

Host contact dynamics shapes richness and dominance of pathogen strains

Francesco Pinotti, Eric Fleury, Didier Guillemot, Pierre-Yves Böelle, Chiara Poletto

▶ **To cite this version:**

Francesco Pinotti, Eric Fleury, Didier Guillemot, Pierre-Yves Böelle, Chiara Poletto. Host contact dynamics shapes richness and dominance of pathogen strains. PLoS Computational Biology, 2019, 15 (5), pp.e1006530. 10.1371/journal.pcbi.1006530 . inserm-02153754

HAL Id: inserm-02153754
https://inserm.hal.science/inserm-02153754

Submitted on 12 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Host contact dynamics shapes richness and dominance of pathogen strains

Francesco Pinotti[1], Éric Fleury[2], Didier Guillemot[3], Pierre-Yves Böelle[1], Chiara Poletto[1]*

1 INSERM, Sorbonne Université, Institut Pierre Louis d'Épidémiologie et de Santé Publique (IPLESP), 75012 Paris, France, **2** INRIA, Paris, France, **3** Inserm, UVSQ, Institut Pasteur, Université Paris-Saclay, Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases (B2PHI), Paris, France

* chiara.poletto@inserm.fr

Check for updates

OPEN ACCESS

Citation: Pinotti F, Fleury É, Guillemot D, Böelle P-Y, Poletto C (2019) Host contact dynamics shapes richness and dominance of pathogen strains. PLoS Comput Biol 15(5): e1006530. https://doi.org/10.1371/journal.pcbi.1006530

Editor: James Lloyd-Smith, University of California, Los Angeles, UNITED STATES

Received: September 19, 2018

Accepted: April 29, 2019

Published: May 21, 2019

Copyright: © 2019 Pinotti et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The contact network and carriage data are available from the http://dx.doi.org/10.6084/m9.figshare.1290932 database.

Funding: FP received funding from "Pierre Louis" School of Public Health of UPMC (www.ed393.upmc.fr). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The interaction among multiple microbial strains affects the spread of infectious diseases and the efficacy of interventions. Genomic tools have made it increasingly easy to observe pathogenic strains diversity, but the best interpretation of such diversity has remained difficult because of relationships with host and environmental factors. Here, we focus on host-to-host contact behavior and study how it changes populations of pathogens in a minimal model of multi-strain interaction. We simulated a population of identical strains competing by mutual exclusion and spreading on a dynamical network of hosts according to a stochastic susceptible-infectious-susceptible model. We computed ecological indicators of diversity and dominance in strain populations for a collection of networks illustrating various properties found in real-world examples. Heterogeneities in the number of contacts among hosts were found to reduce diversity and increase dominance by making the repartition of strains among infected hosts more uneven, while strong community structure among hosts increased strain diversity. We found that the introduction of strains associated with hosts entering and leaving the system led to the highest pathogenic richness at intermediate turnover levels. These results were finally illustrated using the spread of *Staphylococcus aureus* in a long-term health-care facility where close proximity interactions and strain carriage were collected simultaneously. We found that network structural and temporal properties could account for a large part of the variability observed in strain diversity. These results show how stochasticity and network structure affect the population ecology of pathogens and warn against interpreting observations as unambiguous evidence of epidemiological differences between strains.

Author summary

Pathogens are structured in multiple strains that interact and co-circulate on the same host population. This ecological diversity affects, in many cases, the spread dynamics and the efficacy of vaccination and antibiotic treatment. Thus understanding its biological and host-behavioral drivers is crucial for outbreak assessment and for explaining trends of

new-strain emergence. We used stochastic modeling and network theory to quantify the role of host contact behavior on strain richness and dominance. We systematically compared multi-strain spread on different network models displaying properties observed in real-world contact patterns. We then analyzed the real-case example of *Staphylococcus aureus* spread in a hospital, leveraging on a combined dataset of carriage and close proximity interactions. We found that contact dynamics has a profound impact on a strain population. Contact heterogeneity, for instance, reduces strain diversity by reducing the number of circulating strains and leading few strains to dominate over the others. These results have important implications in disease ecology and in the epidemiological interpretation of biological data.

Introduction

Interactions between strains of the same pathogen play a central role in how they spread in host populations. [1–7]. In *Streptococcus pneumoniae* and *Staphylococcus aureus*, for instance, several dozen strains can be characterized for which differences in transmissibility, virulence and duration of colonization have been reported in some cases [8, 9]. Strain diversity may also affect the efficacy of prophylactic control measures such as vaccination or treatment. Indeed, strains may be associated with different antibiotic resistance profiles [3, 5, 10, 11], and developed vaccines may only target a subset of strains [2, 3, 12]. With the increasing availability of genotypic information, it has become easy to describe the ecology of population of pathogens and to monitor patterns of extinction and dominance of pathogen variants [13–17]. However, the reasons for multi-strain coexistence patterns (e.g. coexistence between resistant and sensitive strains) or dominance of certain strains (e.g. in response to the selection pressure induced by treatment and preventive measures) remain elusive. One may invoke selection due to different pathogen characteristics, but also environmental and host population characteristics, leading to differences in host behavior, settings and spatial structure may affect the ecology of strains [14–19]. In particular, human-to-human contacts play a central role in infectious disease transmission [20]. This is increasingly well described thanks to extensive high-resolution data—including mobility patterns [21–23], sexual encounters [24], close proximity interactions in schools [25, 26], workplaces [27], hospitals [16, 28–31], etc.—that enable basing epidemiological assessment on contact data with real-life complexity [32, 33]. For instance, the frequency of contacts can be highly heterogeneous leading more active individuals to be at once more vulnerable to infections and acting as super-spreaders after infection [24, 33–35]. Organizational structure of certain settings (school classes, hospital wards, etc.) and other spatial proximity constraints lead to the formation of communities that can delay epidemic spread [36, 37]. Individual turnover in the host population is also described as a key factor in controlling an epidemic [20, 38]. It is likely that, since they impact the spread of single pathogens, the same characteristics could affect the dynamics in multi-strain populations. It was shown, indeed, that network structure impacts transmission with two interacting strains [39–46], the evolution of epidemiological traits [47–49] and the effect of cross-immunity [50, 51]. Yet in these cases, complex biological mechanisms—such as mutation, variations in transmissibility and infectious period, cross immunity—were used to differentiate between pathogens, thereby making the role of network characteristics difficult to assess in its own right.

For this reason, we focused on the dynamical pattern of human contacts and examined whether it contributes to shaping the population ecology of interacting strains under minimal epidemiological assumptions regarding transmission. We described a neutral situation where

all strains have the same epidemiological traits and compete via mutual exclusion (concurrent infection with multiple strains is assumed to be impossible) in a Susceptible-Infected-Susceptible (SIS) framework. We studied the spread of pathogens in a host population during a limited time window, disregarding long-term evolution dynamics of pathogens. More precisely, new strains were introduced through host turnover rather than *de novo* mutation or recombination in pathogens. We quantified the effect of network properties on the ecological diversity in strain populations with richness and dominance indicators. We assessed in turn heterogeneities in contact frequency, community structure and host turnover by comparing simulation results obtained with network models exhibiting a specific feature. We then interpreted *S. aureus* carriage in patients of a long-term care facility in the light of these results.

Results

Multi-strain spread on dynamical networks

We simulated the stochastic spread of multiple strains on a dynamical contact network of individuals (nodes of the network). Individuals can be either susceptible or infected with a single strain at a given time, and, for each strain, β and μ indicate the transmission and the recovery rate respectively. We assumed turnover of individuals, who enter the system with rate λ_{in}, and associated injection of previously unseen strains, carried by incoming individuals with probability p_s. We considered synthetic network models, each displaying a specific structural feature, as well as a real network reconstructed from close-proximity-interaction data collected in a hospital facility. We calibrated all network models to the same average quantities—average population size $\bar{V}$, fraction of active nodes $\bar{a}$, average degree $\bar{k}$ and strength of the community repartition p_{IN}, when applicable—that were chosen to correspond with the hospital network used in the application. Epidemiological parameters were motivated by the duration of *S. aureus* carriage in patients. A larger range of values was explored in some cases to address their impact on the dynamics. We analyze the structure of strain population at the dynamic equilibrium by computing, for each network model, ecological diversity measures, including species richness and evenness/dominance indices [52, 53]. All details about network models, numerical simulations and ecological indicators are described in the Materials and methods section.

Effects of contact heterogeneity

In order to probe the effect of contact heterogeneity on strain ecology we compared a homogeneous model (HOM) in which all nodes have the same activity potential, i.e. they have equal rate of activation to establish contacts, with a heterogeneous model (HET), akin to the activity-driven model described in [34], where the activity potential is different across nodes and is drawn from a power-law distribution.

Fig 1 shows the results of numerical simulations comparing HOM and HET models. Sample epidemic trajectories are reported in Fig 1A. Here every strain is indicated with its own color to display its dynamics resulting from the interaction with the other strains. Fig 1B–1D shows summary statistics in varying strain transmissibility β. The prevalence presents a well-known behavior for both static and dynamic networks (Fig 1B): contact heterogeneities lower the transmissibility threshold above which total prevalence is significantly above zero, thus allowing the spread of pathogens with low transmissibility. At the same time, however, heterogeneities hamper the epidemic spread when β is large, reducing the equilibrium prevalence [35]. Fig 1C shows the average richness, i.e. the number of distinct strains co-circulating. For low values of β HET displays larger richness values compared to HOM. This trend reverses as

