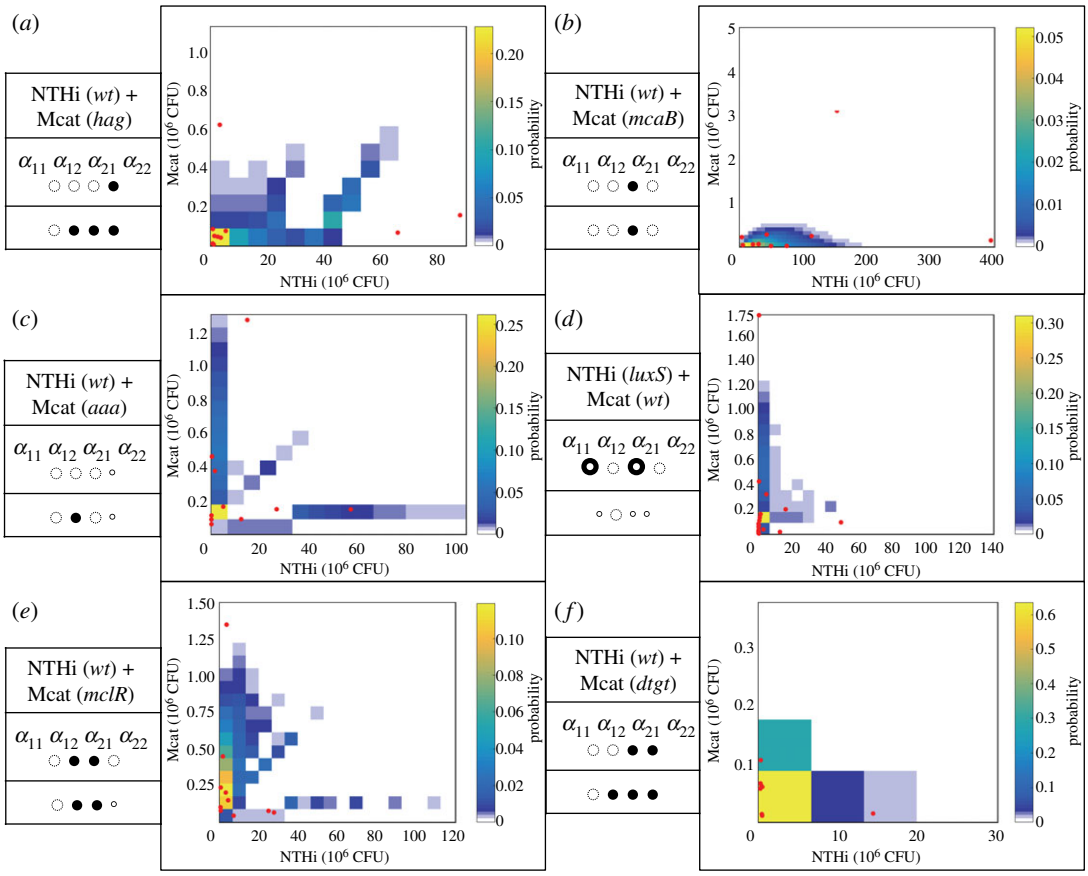


Figure 4. Comparison between the Condorcet winner model and measurement at day 7 post inoculation with mutant strains. The probability distribution function $\hat{p}(N_1, N_2)$ generated by the Condorcet winner model for a co-infection at day 7 post inoculation is shown using a heat map. The measured bacterial loads for the same co-infection for individual chinchillas are shown in red points. The anticipated changes in the LV parameters for a co-infection involving a specific mutant strain are shown in the first row of the table shown on the left of a sub-figure. The changes suggested by the Condorcet winning model are shown in the second row. A filled or a smaller empty circle indicates an increase or decrease of a specific parameter, respectively. The cases where the phenotype is uncertain, i.e. either an increase or decrease, are marked by a bull's-eye symbol (◉). (*a*) NTHi (*wt*)-Mcat (*hag*), (*b*) NTHi (*wt*)-Mcat (*mcaB*), (*c*) NTHi (*wt*)-Mcat (*aaa*), (*d*) NTHi (*luxS*)-Mcat (*wt*), (*e*) NTHi (*wt*)-Mcat (*mclR*) and (*f*) NTHi (*wt*)-Mcat (*dtgt*). Note that we compared our prediction against the measured data in terms of average populations as bacterial measurements were available for only few animals. The individual data points shown on the graphs were not explicitly compared, thus some of the individual measurements (dots) can lie at the boundaries of the predicted distribution (coloured squares) and need not reflect quality of comparison between the average bacterial populations.

infection with NTHi (*wt*) + Mcat (*mcaB*) (figure 4*b*) better than any *strong* model. However, at day 14, a *strong* model (electronic supplementary material, figure S4) described the same data better than any *weak* model. Therefore, unanticipated changes in LV interactions were prevalent in co-infections with mutant bacterial strains both at early and late stages of the co-infection kinetics.

2.4. Host immune responses modulate Mcat–NTHi interactions at later stages of the infection

The mean populations of the wild-type strains of NTHi and Mcat increase as the infection progresses from day 7 to day 14 post inoculation (figure 5). However, the covariance of the NTHi and Mcat populations becomes more negative (approx. twofold change) (electronic supplementary material, table S1). The negative correlation indicates that the populations of the two species are more mutually exclusive; that is, when one species has high abundance, then the other's is low. We hypothesized that the host immune response generated by both the pathogens could lead to a decrease in cooperation (or increase in α_{12} and α_{21}) between Mcat and NTHi. We tested our hypothesis by applying our scheme with the day 7 data as the reference and day 14 data as the test. The model wherein α_{12} and α_{21} increase was the Condorcet winner (figure 5). The agreement demonstrates the role of the host immune response in regulating *passive* inter-species interaction between Mcat and NTHi.

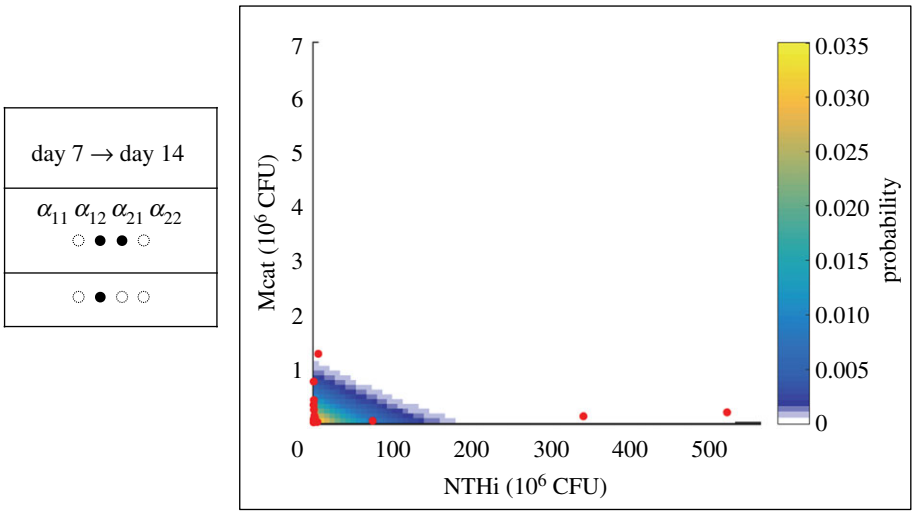


Figure 5. The prediction at day 14 post inoculation for the wild-type strains generated using the day 7 data. The data are displayed using the same visualization scheme as in figure 4.

