Microbiome diversity protects against pathogens by nutrient blocking

Frances Spragge[1,2+] and Erik Bakkeren[1,2+], Martin T. Jahn[1,2], Elizete B. N. Araujo[3], Claire F. Pearson[3], Xuedan Wang[1,2], Louise Pankhurst[1,2], Olivier Cunrath[4]* and Kevin R. Foster[1,2]*

[+]These authors contributed equally to this work

*Corresponding authors: kevin.foster@biology.ox.ac.uk; olivier.cunrath@unistra.fr

1. Department of Biology, University of Oxford, Oxford, UK

2. Department of Biochemistry, University of Oxford, UK

3. Kennedy Institute of Rheumatology, University of Oxford, UK

4. CNRS, UMR7242, Biotechnology and cell signaling, University of Strasbourg, Illkirch, France

Abstract

The human gut microbiome plays an important role in resisting colonization of the host by pathogens, but we lack the ability to predict which communities will be protective. We studied how human gut bacteria influence colonization of two major bacterial pathogens, both *in vitro* and in gnotobiotic mice. While single species alone had negligible effects, colonization resistance greatly increased with community diversity. Moreover, this community-level resistance rested critically upon certain species being present. We explain these ecological patterns via the collective ability of resistant communities to consume nutrients that overlap with those used by the pathogen. Further, we apply our findings to successfully predict communities that resist a novel target strain. Our work provides a reason why microbiome diversity is beneficial and suggests a route for the rational design of pathogen-resistant communities.

One sentence summary

Diverse communities of gut bacteria collectively limit pathogen colonization by blocking nutrient access.

Introduction

The human gut is home to diverse bacterial species collectively known as the gut microbiota. A major health benefit provided by the gut microbiota is protection against pathogen colonization and subsequent infection; a phenomenon known as colonization resistance (*1*). The ability of the microbiota to protect against numerous enteric pathogens is well-documented, with evidence that particular species within the microbiota play a more important role than others (*2-9*). The ways that colonization resistance can arise include competition for nutrients and space, direct antagonism by toxins and other harmful compounds, and promoting host immunity against pathogens (*1, 10, 11*).

While the importance of the microbiota for colonization resistance is clear, however, we currently lack the principles needed to predict, *a priori*, which microbiota species will be effective against a given pathogen. A key challenge is the ecological complexity of the gut. The gut microbiome is a diverse ecological system with many individual species that all have the potential to play a role in colonization resistance. Moreover, these constituent species can also affect each other and interact ecologically in ways that are critical for colonization resistance (*12-16*). This combination of species diversity and the potential for ecological interactions makes colonization resistance a challenging phenotype to understand (*17*).

Here we approached the question of mechanisms of colonization resistance from the perspective of the underlying ecological principles. To do this, we studied colonization resistance provided by a range of human gut bacteria, both alone and in combinations. We performed all experiments in parallel using two species of pathogen, which are both on the WHO priority list: *Klebsiella pneumoniae* and *Salmonella enterica* Serovar Typhimurium (*18*). Both are members of the Enterobacteriaceae found in the human gut microbiome but they have very different lifestyles. *S.* Typhimurium causes acute infection and gastroenteritis (*19, 20*). By contrast, *K. pneumoniae* is a nosocomial, opportunistic pathogen that rarely causes disease in the gut itself, but gut colonization is a major risk factor for antimicrobial resistance associated infections elsewhere in the body (*21*).

Despite these differences, we have identified common principles that underlie colonization resistance to both species. Ecological diversity is important for colonization resistance *in vitro* and in gnotobiotic mice. Moreover, we found that colonization resistance is an ecologically complex trait, whereby the ability of one species to provide colonization resistance can depend entirely upon the presence of other species (*22*). Despite the complexity, we find that these ecological patterns are explained by a simple underlying principle, the collective ability of certain communities to consume nutrients and block pathogen growth. Further, we have shown that this principle offers a way to identify sets of bacterial species that will collectively limit the growth of a particular pathogen.

Results

Single species offer little protection in competition with pathogens

Individual members of the microbiota can promote colonization resistance in various contexts (*2-8*), which suggests that some species are more important for colonization resistance than others. To systematically assess this variability, we screened a diverse set of 100 human gut symbionts (**Table S1; also see Methods**) for their ability to limit pathogen growth. Competition in the gut occurs both at the point a pathogen enters the gut and when a pathogen becomes established (*23, 24*). We designed two co-culture assays to reflect these two aspects of competition in the mammalian gut (**Fig. 1a**). In the first assay (ecological invasion assay), we pre-grew the symbiont alone in standard anaerobic media (modified Gifu anaerobic media; mGAM) buffered to human colonic pH before adding the pathogen. In the second assay (competition assay), we inoculated this media with an equal ratio of symbiont to pathogen, which is designed to capture competition once a pathogen has established itself in the gut.

To assess pathogen growth, we built luminescent strains of *K. pneumoniae* and *S.* Typhimurium and compared luminescence when grown in monoculture and when grown in co-culture with each symbiont. With this assay system, we could rank the strains based on their abilities to limit pathogen growth in both the invasion and competition assays (**Fig. 1b-c, S1**). From this ranking, we took the top ten best-performing non-pathogenic symbiont species in the screen (Methods; shown in

orange in **Fig. S1e-f; Table S1**) and subjected them to a more stringent test of colonization resistance designed to capture both phases of competition in the gut in one assay (extended competition assay, **Fig. 1d**). Here, the pathogen is first introduced into a pre-grown culture of a given symbiont strain and then, after 24 hours, the mixture is passaged into fresh media and allowed to grow for 24 hours, whereafter pathogen abundance is assessed via flow cytometry (**Fig. 1d)**. Despite choosing the best-ranked species from the luminescence screen, all symbionts performed poorly under extended competition, with the majority offering no discernible colonization resistance (**Fig. 1e-f)**. The best performer was *Escherichia coli*, a known competitor of *S.* Typhimurium, and also a member of the Enterobacteriaceae, but even here the protection offered was very limited with the pathogens still able to reach $10^8 – 10^9$ cells/ml.

The outcome of the assay differed greatly when we pooled all ten strains together (**Fig. 1e-f**). Now, the final abundance of both pathogens was strongly suppressed by over three orders of magnitude for *K. pneumoniae* and about two orders of magnitude for *S.* Typhimurium. By contrast, a community made up of the ten worst-performing strains from the luminescence screen (shown in blue in **Fig. S1e-f**) provided little or no colonization resistance (**Fig. 1e-f**). These results, therefore, suggest that strain identity is important for colonization resistance only in the context of a diverse community.

Ecological diversity and complexity drive colonization resistance *in vitro*

Our results indicated that microbiota diversity is important for colonization resistance. This finding fits well with the general idea that microbial diversity is beneficial for microbiome functioning, whereas a loss of diversity, or dysbiosis, can be associated with poor health and disease (*25-27*). While the potential benefits of diversity are clear, cause and effect can be confounded in observational studies (*28*). To systematically test the role of diversity in colonization resistance, we randomly selected communities of increasing diversity from the best ranked 10 strains and competed them against the pathogens in the extended competition assay. To further evaluate the importance of diversity, we also assembled a community of 50 non-pathogenic symbiont species from the strains in

our initial luminescence screen (see Methods). These data indicated a relationship between diversity and colonization resistance. However, we also saw a large variation in colonization resistance across the communities that differed in their composition of two, three and five species. Visual inspection of the data (**Fig. 2c-d)** suggested that a large component of this variability was driven by the composition of the communities.