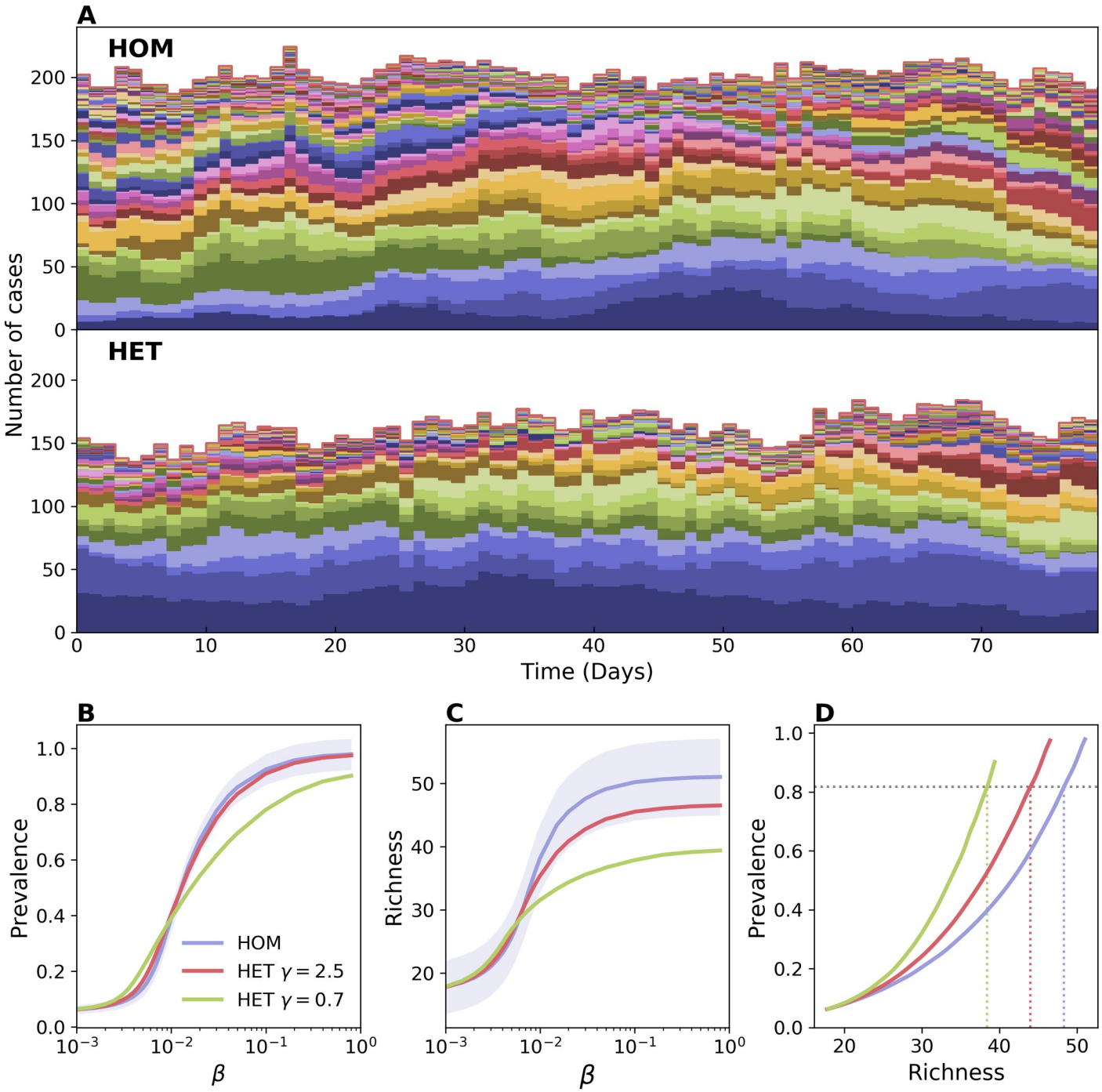


Fig 1. Effect of contact heterogeneity on strain richness. Comparison between a homogeneous (HOM) and a heterogeneous (HET) network. In HOM all nodes have the same activity rate $a_H = 0.285$ and the network average degree is $\bar{k} = 0.89$. In HET the activity rate of each node is drawn randomly from a power-law distribution with support $(\epsilon, 1]$ and the same average value as in HOM. Lower values of the power-law exponent γ correspond to a higher contact heterogeneity. The average degree is the same as in HOM. We chose a population of $\bar{V} = 306$ individuals, average length of stay $\tau = 10$ days, probability of strain injection per incoming individual $p_s = 0.079$, and recovery rate $\mu = 0.00192$ (see Materials and methods). (A) Sample time series of strain abundance for HOM and HET with $\gamma = 0.7$. Each time series is represented with a different color. All abundances are stacked together, so that plot's height represents prevalence. Here $\beta = 0.02$. (B) Average prevalence vs β, and (C) average richness vs β. Two levels of heterogeneity are here considered for HET. For the sake of visualization, the shaded area corresponding to the standard deviation is shown only for HOM. Median and confidence intervals are reported in S1 Fig of the supporting information. (D) Average prevalence vs richness. Dashed lines are shown as a guide to the eye, highlighting variation in richness induced by network topology.

https://doi.org/10.1371/journal.pcbi.1006530.g001

β increases, and the richness is lower in HET consistently with the lower level of prevalence. The relation between richness and prevalence, however, is not straightforward. For instance, the reduction in richness for high β values is important even for the case with limited contact heterogeneity, when prevalence is barely affected. The scaling between prevalence and richness is not linear as β varies (Fig 1D), and the relation between the two quantities varies appreciably among contact networks. In correspondence of a fixed value of prevalence, heterogeneous networks have lower richness—e.g. a prevalence value of ∼0.8 corresponds to ∼20% lower richness in HET with respect to HOM, as highlighted in Fig 1D.

This fact can be explained by the dynamical properties of epidemics on heterogeneous networks. Active nodes, involved in a larger number of contacts, get infected more frequently [35]. Also, a randomly chosen node is likely surrounded by active nodes [33]. As a consequence, injected strains often find their propagation blocked by active infected nodes. In this way, contact heterogeneities enhance the competition induced by mutual exclusion and hamper the wide-spread of emerging strains, similarly to what was found in [46]. This mechanism is further confirmed by looking at the persistence time of strains (S2 Fig in the supporting information). Above the epidemic threshold, it is on average shorter in heterogeneous networks than in homogeneous ones. The distributions are however more skewed in heterogeneous networks, indicating that more strains are going extinct rapidly, while a few others can survive for a long time in the population.

If on the one hand hubs accelerate the extinction of certain strains, on the other they act as super-spreaders, amplifying the propagation of other strains. We find that this impacts profoundly the distribution of strains' abundances, i.e. the strain-specific prevalence. Fig 2A shows that the latter is broader for the HET network, with the most abundant strain reaching a larger proportion of cases. This situation is synthesized by the Berger-Parker index, that quantifies the level of unevenness or dominance of a given ecological system [52, 53]. This is defined as the relative abundance of the most abundant strain (see Materials and methods section). Fig 2B shows that Berger-Parker index increases with β for all networks. This is expected since at low β strains' transmission chains are short and barely interact, while they interfere more at higher values of transmission potential. The Berger-Parker index is always higher in a

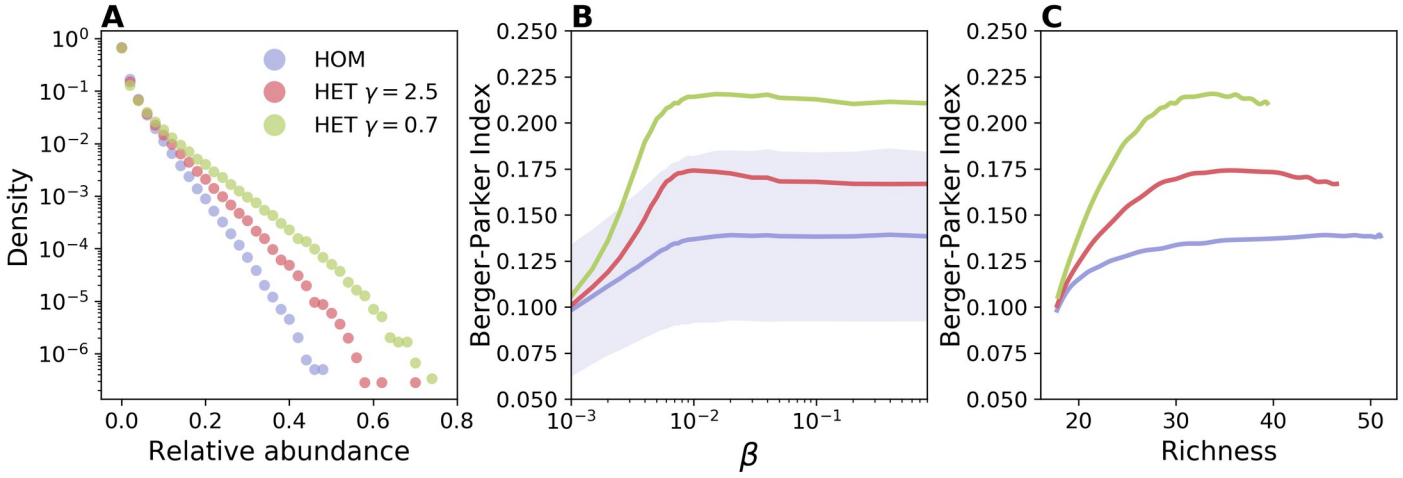


Fig 2. Effect of contact heterogeneity on strain dominance. (A) Distribution of strains' relative abundance, i.e. the frequency of strains infecting a given fraction of the total prevalence, for HOM (blue), HET with $\gamma = 2.5$ (red), and HET with $\gamma = 0.7$ (green). Here $\beta = 0.02$. (B) Berger-Parker index, i.e. the relative abundance of the most abundant strain, as a function of β. For the sake of visualization, the shaded area corresponding to the standard deviation is shown only for HOM. Median and confidence intervals are reported in S1 Fig of the supporting information. (C) Berger-Parker index vs richness. Parameters are the same as in Fig 1.

https://doi.org/10.1371/journal.pcbi.1006530.g002

heterogeneous network, even when the comparison is made at fixed values of richness (Fig 2C). An alternative indicator, the Shannon evenness, shows a similar behavior as displayed in S3 Fig.

The fraction of strains going extinct also depends on stochastic effects in a finite size population. We indeed found that increasing network size, when temporal and topological properties were the same, led to an increase in both persistence time and richness (S4 Fig). This shows that interference among transmission chains is reduced in larger populations. However, the relative abundance distribution remained similar, showing that it is primarily affected by the nodes' activity distribution (S5 Fig).

Eventually, we tested whether additional mechanisms of strain injection were leading to different results. In S6 Fig we assumed new strains to infect susceptible nodes already present in the system with rate q_s, mimicking in this way transmissions originating from an external source, as it can happen in real cases. The plot of S6 Fig shows the same qualitative behavior described here.

Effect of community structure

We considered a community model (COM) with n_C communities in which all nodes are as active as in HOM, but direct a fraction p_{IN} of their links within their community and the rest to nodes in the remaining $n_C - 1$ communities. The closer p_{IN} is to 1, the stronger the repartition in communities is.

Fig 3A and 3B shows that a network with communities displays a higher richness for large β; even when community structure barely affects prevalence (Fig 3B). However, the effect is important only when communities are fairly isolated ($p_{IN} = 0.99$) and the injection from the outside is not so frequent—otherwise the effect is masked by strain injection which occurs uniformly across communities. In particular, for the values of $p_{IN} = 0.78$ and $p_s = 0.079$, chosen to match the hospital application, the difference with the homogeneous case is very small. The limited role of community structure is also confirmed by the fact that once this feature is combined with heterogeneous activation—in a model with the activation scheme of HET and the stub-matching of COM—the latter property has the dominant effect and the richness decreases (S1 Fig).