3. Discussion

Co-infection of animal models with mutant bacterial strains is a powerful tool in probing mechanisms that underlie pathogenesis of polymicrobial infections such as OM. However, the interconnected and variable nature of interactions involving bacterial pathogens within the host makes it challenging to connect specific perturbations, such as a mutation, in these experiments to mechanisms. The data-driven framework developed here provides a systematic method of addressing this challenge. The framework uses bacterial counts measured in animal hosts to quantitatively determine perturbations in the bacterial interactions induced by the replacement of a wild-type bacterial strain with a mutant strain that strongly or weakly regulates bacterial populations in the co-infection. Therefore, using this framework we are able to quantitatively assess the mechanistic role of a specific bacterial phenotype probed by an isogenic mutant strain in affecting the co-infection kinetics. Isogenic mutant strains are used for identification of bacterial determinants of colonization, persistence and virulence. Thus the quantitative information obtained from our framework will be valuable for determining specific targets for diagnostics, development of therapy and potentially vaccination. In addition, our framework also addresses the practical problem of systematically analysing bacterial count data obtained from a small size (approx. 10 animals) of animal cohorts in co-infection experiments.

Application of our framework to co-infection experiments in *Chinchilla lanigera* co-inoculated with wild-type and mutant strains of two major OM pathogens (NTHi and Mcat) found that in a majority of the co-infections the mutant strains gave rise to unanticipated changes in the bacterial interactions, which influenced the bacterial populations substantially. The emergence of unanticipated perturbations of the bacterial interactions is probably caused by the interdependencies between the interactions and the hostile environment of the host [16]. The interdependencies can be caused due to many shared processes such as feeding on common nutrients, exchanging small molecules and the host immune response that regulates the growth of OM pathogens within the host [5,8]. Therefore, when a specific bacterial phenotype is altered in the form of a mutant strain, several other phenotypes in co-infecting OM pathogens are also altered, some of which can be non-intuitive. Our analysis of the synthetic co-infection data lends support to this speculation. We found that correlations between the LV parameters generated non-intuitive changes in several LV interactions when a specific LV interaction was perturbed in a mutant strain, and in many cases these unanticipated changes in the LV interactions were *strong* regulators of the bacterial populations.

Our framework required estimation of LV interactions involving the co-infecting bacterial pathogens using the measured bacterial counts. The estimated interactions between wild-type strains of NTHi and Mcat demonstrate prevalence of cooperative interspecies interactions ($\alpha_{12} < 0$, $\alpha_{21} < 0$) (electronic supplementary material, figure S3). Previous experiments noted several molecular mechanisms regarding the help of NTHi towards Mcat's growth, e.g. the quorum signal AI-2 secreted by NTHi helping Mcat to form a biofilm and thereby helping it to survive within the host [22,25]. Reciprocally, the estimated interactions also suggested cooperation of Mcat towards the growth of NTHi (or $\alpha_{21} < 0$).

Such a cooperative effect can potentially occur through *passive* interactions. For example, Mcat binds and sequesters AI-2 molecules secreted by NTHi; this sequestration could help keep the AI-2 abundances at an optimal level for production of quorum signals by NTHi. Similar optimal regulation of quorum sensing has been found in mutualistic relationship between two human oral bacteria, *Actinomyces naeslundii* T14 V and *Streptococcus oralis* 34, where AI-2 secreted from the latter bacterial species help the former species to form biofilms [26]. Furthermore, Mcat bacteria are known to form large aggregates (or autoagglutination) via the Hag protein. Such aggregates can help NTHi to form biofilms as mixed NTHi-Mcat biofilms within the host [20,21]. Another example of cooperation involves nutrient recruitment. The inflammation caused by Mcat can produce an influx of host serum which also can provide nutrients for the growth of NTHi [27]. All three of the above sources of interaction could contribute towards generating an overall cooperative interaction from Mcat to NTHi.

We found that several LV interactions (e.g. α_{12} and α_{21}) involving NTHi and Mcat are tightly correlated ($|\mathrm{Corr}| > 0.5$) with each other. As the MaxEnt method estimates the most spread out or uniform probability distribution phenotype that is consistent with the measured data [17], the method used here provided the most conservative estimate correlations between the LV interactions. This approach is a major departure from several methods that have been developed in recent years to evaluate interactions [28–30] pertaining to microbiome datasets where the correlations between the interaction parameters are not analysed. Ecological models often assume LV interactions between coexisting species as uncorrelated random variables [13]. This assumption makes the calculations amenable to analytical methods. The presence of strong correlations between LV interaction parameters could have important implications in assessing general principles underlying the diversity of eco-systems [31].

4. Material and methods

4.1. Experiments

Moraxella catarrhalis persistence in the middle ear chambers of chinchillas was assessed essentially as described previously [2]. Animals were purchased on need and allowed to acclimate to the vivarium for more than 10 days before infection. No used animals showed visible sign of illness prior to infection. Chinchillas (five animals per group) were anaesthetized with isofluorane and infected via transbullar injection with both approximately 10^4 CFU of *M. catarrhalis* and approximately 10^3 CFU of *H. influenzae*. All inocula were confirmed by plate counting. Animals were euthanized at 7 and 14 days post infection, and their bullae were aseptically opened to recover possible effusion fluid. Middle ear lavage was performed using sterile PBS and also saved and combined. Bullae were then excised and homogenized in 10 ml of sterile PBS. All of fluid and homogenized samples were serially diluted and plated on brain heart infusion (BHI) agar plates to obtain viable counts of *M. catarrhalis*. Note that *H. influenzae* would not grow on BHI plate due to lack of haemin and NAD. *H. influenzae* bacteria were enumerated on BHI agar supplemented with 10 $\mu g\,ml^{-1}$ haeme and NAD and containing 5 $\mu g\,ml^{-1}$ clarithromycin, which inhibits the growth of *M. catarrhalis*.