One species that appeared to be important for outcomes was *E. coli*. To explore this finding, we randomly selected additional *E. coli*-containing communities and again evaluated colonization resistance (**Fig. 2c-d**). We also performed drop-out experiments, where we made up the 10 and 50 species communities without *E. coli* (**Fig. 2c-d**). These data revealed a strong and clear monotonic increase in colonization resistance as species diversity increases **(Fig. 2c-d, green circles**), but this relationship is much weaker or disappears entirely in the absence of *E. coli* (**Fig. S2a-b**). In ecological terms, these data show that colonization resistance rests upon a strong higher-order effect involving other community members and *E. coli (22)*. By higher-order effects here, we mean cases where the effect of one species on another is changed by the presence of a third-party species in a community *(22)*. That is, while *E. coli* alone or the rest of the community alone each have little impact on pathogen growth, together they have a strong effect on pathogen growth. Such higher-order effects are considered important in ecology as they imply context dependence, which can make a system difficult to understand and predict (*22, 29-32*). Another way to illustrate the effect of diversity on colonization resistance is to compare our data to a simple null model. Consider, for example, a model where each additional species proportionally improves colonization resistance to the pathogen. Specifically, we can compare our experimental data for *E. coli* containing communities to a null model where pathogen abundance scales according to $1/n$, where n is the number of species present. This analysis shows that the deviation from such a null model increases as diversity increases, where colonization resistance is again greater than expected for diverse communities (**Fig. S3**).

We also asked whether the role of *E. coli* within communities was a strain-specific effect. We replaced *E. coli* strain IAI1, identified in our screen, by each of four other *E. coli* strains historically isolated from the human gut (*33-35*). The effect of *E. coli* was similar when *E. coli* IAI1 was

129 substituted by most *E. coli* strains (**Fig. 2e-f**), which indicates that the higher-order effect involving *E.*
130 *coli* is a general property of closely-related strains.

131 Further inspection of the data pointed to other species that were important for colonization
132 resistance in diverse communities. In *E. coli* containing communities, the presence of *Bifidobacterium*
133 *breve* appeared to be important in excluding *K. pneumoniae* (**Fig. S4a**), and *Lacrimispora*
134 *saccharolyticum* and *Phocaeicola vulgatus* in the exclusion of *S.* Typhimurium (**Fig. S4b**). We
135 confirmed these patterns through a series of systematic drop-out experiments (**Fig. S4c-d**). However,
136 it was still possible to achieve equivalent colonization resistance in more diverse communities that
137 lack these species (**Fig. S4c-d**), which again points to the underlying benefits of a diverse microbiota.

138

139 **Ecological diversity and complexity also drive colonization resistance *in vivo***

140 To validate our *in vitro* methods, we tested the ability of symbiont communities to resist pathogen
141 colonization in gnotobiotic mice (**Fig. 3a**). Germ-free mice were colonized with symbiont
142 communities differing in diversity, and the presence or absence of *E. coli*. Successful colonization by
143 *S.* Typhimurium causes an acute infection and massive gut inflammation, which is a major
144 confounding effect for studying the effects of community composition on pathogen growth. Animals
145 with a less protective microbiota can rapidly succumb to the infection, such that one cannot follow
146 ecological dynamics over time in a comparable way across treatments. We, therefore, chose to use an
147 avirulent variant of *S.* Typhimurium to eliminate the effect of gut inflammation on pathogen and host,
148 where one can use pathogen abundance as a measure of disease risk (*19, 36, 37*). We introduced
149 communities across the same range of diversities as before but, in contrast to *in vitro* assays, not all
150 symbiont species will reliably colonize germ-free mice (*38*). Therefore, we used metagenomic
151 sequencing to confirm that introducing a higher diversity community to the mice did indeed result in a
152 higher diversity of strains colonizing the gut as measured by two metrics of alpha diversity, and to
153 identify the relative abundance of all members (**Fig. 3b-c, S5**).

These experiments revealed that, as observed *in vitro*, microbiome diversity is negatively correlated with pathogen abundance in feces for both pathogens (compare 10 vs 50-member communities; **Fig. 3d-e; Fig. S6**). Moreover, drop-out experiments revealed again the importance of the combination of *E. coli* and other community members for colonization resistance (**Fig. 3d-e**). We also observe that higher diversities are needed for efficient colonization resistance in the mammalian gut than in our *in vitro* assays, which is likely to be explained by the higher degree of environmental and spatial heterogeneity in the gut compared to a test tube. Nevertheless, the key patterns remain the same between the gnotobiotic mouse experiments and our *in vitro* assays. Both ecological diversity and higher-order interactions are important for colonization resistance to both pathogens. As before, we saw a strong deviation from a simple null model of ecological competition at high levels of diversity (**Fig. S3**). In addition to showing the generality of these patterns, this fit between the *in vitro* and *in vivo* methods validates our extended competition assay as an approach to interrogate the ecology of colonization resistance.

A simple principle explains the roles of diversity and complexity in colonization resistance

The discovery of such higher-order effects in colonization resistance indicates that colonization resistance is an ecologically complex trait (*22*), which can be challenging to work with owing to high levels of context dependence (*17, 22, 29, 30*). Nevertheless, we sought to understand the mechanisms underpinning colonization resistance by returning to our *in vitro* data gathered from large numbers of different communities. The genomic data enabled us to assess functional similarity between symbiont communities and pathogens from overlap in protein compositions. Specifically, we calculated the percentage of all protein families carried by a pathogen that were also present in each community investigated (see Methods). We reasoned that this measure of functional similarity may map to niche overlap and, therefore, the strength of ecological competition between symbionts and pathogens. We first confirmed that the number of encoded protein families covered by our experimental communities increases proportionally with the number of added strains (**Fig. S7**). Permutation analyses also

confirmed that the randomly selected communities we have studied experimentally are a good representation of all possible communities that we could have studied (**Fig. S8**).

The potential importance of protein family overlap was already clear from the effects of *E. coli* in our experimental data (**Figs. 2-3**). *E. coli* is in the same family of bacteria as *K. pneumoniae* and *S.* Typhimurium and can be seen to contribute greatly to the overlap in the protein families carried by a given community and either of the pathogens in our experimental communities (**Fig. S9a-d**). However, by taking only the communities that contain *E. coli* to control for this effect, we also see a strong correlation between a community's protein family overlap with the pathogen and its colonization resistance in our *in vitro* assays (**Fig. 4a-b; Fig. S9**). In other words, if the symbiont strain or community encoded many of the same (or similar) proteins as the pathogen, it provides better colonization resistance. The same analysis for communities that lack *E. coli* is not informative as colonization resistance is consistently so low across all communities (**Fig. 2**).

Altogether, our genomic analysis suggests that communities that overlap highly with the pathogens in encoded functions provide the best colonization resistance. These analyses support our hypothesis that niche overlap is important for our observed ecological patterns in colonization resistance. One of the key drivers of niche overlap is resource competition (*39, 40*), which is a known contributor to colonization resistance to *K. pneumoniae* and *S.* Typhimurium *(12, 13)*. We, therefore, explored the role of nutrient competition by generating metabolic profiles for the two sets of 10 key symbiont species identified in the original screen against each pathogen (**Fig. 1**) using AN Biolog MicroPlates that profile the metabolic activity of each strain on 95 carbon sources (**Fig. S10**). Note that to cover the two sets of 10 species, we only profiled 16 strains in practice because there were some parallels between the two sets of top-ranked strains in the luminescence screen. We first established that there was a strong positive association between the protein family (genomic) and metabolic (Biolog) overlap of communities with the pathogens (**Fig. S11**). We then assessed the ability of metabolic overlap to predict colonization resistance (**Fig. 4c-d**). Colonization resistance was only observed once communities shared sufficiently high overlap in their carbon source utilisation profile with a pathogen. Moreover, communities with the greatest metabolic overlap with a pathogen

provided the greatest colonization resistance. An important observation from these data is that it is not diversity *per se* that predicts colonization resistance, it is the overlap between the pathogen and the communities. This pattern is made clear by the observation that communities of different diversities, but the same overlap, appear proximally in the plots (neighbouring points of different color in **Fig. 4c-d**).