The relation between richness and prevalence remains the same when adding the injection of new strains due to the transmission from an external source. This mechanism further increases the richness. When β is high and the fraction of infected nodes is close to one, however, such a mechanism is hindered by the fact that susceptible nodes, that can get infected from the external source, are rare (see S6 Fig). This is why richness starts to decrease for high values of β.

We tested the consequences of communities in strain dominance by plotting the Berger-Parker index in Fig 3C. For low β, the behavior of the Berger-Parker index follows the trend in richness. The initial decrease in this indicator is due to the increase in richness, that occurs at constant prevalence and is thus associated to a decrease in the average abundance [54]—green curve corresponding to $p_{IN} = 0.99$ and $p_s = 0.01$. At larger values of β, instead, increased competition levels induced higher dominance levels.

The increase in strain diversity is due to the reduced competition among strains introduced in different communities. When coupling among communities is low, indeed, strains may spend the majority of time within the community they were injected in, thus avoiding strains injected in other communities. Fig 3D confirms this hypothesis by showing the Inverse Participation Ratio (*IPR*) [55] that quantifies uniformity in the repartition of abundance across communities. Values close to zero indicate uniform repartition, while, conversely, values close to 1

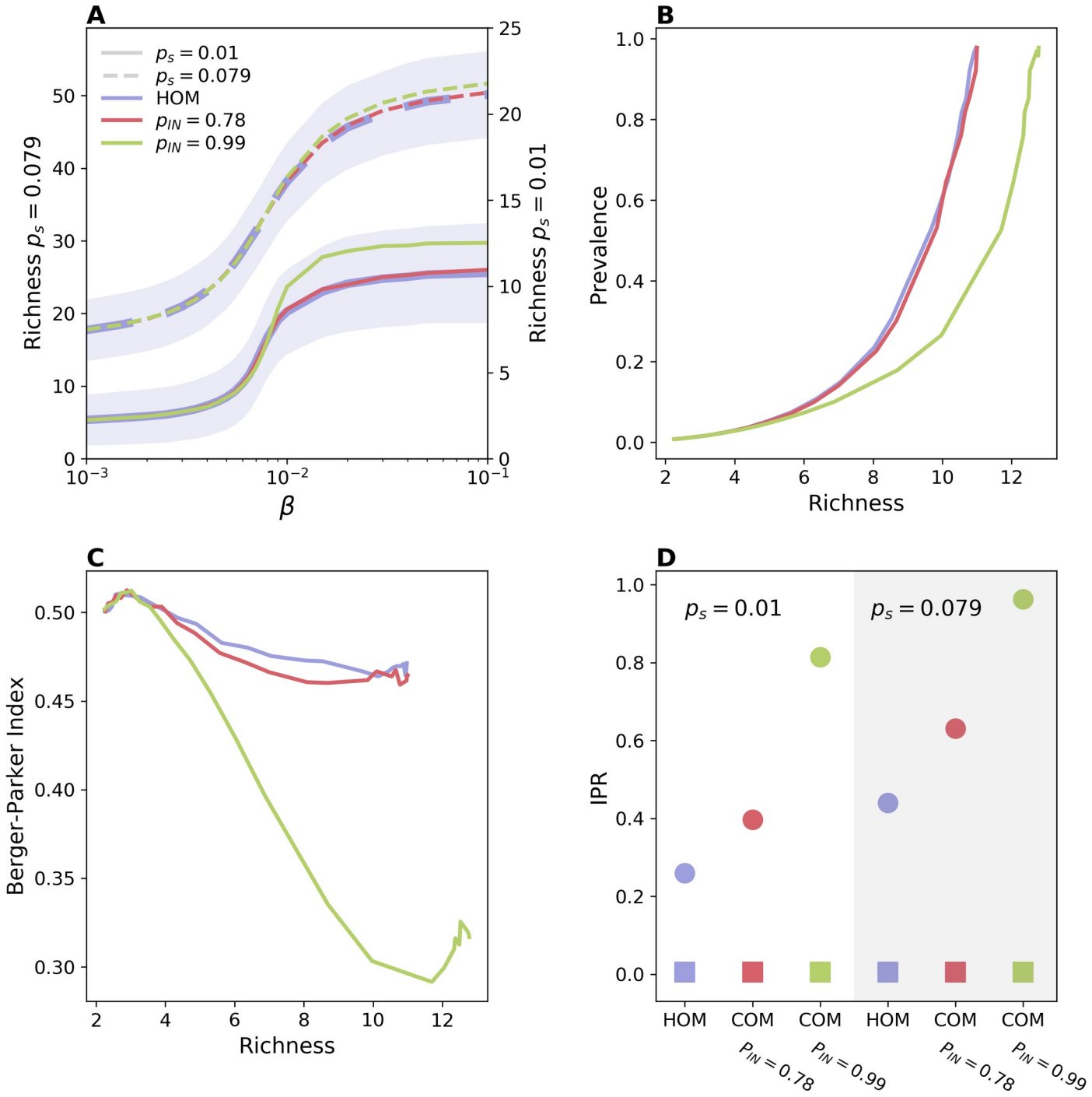


Fig 3. Impact of community structure. (A) Richness vs β for HOM (blue), COM with $p_{IN} = 0.78$ (red), and COM with $p_{IN} = 0.99$ (green). For both COM models we have set $n_C = 6$. Solid lines correspond to $p_s = 0.01$, while dashed lines correspond to $p_s = 0.079$. Solid lines refer to the right y-axis, while dashed ones to the left y-axis. For the sake of visualization, the shaded area corresponding to the standard deviation is shown only for HOM. Median and confidence intervals are reported in S1 Fig of the supporting information. (B) Prevalence vs richness and (C) Berger-Parker index vs richness, for $p_s = 0.01$. (D) Average *IPR* for both $p_s = 0.01$ (white background) and $p_s = 0.079$ (gray background). Here $\beta = 0.02$. Squares correspond to *IPR* obtained from total prevalence while circles correspond to *IPR* obtained from strains' abundances. A value of the *IPR* close to 1 indicates localization over one community. Here no injection due to transmission from an external source is assumed ($q_s = 0$). The effect of this second mechanism is shown in S6 Fig.

https://doi.org/10.1371/journal.pcbi.1006530.g003

indicate that, on average, a strain is confined within a single community for most of the time (more details are reported in the Materials and methods section). The strength of the community structure does not affect the repartition of the total prevalence (squares in the plot), however it alters the average *IPR* value computed from the abundance of single strains, thus strains become more localized as p_{IN} increases. Notice that a certain degree of localization is present also in the homogeneous network, due to those strains causing very few generations before going extinct.

As a sensitivity analysis we tested whether the main results obtained so far are the same in a more realistic situation where additional heterogeneous properties of nodes are accounted for. We consider the case in which infectious duration varies across individuals, as happens for *S. aureus* colonization. S7 Fig shows that the inclusion of three classes differing in recovery rate reduces richness and increases the Berger-Parker index with respect to the homogeneous recovery. However, the effects discussed so far—e.g. reduction and amplification of richness in HET and COM, respectively—are still present.

Effect of turnover of individuals

Node turnover represents another important property of a network that may impact the ecological dynamics of strains for two reasons: incoming individuals contribute to richness by injecting new strains; on the other hand, the removal from the population of infected nodes breaks transmission chains and hampers the persistence of strains. The result of the interplay between these two mechanisms is summarized by the plot of richness as a function of β and node length of stay, τ,—Fig 4A. The figure, obtained with the HOM model, shows two distinct regimes. In the former case, richness decreases as τ increases, because replacement of individuals becomes slower and injections less frequent. In the high β regime, instead, the average richness at fixed β does not depend monotonically on the node turnover but it is instead maximized at intermediate τ. Interestingly, the optimal value of τ decreases as β increases. This

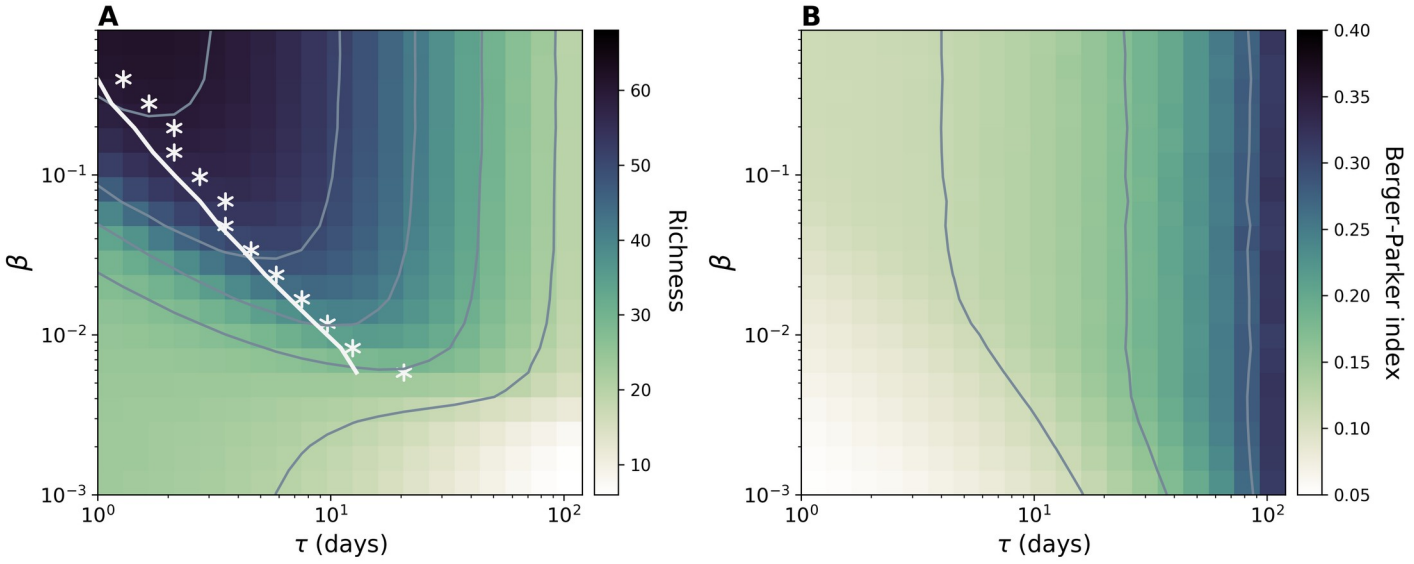


Fig 4. Effects of node length of stay on strain diversity. (A) Average richness and (B) Average Berger-Parker index for simulations on HOM model. Contour plots are shown in both figures. While exploring τ we also set the value of the average network size $\bar{V}$ to 306, thus the injection rate can be computed by the relation $\lambda_{in} = \bar{V}/\tau$. For each value of β we highlight in panel (A) the value of the length of stay corresponding to the maximum richness (white asterisks) together with the analytical prediction (white line) obtained by using Eq (1). Here $\mu = 0.00192$, $\bar{k} = 0.89$, $a_H = 0.28$.

https://doi.org/10.1371/journal.pcbi.1006530.g004

behavior can be explained by looking at the balance between injection and extinction that determines the equilibrium value of richness, $\bar{N}_S$. This reads [56]:

$$\bar{N}_S = \lambda_{\text{in}} p_s T_{\text{pers}}(\beta, \tau) = \bar{V}\, p_s\, \frac{T_{\text{pers}}(\beta, \tau)}{\tau}, \quad (1)$$

where $\lambda_{\text{in}} p_s$ is the rate at which new strains are introduced and T_{pers} is the average persistence time of a strain. The trade-off between injection and extinction appears as the ratio between the two time scales, T_{pers} and τ. In the limit $\tau \to 0$ the spread plays no role, even for high β. As τ increases, newly introduced infectious seeds have a higher probability to spread, thus the average extinction time initially increases super-linearly with τ (see S8 Fig in the supporting information) resulting in an increase of richness. However, past a certain value of τ, T_{pers} does not grow super-linearly anymore, thus a further increase in τ is detrimental for pathogen diversity because it is associated to fewer introductions. This general behavior was not altered by the accounting for introductions by transmissions from an external source as shown in S6 Fig.