4.1.1. Bacterial strains and growth conditions

Moraxella catarrhalis strain O35E, as a WT and parent strain in this study, is a commonly used laboratory strain [32]. *Moraxella catarrhalis* O35E *hag*::Sp containing a spectinomycin resistance cassette disrupting the *hag* gene is a kind gift from Dr Eric Hansen. Non-typeable *H. influenzae* strain 86-028NP is a nasopharyngeal isolate from a child with chronic OM [33], and its *luxS*::Kn mutant with a kanamycin resistance cassette disrupting the AI-2 synthase was described previously [22]. *Moraxella catarrhalis* strains were cultured in BHI medium (Difco), and *H. influenzae* strains were cultivated in BHI medium supplemented with 10 $\mu g\,ml^{-1}$ of haemin chloride (MP Biomedicals) and 10 $\mu g\,ml^{-1}$ of NAD (Sigma), referred as supplemented BHI (sBHI). Mixture cultures of both *M. catarrhalis* and *H. influenzae* used sBHI.

4.1.2. Mutant bacterial strains

M. catarrhalis mclR::spec was generated by insertional mutagenesis using the following approach. Genomic DNA was purified from *M. catarrhalis* O35E using the Wizard genomic DNA purification kit (Promega), essentially according to the manufacturer's instructions. Portions of the *mclR* (allele MCR_1062) open reading frame were amplified using the PCR and primers specific for intragenic

regions (luxRUPF: CATCATGACTTGGTAACTTGCTG, luxRUPR: GCTGATCGGCAATTTGCCCCC GGGGTCGAGTGGCTTCTACACC). The resulting amplicons were cloned and a spectinomycin resistance cassette was introduced into a SmaI restriction site within the intergenic primers. This mutant allele was introduced into the parental strain by natural transformation and the resistant derivatives were confirmed by PCR and DNA sequence analysis.

4.1.3. Hag promoter mutants

Deletion of a tgt sequence and insertion of an *aaa* sequence within a predicted lux box within the hag promoter was achieved by overlap PCR; the resulting allele was introduced by natural transformation using a linked spectinomycin resistance marker (Li *et al.*, unpublished data).

4.2. Estimation of probability distribution function $q(\{\alpha_{ij}\})$ of Lotka–Volterra interactions $\{\alpha_{ij}\}$

4.2.1. Lotka–Volterra modelling of the population kinetics

We described the population kinetics of populations of NTHi and Mcat in the middle ear of the host (*Chinchilla lanigera*) using two coupled ordinary differential equations (ODEs) following the Lotka–Volterra (LV) model [12]:

$$\left.\begin{aligned}\frac{\mathrm{d}N_1^{(\mathrm{p})}}{\mathrm{d}t} &= r_1^{(\mathrm{p})}N_1^{(\mathrm{p})}(K_1^{(\mathrm{p})} - a_{11}^{(\mathrm{p})}N_1^{(\mathrm{p})} - a_{12}^{(\mathrm{p})}N_2^{(\mathrm{p})})\\ \text{and}\quad \frac{\mathrm{d}N_2^{(\mathrm{p})}}{\mathrm{d}t} &= r_2^{(\mathrm{p})}N_2^{(\mathrm{p})}(K_2^{(\mathrm{p})} - a_{21}^{(\mathrm{p})}N_1^{(\mathrm{p})} - a_{22}^{(\mathrm{p})}N_2^{(\mathrm{p})}).\end{aligned}\right\} \tag{4.1}$$

$N_1^{(\mathrm{p})}$ and $N_2^{(\mathrm{p})}$ denote the populations of the NTHi and Mcat strains, respectively, in the middle ear of an individual host. Individual hosts are indexed by the superscript p. The parameters r_1K_1 and r_2K_2 denote the growth rates of species 1 and 2, respectively, where K_1 and K_2 denote the corresponding carrying capacities. The parameters $\{a_{ij}\}$ describe effective interactions involving the bacterial species. a_{11} and a_{22} denote self-competition for growth for species 1 and 2, respectively, and a_{12} and a_{21} denote the influence (competition or cooperation) of species 2 and 1 on the growth of the other species, 1 and 2, respectively. Note, a_{12} and a_{21} can assume positive (indicating competition) or negative (indicating cooperation) values, whereas a_{11} and a_{22} can only possess non-negative values. The carrying capacities and the interaction parameters $\{a_{ij}\}$ determine the maximum bacterial load that can be sustained in the local environment. The above simple description of the bacterial infection kinetics within a host provides a coarse-grained and effective description of the kinetics, where the bacterial populations represent an average over spatial length scales including spatial structures such as biofilms. The effects of the immune response, nutrients, protective effects of biofilm formation and quorum sensing are effectively described in terms of the interaction parameters and the carrying capacities (table 1). The above model describes the interactions between the bacterial species and the host minimally where the LV interaction parameters provide a clear description of the inter- and intra-species interactions between the co-infecting bacterial species. These effects can vary from host to host giving rise to host-dependent values of the effective parameters; therefore, we consider host–host variations of the model parameters ($\{a_{ij}\}$, $\{K_i\}$). As the NTHi and Mcat replication rates are approximately $1\,\mathrm{h}^{-1}$, it is reasonable to assume that in a time scale of days the kinetics in equation (4.1) reaches a steady state, i.e.

$$\left.\begin{aligned}r_1^{(\mathrm{p})}N_1^{(\mathrm{p})}(K_1^{(\mathrm{p})} - a_{11}^{(\mathrm{p})}N_1^{(\mathrm{p})} - a_{12}^{(\mathrm{p})}N_2^{(\mathrm{p})}) &= 0\\ \text{and}\quad r_2^{(\mathrm{p})}N_2^{(\mathrm{p})}(K_2^{(p)} - a_{21}^{(\mathrm{p})}N_1^{(\mathrm{p})} - a_{22}^{(\mathrm{p})}N_2^{(\mathrm{p})}) &= 0.\end{aligned}\right\} \tag{4.2}$$

The steady-state equations (equation (4.2)) help reduce the number of parameters in determining populations of NTHi and Mcat.

$$\left.\begin{aligned}N_1^{(\mathrm{p})}(1 - \alpha_{11}^{(\mathrm{p})}N_1^{(\mathrm{p})} - \alpha_{12}^{(\mathrm{p})}N_2^{(\mathrm{p})}) &= 0\\ \text{and}\quad N_2^{(\mathrm{p})}(1 - \alpha_{21}^{(\mathrm{p})}N_1^{(\mathrm{p})} - \alpha_{22}^{(\mathrm{p})}N_2^{(\mathrm{p})}) &= 0.\end{aligned}\right\} \tag{4.3}$$