Overall, our data point to the importance of nutrient competition, and specifically nutrient use overlap between a community and a pathogen, as an explanation for the patterns we observe in colonization resistance. To further support this conclusion, we performed experiments where the pathogens were grown in cell-free (spent) media collected from different communities, which excluded cell-cell contact mechanisms as explanations for colonization resistance. Growing the pathogen in the spent media of *E. coli* and the 10 species communities recapitulated the patterns seen in the competition experiments, consistent with the effect of nutrient competition (**Fig. S12**). As a final test, we sought a nutrient that can be used by the pathogens only and used it to perform nutrient supplementation experiments (**Fig. 4e-f**). We identified galactitol from the Biolog plates (**Fig. S10**). The pathogens can use this sugar alcohol but it has the desirable property that it cannot be used by any of the symbionts in our focal 10-species communities, except for *E. coli.* We engineered a strain of *E. coli* that lacks the transporter for cell import (*E. coli gatABC* deletion mutant). By adding in galactitol to our standard media, we found that colonization resistance in a diverse community is lost if the pathogens can use the nutrient but *E. coli* cannot (**Fig. 4e-f**). However, colonization resistance is restored when *E. coli* can use the nutrient. Further, if a pathogen is engineered so that it cannot use galactitol (*S. Typhimurium gatABC* deletion mutant), colonization resistance is restored. These outcomes are exactly as expected if nutrient competition is the cause of colonization resistance.

Our data show that the ability of a microbiota community to consume nutrients required by a pathogen for growth underlies the colonization resistance we observe. Importantly, the nutrient blocking effect is a property of the whole community rather than any one species alone. That colonization resistance is a community-level trait explains the importance of the ecological diversity, and complexity (*22*) we observed in our experiments. Despite considerable genomic and metabolic

234 overlap with the pathogens, a species like *E. coli* cannot alone block enough nutrients to provide
235 colonization resistance. It is only in combination with other species, that *E. coli* becomes effective at
236 limiting pathogen growth.

237

238 **Nutrient blocking can identify protective communities**

239 Our experiments indicate that colonization resistance is an ecologically complex trait, but that this
240 complexity can be understood and predicted via a simple underlying principle. As an additional test of
241 these findings, we used the nutrient blocking to predict community compositions that provide
242 colonization resistance to a bacterial strain that was not present in our initial experiments. For this
243 test, we chose an antimicrobial resistant (AMR) clinical *E. coli* strain, which was isolated from the
244 urine of a patient. AMR *E. coli* strains are a major current target for alternatives to antibiotics because
245 members of this species have recently been found to be responsible for the most AMR-associated
246 deaths of any bacterial species (*41*).

247 We first analysed the AMR *E. coli* isolate on AN Biolog MicroPlates to assess its carbon
248 source utilisation and compared this to the top ranked strains from our initial luminescence screen
249 (**Fig. 1**). We reasoned that these top-ranked strains were a good place to start as *E. coli* is also a
250 member of the Enterobacteriaceae, like the two pathogens that were used to select the top-ranked
251 strains. As expected, the AMR *E. coli* had the greatest protein overlap with the symbiont *E. coli* in our
252 16 strains but, importantly, additional strains were predicted to be required to block nutrient
253 availability based on the overlap needed to suppress the two pathogens (**Fig. S13)**. We next used the
254 Biolog data to computationally assemble all possible communities of one, two, three and five species
255 from the 16 strains and calculated their resource utilisation overlap with the AMR *E. coli* (**Fig. 5a**).
256 Again, in line with our findings, diversity improved the median resource utilisation overlap, but this
257 depended strongly on the presence of the symbiont *E. coli*.

258 The simplest test of the importance of nutrient blocking is to remove the symbiont *E. coli*
259 from a community and test the impact. Doing this for the community of all 16 strains confirmed the

importance of *E. coli* for colonization resistance (**Fig. 5b**). However, we also tested our ideas on communities that contain *E. coli*. In these experiments we identified communities predicted to have the highest and lowest overlap with the target strain at each diversity level, where communities were randomly chosen if there were ties in rank. We then used our extended competition assay (**Fig. 1d**) to test the ability of the AMR *E. coli* to invade the communities. As predicted by nutrient blocking, this revealed that increasing diversity leads to increased colonization resistance and, critically, for each diversity level, the community predicted to resist the AMR *E. coli* consistently performed better in colonization resistance than the community predicted to do poorly (**Fig. 5b).** This result was clearest for the two and three species communities. For the five species community, the best performing community was only marginally better than the worst. We reasoned that this was because, in these experiments, we are limited to choosing from 16 strains that were preselected for being relatively good competitors to Enterobacteriaceae (**Fig. 1b-c, S1**).

To test the nutrient blocking principle more robustly we selected from a wider range of possible strains from our set of 50 strains that we used in our *in vitro* and *in vivo* experiments (**Fig. 2-3**). Most of these strains had not been characterised for their functioning in community-level colonization resistance, other than in the 50 species treatment. We also used this set of experiments to test the power of the nutrient blocking principle to predict colonization resistance based upon genomic data alone. Rather than using the Biolog phenotypic assay, therefore, we returned to our measure of protein family overlap, which calculated the overlap in all protein types between an invading strain and different communities. To do this we only had to sequence the AMR *E. coli* clinical isolate because all 50 other strains were sequenced. Using the same approach as for the Biolog predictions, we then assembled communities *in silico* that all contained the symbiont *E. coli* strain and, in each case, calculated their protein family overlap with the AMR *E. coli* (**Fig. 5c-d**). As before, we chose communities with the lowest and highest overlap to the AMR *E. coli* across a range of diversities (randomly choosing communities if there were ties in rank) and experimentally assessed colonization resistance using the extended competition assay.

We again see the importance of community diversity in these experiments. Moreover, despite using only genomic information and a much larger set of possible communities, we observed improvement in colonization resistance from the worst to best communities at each diversity level (**Fig. 5e**). Finally, we evaluated our ability to select highly and poorly performing communities by assessing colonization resistance in additional five-species communities. Notably, at the five-species level, more than 200,000 communities with *E. coli* can be assembled from 50 strains. We used our algorithm to sample approximately 50,000 and, from these, we identified four additional community compositions predicted to perform well and four predicted to perform poorly (to give five of each class). In line with our predictions, the communities predicted to be colonization resistant showed a median 100-fold reduction in the abundance of the AMR *E. coli* compared to those predicted to be permissive (**Fig. 5f**).

Discussion

A key benefit of the microbiome is its ability to reduce the probability of infection via colonization resistance (*1, 2, 10, 42*). Here we have used an ecological approach to understand the principles of colonization resistance in the gut microbiome. By screening a collection of human gut symbionts, we found that individual strains were unable to provide effective resistance to pathogens (**Fig. 1**), but that colonization resistance increases monotonically with ecological diversity (**Fig. 2-3**). Our work, therefore, supports the general hypothesis that a more diverse microbiome can carry health benefits (*28, 43-45*). While much discussed, evidence for this hypothesis is typically based upon correlations between microbiome diversity and health outcomes (*28, 45, 46*). Here, we provide experimental evidence that microbiome diversity can provide health benefits via an increased ability to protect against pathogens. Moreover, we explained this pattern in terms of the importance of the overlap between the nutrient requirements of an invading pathogen and the resident community (**Fig. 4**).