We derive an approximate formula for T_{pers} considering an emerging strain competing with a single effective strain formed by all other strains grouped together. This formulation, enabled by the neutral hypothesis, makes it possible to write the master equation describing the dynamics and to use the Fokker-Planck approximation to derive persistence times (see Materials and methods section). Analytical results well reproduce the behavior observed in the simulations, and, in particular, the value of the length of stay maximizing richness for different β as shown by the comparison between white stars and continuous line in Fig 4B. The quantitative match for other values of p_s is reported in S9 Fig.

Unlike richness, Berger-Parker index always increases monotonically with the length of stay—Fig 4B. This behavior is due to the correlation of this indicator with average abundance, similarly to what we discussed in the previous section.

Spread of *S. aureus* in a hospital setting

We conclude by analyzing the real-case example of the *S. aureus* spread in a hospital setting [10, 57]. We used close-proximity-interaction (CPI) data recorded in a long-term health-care facility during 4 months by the i-Bird study [16, 28, 31]. These describe a high-resolution dynamical network whose complex structure reflects the hospital organization, the subdivision in wards and the admission and discharge of patients [58]. Together with the measurements of contacts, weekly nasal swabs were routinely performed to monitor the *S. aureus* carriage status of the participants and to identify the spa-type and the antibiotic resistance profile of the colonizing strains.

The modeling framework considered here well applies to this case. The SIS model is widely adopted for modeling the *S. aureus* colonization [59, 60], and the assumption of mutual exclusion is made by the majority of works to model the high level of cross-protection recognized by both epidemiological and microbiological studies [61, 62]. The dynamic CPI network was previously shown to be associated with paths of strain propagation [16]. Consistently, we assumed that transmission is mediated by network links with transmissibility β. In addition, new strains are introduced in the population carried by incoming patients, or through contacts with persons not taking part in the study.

Fig 5A shows weekly carriage and its breakdown in different strains. Prevalence and richness fluctuate around the average values 87,3 ± 6,3 cases and 39,8 ± 2 strains, respectively. Simulation results are reported in Fig 5B, that displays the impact of transmission and introduction rate on richness and prevalence. When q_s is low we find a positive trend between

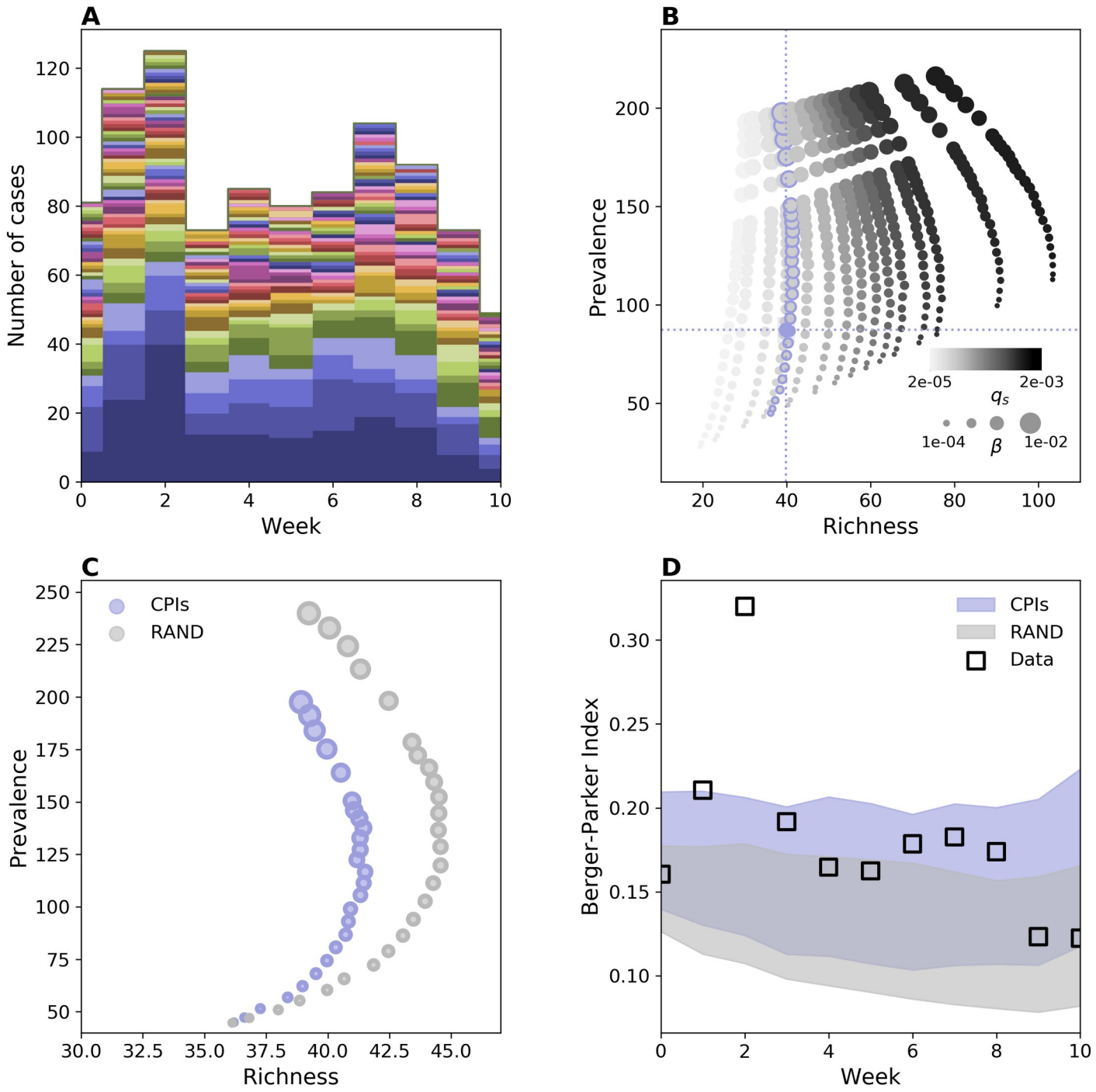


Fig 5. *S. aureus* population structure on hospital network. (A) Weekly carriage data measured during the i-Bird experiment. Each *S. aureus* strain's abundance time series is represented by a different color. All time series have been stacked as in Fig 1A. (B) Prevalence vs. richness from simulations on the hospital network for different β and different rates of introduction, here tuned by the parameter q_s. Blue dashed lines represent the average empirical values. (C) Prevalence vs richness for hospital contact data (blue dots) and RAND null model (grey dots). In RAND contacts are randomized by preserving the first and the last contact of every individual. Markers' size in (B) and (C) is proportional to the value of β. The hospital curve corresponds to the curve in (B) with blue-contour markers. (D) Weekly value of Berger-Parker index in carriage data (squares) along with the same quantity from the simulations. The shaded areas indicate the average plus/minus the standard deviation obtained from 1000 stochastic runs. For each network, parameter values are the ones that reproduce empirical prevalence and richness. Duration of colonization is assumed here to be 35 days [63]. Alternative values of this parameter led to the same qualitative results (S10 Fig).

https://doi.org/10.1371/journal.pcbi.1006530.g005

richness and prevalence, consistently with the synthetic case. For larger values of q_s the trend appears instead different. As transmissibility increases, richness initially grows with prevalence and then decreases after a certain point. This behavior is the same as observed in S6 Fig and stems from the reduction of susceptible nodes, that causes a decline in the expected injection rate—see Materials and methods section.

To quantify the effect of contact patterns on *S. aureus* population ecology we compared simulation results with the ones on a network null model. Specifically, we built the RAND null model that randomizes contacts while preserving just the first and the last contact of every individual. The randomization preserves node turnover, the number of active nodes and links and destroys contact heterogeneities and community structure along with other correlations. Fig 5C shows the comparison for different transmissibility values. The effect of the network is consistent with the theoretical results described for a heterogeneous network, i.e. smaller richness values correspond to the same prevalence in the real network compared to the homogeneous one. We then quantified the level of dominance of the multi-strain distribution by means of the Berger-Parker index. We chose for each network the values of q_s and β that better reproduce empirical richness and prevalence and, interestingly, we found that, for the two cases, same average richness and prevalence correspond to different levels of Berger-Parker index. The Berger-Parker index obtained with the real network is the highest and the one that better matches the empirical values—i.e. the empirical values are within one standard deviation of the mean for almost all weeks. Based on this result we argue that contact heterogeneities, along with the other properties of the contact network, contribute to the increased dominance of certain strains.

Discussion

Multiple biological and environmental factors concur in shaping pathogen diversity. We focused here on the host contact network and we used a minimal transmission model to assess the impact of this ingredient on strain population ecology, quantifying the effects of three main network properties, i.e. heterogeneous activity potential, presence of communities and turnover of individuals. Results show that the structure and dynamics of contacts can alter profoundly strains' co-circulation. Contact heterogeneities were found to shape the distribution of strains' abundances. Highly active nodes are known to play an important role in outbreak dynamics by acting as super-spreaders [33]. At the same time, however, they were found to enhance the interference between the transmission chains of different strains, thus hindering the spread of an emerging variant [46]. Here we showed that the combination of these two dynamical mechanisms reduces the richness and increases the level of heterogeneity in strains' abundances. In particular, hubs could allow strains with no biological advantage to generate a large number of cases and outcompete other equally fit strains. This mechanism may potentially bias the interpretation of biological data. Dynamical models that do not properly account for contact structure could overestimate the difference in strains' epidemiological traits in the attempt to explain observed fluctuations in strain abundance induced in reality by super-spreading events. Moreover, these models could provide biased assessment of transmission vs. introduction rates.