We defined $\alpha_{ij}^{(\mathrm{p})} = a_{ij}^{(\mathrm{p})}/K_i^{(\mathrm{p})}$ in equation (4.3). Thus, the dependences of the carrying capacities $\{K_i\}$ are effectively contained in the scaled variables $\{\alpha_{ij}\}$. The *LV* interaction parameters $\alpha_{11}^{(\mathrm{p})}$, $\alpha_{12}^{(\mathrm{p})}$, $\alpha_{21}^{(\mathrm{p})}$, $\alpha_{22}^{(\mathrm{p})}$

determine the bacterial abundances at the steady state (or the stable fixed points). The above equations produce four fixed points, $(N_1^{(\mathrm{p})}=0, N_2^{(\mathrm{p})}=0)$, $(N_1^{(\mathrm{p})}=0, N_2^{(\mathrm{p})}=1/\alpha_{22}^{(\mathrm{p})})$, $(N_1^{(\mathrm{p})}=1/\alpha_{11}^{(\mathrm{p})}, N_2^{(\mathrm{p})}=0)$ and $(N_1^{(\mathrm{p})}=([\boldsymbol{\alpha}^{(\mathrm{p})}]^{-1})_{11}+([\boldsymbol{\alpha}^{(\mathrm{p})}]^{-1})_{12}$, $N_2^{(\mathrm{p})}=([\boldsymbol{\alpha}^{(\mathrm{p})}]^{-1})_{21}+([\boldsymbol{\alpha}^{(\mathrm{p})}]^{-1})_{22})$, where, $\boldsymbol{\alpha}^{(\mathrm{p})}$ and $[\boldsymbol{\alpha}^{(\mathrm{p})}]^{-1}$ denote the matrix, $\{\alpha_{ij}^{(\mathrm{p})}\}$ and its inverse. The stability of the fixed points is determined by the linear stability analysis (electronic supplementary material, §S3). We consider only the stable fixed points, where $N_1^{(\mathrm{p})}>0$ and $N_2^{(\mathrm{p})}>0$. The parameter values yielding any other type of solutions (e.g. $N_1^{(\mathrm{p})}=N_2^{(\mathrm{p})}=0$, $N_1^{(\mathrm{p})}<0$ or $N_2^{(\mathrm{p})}<0$, $N_1^{(\mathrm{p})}\to\infty$ or $N_2^{(\mathrm{p})}\to\infty$) are assumed not to occur in the bacterial kinetics. Thus, we consider the solutions,

$$\left.\begin{aligned} N_1^{(\mathrm{p})} &= ([\alpha^{(\mathrm{p})}]^{-1})_{11}+([\alpha^{(\mathrm{p})}]^{-1})_{12} \\ \text{and}\quad N_2^{(\mathrm{p})} &= ([\alpha^{(\mathrm{p})}]^{-1})_{21}+([\alpha^{(\mathrm{p})}]^{-1})_{22}, \end{aligned}\right\} \tag{4.4}$$

when they are stable and are positive. Next, we estimated the LV interaction parameters $\{\alpha_{ij}^{(\mathrm{p})}\}$ from the bacterial loads ($N_1^{(\mathrm{p})}$ and $N_2^{(\mathrm{p})}$) for an individual animal (indexed by p). For a measured value of $N_1^{(\mathrm{p})}$ and $N_2^{(\mathrm{p})}$, it is not possible to estimate the four parameters $\{\alpha_{ij}^{(\mathrm{p})}\}$ uniquely using the above equations. Therefore, we developed a MaxEnt-based inference scheme to estimate the parameters in the chinchilla population using the measured bacterial loads. The MaxEnt-based method estimates parameters based on the measured data without any additional prior assumption. This also implies that MaxEnt estimates the 'flattest' distribution that is consistent with the measured data. A recent work [34] used maximum caliber inference, which is an extension of MaxEnt for analysing time-dependent data [15,17], to estimate parameters in a gene regulatory reaction network. Parameter estimation in gene regulatory reaction networks using sparse time-dependent data represents a problem of similar spirit as the problem investigated here. First, we estimated the probability distribution function of N_1 and N_2, $\hat{p}(N_1, N_2)$, in the chinchilla population using MaxEnt (electronic supplementary material, figure S1). Then, we estimated the joint probability distribution function $\hat{q}(\{\alpha_{ij}\})$ in the interaction parameters $\{\alpha_{ij}\}$ using the estimated $\hat{p}(N_1, N_2)$ by applying MaxEnt the second time (electronic supplementary material, figures S2–S3). The details regarding the implementation of the method for the *in vivo* data are provided below and in the electronic supplementary material, §S2.

4.2.2. MaxEnt estimation of $\hat{p}(N_1, N_2)$

The bacterial loads for wild-type strains of NTHi and Mcat in 10 adult chinchillas (10×2 ears $=20$ samples) were used to calculated the mean bacterial loads ($E(N_1)$ and $E(N_2)$), variances ($\sigma^2(N_1)$ and $\sigma^2(N_2)$) and the covariance, $\mathrm{Cov}(N_1, N_2)$. $\hat{p}(N_1, N_2)$ was estimated using a MaxEnt procedure where the space in N_1 and N_2 was discretized on an equally spaced 81×81 lattice ($\{\mathrm{I}, \mathrm{J}\}$). The ranges of N_1 and N_2 were $0–530\times 10^6$ CFUs and $0–7\times 10^6$ CFUs, respectively. The MaxEnt method involved maximizing Shannon entropy [16,17], $S=-\sum_{I,J} p(N_1^{(\mathrm{I})}, N_2^{(\mathrm{J})})\ln[p(N_1^{(\mathrm{I})}, N_2^{(\mathrm{J})})]$ subject to the constraints, $E(N_1)=\sum_{I,J} N_1^{(\mathrm{I})} p(N_1^{(\mathrm{I})}, N_2^{(\mathrm{J})})$, $E(N_2)=\sum_{I,J} N_2^{(\mathrm{J})} p(N_1^{(\mathrm{I})}, N_2^{(\mathrm{J})})$, $\sigma^2(N_1)\equiv\sum_{I,J} N_1^{(\mathrm{I})} N_1^{(\mathrm{I})} p(N_1^{(\mathrm{I})}, N_2^{(\mathrm{J})})$, $\sigma^2(N_2)\equiv\sum_{I,J} N_2^{(\mathrm{I})} N_2^{(\mathrm{I})} p(N_1^{(\mathrm{I})}, N_2^{(\mathrm{J})})$ and $\mathrm{Cov}(N_1, N_2)=\sum_{I,J} N_1^{(\mathrm{I})} N_2^{(\mathrm{J})} p(N_1^{(\mathrm{I})}, N_2^{(\mathrm{J})})$. The solution that maximizes S is given by $\hat{p}(N_1^{(\mathrm{I})}, N_2^{(\mathrm{J})})\propto\exp(\lambda_1 N_1^{(\mathrm{I})}+\lambda_2 N_2^{(\mathrm{J})}+\lambda_3 N_1^{(\mathrm{I})} N_1^{(\mathrm{I})}+\lambda_4 N_2^{(\mathrm{J})} N_2^{(\mathrm{J})}+\lambda_5 N_1^{(\mathrm{I})} N_2^{(\mathrm{J})})$. The five Lagrange's multipliers ($\lambda_1, \ldots, \lambda_5$) were calculated by solving the constraint equations in Matlab using the built-in function *fsolve*. We will denote the discrete probability distribution, $\hat{p}(N_1^{(\mathrm{I})}, N_2^{(\mathrm{J})})$ by $\hat{p}(N_1, N_2)$, hereafter, to keep the notation simple.