We found that certain combinations of species display much greater colonization resistance together than when alone. These non-additive effects mean that colonization resistance is formally a

312 complex ecological trait in the canon of ecology (*22*). Such effects are often assumed to imply a
313 complex network of interactions between species where, for example, one symbiont species affects a
314 second symbiont species and changes the way this second species interacts with a pathogen. However,
315 consideration of nutrient competition and particularly the level of overlap between a pathogen and
316 community revealed that much simpler processes explain the complexity we see. One species alone is
317 not sufficient to strongly impact pathogen growth, but rather a combination of species is required to
318 block nutrient access. Interestingly, the combinations of species that make colonization-resistant
319 communities are often very phylogenetically diverse. Generating resistant communities does not rest
320 upon simply finding closely related species to a given pathogen, in our case against
321 Enterobacteriaceae members. Instead, a mixture of gram-positive and gram-negative species is often
322 what performs best (**Fig. 5f, Table S7**). While both of our pathogens are members of the same family,
323 they have different life histories (*19-21*). *S.* Typhimurium is a specialist gut pathogen while *K.*
324 *pneumoniae* is an opportunistic pathogen that typically causes no pathology in the gut itself, instead
325 causing infections in other parts of the body. Consistent with this, we find it is consistently easier to
326 generate colonization resistance to *K. pneumoniae* than *S.* Typhimurium. Nevertheless, the importance
327 of ecological diversity and complexity is observed for both pathogens. We anticipate that the
328 importance of ecological diversity and the principle of nutrient blocking will apply generally, given
329 the widespread evidence that nutrient competition is important for diverse species in the microbiome,
330 including many other pathogens (*9, 47-55*).

331 By assembling a wide range of communities of defined compositions, we have been able to
332 establish links between ecological diversity, complexity and nutrient competition in colonization
333 resistance. However, a limitation of this approach is that we have focused on relatively low diversity
334 communities (up to 50 species). A question for the future is whether our findings will hold for the
335 higher levels of diversity that can naturally occur in the human microbiome. Consistent with our
336 findings, colonization risk with species like *K. pneumoniae* is increased after antibiotic treatment that
337 can lower species diversity in the microbiome (*56*). Nutrient competition is central to the patterns that
338 we have described here. Colonization resistance can develop via additional mechanisms, which

include toxin-mediated bacterial competition and effects via the host immune system (*1, 10*). Our work does not exclude the potential importance of these additional mechanisms such as direct killing of invading strains by members of the community, which can act in parallel to nutrient competition (*12, 13, 15, 39*). Moreover, our predictions of resistant and permissive communities are not perfect (e.g., high overlap communities in **Fig. 4c**; the outlier in **Fig. 5f**). However, it is notable that these deviations are in the direction of a community being more resistant than expected. This suggests that predictions based upon nutrient blocking may often be conservative with errors resulting in communities performing better than expected whenever other mechanisms of colonization resistance are at play.

Our work shows that colonization resistance is not the property of single microbiome species but instead the collective property of multiple species. Specifically, we find that the effect of a given symbiont species on a pathogen can be strongly dependent on whether other symbiont species are also present to consume nutrients that the pathogen needs. This finding suggests that one can use the idea of nutrient blocking to identify sets of microbiome species that will limit the growth of a target strain. As a proof-of-principle, we tested this idea for an AMR *E. coli* strain, which revealed that species sets can be successfully identified that collectively suppress an incoming strain (**Fig. 5**). Importantly, we find that this can be done without specific information on available nutrients or the metabolism of the species under study. Instead, a measure of genomic overlap can be used as a proxy for niche overlap to assemble communities that perform nutrient blocking (**Fig. 5c-f**). The human microbiome is dauntingly complex and has great potential for context-dependent effects. However, we found that microbiome complexity can arise via simple underlying principles, which gives promise to the goal of rationally designing microbiomes for better health.

Materials and methods

Bacterial strains and plasmids A full list of bacterial strains used in this study is provided in **Table S1**. The plasmids used in this study are listed in **Table S2**. *Klebsiella pneumoniae* subsp. *pneumoniae* purchased from DSMZ (stock number 30104) was used in all experiments containing *K. pneumoniae*.

Salmonella enterica serovar Typhimurium (strain SL1344 (*57*)) was used in all experiments containing *S.* Typhimurium. The 100 strains were chosen for being abundant and/or important members of the human gut microbiota that cover the key phylogenetic groups found in the human gut microbiota. In addition, all are type strains, have had their genome sequenced (which we rely on for the protein family analysis), and are known to be culturable in nutrient-rich, anaerobic medium (mGAM) (*58, 59*), so were conducive to experimental work. Strains were kindly provided by Nassos Typas (EMBL Heidelberg, Germany), or were ordered from the German Collection of Microorganisms, DSMZ or the American Type Culture Collection (ATCC; Staph epidermidis). The *Lactobacillus plantarum* strain was provided by the Department of Food and Nutritional Sciences, University of Reading. *Klebsiella pneumoniae* ATCC 700721 was provided by the Modernising Medical Microbiology research group, Nuffield Department of Medicine, University of Oxford. *Escherichia coli* strains Z1331 and Z1269 were isolated from the feces of two healthy human donors in Switzerland (*33*). The ampicillin-resistant AMR *E. coli* strain is a urine clinical isolate provided by the Pathogen Bank at Nottingham University Hospitals NHS Trust.

Bacterial growth conditions For engineering of *E. coli*, *K. pneumoniae*, or *S.* Typhimurium, strains were grown aerobically in Lysogeny broth (LB; Fisher Scientific) with the appropriate antibiotics (**Table S1**) at 37°C, shaking at 220 rpm. All symbionts were cultured under anaerobic conditions (5% H_2, 5% CO_2, 90% N_2, <20ppm O_2) in modified Gifu Anaerobic Medium (mGAM; Nissui Pharmaceuticals) broth buffered to pH 6.2 with 100mM 2-(N-morpholino)ethanesulfonic acid (MES; Sigma-Aldrich). The redox indicator dye Resazurin (100µg/L media; Sigma-Aldrich) was added as a quality control check for the presence of oxygen (turns red when conditions are not sufficiently anaerobic). To prepare glycerol stocks, individual strains were first streaked onto mGAM agar (Nissui Pharmaceuticals) and grown under anaerobic conditions (if possible, as not all strains can grow as single colonies on agar-based media). The identity of each strain was confirmed using 16S rRNA sequencing (Sanger sequencing; Source Biosciences) using the primers oOPC-953 and oOPC-954 for most species, or g-Bifid-F and g-Bifid-R for Bifidobacteria (**Table S3**). Single colonies were then inoculated in mGAM broth and stored at -70 degrees Celsius in mGAM with a final concentration of 25% glycerol.

Genetic engineering of bacterial strains Luminescence and fluorescence plasmids were transformed into *K. pneumoniae* and *S.* Typhimurium using electroporation. Briefly, 5mL overnight culture was washed 3x with cold Milli-Q water, before being concentrated in 500µl cold Milli-Q water. 2µl of plasmid was mixed with 100µl of concentrated cells and electroporated (1.8kV; 0.1cm gap cuvettes) before recovery in 1ml pre-warmed LB (1h at 37°C, shaking at 220rpm) and plating on LB with the appropriate antibiotics.

Gene deletions were generated as described in (*60*). Briefly, 700 base pairs upstream and downstream of region to be deleted were PCR amplified (Phusion®, NEB) and inserted into the suicide vector pFOK,

which was linearised using the restriction enzymes BamHI and EcoRI, using the NEBuilder® HiFi DNA Assembly (NEB). The plasmid was introduced into a diaminopimelic acid auxotroph E. coli strain (JKe201). After 6h of mating between the plasmid-containing donor *E. coli* strain and the recipient *E. coli*, *S.* Typhimurium or *K. pneumoniae* strain, trans-conjugants were selected on LB plates containing 50µg/mL kanamycin. Counter-selection was performed on no-salt LB plates supplemented with 0.5µg/mL of anhydrous tetracycline and 20% of sucrose at 30ºC. Mutants were screened by colony PCR and the sequence was verified using Sanger sequencing (Eurofins). Primers used for genetic engineering in this study are listed in **Table S3**.