The presence of communities causes the separation of strains and mitigates the effect of competition thus enhancing co-existence. A similar behavior was already pointed out before [46, 51, 59, 64], e.g. for the spread of *S. pneumoniae*, as induced by age assortativity [64], for the case of *S. aureus* where distinct settings were considered [59], and for a population of antigenic distinct strains in presence of cross-immunity [51]. We found that the impact of community structure is not so strong, and it is likely minor when individuals of different

communities have frequent contacts. No appreciable variation was observed, indeed, for p_{IN} = 0.78, chosen to match the inter-ward coupling of the hospital network. Similar results can be expected for school classes or workplace departments presenting a similar level of community mixing. The effect on richness becomes appreciable for low community coupling (e.g. p_{IN} = 0.99 in Fig 3). This is consistent with a certain degree of diversity observed among strains belonging to separated communities, as it is the case of different hospitals [15].

Eventually, the analysis of turnover of individuals revealed major effects on strain diversity, when this mechanism is also the main driver of the introduction of strains in the population. When transmissibility is low richness decreases with host length of stay. When transmissibility is above the epidemic threshold we showed the existence of an optimal value of the length of stay that maximizes strain richness as a result of the interplay between two competing time scales, namely the typical inter-introduction time and the average persistence time of a strain. This provides insights for the spread of bacterial infections in transmission settings, such as hospitals or farms, that are of particular relevance for the spread of antimicrobial resistance and that are characterized by a rapid host turnover [15, 31, 65]. For the case of hospitals, for instance, they suggest that variations in patients' length of stay, as induced by a change of policy, could have appreciable effects on the population structure of nosocomial pathogens.

We adopted a neutral model to better disentangle the relative role of the different network properties. A wide disease-ecology literature addressed the consequences of neutral hypotheses on multi-strain balance in order to provide a benchmark for interpreting the observed co-existence patterns and gauging the effect of selective forces potentially at play [11, 18, 66, 67]. Many of these works addressed, for instance, the co-existence between susceptible and resistant strains of *S. pneumoniae* [11, 66]. However, this assumption was rarely adopted in network models, that consider for the majority strains with different epidemiological traits with the aim of describing pathogen selection and evolution [47–49, 68]. Strains were assumed to have the same infection parameters in [50, 51], where the role of community structure and clustering was analyzed in conjunction with cross-immunity. With respect to these works, the minimal transmission model used here enabled a transparent comprehension of the role of the network. Multiple identical SIS processes can be mapped, in fact, on a single SIS process, in such a way that the wide literature of single SIS processes allows for a better understanding of the behavior recovered in the simulations [32, 33]. Strains can be also grouped in two macro-strains. This strategy allowed us to adopt the viewpoint of an emerging strain and study its competition with the others seen as a unique macro-strain. The associated master equation and Fokker-Planck approximation allowed computing the average extinction time, capturing the key aspects of the dynamics. In a future work this theoretical framework could be extended to consider other network topologies. It could, for instance, be coupled with the activity-block approximation to describe heterogeneous networks. Additional numerical analyses, based on a similar transmission model, could also address other properties known to alter spreading dynamics, such as heterogeneous inter-contact time distribution or topological and temporal correlations.

As a case study, we analyzed the spread of *S. aureus* in a hospital taking advantage of the simultaneous availability of contact and carriage information [16]. The temporal and topological features of the network lead to a lower prevalence and richness with respect to the homogeneous mixing (although the effect was quite small). In addition, similar prevalence and richness values are associated to different dominance levels in different networks—i.e. different values of the Berger-Parker index—with the real network leading to a higher dominance as observed in reality. This behavior can be explained by the theoretical results and can be attributed essentially to the effect of contact heterogeneities, considering that the community structure does not have appreciable effects for this network, as discussed above. The importance of

accounting for host contacts and hospital organization in the assessment of bacterial spread and designing interventions has been recognized by several studies [16, 28–31, 63]. Here we show that this element may be critical also for understanding the population ecology of the bacterium. It is important to note however that, while the realistic network provides results that are closer to the data, this ingredient explains only part of the heterogeneity observed in the abundance. This shows that the contact network is a relevant factor, but other factors should be considered as well. The approach used here is intentionally simplified, as we focused on the main dynamical consequences of the contact network. Clearly, more detailed models can be designed to reproduce more closely the data. A certain degree of variation in the epidemiological traits could be at play, as for example the fitness cost of resistance [8]. Role of hosts in the network (e.g. patients vs. health-care workers), and heterogeneities in health conditions, antibiotic treatment and hygiene practices are also known to affect duration of carriage and chance of transmission [16, 28, 31, 63]. Eventually, we must consider that the comparison of model output with carriage data is also affected by the limitation of the dataset itself, already described in [16]. In particular, the weekly swabs may leave transient colonization undetected. Moreover, while the relevance of CPIs as proxies for epidemiological links has been demonstrated [16], the transmission through the environment (e.g. in the form of fomites) is also possible.

The understanding provided here can be relevant for other population settings, temporal scales and geographical levels. In addition, the modeling framework could be applied to pathogens other than *S. aureus*, such as *human papillomavirus*, *S. pneumoniae* and *Neisseria meningitidis*, for which the strong interest in the study of the strain ecology is justified by the public health need for understanding and anticipating trends in antibiotic resistance, or the long-term effect of vaccination [1, 2, 4, 5]. With this respect, if the simple framework introduced here increases our theoretical comprehension of the multi-strain dynamics, more tailored models may become necessary according to the specific case. In particular, we have considered complete mutual exclusion as the only mechanism for competition. In reality, a secondary inoculation in a host that is already a carrier may give rise to alternative outcomes, such as co-infection or replacement [69]. In addition, infection or carriage may confer a certain level of long-lasting strain-specific protection and/or a short-duration transcendent immunity [11, 50]. Eventually mechanisms of mutation and/or recombination are at play and their inclusion into the model can be important according to the time scale of interest.

Materials and methods

Network models

We provide here details of the generative algorithms used for the contact network models. Network dynamics is implemented in discrete time according to the following rules common to all models:

Turnover dynamics: new nodes arrive according to a Poisson process with rate λ_{in} and leave after a random time drawn from an exponential probability distribution with average τ. After a short initial transient, population size is Poisson distributed with average $\bar{V} = \lambda_{in}\tau$. Upon admission, a node i is assigned with an activity potential a_i, i.e. an activation rate, drawn at random from a given probability distribution $P(a)$. Any node retains this property throughout its whole lifespan.

Activation Pattern: each node i becomes active with rate a_i. It then receives a number of stubs drawn from a zero-truncated Poisson distribution with parameter κ—we require active nodes to engage in at least one contact. The average number of stubs, computed among active

nodes, is thus given by $\kappa/(1 - e^{-\kappa})$, and the average degree can be computed by the latter quantity multiplied by the average activity potential. The active status lasts for a single time step.

Stub-matching: stubs are then matched according to the actual model considered.

We now describe in detail each network model:

HOM: in this model each node has the same probability a_H to be active during each time step; the activity distribution is thus $P(a) = \delta(a - a_H)$, where $\delta(x)$ is the Dirac's delta function. Stubs are matched completely at random in order to form links, according to a configuration model [33]. We discard eventual self-links and multiple links that may occur during the matching procedure.

HET: here each node i has its own activity rate a_i, drawn from a power-law distribution $P(a) \propto a^{-\gamma}$, with $a \in (\epsilon, 1]$. We tune the variance by varying γ—lower γ higher variance. We then set ϵ to have the average activity $\bar{a}$ equal to a_H in HOM. Stub-matching procedure is the same as in HOM. HET model is thus a variant of the activity driven model introduced in [34] with the difference that here contacts are created only among active individuals.

COM: incoming nodes are assigned to one among n_C communities with equal probability—so that communities have the same size on average—and belong to the same community throughout their whole lifespan. Stubs are matched according to the community each node belongs to. Precisely, any stub is matched either with another stub of the same community, with probability p_{IN}, or with a stub of a different community, with complementary probability. Here the stub-matching procedure results in a larger number of lost links—to eliminate multiple links and self-loops—compared to HOM and HET, due to the difficulty in matching stubs within small groups. Thus, the parameter κ has to be adjusted manually to recover the same average degree as in HOM and HET. Each node has the same activity potential a_H as in HOM.

Hospital network and null model

We use a dynamical contact network obtained from CPI data collected during the i-Bird study in a French hospital. Details of the network are already reported in [16]. Briefly, the dataset describes contacts occurring between 592 individuals from July to November 2009. The study involved both patients and health-care workers, distributed in 5 wards, as well as hospital service staff. Every participant wore a wireless device designed to broadcast a signal every 30 s containing information about its ID. Signal strength was tuned so that only devices within a small distance (around 1.5 m) were able to register a contact. CPIs were finally aggregated daily, keeping the information about their cumulative duration within each day.

We discard CPIs relative to the first 2 weeks and the last 4 weeks of dataset, corresponding to a period of adjustments in the measurements and progressive dismissal of the experiment, respectively. Simulations conducted with the CPIs network were compared with results obtained with a null model which we refer to as RAND. According to this randomization scheme the activity of a node is randomized while respecting the constraint that removal and addition of contacts must not alter the time of the first and the last contact of each node (t_S and t_L respectively). Notice that RAND preserves the number of nodes that are present at any time in the network by preserving their first contact t_S and their length of stay $t_L - t_S$. Null models randomizing the latter properties lead to misleading results when node length of stay is heterogeneous and node turnover occurs [70]. RAND also sets all contact weights equal to the average weight value.

Spreading simulations

Spreading dynamics is stochastic and is performed in discrete time. At each time step of duration Δt, we update the state of each node: each infected node transmits the strain it is carrying

to a susceptible neighbor with probability $\beta\Delta t$ and it turns susceptible with probability $\mu\Delta t$. Notice that due to mutual exclusion, an individual can be infected by a single strain at a time [71]. Strain injection is given by the combination of two processes: incoming individuals bring a new strain with probability p_s, and susceptible individuals turn infectious with a new strain with probability $q_s\Delta t$. The two mechanisms mimic respectively incoming infectious individuals (e.g. admission of colonized patients) and transmission from an external source (in the hospital example this corresponds to contacts with individuals that were not participating in the study). The expected injection rate, which accounts for both introduction mechanisms, is thus given by $\iota = \lambda_{\text{in}} p_s + \bar{S} q_s$, where $\bar{S}$ is the average number of susceptible individuals at the equilibrium. In the theoretical analysis in the main paper we assumed $q_s = 0$ for simplicity, thus variations in ι were induced by variations in λ_{in} and p_s. The case $q_s > 0$ was considered in the supporting information.