4.2.3. Estimation of $\hat{q}(\{\alpha_{ij}\})$

We estimated $\hat{q}(\{\alpha_{ij}\})$ using $\hat{p}(N_1, N_2)$ by applying the MaxEnt inference the second time. We discretize the four-dimensional space spanned by α_{11}, α_{12}, α_{21} and α_{22} on a grid. α_{11} and α_{22} assume only positive real values, and α_{21} and α_{12} can assume both positive and negative real values. Specifically, we discretized α_{11} from 0.027 to 2×10^6 CFU and α_{22} from 18.9189 to 1400×10^6 CFU into 74 bins each. We discretized α_{12} from -2000 to 50×10^6 CFU and α_{21} from -50 to 1×10^6 CFU into 201 bins each. We have varied the bounds and the lattice sizes of the grids and there was no change in the qualitative results (electronic supplementary material, figure S5). In this case, the Shannon's entropy [16,17], $S_Q=-\sum_{\{\alpha ij\}} q(\{\alpha_{ij}\})\ln[q(\{\alpha_{ij}\})]$ was maximized subject to the constraint that the estimated

$q(\{\alpha_{ij}\})$ should reproduce $\hat{p}(N_1, N_2)$, i.e.

$$\hat{p}(N_1, N_2) = \sum_{\{\alpha_{ij} | \text{equation(4.4) and stability conditions}\}} \hat{q}(\{\alpha_{ij}\}). \tag{4.5}$$

The stability conditions (electronic supplementary material, §S3) make sure that the fixed points in equation (4.4) are stable solutions. Equation (4.5) was inverted to obtain $\hat{q}(\{\alpha_{ij}\})$. The solution is given by Das *et al.* [14],

$$\hat{q}(\{\alpha_{ij}\}) = \frac{\hat{p}(N_1(\{\alpha_{ij}\}), N_2(\{\alpha_{ij}\}))}{\kappa(\{\alpha_{ij}\})}, \tag{4.6}$$

$\kappa(\{\alpha_{ij}\})$ in the above equations denotes the degeneracy factor or the number of distinct points in the α space that produce the same value of N_1 and N_2. We calculated $\kappa(\{\alpha_{ij}\})$ numerically by counting the number of lattice points in the α space that map to the same lattice point in the N-space.

4.3. Determination of the role ($x = \textit{weak}$ or $x = \textit{strong}$) of the mutant strain in co-infection kinetics

Our framework to evaluate the role of the mutant strain is divided into two main steps as outlined in the main text. *Step 1*. We estimated $\hat{q}(\{\alpha_{ij}\})$ using the reference dataset ${}_nD_r$ as described in §4.2. We used $\hat{q}(\{\alpha_{ij}\})$ to generate models that belong to the *weak* or the *strong* category. The details regarding how these models were generated are given below. *Weak models*: A specific mutant strain is hypothesized to possess loss or gain of phenotype(s) at the design stage of the co-infection experiments (see table 1 for details). The hypothesized changes in the specific phenotypes for the mutant strain will result in changes in a subset of LV interaction parameters pertaining to the co-infection kinetics of the populations of the mutant strain and another bacterial species within the chinchilla host. This subset of LV interaction parameters is denoted by $\{\alpha_{pq}\} \subset \{\alpha_{\text{ij}}\}$, where $p = \text{i}$ and $q = \text{j}$ for each α_{ij} in $\{\alpha_{pq}\}$. E.g. for the co-infection with NTHi (*wt*) + Mcat (*mcaB*), the *mcaB* mutant strain increases the value of α_{21}, thus the set $\{\alpha_{pq}\}$ contains only one parameter α_{21}, whereas for co-infection with NTHi (wt) + Mcat (*dtgt*), the *dtgt* mutant strain increases both α_{21} and α_{22}, and $\{\alpha_{pq}\}$ will contain two parameters, α_{21} and α_{22}. The models in the *weak* category are defined by the probability distribution function of the interactions parameters ($q_w(\alpha;\mathbf{a})$) in the models. The *weak* models are parametrized by $\{a_k\}$ which quantifies the extent by which each of the interaction parameters in $\{\alpha_{pq}\}$ is perturbed in $\hat{q}(\{\alpha_{ij}\})$ to generate $q_w(\alpha;\mathbf{a})$, i.e.

$$q_w(\alpha; \mathbf{a}) = \hat{q}(\{\alpha_{ij}\}) \prod_{k \in K} \Theta(a_k, \alpha_{pq}), \tag{4.7}$$

where $K = \{(p, q)\}$. $\Theta(a_k, \alpha_{pq})$ denotes the Heaviside theta function $\Theta(a_k - \alpha_{pq})$ or $\Theta(\alpha_{pq} - a_k)$. $\Theta(a_k - \alpha_{pq})$ is chosen when the mutation decreases the value of α_{pq} such that the values $\alpha_{pq} > a_k$ are absent in the mutant, or $\Theta(\alpha_{pq} - \text{a}_k)$ is chosen when the mutation increases the values of α_{pq} such that the lower values $\alpha_{pq} < a_k$ are absent in the mutant. To illustrate, the weak model for the co-infection NTHi (*wt*) + Mcat (*mcaB*) is parametrized by a_1 and is generated using $q_w(\alpha; a_1) = \hat{q}(\{\alpha_{ij}\})\Theta(\alpha_{21} - a_1)$. *Strong models*: The models in the *strong* category considered unanticipated changes in the LV interaction parameters for co-infection kinetics where a wild-type strain is replaced by a mutant strain. If a mutant strain in a co-infection is hypothesized to change a subset of LV interaction parameters $\{\alpha_{pq}\}$, then the strong models consider changes in $\{\alpha_{pq}\}$ and additional LV interactions (or unanticipated changes) outside of $\{\alpha_{pq}\}$. Similar to the weak models, the strong models are defined by the probability distribution function of the interactions parameters ($q_s(\alpha; \mathbf{a})$) and are parametrized by $\{a_k\}$ which quantifies the extent by which each of the interaction parameters is perturbed in $\hat{q}(\{\alpha_{ij}\})$ to generate $q_s(\alpha; \mathbf{a})$, i.e.