Luminescence screen *K. pneumoniae* (DSM 30104) and *S.* Typhimurium (SL1344) carrying a low copy number plasmid with P_{nptII} promoter-driven expression of *lux*CDBAE-*frp* (pRSJ-p_{nptII}::ilux; also known as the improved lux operon plasmid (*61*)) were used for the luminescence screen. Symbiont strains were tested in pairwise co-culture with the pathogens. All strains tested were first grown in monoculture anaerobically in static mGAM broth buffered to pH 6.2 with 100mM MES at 37ºC. *K. pneumoniae* WT and *S.* Typhimurium WT carrying the luminescence plasmid pRSJ-p_{nptII}::ilux were separately incubated in the same media overnight.

Competition assay: once all strains had reached stationary phase (12-72 hours of growth depending on the species), 20µL of each symbiont strain was dispensed in technical triplicates in a 96-well plate containing 160µL of mGAM broth. 20µL of *K. pneumoniae* pRSJ-p_{nptII}::ilux overnight culture or S. Typhimurium pRSJ-p_{nptII}::ilux overnight culture was then added to each well.

Ecological invasion assay: 180µL of each symbiont strain was dispensed in technical triplicates in a 96-well plate and 20µL *K. pneumoniae* pRSJ-p_{nptII}::ilux overnight culture or *S.* Typhimurium pRSJ-p_{nptII}::ilux overnight culture was added to each well.

96-well plates from both assays were incubated anaerobically for 6 hours at 37ºC. Before luminescence measurement, 96-well plates were brought out of the anaerobic chamber and exposed to O_2 for 10 minutes because, critically, O_2 is needed for light production catalysed by the enzyme luciferase. Luminescence at 515-575nm and OD-600nm were measured using a CLARIOstar Plus spectrometer (BMG Labtech). Each data point (**Fig. S1)** represents the median of at least 3 independent biological replicates (that is, using different overnight cultures of the same strain in different 96-well plates).

Selection of the top and bottom ten symbiont strains from the screen For each of the ecological invasion and competition assays, symbiont strains were ordered from best- to worst-performing. Symbiont strains were given two rankings: one from the competition assay and one from the invasion assay. The two rankings were summed and ordered from lowest sum (best overall performing) to highest sum (worst overall performing). From these ranks, the top ten and bottom ten strains were chosen. In

order to focus on the ability of non-pathogenic strains to limit colonization by pathogens, we decided to focus for the community experiments on strains that have a category 1 safety level. Many common gut bacteria cause occasional infections and therefore categorised as opportunistic pathogens with a category 2 safety level, which can blur the line between being protective and causing disease. Removing these strains took us from 100 to 54 strains. From these, we selected 50 species as our core set, removing some strains that were duplicates of the same species.

Construction of phylogenomic tree The phylogenomic tree of strains in the luminescence screen (n=100; **Fig. 1b-c; Table S1**) was inferred using default settings based on 67 single-copy core genes using anvio v7.1 (*62*) and plotted using iTOL v6.7.5 (*63*).

Extended competition assay All culturing was performed in a shaking incubator at 225rpm and 37ºC, and under anaerobic conditions. Symbiont strains were grown to stationary phase (12-72 hours, depending on the species), in Hungate tubes containing 5mL mGAM broth and 100mM MES and buffered to pH 6.2. *K. pneumoniae* and *S.* Typhimurium carrying the fluorescence plasmid pBC11 (YPet) (*64*) were incubated overnight, with the appropriate antibiotics (**Table S1**).

Once in stationary phase, monocultures of symbionts were passaged (100µL culture into a new 5mL tube of media) and grown for ~17 hours. Communities were assembled under anaerobic conditions into new 5mL Hungate tubes (see section "community preparation") and grown for 24 hours. Once grown, communities were invaded with 100µl of the pathogen (10^6 cells/mL final concentration) in each Hungate tube containing 5mL culture. Samples were taken immediately after addition of the pathogen and prepared for measurement using flow cytometry (day 0). After 24 hours growth, the invaded communities were sampled for flow cytometry (day 1) and 100µL was passaged into new tubes of 5mL mGAM broth. After a further 24 hours, the end-point communities were sampled and measured a final time (day 2).

For nutrient supplementation experiments (galactitol), 2.5mL mGAM buffered with 100mM MES to pH 6.2 was prepared and autoclaved in Hungate tubes. A 2x stock solution of filter-sterilised sugar in Milli-Q water (e.g. 0.2% or 2% galactitol) was prepared and de-oxygenated in the anaerobic chamber. 2.5mL of the stock solution was added into the Hungate tube to generate a 5ml total volume with the appropriate concentration of the sugar. The extended competition experiment is performed as described above, except samples were passaged again into tubes with the appropriate concentration of the sugar. Samples were left to grow for 48h instead of the standard 24h to allow utilisation of low priority sugars, such as galactitol.

For experiments with the AMR *E. coli* clinical isolate, the experiment was performed as described above, except samples were analysed using selective plating instead of flow cytometry (LB + 100µg/mL ampicillin to select for the AMR *E. coli* strain; MacConkey agar to enumerate total *E. coli* densities).

Community preparation All culturing was performed in a shaking incubator at 225rpm and 37ºC, and under anaerobic conditions. Constituent monocultures were grown to stationary phase, passaged into new media and grown overnight for approximately ~17 hours, and then combined into communities. The OD-600nm of the single strains was measured and the cultures were streaked anaerobically on mGAM agar to check for contamination. Communities ranged from 2 to 50 strains in size and were assembled under anaerobic conditions.

In vitro: For the communities containing 2-10 strains, equal volumes (100µL) of each overnight culture were combined in a Hungate tube containing 5mL mGAM to form a community. For the 49- and 50-strain communities, 20µL of each single strain was added, equating to a total volume of 1mL. The communities were made by adding 1mL of each strain anaerobically to a 50mL falcon tube, mixing, and removing 1mL to add to a fresh Hungate tube of 5mL mGAM broth. Communities were grown in the shaking incubator (37ºC, 225rpm) for 24 hours before being invaded with 10^6 cells/mL of the pathogen.

In vivo: Strains were grown separately, then OD-600nm of each culture was measured. For communities containing 2-10 strains, the volume of culture containing 10^9 cells was calculated for each individual strain, based on a standard curve of OD-600nm and flow cytometry quantification. This was added to a large falcon tube. The community mix was centrifuged (10 minutes, room temperature, 14000rpm) and the resulting pellet resuspended in 2mL anaerobic PBS and kept on ice. 200µl were gavaged into each mouse corresponding to 10^8 cells for each strain. For the 49- or 50-strain communities, we used less cells for each strain to avoid gavaging a large total cell density into mice (as done for the *in vitro* experiments). The volume of culture containing $2*10^8$ cells was calculated for each constituent strain and they were combined. The mixture was centrifuged and the pellet resuspended in 2mL PBS. Mice were gavaged with 200µL of the inoculum. Therefore, the density of bacteria in the 200µL gavage was equivalent for the 50-strain community as for the 10-strain community (10^9 total cells). For all inoculums, samples were taken for plating anaerobically on mGAM agar to confirm the cell density.

Flow cytometry Hungate tubes containing culture to be sampled were de-pressurised under anaerobic conditions using needles. Samples were removed using needles and syringes. 10µL of each sample was diluted in 90µL PBS + chloramphenicol 200µg/ml then 10µL added to a 96-well flat-bottomed plate already containing 80µL PBS + chloramphenicol 200µg/ml. The 96-well plate was left to shake at room temperature on the bench for 15 minutes to allow oxygen-dependent folding of YPet, while preventing further growth of the sample with the bacteriostatic antibiotic chloramphenicol. Next, all bacterial cells

were fixed, permeabilised and fluorescently stained by adding 90µL PBS containing 4% paraformaldehyde, 0.4% Triton X-100 and 1µg/mL 4',6-diamidino-2-phenylindole (DAPI) to each sample well and incubating the plate in the dark for 1 hour at room temperature on a microplate shaker. Prior to the measurement, 20µL AccuCheck counting beads (ThermoFisher- PBC100) was added to every sample well. Two wells containing 200µL PBS were left between sample wells, in order to prevent cross-contamination between samples. 96-well plates were run in a flow cytometer (Attune NxT Autosampler with Attune NxT Flow Cytometer, ThermoFisher) set to: acquisition volume 50µL, total draw volume 80µl, total sample volume 200µL, at a speed of 100µL/minute. Relevant spectral parameters were recorded by the flow cytometer, equipped with 405nm, 488nm and 561nm lasers. The following channels were used: DAPI staining, excitation 405nm, emission 415-465nm ('blue'); YPet fluorescence, excitation 488nm, emission 505-515nm and 525-555nm ('green').