Simulations on synthetic networks differ from those on the hospital network in the combination of the spreading and network dynamics. In the synthetic network case, at each time step of duration Δt = 1h, both network and spreading dynamics are simulated one after the other. On average, $\lambda_{in}\Delta t$ new nodes enter in the population per time step, while existing nodes can leave with probability $\Delta t/\tau$. Nodes then form contacts according to the specific generative network algorithm. Eventually, transmission and recovery are simulated as explained above. In order to reconstruct the equilibrium dynamics we run simulations for a sufficiently long time span, discarding a transient time of $4 \cdot 10^4$ time steps. We verified that the dynamical properties at the equilibrium are unaffected by initial conditions.

For the hospital example, the network is an external parameter fed into the simulations. Contacts were aggregated daily keeping the information of their total duration. We used this information by considering a weighted network with the link weight, w_{ij}, representing the number of contacts of duration 30 s registered during the day between i and j. We then assumed Δt = 1 day and computed the probability of infection depending on the weight as $1 - (1 - \beta\delta)^{w_{ij}}$, with δ = 30 s. We initialized the system with the same configuration observed in the data, i.e. the initial status for each node is set according to *S. aureus* carriage during the starting week. Simulation length is bound to the hospital contact network duration.

In order to facilitate the comparison between the synthetic and the real scenarios, parameters of the network models were set based on the properties of the hospital network. The average size, the average activity potential and the average degree were set equal to the values estimated from the hospital network, i.e. $\bar{V} = 306$, $\bar{a} = 0.28$, $\bar{k} = 0.89$ respectively. For the COM model the number of communities (n_C = 6) and one of the two explored values of p_{IN} (p_{IN} = 0.78) were also informed by the data. Additional values of $\bar{V}$ and p_{IN} were also tested. Epidemiological parameters were informed by the data in some cases—p_s = 0.079 as computed from carriage data -, or chosen among plausible values for the *S. aureus* colonization—i.e. μ^{-1}, that was set equal to either 21 or 35 days with other values from 14 to 49 days explored in the supporting information. Values of β were explored systematically. For consistency, values of rates throughout the manuscript were always expressed per hour.

Analysis of carriage data

Carriage data was obtained from weekly swabs in multiple body areas, including the nares. Swabs that resulted positive to *S. aureus* were further examined. Spa-type and antibiotic resistance profiles (MSSA or MRSA) were then determined. In this work we regard two strains as different if they differ in spa-type and/or antibiotic resistance profile. We considered carriage data obtained from nasal swabs dismissing other body areas since the anterior nares represent the most important niche for *S. aureus* [72].

Ecological measures and other indicators

We described strain population diversity through standard ecological indicators. The abundance of a strain i, N_i, is the strain-associated prevalence. From this quantity we computed the relative abundance, $f_i = \frac{N_i}{\sum_i N_i}$, and the relative abundance distribution, being the frequency of strains with relative abundance f. The Berger-Parker index is the relative abundance of the dominant strain, i.e. $\max_i f_i$.

To analyze repartition of strains across communities we use the Inverse Participation Ratio (*IPR*) [55]. The general definition of this quantity is the following. Given a vector $\vec{v}$ with l components $\{v_i\}_{i=1,\ldots,l}$, all within the range [0, 1], the *IPR* is given by:

$$IPR = \sum_{i=1}^{l} v_i^4. \tag{2}$$

If all the components are of the order (l^{-1}) then the *IPR* is small. Instead if one component $v_i \sim 1$ then $IPR \sim 1$ too, reflecting localization of $\vec{v}$. The IPR for total prevalence is computed by setting v_i equal to the fraction of infected individuals belonging to community $i = 1, \ldots, l = n_C$, while the *IPR* for a single strain is computed by setting v_i equal to the fraction of individuals infected by that particular strain and belonging to community i. We can extend the *IPR* computation to HOM case by assigning nodes to different groups as in COM but without affecting the stub-matching scheme.

Analytical results for the homogeneous network

In order to estimate the value of the length of stay maximizing the average richness for a given value of β when the contact structure is given by the HOM network we consider a homogeneous mixing version of our system.

Due to Eq (1) the calculation of the average richness reduces to the calculation of the average persistence time. In order to estimate such quantity we focus on a particular strain, labelled as "strain A", which is injected at $t = 0$ and we group all other strains under the label "strain B". We are allowed to do so because all strains have identical parameters. We therefore reduce our initial, multi-strain problem, to a two-strain problem. Since all new strains that will be injected after $t = 0$ will be labeled as strain B, it is clear that A is doomed to extinction since there exists an infinite reservoir of B. The average time to extinction is therefore the average time to extinction of strain A.

Since HOM network realizes quite well homogeneous mixing conditions we regard our system as homogeneously mixed. Within this framework it is sufficient to specify the numbers of hosts infected by strain A (n_A), hosts infected by strain B (n_B) and susceptible hosts (n_s). The master equation for the joint probability distribution $P(n_A, n_B, n_s)$ is given by [73]:

$$\begin{aligned}
\dot{P}(n_A, n_B, n_s) &= \beta' \bar{V}^{-1}(n_A - 1)(n_s + 1)P(n_A - 1, n_B, n_s + 1) \\
&+ \beta' \bar{V}^{-1}(n_B - 1)(n_s + 1)P(n_A, n_B - 1, n_s + 1) \\
&+ \mu(n_A + 1)P(n_A + 1, n_B, n_s - 1) + \mu(n_B + 1)P(n_A, n_B + 1, n_s - 1) \\
&+ \lambda_{out}(n_A + 1)P(n_A + 1, n_B, n_s) \\
&+ \lambda_{out}(n_B + 1)P(n_A, n_B + 1, n_s) + \lambda_{out}(n_s + 1)P(n_A, n_B, n_s + 1) \\
&+ \lambda_{out}\bar{V}p_s P(n_A, n_B - 1, n_s) + \lambda_{out}\bar{V}(1 - p_s)P(n_A, n_B, n_s - 1) \\
&- [(n_A + n_B)(\beta' \bar{V}^{-1} n_s + \mu) + \lambda_{out}(n_A + n_B + n_s) + \lambda_{out}\bar{V}]P(n_A, n_B, n_s),
\end{aligned} \tag{3}$$

Where $\beta' = \beta\bar{k}$. Terms appearing on the right-hand side of the equation represent the

probability flow associated to each transition event. The first four terms describe, in order, the infection due to strain A, the infection due to strain B, the recovery from A and the recovery from B. The remaining terms are then associated to the discharge of either one of the three types of individuals—infected with A, infected with B and susceptibles—and to the admission of infected of type B and susceptibles respectively. In order to obtain some approximate solution to this equation we assume that the average number of individuals $n_A + n_B + n_s$ and the total prevalence $n_A + n_B$ do not fluctuate in time and are therefore equal to $\bar{V}$ and $i(\infty)\bar{V}$ respectively, where $i(\infty)$ is given by:

$$i(\infty) = \frac{\beta' - \mu - \lambda_{out} + \sqrt{(\beta' - \mu - \lambda_{out})^2 + 4\beta'\lambda_{out}p_s}}{2\beta'}. \quad (4)$$

After performing the Van-Kampen size expansion we are left with a Fokker-Planck equation for the density of A $f\left(x = \frac{n_A}{\bar{V}}\right) = P(n_A)$:

$$\partial_t f = -\partial_x(D_1(x)f) + \frac{1}{2\bar{V}}\partial_x^2(D_2(x)f), \quad (5)$$

where $D_1 = \beta'\ (1 - i(\infty))\ x - \mu - \lambda_{out}$ and $D_2 = \beta'\ (1 - i(\infty))\ x + \mu + \lambda_{out}$ are the so-called drift and diffusion coefficients respectively.

According to the theory of stochastic processes [73] the average extinction time $T_{\text{pers}}(x_0)$ (where x_0 represents the initial density of strain A) satisfies:

$$D_1(x_0)\frac{d}{dx_0}\ T_{\text{pers}} + \frac{1}{2\bar{V}}D_2(x_0)\frac{d^2}{dx_0^2}T_{\text{pers}} = -1, \quad (6)$$

with boundary conditions $T_{\text{pers}}(0) = 0$ and $\frac{d}{dx_0}T_{\text{pers}}(i(\infty)) = 0$. The solution is finally given by:

$$T_{\text{pers}}(x_0) = \frac{i(\infty)}{\lambda_{out}p_s}[Ei(-\alpha i(\infty))(e^{\alpha x_0} - 1) - e^{\alpha x_0}Ei(-\alpha x_0) + ln(\alpha x_0) + \gamma_E], \quad (7)$$

where $Ei(x)$ is the exponential integral function and γ_E is Euler-Mascheroni constant. When a new strain is introduced its prevalence is just 1, therefore we estimate the average extinction time using $T_{\text{pers}}(x_0 = \bar{V}^{-1})$.