$$q_s(\alpha; \mathbf{a}) = \hat{q}(\{\alpha_{ij}\}) \prod_{k \in K} \Theta(a_k, \alpha_{pq}) \prod_{h \in H} \Theta(a_h, \alpha_{uv}). \tag{4.8}$$

In the above equations, $\text{H} \subset \{\alpha_{ij}\} \backslash \text{K}$, where H defines the set of LV interactions outside $\{\alpha_{pq}\}$. For example, the *strong* models for co-infection NTHi (*wt*) + Mcat (*mcaB*) were generated from in total six types of models: three types that vary pairs of interactions simultaneously, namely, $q_s(\alpha; a_1, a_2) = \hat{q}(\{\alpha_{ij}\})\Theta(\alpha_{21} - a_1)\Theta(\alpha_{11}, a_2)$, $q_s(\alpha; a_1, a_3) = \hat{q}(\{\alpha_{ij}\})\Theta(\alpha_{21} - a_1)\Theta(\alpha_{12}, a_3)$, $q_s(\alpha; a_1, a_4) = \hat{q}(\{\alpha_{ij}\})\Theta(\alpha_{21} - a_1)\Theta(\alpha_{22}, a_4)$, and three types that vary a triplet of interactions simultaneously, namely,

$q_s(\alpha; a_1, a_2, a_3) = \hat{q}(\{\alpha_{ij}\})\Theta(\alpha_{21} - a_1)\Theta(\alpha_{11}, a_2)\Theta(\alpha_{12}, a_3)$, $q_s(\alpha; a_1, a_2, a_4) = \hat{q}(\{\alpha_{ij}\})\Theta(\alpha_{21} - a_1)\Theta(\alpha_{11}, a_2)\Theta(\alpha_{22}, a_4)$ and $q_s(\alpha; a_1, a_3, a_4) = \hat{q}(\{\alpha_{ij}\})\Theta(\alpha_{21} - a_1)\Theta(\alpha_{12}, a_3)\Theta(\alpha_{22}, a_4)$. Note, in each of the six types of models α_{21} is always included and varied in the same way (increasing α_{21}). We consider changes up to the triplet of LV interactions for generating the *strong* models.

Step 2. The models $q_w(\alpha; \mathbf{a})$ and $q_s(\alpha; \mathbf{a})$ were used to generate means of bacterial populations $(\theta_{w/s} \equiv \{\bar{N}_1(\{a_k\}), \bar{N}_2(\{a_k\})\}_{w/s} \equiv \{\theta^1_{w/s}, \theta^2_{w/s}\})$ in the hosts using the steady-state equations (equation (4.1)). We assumed that each of the t samples of the test dataset $({}^{(1)}_{n'}D_x \ldots, {}^{(t)}_{n'}D_x)$ is distributed as a bivariate normal distribution,

$$P({}^{(i)}_{n'}D_x; \mu_1, \mu_2, \Sigma_1, \Sigma_2) = A\exp(-{}^{(i)}\chi^2), \tag{4.9}$$

where

$${}^{(i)}\chi^2 = n'\frac{[\mu_1 - {}^{(i)}_{n'}D^1_x]^2}{2\Sigma_1^2} + n'\frac{[\mu_2 - {}^{(i)}_{n'}D^2_x]^2}{2\Sigma_2^2}, \tag{4.10}$$

$\mu_{1,2}$ and $\Sigma_{1,2}$ can be estimated from ${}^{(1)}_{n'}D_x$, …, ${}^{(t)}_{n'}D_x$ as,

$$\hat{\mu}_1 = \frac{1}{t}\sum_{i=1}^{t}{}^{(i)}_{n'}D^1_x, \quad \hat{\mu}_2 = \frac{1}{t}\sum_{i=1}^{t}{}^{(i)}_{n'}D^2_x,$$
$$\hat{\Sigma}^2_1 = \frac{1}{t}\sum_{i=1}^{t}{}^{(i)}_{n'}D^3_x \quad \hat{\Sigma}^2_2 = \frac{1}{t}\sum_{i=1}^{t}{}^{(i)}_{n'}D^4_x.$$

Note the j index in the superscript of ${}^{(i)}_{n'}D^j_x$ denotes the j^{th} element of the set ${}^{(i)}_{n'}D_x$ and should not be confused as a power. We demand that the means of the bacterial populations in the models describe the means in $P({}^{(i)}_{n'}D_x; \mu_1, \mu_2, \Sigma_1, \Sigma_2)$, i.e.

$$\theta_{w/s}(\{a_k\})^1 = \mu_1, \ \theta_{w/s}(\{a_k\})^2 = \mu_2. \tag{4.11}$$

Next, we determined the parameters $(\{a_k\})$ by minimizing ${}^{(i)}\chi^2$ (or maximizing the corresponding likelihood),

$${}^{(i)}\chi^2_{w/s} = n'\frac{[\theta^1_{w/s} - {}^{(i)}_{n'}D^1_x]^2}{2\Sigma_1^2} + n'\frac{[\theta^2_{w/s} - {}^{(i)}_{n'}D^2_x]^2}{2\Sigma_2^2} \tag{4.12}$$

and then computed AIC for the ith sample of the test dataset after the minimization as,

$${}^{(i)}\text{AIC}_{w/s} = -2{}^{(i)}\chi^2(\mu = \hat{\theta}_{w/s})) + g(K), \tag{4.13}$$

where $\hat{\theta}_{w/s}$ is the mean bacterial population in the model evaluated at the $\{a_k\}$ values that minimized equation (4.10), and K is the number of parameters in $\{a_k\}$. We used $g(K) = 2K$ for our calculations.

4.3.1. Evaluation of the Condorcet winner

We made head-to-head comparisons for all possible pairwise combinations of models using AIC values as the metric. For a given pair, the winning model was the one which had a lower AIC value for a majority of the t samples. The model that won all of its head-to-head comparisons was declared the overall winner; this model is also known as the Condorcet winner, because it is preferred more than all others in pairwise comparisons. Throughout our study, both for the *in silico* and *in vivo* portions, we always found a Condorcet winner. The category under which the Condorcet winning model fell (*weak* or *strong*) was then assigned to x.

4.4. Generation of the synthetic data

The purpose of the synthetic data was to evaluate the framework with data that mimics *in vivo* data but has known levels of mutation strength and host immune response. In addition to the reference dataset, we generated 36 total *in silico* mutation datasets. Specifically, we generated a mutation on each of the four parameters (α_{11}, α_{12}, α_{21}, α_{22}); a mutation had one of three levels of severity (low, moderate, large), and was paired with one of three levels of host's immune response (none, weak, strong). Below, we outline how each was generated.