Flow cytometry data was processed using FlowJo v10.8.1 software. 'Beads' were gated on a linear (FSC-H; SSC-H) axis. 'Bacteria' were gated on a logarithmic (FSC-H; SSC-H) axis. DAPI was gated within the 'Bacteria' gate, on a logarithmic (Ex405-417; Ex405-495) axis. Within the DAPI gate, YPet was gated on a logarithmic (Ex488-503; Ex488-555) axis. Count statistics for beads, DAPI and YPet were exported to Microsoft Excel and pathogen density (cells/mL) was calculated.

Mouse husbandry and experiments Mouse experiments were performed with 6-8 weeks old germ-free wild-type C57BL/6J female mice (RRID IMSR_JAX:000664) that were bred and maintained in germ-free isolators at the Kennedy Institute, University of Oxford for more than five years. During experiments, mice were housed as pairs or trios in sterile, individually ventilated Sentry SPP cages (Allentown) with enrichment, irradiated food and autoclaved water. Cages of mice were disinfected in TecCare Ultra Hydrogen peroxide, peracetic acid solution before mice were handled using individually sterilised gloves (single use for each cage) in a biosafety cabinet that was disinfected with the same acid solution in between each treatment group. Cages of mice were randomly assigned to a group, and cages were opened in the same order each time. Researchers were not blinded because of the unbiased quantification methods of colony counting and metagenomic sequencing. Samples sizes were determined using standard practices of 7-8 mice per group distributed over 2-3 cages, and repeated the experiment three times to allow experimental independence and sufficient power for nonparametric testing. Mouse experiments were performed in accordance with the UK Animal Scientific Procedures Act (1986) under a UK Home Office licence (PPL 9127884) assessed by the Medical Sciences Division Animal Welfare and Ethical Review Body (AWERB) at the University of Oxford. Due to strictly defined humane endpoints, one mouse needed to be terminated early. In this case, we excluded data from that mouse.

Germ-free mice were gavaged with a 200µL inoculum of symbionts (see "community preparation") at day -14 of the experiment and again on day -12 of the experiment (see scheme **Fig. 3a)**. On day 0,

overnight cultures of the pathogens (*K. pneumoniae* and *S.* Typhimurium) grown aerobically in LB containing the appropriate antibiotics were washed 3x with PBS and diluted. 10^6 CFU of the pathogen was given to each mouse as a 100µl gavage. Feces were sampled daily after gavage of the pathogen. Fecal pellets were weighed and homogenised with a 5mm stainless steel bead in 1ml PBS (samples were shaken vigorously in 2ml Eppendorf tubes). CFUs were enumerated using selective plating (MacConkey or LB agar for *E. coli*; MacConkey + 50 µg/mL streptomycin for *S.* Typhimurium; LB + 50µg/mL carbenicillin for *K. pneumoniae*). *E. coli*, *K. pneumoniae* and *S.* Typhimurium, can be differentiated by color on MacConkey agar (*S.* Typhimurium cannot utilise lactose) or morphology (*K. pneumoniae* produces a capsule). Mice were euthanised 4 days after infection with the pathogen.

S. Typhimurium triggers inflammation to bloom in the gut (*19*). To allow us to focus on early colonization events prior to triggering inflammation (i.e., colonization resistance), we used an avirulent strain (SL1344 Δ*invG* Δ*ssaV*) to avoid triggering inflammation, disease and mortality (*65*).

DNA isolation from feces and metagenomic sequencing Aliquots of the feces were frozen immediately after collection (prior to homogenisation in PBS) and stored at -80°C until use. Fecal samples were thawed and resuspended in nuclease-free water, before being transferred to a lysing matrix B tube (MP Biomedicals). 3 rounds of bead beating were performed at 6 m/s for 40s. Samples were centrifuged at high speed and DNA in the supernatant was precipitated by adding sodium acetate (1/10 volume) and ice-cold ethanol (96-100%; equal volume), and left at -20°C overnight. Samples were centrifuged at high speed and the pellet was washed twice with 70% ethanol, before being dried and resuspended in nuclease-free water.

Samples were further purified using an AMPure clean-up protocol. The samples were mixed with AMpure XP beads (Beckman Coulter) and incubated at room temperature for 5 minutes, before being washed twice with 70% ethanol, using a magnet to avoid removing the DNA bound to the AMPure beads. After air drying, nuclease-free water was added to the beads, to allow collection of the purified DNA in the supernatant.

CosmosID performed metagenomic sequencing on isolated DNA. DNA libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina) and IDT Unique Dual Indexes. Prepared libraries were sequenced on the Illumina NovaSeq 6000 platform 2x150bp (3M reads per sample). CosmosID performed bioinformatic analysis according to proprietary methods on the raw data to generate fine-grained taxonomic and relative abundance estimates (**Fig. S5**).

Spent media assay Strains were anaerobically inoculated, grown, passaged and assembled into communities as described above. Communities were grown anaerobically in the shaking incubator (225rpm, 37ºC) for 96 hours to maximise depletion of nutrients utilisable by the community. Supernatants were prepared by centrifuging the 96-hour cultures (10 minutes, 4400rpm, room

temperature) and filter-sterilising the supernatant (0.2µm filter). All preparation of supernatants were done under anaerobic conditions. Two of each of the communities were assembled so that, for each community, one Hungate tube containing 5mL supernatant and another Hungate tube containing 2.5mL supernatant and 2.5mL mGAM (the re-supplemented treatment) were prepared. The supernatants were invaded with 10^6 cells/mL of fluorescent pathogen, sampled and incubated for 24 hours as before. Samples were taken on days 0 and 1.

Nutrient utilisation overlap using biolog assays Utilisation of carbon sources by symbionts and the pathogens were assessed using AN Biolog Microplates™ (Biolog) according to the protocol of the manufacturer. Briefly, strains were grown in mGAM and passaged (as detailed above). The cultures were centrifuged anaerobically and washed twice in anaerobic PBS. The samples were concentrated and an aliquot was taken to measure the OD-600nm. The cells were diluted or concentrated such that a 200µl aliquot of the cells would correspond to 65% transmittance (OD 0.187; the density outlined by the manufacturer) in the 14ml volume of AN inoculating fluid provided. The concentrated cells were added to the inoculating fluid aerobically (a small amount of O_2 is needed to oxidise the buffer) and 100µl was aliquoted into each well of the microplate. After 10min in aerobic conditions, the plates were put in an air-tight container with a GasPak EZ Container System Sachet to generate hydrogen-free anaerobic conditions. Samples were incubated at 35°C for 24h before measurement at 590nm. Since *S.* Typhimurium SL1344 is a histidine auxotroph, a prototrophic strain was made using P22 transduction (*66*) of the allele from a prototrophic strain of *S.* Typhimurium (ATCC 14028S). Positive clones were selected for on M9 minimal media with 50µg/ml streptomycin and re-streaked 3 times on LB plates to ensure phages are removed. Growth in the absence of exogenous histidine was validated by streaking on M9 media.