Supporting information

S1 Text. Multi-strain model with heterogeneous recovery classes. This file contains additional information about simulations with individuals grouped into classes with different recovery rates.
(PDF)

S1 Fig. Richness and Berger-Parker index as a function of transmissibility for HOM, HET, COM and COM+HET models. Each model is displayed on a different column. HET is characterized by activity distribution exponent $\gamma = 0.7$. COM+HET model is simulated using the same activation pattern as in HET with $\gamma = 0.7$ and the same stub-matching procedure as in COM. We consider the case $p_{IN} = 0.99$. The first two rows correspond to $p_s = 0.01$, whereas the last two to $p_s = 0.079$. For each scenario we show the median (solid line), as well as 50% and 95% CI (shaded areas).
(PNG)

S2 Fig. Summary indicators of the persistence time distribution as a function of transmissibility for both HOM and HET models. HOM and HET are displayed in blue and green

respectively. (A), (B) and (C) display distribution's average, coefficient of variation and skewness, respectively. Other parameters are as in Fig 2 in the main text.
(PNG)

S3 Fig. Shannon evenness for HOM model and two instances of HET model. HOM is depicted in blue, whereas instances of HET model with activity distribution exponent $\gamma = 2.5$ and $\gamma = 0.7$ are depicted in orange and green respectively. Shaded blue area represents standard deviation for HOM. We introduce the relative abundance of the i-th strain: $n_i = N_i \Big/ \sum_i N_i$, with N_i the abundance of the strain i (i.e. the number of infected with strain i). Shannon evenness is defined as the normalized Shannon entropy $S(\{n_i\}) = -\mathcal{N}^{-1} \sum_i n_i \ln n_i$, with $\mathcal{N} = \ln N_S$. Parameters are the same as in Fig 1 in the main text.
(PNG)

S4 Fig. Impact of network size. Richness (A,B), average persistence time (C,D), prevalence (E,F) and Berger-Parker index (G,H) as a function of transmissibility for both HOM and HET models (first and second columns respectively). For each value of $\bar{V}$ we compute p_s to have the strain injection rate, $\bar{V} p_s$, the same across the different networks. Other parameters are as in Fig 2 of the main paper. Increasing network size results in a larger number of co-circulating strains, while the re-scaled prevalence and the Berger-Parker index are almost independent of $\bar{V}$. Notice that increasing network size does not lead to any qualitative change in the relation between HOM and HET.
(PNG)

S5 Fig. Relative abundance distribution in varying network size for HOM and HET models. HOM and HET are depicted in blue and green respectively. For each value of $\bar{V}$ we compute p_s to have the strain injection rate, $\bar{V} p_s$, the same across the different networks. Other parameters are as in Fig 2 of the main paper.
(PNG)

S6 Fig. Richness for the different network models with transmission from an external source. The frequency of transmissions from an external source is tuned by q_s, which we set here to 0.0002. (A) Richness for HOM model (blue markers) and HET model with activity distribution exponent $\gamma = 0.7$ (green markers). Here $p_s = 0.079$. (B) Richness index for HOM model (blue markers) and COM model with within-community connection probability $p_{IN} = 0.99$ (green markers). Here $p_s = 0.01$. (C) Richness as a function of β and τ for HOM model. Here $p_s = 0.079$.
(PNG)

S7 Fig. Multi-strain dynamics when recovery rate is heterogeneous across individuals. Here, each node belongs to one out of three classes according to its recovery rate—see description in the dedicated section of this supporting information. We compare HOM (blue), HET (green), COM (red) models with and without heterogeneity in the recovery rate (triangles and circles respectively). Panels show prevalence (A), richness (B) and Berger-Parker index (C). Other parameters are like in Figs 1, 2 and 3 in the main paper ($\gamma = 0.7$ for HET and $p_{IN} = 0.99$ for COM).
(PNG)

S8 Fig. Average persistence time for HOM in varying transmissibility and length of stay. The quantity is computed from the simulations. The dashed gray line represents a linear trend as a guide to the eye. Parameters are the same as in Fig 4 in the main text.
(PNG)

S9 Fig. Comparison between simulations for HOM model and analytical predictions obtained using the Fokker-Planck framework. Solid lines represent average richness obtained by using Eqs (1) and (7) from the main text while dots represent simulations results. Here $\beta = 0.04$ while other parameters are the same as in Fig 4 in the main text.
(PNG)

S10 Fig. Prevalence vs richness for several values of the infectious period and using the CPI network. The value of q_s is the same for the curve highlighted in Fig 5B in the main text, $q_s = 0.00018$. Here dot size is proportional to the magnitude of β.
(PNG)

Acknowledgments

Authors would like to thank Vittoria Colizza, Lulla Opatowski and Laura Temime for useful discussion.

Author Contributions

Conceptualization: Francesco Pinotti, Pierre-Yves Böelle, Chiara Poletto.

Data curation: Francesco Pinotti, Éric Fleury, Didier Guillemot, Pierre-Yves Böelle.

Investigation: Éric Fleury, Didier Guillemot, Pierre-Yves Böelle.

Methodology: Francesco Pinotti, Chiara Poletto.

Supervision: Pierre-Yves Böelle, Chiara Poletto.

Writing – original draft: Francesco Pinotti, Pierre-Yves Böelle, Chiara Poletto.

Writing – review & editing: Francesco Pinotti, Éric Fleury, Didier Guillemot, Pierre-Yves Böelle, Chiara Poletto.

References

1. Schiffman M, Castle PE. Human Papillomavirus: Epidemiology and Public Health. Archives of Pathology & Laboratory Medicine. 2003; 127(8):930–934.
2. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. The Lancet. 2011; 378(9807):1962–1973. https://doi.org/10.1016/S0140-6736(10)62225-8
3. Bogaert D, de Groot R, Hermans P. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. The Lancet Infectious Diseases. 2004; 4(3):144–154. https://doi.org/10.1016/S1473-3099(04)00938-7 PMID: 14998500
4. Atkins KE, Lafferty EI, Deeny SR, Davies NG, Robotham JV, Jit M. Use of mathematical modelling to assess the impact of vaccines on antibiotic resistance. The Lancet Infectious Diseases. 2017. https://doi.org/10.1016/S1473-3099(17)30478-4 PMID: 29146178
5. Chambers HF, Deleo FR. Waves of resistance: Staphylococcus aureus in the antibiotic era. Nat Rev Microbiol. 2009; 7(9):629–641. https://doi.org/10.1038/nrmicro2200 PMID: 19680247
6. Reich NG, Shrestha S, King AA, Rohani P, Lessler J, Kalayanarooj S, et al. Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. J R Soc Interface. 2013; 10(86). https://doi.org/10.1098/rsif.2013.0414 PMID: 23825116
7. Cohen T, Colijn C, Murray M. Modeling the effects of strain diversity and mechanisms of strain competition on the potential performance of new tuberculosis vaccines. Proc Natl Acad Sci U S A. 2008; 105(42):16302–16307. https://doi.org/10.1073/pnas.0808746105 PMID: 18849476
8. Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. Evol Appl. 2015; 8(3):273–283. https://doi.org/10.1111/eva.12196 PMID: 25861385
9. Opatowski L, Varon E, Dupont C, Temime L, van der Werf S, Gutmann L, et al. Assessing pneumococcal meningitis association with viral respiratory infections and antibiotics: insights from statistical and

mathematical models. Proceedings Biological Sciences. 2013; 280(1764):20130519. https://doi.org/10.1098/rspb.2013.0519 PMID: 23782877

10. Boucher HW, Corey GR. Epidemiology of Methicillin-Resistant Staphylococcus aureus. Clin Infect Dis. 2008; 46(Supplement_5):S344–S349. https://doi.org/10.1086/533590 PMID: 18462089
11. Cobey S, Lipsitch M. Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. Science. 2012; 335(6074):1376–1380. https://doi.org/10.1126/science.1215947 PMID: 22383809
12. Murall CL, Bauch CT, Day T. Could the human papillomavirus vaccines drive virulence evolution? Proc R Soc B. 2015; 282(1798):20141069. https://doi.org/10.1098/rspb.2014.1069 PMID: 25429011
13. Aanensen DM, Feil EJ, Holden MTG, Dordel J, Yeats CA, Fedosejev A, et al. Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive Staphylococcus aureus in Europe. mBio. 2016; 7(3):e00444–16. https://doi.org/10.1128/mBio.00444-16 PMID: 27150362
14. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. PLoS Pathog. 2014; 10(2):e1003932. https://doi.org/10.1371/journal.ppat.1003932 PMID: 24586153
15. Donker T, Reuter S, Scriberras J, Reynolds R, Brown NM, Török ME, et al. Population genetic structuring of methicillin-resistant Staphylococcus aureus clone EMRSA-15 within UK reflects patient referral patterns. Microbial Genomics. 2017; 3(7). https://doi.org/10.1099/mgen.0.000113 PMID: 29026654
16. Obadia T, Silhol R, Opatowski L, Temime L, Legrand J, Thiébaut ACM, et al. Detailed Contact Data and the Dissemination of Staphylococcus aureus in Hospitals. PLOS Computational Biology. 2015; 11(3): e1004170. https://doi.org/10.1371/journal.pcbi.1004170 PMID: 25789632
17. Nekkab N, Astagneau P, Temime L, Crépey P. Spread of hospital-acquired infections: A comparison of healthcare networks. PLOS Computational Biology. 2017; 13(8):e1005666. https://doi.org/10.1371/journal.pcbi.1005666 PMID: 28837555
18. Fraser C, Hanage WP, Spratt BG. Neutral microepidemic evolution of bacterial pathogens. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(6):1968–1973. https://doi.org/10.1073/pnas.0406993102 PMID: 15684071
19. Robinson K, Fyson N, Cohen T, Fraser C, Colijn C. How the Dynamics and Structure of Sexual Contact Networks Shape Pathogen Phylogenies. PLOS Computational Biology. 2013; 9(6):e1003105. https://doi.org/10.1371/journal.pcbi.1003105 PMID: 23818840
20. Matt J Keeling PR. Modeling Infectious Diseases in Humans and Animals. Princeton University Press; 2008.
21. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the impact of human mobility on malaria. Science (New York, NY). 2012; 338(6104):267–270. https://doi.org/10.1126/science.1223467
22. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(51):21484–21489. https://doi.org/10.1073/pnas.0906910106 PMID: 20018697
23. Louail T, Lenormand M, Picornell M, Cantú OG, Herranz R, Frias-Martinez E, et al. Uncovering the spatial structure of mobility networks. Nature Communications. 2015; 6:6007. https://doi.org/10.1038/ncomms7007 PMID: 25607690
24. Rocha LEC, Liljeros F, Holme P. Information dynamics shape the sexual networks of Internet-mediated prostitution. Proceedings of the National Academy of Sciences. 2010; 107(13):5706–5711. https://doi.org/10.1073/pnas.0914080107
25. Mastrandrea R, Fournet J, Barrat A. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. PLOS ONE. 2015; 10(9): e0136497. https://doi.org/10.1371/journal.pone.0136497 PMID: 26325289
26. Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A high-resolution human contact network for infectious disease transmission. Proceedings of the National Academy of Sciences. 2010; 107(51):22020–22025. https://doi.org/10.1073/pnas.1009094108
27. Génois M, Vestergaard CL, Fournet J, Panisson A, Bonmarin I, Barrat A. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. Network Science. 2015; 3(3):326–347. https://doi.org/10.1017/nws.2015.10
28. Duval A, Obadia T, Martinet L, Boëlle PY, Fleury E, Guillemot D, et al. Measuring dynamic social contacts in a rehabilitation hospital: effect of wards, patient and staff characteristics. Scientific Reports. 2018; 8(1):1686. https://doi.org/10.1038/s41598-018-20008-w PMID: 29374222

29. Vanhems P, Barrat A, Cattuto C, Pinton JF, Khanafer N, Régis C, et al. Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. PLOS ONE. 2013; 8(9): e73970. https://doi.org/10.1371/journal.pone.0073970 PMID: 24040129