4.4.1. The reference set ${}_nD_r$

We first created a reference dataset analogous to a NTHi (*wt*) + Mcat (*wt*) co-infection experiment. Using the simplified LV two-species model (equation (4.4)), we set the following ranges for each of the interaction parameters: $\alpha_{11} \in [2.74 \times 10^{-3}, 0.2]$, $\alpha_{12} \in [-200,5]$, $\alpha_{21} \in [-5, 0.1]$ and $\alpha_{22} \in [1.89189, 140]$. These ranges were guided by the values observed from the *in vivo* data so that the *in silico* reference set was roughly similar to *in vivo* data. To calculate a single data point of a (N_1, N_2) pair of abundances, we randomly drew a value for each parameter (assuming a uniform distribution) and calculated the steady-state abundances using equation (4.4). For all the datasets, we generated $n = 10^5$ points per dataset. The reference dataset ${}_nD_r$ was calculated by collating all the (N_1, N_2) pairs obtained by drawing the α parameters from random distributions as described above.

4.4.2. Mutation sets

All 'mutations' of the reference dataset were done by increasing the value of exactly one interaction parameter; specifically, by raising the minimum value in the parameter's range. The three levels of the mutation indicated the severity of the increase. A low level α_{11} mutation has $\alpha_{11} \in [0.027, 0.2]$; at moderate severity: $\alpha_{11} \in [0.1, 0.2]$ and at large severity: $\alpha_{11} \in [0.18, 0.2]$. The ranges for all twelve possibilities are shown in electronic supplementary material, figures S9–S12.

We also aimed to evaluate the framework with respect to different levels of a host's immune response. In cases with no host response, we use the above method. For a non-zero immune response, we used a simplified model wherein only N_2 induces the immune response, and only N_1 is affected (detrimentally). Specifically, we introduced a Michaelis–Menten term to the first ODE in the standard LV two-species model:

$$\left.\begin{aligned} \frac{\mathrm{d}N_1^{(\mathrm{p})}}{\mathrm{d}t} &= N_1^{(\mathrm{p})}(1 - a_{11}^{(\mathrm{p})} N_1^{(\mathrm{p})} - a_{12}^{(\mathrm{p})} N_2^{(\mathrm{p})}) - k_d\, N_1^{(\mathrm{p})} \left(\frac{N_2^{(\mathrm{p})}}{K_m + N_2}\right) \\ \text{and} \quad \frac{\mathrm{d}N_2^{(\mathrm{p})}}{\mathrm{d}t} &= N_2^{(\mathrm{p})}(1 - a_{21}^{(\mathrm{p})} N_1^{(\mathrm{p})} - a_{22}^{(\mathrm{p})} N_2^{(\mathrm{p})}). \end{aligned}\right\} \tag{4.14}$$

In that term, K_m is the amount of N_2 necessary for the immune response to be at half-max. The susceptibility of N_1 to an immune attack is governed by the k_d parameter. So, with no immune response, we set $k_d = 0$. For a weak immune response $k_d \in [0,1]$. For a strong immune response, $k_d \in [0,10]$. As before, when choosing a value for k_d, we assumed a uniform distribution.

Animal ethics. All chinchilla infections were performed according to the protocols approved by the Wake Forest Animal Care and Use Committee. This study was conducted according to the guidelines outlined by National Science Foundation Animal Welfare Requirements and the Public Health Service Policy on the Humane Care and Use of Laboratory Animals. The Institutional Animal Care and Use Committees (IACUC) at Wake Forest School of Medicine approved these animal studies. Wake Forest's IACUC oversees the welfare, well-being and proper care and use of all vertebrate animals used for research and educational purposes at Wake Forest School of Medicine. The approved protocol number for the project is A13-140 and the Wake Forest School of Medicine animal welfare Assurance Number is D16-00248.

Data accessibility. The *in vivo* datasets and the Matlab code used to analyse the data are available at Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.j89d064 [35].

Authors' contribution. V.L., W.C.L.S., W.E.S. and J.D. planned research and wrote the manuscript. L.T. and W.E.S. carried out experiments. V.L., W.C.L.S., S.M. and J.D. carried out *in silico* simulations and mathematical calculations, and analysed *in silico* results. V.L., W.E.S. and J.D. analysed experimental data.

Competing interests. The authors declare no competing interests.

Funding. This work was supported by grants (R01GM103612) from NIGMS to J.D. and (RO1DC10051) to W.E.S.

Acknowledgements. We thank Ali Snedden at the High Performance Computing (HPC) Center at the Research Institute at the Nationwide Children's Hospital for the technical help with the computation. The computation time provided by the Baker cluster at HPC is also acknowledged. We thank Veronica Vieland for a critical reading of the manuscript.

References

1. Bakaletz LO. 2010 Immunopathogenesis of polymicrobial otitis media. *J. Leukoc. Biol.* **87**, 213–222. (doi:10.1189/jlb.0709518)

2. Armbruster CE, Hong W, Pang B, Weimer KE, Juneau RA, Turner J, Swords WE. 2010 Indirect pathogenicity of *Haemophilus influenzae* and *Moraxella catarrhalis* in polymicrobial otitis media occurs via interspecies quorum signaling. *MBio* **1**, e00102-10. (doi:10.1128/mBio.00102-10)