For each plate, the absorbance reading at 590nm was subtracted from the blank (no carbon source control in well A1). Each strain generates a different background signal so it is important to do this on each plate individually. Each strain was measured as three independent biological replicates and the median absorbance value was taken for each carbon source. We used a thresholding approach to determine if a strain can metabolise a given carbon source (defined as Abs 590nm >0.1 median, after blank subtraction; **Fig. S10**).

To predict overlap with the pathogens, whether a given strain can use a carbon source was compared to the pathogen. We assessed which percentage of carbon sources that the pathogen can use can also be used by the symbiont. To calculate overlap between pathogens and communities, we used an additive calculation approach, where if a strain is contained in a community that can use a carbon source, the entire community can use the carbon source. For simplicity, we did not treat cases where multiple species use the same nutrients within the community differently than if a given nutrient is only covered by one species.

Genomic analysis using protein overlap Genomic information from 50 symbiont strains and two pathogens (*S.* Typhimurium, *K. pneumoniae*) were retrieved from the PATRIC database (*67*) (**Table S1**). On this set, we applied a cluster-based analysis that groups proteins encoded in genomes into PATRIC global protein families (*68*). A majority of symbiont strain proteins obtained a protein family designation (clustering rate of 97.83%; 172 848 of 176 675) with protein families being populated by proteins from on average 2.44 genomes (**Fig. S7**).

Based on this set, we drew protein family designations for focal symbiont communities and calculated pathogen/community overlap using the same approach as for the Biolog plates above. That is, for each protein family encoded in the pathogen's genome, we checked if it overlapped with any protein families encoded in the community.

Whole-genome sequencing An overnight culture of *E. coli* 19Y000018 (the AMR clinical urine isolate) was prepared in LB containing 100µg/ml ampicillin. A 1mL pellet was taken and DNA was extracted using the same ethanol precipitation and AMPure clean-up protocol detailed above in the section "DNA isolation from feces". Source BioScience performed whole-genome Illumina sequencing on a NovaSeq 6000 to generate 10M 150bp paired end reads. Genome assembly was performed with Unicycler v0.4.8 using default settings and subjected to protein family annotation as described in "Genomic analysis using protein overlap".

Predictions based on biologs and protein families *Biolog-based predictions:* The carbon source utilisation profile of the AMR *E. coli* clinical isolate was assessed using AN Biolog Microplates™ as detailed above. Next, all possible communities made of 1, 2, 3, and 5 species were identified from the 16 symbiont strains that were previously analysed for carbon source utilisation. For each community, the carbon source utilisation overlap was calculated (as detailed above; **Fig. 5a**). At each diversity level, communities were sorted by overlap to the AMR *E. coli* clinical isolate. Given that the symbiont *E. coli* is needed for colonization resistance in our experiments, we only included communities where the symbiont *E. coli* is present. At each diversity level, we chose the highest ranked community (defined "predicted best") and lowest ranked community (defined "predicted worst"). If there were ties in rank, one community was randomly chosen. These communities were experimentally assessed for colonization resistance using the extended competition *in vitro* assay as detailed above.

Protein family-based predictions: We used the whole-genome sequence of the AMR *E. coli* isolate to calculate protein family overlap to communities of symbionts drawn from our 50-species community. As for the biolog-based predictions, we restricted the analysis to communities that contained the symbiont *E. coli* strain. Specifically, we generated 100,000 different communities randomly, each containing the symbiont *E. coli* and a total of 2, 3, 5, or 10 species from the 50 species pool. We ranked the communities according to their pathogen overlap at each diversity level and selected the highest and

lowest ranked communities for experimental validation. If ties occurred, we chose the communities at random, as for the biolog-based predictions.

Statistical analysis All graphs were made and statistical analysis carried out in Prism v9.4.1 (GraphPad). Protein family overlap analyses and protein family-based predictions were computed in R version v4.0.5 (*69*). Figure legends indicate the statistical test used and the sample sizes. Nonparametric tests were used to avoid the assumptions of normality.

Ethical statement Mouse experiments were performed in accordance with the UK Animal Scientific Procedures Act (1986) under a UK Home Office licence (PPL 9127884) assessed by the Medical Sciences Division Animal Welfare and Ethical Review Body (AWERB) at the University of Oxford. Mice were euthanised using cervical dislocation followed by exsanguination as a confirmation of death.

References

1. M. T. Sorbara, E. G. Pamer, Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them. *Mucosal Immunol* **12**, 1-9 (2019).
2. A. Jacobson *et al.*, A Gut Commensal-Produced Metabolite Mediates Colonization Resistance to Salmonella Infection. *Cell Host Microbe* **24**, 296-307 e297 (2018).
3. E. M. Velazquez *et al.*, Endogenous Enterobacteriaceae underlie variation in susceptibility to Salmonella infection. *Nat Microbiol* **4**, 1057-1064 (2019).
4. S. Y. Wotzka *et al.*, Escherichia coli limits Salmonella Typhimurium infections after diet shifts and fat-mediated microbiota perturbation in mice. *Nat Microbiol* **4**, 2164–2174 (2019).
5. R. P. Sequeira, J. A. K. McDonald, J. R. Marchesi, T. B. Clarke, Commensal Bacteroidetes protect against Klebsiella pneumoniae colonization and transmission through IL-36 signalling. *Nat Microbiol* **5**, 304-313 (2020).
6. C. G. Buffie *et al.*, Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517**, 205-208 (2015).
7. S. G. Kim *et al.*, Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. *Nature* **572**, 665-669 (2019).
8. M. X. Byndloss *et al.*, Microbiota-activated PPAR-gamma signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science* **357**, 570-575 (2017).
9. R. A. Oliveira *et al.*, Klebsiella michiganensis transmission enhances resistance to Enterobacteriaceae gut invasion by nutrition competition. *Nat Microbiol* **5**, 630-641 (2020).
10. G. Caballero-Flores, J. M. Pickard, G. Nunez, Microbiota-mediated colonization resistance: mechanisms and regulation. *Nat Rev Microbiol* **21**, 347–360 (2023).
11. E. T. Granato, T. A. Meiller-Legrand, K. R. Foster, The Evolution and Ecology of Bacterial Warfare. *Curr Biol* **29**, R521-R537 (2019).
12. C. Eberl *et al.*, E. coli enhance colonization resistance against Salmonella Typhimurium by competing for galactitol, a context-dependent limiting carbon source. *Cell Host Microbe* **29**, 1680-1692 e1687 (2021).
13. L. Osbelt *et al.*, Klebsiella oxytoca causes colonization resistance against multidrug-resistant K. pneumoniae in the gut via cooperative carbohydrate competition. *Cell Host Microbe* **29**, 1663-1679 e1667 (2021).