30. Voirin N, Payet C, Barrat A, Cattuto C, Khanafer N, Régis C, et al. Combining High-Resolution Contact Data with Virological Data to Investigate Influenza Transmission in a Tertiary Care Hospital. Infection Control & Hospital Epidemiology. 2015; 36(3):254–260. https://doi.org/10.1017/ice.2014.53

31. Assab R, Nekkab N, Crépey P, Astagneau P, Guillemot D, Opatowski L, et al. Mathematical models of infection transmission in healthcare settings: recent advances from the use of network structured data. Current Opinion in Infectious Diseases. 2017; 30(4):410–418. https://doi.org/10.1097/QCO.0000000000000390 PMID: 28570284

32. Holme P, Saramäki J, editors. Temporal Networks. Springer; 2013.

33. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. Reviews of Modern Physics. 2015; 87(3):925–979. https://doi.org/10.1103/RevModPhys.87.925

34. Perra N, Gonçalves B, Pastor-Satorras R, Vespignani A. Activity driven modeling of time varying networks. Scientific Reports. 2012; 2:469. https://doi.org/10.1038/srep00469 PMID: 22741058

35. Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. Journal of Theoretical Biology. 2005; 235(2):275–288. https://doi.org/10.1016/j.jtbi.2005.01.011 PMID: 15862595

36. Salathé M, Jones JH. Dynamics and Control of Diseases in Networks with Community Structure. PLOS Computational Biology. 2010; 6(4):e1000736. https://doi.org/10.1371/journal.pcbi.1000736 PMID: 20386735

37. Sah P, Leu ST, Cross PC, Hudson PJ, Bansal S. Unraveling the disease consequences and mechanisms of modular structure in animal social networks. Proceedings of the National Academy of Sciences. 2017; 114(16):4165–4170. https://doi.org/10.1073/pnas.1613616114

38. Holme P, Liljeros F. Birth and death of links control disease spreading in empirical contact networks. Scientific Reports. 2014; 4:4999. https://doi.org/10.1038/srep04999 PMID: 24851942

39. Darabi Sahneh F, Scoglio C. Competitive epidemic spreading over arbitrary multilayer networks. Physical Review E. 2014; 89(6):062817. https://doi.org/10.1103/PhysRevE.89.062817

40. Poletto C, Meloni S, Van Metre A, Colizza V, Moreno Y, Vespignani A. Characterising two-pathogen competition in spatially structured environments. Scientific Reports. 2015; 5:7895. https://doi.org/10.1038/srep07895 PMID: 25600088

41. Poletto C, Meloni S, Colizza V, Moreno Y, Vespignani A. Host Mobility Drives Pathogen Competition in Spatially Structured Populations. PLOS Computational Biology. 2013; 9(8):e1003169. https://doi.org/10.1371/journal.pcbi.1003169 PMID: 23966843

42. Sanz J, Xia CY, Meloni S, Moreno Y. Dynamics of Interacting Diseases. Phys Rev X. 2014; 4(4):041005.

43. Karrer B, Newman MEJ. Competing epidemics on complex networks. Phys Rev E. 2011; 84(3):036106. https://doi.org/10.1103/PhysRevE.84.036106

44. Cai W, Chen L, Ghanbarnejad F, Grassberger P. Avalanche outbreaks emerging in cooperative contagions. Nature Physics. 2015; 11(11):936–940. https://doi.org/10.1038/nphys3457

45. Hébert-Dufresne L, Althouse BM. Complex dynamics of synergistic coinfections on realistically clustered networks. Proceedings of the National Academy of Sciences. 2015; 112(33):10551–10556. https://doi.org/10.1073/pnas.1507820112

46. Leventhal GE, Hill AL, Nowak MA, Bonhoeffer S. Evolution and emergence of infectious diseases in theoretical and real-world networks. Nature Communications. 2015; 6:6101. https://doi.org/10.1038/ncomms7101 PMID: 25592476

47. Eames KTD, Keeling MJ. Coexistence and Specialization of Pathogen Strains on Contact Networks. The American Naturalist. 2006; 168(2):230–241. https://doi.org/10.1086/505760 PMID: 16874632

48. Ballegooijen WMv, Boerlijst MC. Emergent trade-offs and selection for outbreak frequency in spatial epidemics. Proceedings of the National Academy of Sciences. 2004; 101(52):18246–18250. https://doi.org/10.1073/pnas.0405682101

49. Lion S, Gandon S. Spatial evolutionary epidemiology of spreading epidemics. Proceedings Biological Sciences. 2016; 283 (1841). https://doi.org/10.1098/rspb.2016.1170 PMID: 27798295

50. Buckee CO, Koelle K, Mustard MJ, Gupta S. The effects of host contact network structure on pathogen diversity and strain structure. PNAS. 2004; 101(29):10839–10844. https://doi.org/10.1073/pnas.0402000101 PMID: 15247422

51. Buckee C, Danon L, Gupta S. Host community structure and the maintenance of pathogen diversity. Proceedings of the Royal Society of London B: Biological Sciences. 2007; 274(1619):1715–1721. https://doi.org/10.1098/rspb.2007.0415

52. Morris EK, Caruso T, Buscot F, Fischer M, Hancock C, Maier TS, et al. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. Ecol Evol. 2014; 4(18):3514–3524. https://doi.org/10.1002/ece3.1155 PMID: 25478144

53. Azaele S, Suweis S, Grilli J, Volkov I, Banavar JR, Maritan A. Statistical mechanics of ecological systems: Neutral theory and beyond. Rev Mod Phys. 2016; 88(3):035003. https://doi.org/10.1103/RevModPhys.88.035003

54. DeBenedictis PA. On the Correlations between Certain Diversity Indices. The American Naturalist. 1973; 107(954):295–302. https://doi.org/10.1086/282831

55. Pastor-Satorras R, Castellano C. Distinct types of eigenvector localization in networks. Scientific Reports. 2016; 6:18847. https://doi.org/10.1038/srep18847 PMID: 26754565

56. Suweis S, Bertuzzo E, Mari L, Rodriguez-Iturbe I, Maritan A, Rinaldo A. On species persistence-time distributions. Journal of Theoretical Biology. 2012; 303:15–24. https://doi.org/10.1016/j.jtbi.2012.02.022 PMID: 22763130

57. van Kleef E, Robotham JV, Jit M, Deeny SR, Edmunds WJ. Modelling the transmission of healthcare associated infections: a systematic review. BMC Infect Dis. 2013; 13:294. https://doi.org/10.1186/1471-2334-13-294 PMID: 23809195

58. Martinet L, Crespelle C, Fleury E, Boëlle PY, Guillemot D. The Link Stream of Contacts in a Whole Hospital. arXiv:180505752 [cs]. 2018;.

59. Kouyos R, Klein E, Grenfell B. Hospital-Community Interactions Foster Coexistence between Methicillin-Resistant Strains of Staphylococcus aureus. PLOS Pathogens. 2013; 9(2):e1003134. https://doi.org/10.1371/journal.ppat.1003134 PMID: 23468619

60. Bonten MJ, Austin DJ, Lipsitch M. Understanding the spread of antibiotic resistant pathogens in hospitals: mathematical models as tools for control. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America. 2001; 33(10):1739–1746. https://doi.org/10.1086/323761

61. Margolis E, Yates A, Levin BR. 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. 2010; 10:59. https://doi.org/10.1186/1471-2180-10-59 PMID: 20178591

62. Dall'Antonia M, Coen PG, Wilks M, Whiley A, Millar M. Competition between methicillin-sensitive and -resistant Staphylococcus aureus in the anterior nares. J Hosp Infect. 2005; 61(1):62–67. https://doi.org/10.1016/j.jhin.2005.01.008 PMID: 15893854

63. Chang Q, Lipsitch M, Hanage WP. Impact of Host Heterogeneity on the Efficacy of Interventions to Reduce Staphylococcus aureus Carriage. Infect Control Hosp Epidemiol. 2016; 37(2):197–204. https://doi.org/10.1017/ice.2015.269 PMID: 26598029

64. Cobey S, Baskerville EB, Colijn C, Hanage W, Fraser C, Lipsitch M. Host population structure and treatment frequency maintain balancing selection on drug resistance. Journal of The Royal Society Interface. 2017; 14(133):20170295. https://doi.org/10.1098/rsif.2017.0295

65. van Duijkeren E, Ikawaty R, Broekhuizen-Stins MJ, Jansen MD, Spalburg EC, de Neeling AJ, et al. Transmission of methicillin-resistant Staphylococcus aureus strains between different kinds of pig farms. Veterinary Microbiology. 2008; 126(4):383–389. https://doi.org/10.1016/j.vetmic.2007.07.021 PMID: 17765409

66. Gjini E, Valente C, Sá-Leão R, Gomes MGM. How direct competition shapes coexistence and vaccine effects in multi-strain pathogen systems. Journal of Theoretical Biology. 2016; 388:50–60. https://doi.org/10.1016/j.jtbi.2015.09.031 PMID: 26471070

67. Lipsitch M, Colijn C, Cohen T, Hanage WP, Fraser C. No coexistence for free: Neutral null models for multistrain pathogens. Epidemics. 2009; 1(1):2–13. https://doi.org/10.1016/j.epidem.2008.07.001 PMID: 21352747

68. Buendía V, Muñoz MA, Manrubia S. Limited role of spatial self-structuring in emergent trade-offs during pathogen evolution. arXiv:180408463 [q-bio]. 2018;.

69. Sofonea MT, Alizon S, Michalakis Y. Exposing the diversity of multiple infection patterns. Journal of Theoretical Biology. 2017; 419:278–289. https://doi.org/10.1016/j.jtbi.2017.02.011 PMID: 28193485

70. Li M, Rao VD, Gernat T, Dankowicz H. Lifetime-preserving reference models for characterizing spreading dynamics on temporal networks. Scientific Reports. 2018; 8(1):709. https://doi.org/10.1038/s41598-017-18450-3 PMID: 29335422

71. Stanoev A, Trpevski D, Kocarev L. Modeling the Spread of Multiple Concurrent Contagions on Networks. PLOS ONE. 2014; 9(6):e95669. https://doi.org/10.1371/journal.pone.0095669 PMID: 24922541

72. Lauderdale TLY, Wang JT, Lee WS, Huang JH, McDonald LC, Huang IW, et al. Carriage rates of methicillin-resistant *Staphylococcus aureus* (MRSA) depend on anatomic location, the number of sites

cultured, culture methods, and the distribution of clonotypes. Eur J Clin Microbiol Infect Dis. 2010; 29 (12):1553–1559. https://doi.org/10.1007/s10096-010-1042-8 PMID: 20820833

73. Gardiner C. Stochastic Methods: A Handbook for the Natural and Social Sciences. 4th ed. Springer Series in Synergetics. Berlin Heidelberg: Springer-Verlag; 2009.