3. Brogden KA, Guthmiller JM, Taylor CE. 2005 Human polymicrobial infections. *Lancet* **365**, 253–255. (doi:10.1016/S0140-6736(05)70155-0)
4. Duan K, Sibley CD, Davidson CJ, Surette MG. 2009 Chemical interactions between organisms in microbial communities. In *Bacterial sensing and signaling* (eds M Collin, R Schuch), vol. 16, pp. 1–17. Basel, Switzerland: Karger. (doi:10.1159/000219369)
5. Lysenko ES, Ratner AJ, Nelson AL, Weiser JN. 2005 The role of innate immune responses in the outcome of interspecies competition for colonization of mucosal surfaces. *PLoS Pathog.* **1**, e1. (doi:10.1371/journal.ppat.0010001)
6. Ratner AJ, Lysenko ES, Paul MN, Weiser JN. 2005 Synergistic proinflammatory responses induced by polymicrobial colonization of epithelial surfaces. *Proc. Natl Acad. Sci. USA* **102**, 3429–3434. (doi:10.1073/pnas.0500599102)
7. Margolis E, Yates A, Levin BR. 2010 The ecology of nasal colonization of *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Staphylococcus aureus*: the role of competition and interactions with host's immune response. *BMC Microbiol.* **10**, 59. (doi:10.1186/1471-2180-10-59)
8. Murray JL, Connell JL, Stacy A, Turner KH, Whiteley M. 2014 Mechanisms of synergy in polymicrobial infections. *J. Microbiol.* **52**, 188–199. (doi:10.1007/s12275-014-4067-3)
9. Armbruster CE, Swords WE. 2010 Interspecies bacterial communication as a target for therapy in otitis media. *Expert Rev. Anti Infect. Ther.* **8**, 1067–1070. (doi:10.1586/eri.10.109)
10. Lysenko ES, Lijek RS, Brown SP, Weiser JN. 2010 Within-host competition drives selection for the capsule virulence determinant of *Streptococcus pneumoniae*. *Curr. Biol.* **20**, 1222–1226. (doi:10.1016/j.cub.2010.05.051)
11. Bakaletz LO. 2004 Developing animal models for polymicrobial diseases. *Nat. Rev. Microbiol.* **2**, 552–568. (doi:10.1038/nrmicro928)
12. Kot M. 2001 *Elements of mathematical ecology*. Cambridge, UK: Cambridge University Press.
13. May RM. 1981 *Theoretical ecology: principles and applications*, 2nd edn. Oxford, UK: Blackwell Scientific Publications.
14. Das J, Mukherjee S, Hodge SE. 2015 Maximum entropy estimation of probability distribution of variables in higher dimensions from lower dimensional data. *Entropy* **17**, 4986–4999. (doi:10.3390/e17074986)
15. Mukherjee S, Seok SC, Vieland VJ, Das J. 2013 Cell responses only partially shape cell-to-cell variations in protein abundances in *Escherichia coli* chemotaxis. *Proc. Natl Acad. Sci. USA* **110**, 18 531–18 536. (doi:10.1073/pnas.1311069110)
16. Mukherjee S, Weimer KE, Seok SC, Ray WC, Jayaprakash C, Vieland VJ, Swords WE, Das J. 2014 Host-to-host variation of ecological interactions in polymicrobial infections. *Phys. Biol.* **12**, 016003. (doi:10.1088/1478-3975/12/1/016003)
17. Presse S, Ghosh K, Lee J, Dill KA. 2013 Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **85**, 1115–1141. (doi:10.1103/RevModPhys.85.1115)
18. Burnham KP, Anderson DR. 2003 *Model selection and multimodel inference: a practical information-theoretic approach*. Berlin, Germany: Springer Science & Business Media.
19. Balder R, Krunkosky TM, Nguyen CQ, Feezel L, Lafontaine ER. 2009 Hag mediates adherence of *Moraxella catarrhalis* to ciliated human airway cells. *Infect. Immun.* **77**, 4597–4608. (doi:10.1128/IAI.00212-09)
20. Holm MM, Vanlerberg SL, Sledjeski DD, Lafontaine ER. 2003 The Hag protein of *Moraxella catarrhalis* strain O35E is associated with adherence to human lung and middle ear cells. *Infect. Immun.* **71**, 4977–4984. (doi:10.1128/IAI.71.9.4977-4984.2003)
21. Pearson MM, Lafontaine ER, Wagner NJ, St Geme III JW, Hansen EJ. 2002 A *hag* mutant of *Moraxella catarrhalis* strain O35E is deficient in hemagglutination, autoagglutination, and immunoglobulin D-binding activities. *Infect. Immun.* **70**, 4523–4533. (doi:10.1128/IAI.70.8.4523-4533.2002)
22. Armbruster CE, Hong W, Pang B, Dew KE, Juneau RA, Byrd MS, Love CF, Kock ND, Swords WE. 2009 LuxS promotes biofilm maturation and persistence of nontypeable *Haemophilus influenzae in vivo* via modulation of lipooligosaccharides on the bacterial surface. *Infect. Immun.* **77**, 4081–4091. (doi:10.1128/IAI.00320-09)
23. Press WH. 1992 *Numerical recipes in FORTRAN: the art of scientific computing*, 2nd edn. Cambridge, UK: Cambridge University Press.
24. Young HP. 1988 Condorcet's theory of voting. *Am. Political Sci. Rev.* **82**, 1231–1244. (doi:10.2307/1961757)
25. Armbruster CE, Pang B, Murrah K, Juneau RA, Perez AC, Weimer KE, Swords WE. 2011 RbsB (NTHI_0632) mediates quorum signal uptake in nontypeable *Haemophilus influenzae* strain 86-028NP. *Mol. Microbiol.* **82**, 836–850. (doi:10.1111/j.1365-2958.2011.07831.x)
26. Rickard AH, Palmer Jr RJ, Blehert DS, Campagna SR, Semmelhack MF, Egland PG, Bassler BL, Kolenbrander PE. 2006 Autoinducer 2: a concentration-dependent signal for mutualistic bacterial biofilm growth. *Mol. Microbiol.* **60**, 1446–1456. (doi:10.1111/j.1365-2958.2006.05202.x)
27. Szelestey BR, Heimlich DR, Raffel FK, Justice SS, Mason KM. 2013 *Haemophilus* responses to nutritional immunity: epigenetic and morphological contribution to biofilm architecture, invasion, persistence and disease severity. *PLoS Pathog.* **9**, e1003709. (doi:10.1371/journal.ppat.1003709)
28. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. 2012 Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* **8**, e1002606. (doi:10.1371/journal.pcbi.1002606)
29. Fisher CK, Mehta P. 2014 Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS ONE* **9**, e102451. (doi:10.1371/journal.pone.0102451)
30. Stein RR, Bucci V, Toussaint NC, Buffie CG, Ratsch G, Pamer EG, Sander C, Xavier JB. 2013 Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput. Biol.* **9**, e1003388. (doi:10.1371/journal.pcbi.1003388)
31. Barbier M, Arnoldi J-F, Bunin G, Loreau M. 2018 Generic assembly patterns in complex ecological communities. *Proc. Natl Acad. Sci. USA* **115**, 2156. (doi:10.1073/pnas.1710352115)
32. Unhanand M, Maciver I, Ramilo O, Arencibiamireles O, Argyle JC, Mccracken GH, Hansen EJ. 1992 Pulmonary clearance of *Moraxella catarrhalis* in an animal model. *J. Infectious Dis.* **165**, 644–650. (doi:10.1093/infdis/165.4.644)
33. Bakaletz LO, Tallan BM, Hoepf T, Demaria TF, Birck HG, Lim DJ. 1988 Frequency of fimbriation of nontypable *Haemophilus influenzae* and its ability to adhere to chinchilla and human respiratory epithelium. *Infect. Immun.* **56**, 331–335.
34. Firman T, Wedekind S, McMorrow T, Ghosh K. 2018 Maximum caliber can characterize genetic switches with multiple hidden species. *J. Phys. Chem. B* **122**, 5666–5677. (doi:10.1021/acs.jpcb.7b12251)
35. Lakhani V, Tan L, Mukherjee S, Stewart WCL, Swords WE, Das J. 2018 Data from: Mutations in bacterial genes induce unanticipated changes in the relationship between bacterial pathogens in experimental otitis media. Dryad Digital Repository. (doi:10.5061/dryad.j89d064)