14. S. Caballero *et al.*, Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. *Cell Host Microbe* **21**, 592-602 e594 (2017).
15. S. Brugiroux *et al.*, Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. *Nat Microbiol* **2**, 16215 (2016).
16. A. G. Cheng *et al.*, Design, construction, and in vivo augmentation of a complex gut microbiome. *Cell* **185**, 3617-3636 e3619 (2022).
17. S. Widder *et al.*, Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J* **10**, 2557-2568 (2016).
18. E. Tacconelli *et al.*, Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis* **18**, 318-327 (2018).
19. B. Stecher *et al.*, Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. *PLoS Biol* **5**, 2177-2189 (2007).
20. S. E. Majowicz *et al.*, The global burden of nontyphoidal Salmonella gastroenteritis. *Clin Infect Dis* **50**, 882-889 (2010).
21. C. L. Gorrie *et al.*, Gastrointestinal Carriage Is a Major Reservoir of Klebsiella pneumoniae Infection in Intensive Care Patients. *Clin Infect Dis* **65**, 208-215 (2017).
22. M. M. Mayfield, D. B. Stouffer, Higher-order interactions capture unexplained complexity in diverse communities. *Nat Ecol Evol* **1**, 62 (2017).
23. Y. Litvak, A. J. Baumler, The founder hypothesis: A basis for microbiota resistance, diversity in taxa carriage, and colonization resistance against pathogens. *PLoS Pathog* **15**, e1007563 (2019).
24. K. Z. Coyte, S. Rakoff-Nahoum, Understanding Competition and Cooperation within the Mammalian Gut Microbiome. *Curr Biol* **29**, R538-R544 (2019).
25. O. Manor *et al.*, Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat Commun* **11**, 5206 (2020).
26. C. A. Lozupone, J. I. Stombaugh, J. I. Gordon, J. K. Jansson, R. Knight, Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220-230 (2012).
27. C. Tropini *et al.*, Transient Osmotic Perturbation Causes Long-Term Alteration to the Gut Microbiota. *Cell* **173**, 1742-1754 e1717 (2018).
28. K. V. Johnson, P. W. Burnet, Microbiome: Should we diversify from diversity? *Gut Microbes* **7**, 455-458 (2016).
29. A. Baichman-Kass, T. Song, J. Friedman, Competitive interactions between culturable bacteria are highly non-additive. *Elife* **12**, e83398 (2023).
30. T. J. C. Ian Billick, Higher Order Interactions in Ecological Communities: What Are They and How Can They be Detected? *Ecology* **75**, 15 (1994).
31. A. L. Gould *et al.*, Microbiome interactions shape host fitness. *Proc Natl Acad Sci U S A* **115**, E11951-E11960 (2018).
32. W. B. Ludington, Higher-order microbiome interactions and how to find them. *Trends Microbiol* **30**, 618-621 (2022).
33. S. Y. Wotzka *et al.*, Microbiota stability in healthy individuals after single-dose lactulose challenge-A randomized controlled study. *PLoS One* **13**, e0206214 (2018).
34. S. B. Formal, G. J. Dammin, E. H. Labrec, H. Schneider, Experimental Shigella infections: characteristics of a fatal infection produced in guinea pigs. *J Bacteriol* **75**, 604-610 (1958).
35. F. R. Blattner *et al.*, The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-1462 (1997).
36. M. Ackermann *et al.*, Self-destructive cooperation mediated by phenotypic noise. *Nature* **454**, 987-990 (2008).
37. E. Gul *et al.*, The microbiota conditions a gut milieu that selects for wild-type Salmonella Typhimurium virulence. *PLoS Biol* **21**, e3002253 (2023).

38. R. J. Gibbons, S. S. Socransky, B. Kapsimalis, Establishment of Human Indigenous Bacteria in Germ-Free Mice. *J Bacteriol* **88**, 1316-1323 (1964).
39. R. Freter, H. Brickner, M. Botney, D. Cleven, A. Aranki, Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. *Infect Immun* **39**, 676-685 (1983).
40. A. Wagner, Competition for nutrients increases invasion resistance during assembly of microbial communities. *Mol Ecol* **31**, 4188-4203 (2022).
41. C. Antimicrobial Resistance, Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629-655 (2022).
42. M. R. McLaren, B. J. Callahan, Pathogen resistance may be the principal evolutionary advantage provided by the microbiome. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190592 (2020).
43. M. J. Blaser, The theory of disappearing microbiota and the epidemics of chronic diseases. *Nat Rev Immunol* **17**, 461-463 (2017).
44. M. Fassarella *et al.*, Gut microbiome stability and resilience: elucidating the response to perturbations in order to modulate gut health. *Gut* **70**, 595-605 (2021).
45. H. C. Wastyk *et al.*, Gut-microbiota-targeted diets modulate human immune status. *Cell* **184**, 4137-4153 e4114 (2021).
46. E. Le Chatelier *et al.*, Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541-546 (2013).
47. S. E. Winter *et al.*, Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature* **467**, 426-429 (2010).
48. K. M. Pruss, J. L. Sonnenburg, C. difficile exploits a host metabolite produced during toxin-mediated disease. *Nature* **593**, 261-265 (2021).
49. L. Maier *et al.*, Microbiota-derived hydrogen fuels Salmonella typhimurium invasion of the gut ecosystem. *Cell Host Microbe* **14**, 641-651 (2013).
50. M. L. Jenior, J. L. Leslie, V. B. Young, P. D. Schloss, Clostridium difficile Colonizes Alternative Nutrient Niches during Infection across Distinct Murine Gut Microbiomes. *mSystems* **2**, e00063-00017 (2017).
51. K. M. Ng *et al.*, Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96-99 (2013).
52. A. W. Hudson, A. J. Barnes, A. S. Bray, D. A. Ornelles, M. A. Zafar, Klebsiella pneumoniae l-Fucose Metabolism Promotes Gastrointestinal Colonization and Modulates Its Virulence Determinants. *Infect Immun* **90**, e0020622 (2022).
53. A. G. Jimenez, M. Ellermann, W. Abbott, V. Sperandio, Diet-derived galacturonic acid regulates virulence and intestinal colonization in enterohaemorrhagic Escherichia coli and Citrobacter rodentium. *Nat Microbiol* **5**, 368-378 (2020).
54. F. C. Pereira *et al.*, Rational design of a microbial consortium of mucosal sugar utilizers reduces Clostridiodes difficile colonization. *Nat Commun* **11**, 5104 (2020).
55. E. Gul *et al.*, Differences in carbon metabolic capacity fuel co-existence and plasmid transfer between Salmonella strains in the mouse gut. *Cell Host Microbe* **31**, 1140-1153 e1143 (2023).
56. N. Raffelsberger *et al.*, Gastrointestinal carriage of Klebsiella pneumoniae in a general adult population: a cross-sectional study of risk factors and bacterial genomic diversity. *Gut Microbes* **13**, 1939599 (2021).
57. S. K. Hoiseth, B. A. Stocker, Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. *Nature* **291**, 238-239 (1981).
58. L. Maier *et al.*, Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623-628 (2018).
59. M. Tramontano *et al.*, Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nat Microbiol* **3**, 514-522 (2018).

60. F. R. Cianfanelli, O. Cunrath, D. Bumann, Efficient dual-negative selection for bacterial genome editing. *BMC Microbiol* **20**, 129 (2020).
61. R. Soldan *et al.*, From macro to micro: a combined bioluminescence-fluorescence approach to monitor bacterial localization. *Environ Microbiol* **23**, 2070-2085 (2021).
62. A. M. Eren *et al.*, Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* **6**, 3-6 (2021).
63. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W296 (2021).
64. O. Cunrath, D. Bumann, Host resistance factor SLC11A1 restricts Salmonella growth through magnesium deprivation. *Science* **366**, 995-999 (2019).
65. S. Hapfelmeier *et al.*, The Salmonella pathogenicity island (SPI)-2 and SPI-1 type III secretion systems allow Salmonella serovar typhimurium to trigger colitis via MyD88-dependent and MyD88-independent mechanisms. *J Immunol* **174**, 1675-1685 (2005).
66. N. L. Sternberg, R. Maurer, Bacteriophage-mediated generalized transduction in Escherichia coli and Salmonella typhimurium. *Methods Enzymol* **204**, 18-43 (1991).
67. A. R. Wattam *et al.*, Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* **45**, D535-d542 (2017).
68. J. J. Davis *et al.*, PATtyFams: Protein Families for the Microbial Genomes in the PATRIC Database. *Front Microbiol* **7**, 118 (2016).
69. R Development Core Team*,* R. F. f. S. Computing, Ed. (Vienna, Austria, 2021).
70. F. Spragge and E. Bakkeren, M.T. Jahn, E.B.N. Araujo, C.F. Pearson, X. Wang, L. Pankhurst, O. Cunrath, K.R. Foster. Data for: Microbiome diversity protects against pathogens by nutrient blocking, Dryad (https://doi.org/10.5061/dryad.pnvx0k6v8) (2023).
71. F. Spragge and E. Bakkeren, M.T. Jahn, E.B.N. Araujo, C.F. Pearson, X. Wang, L. Pankhurst, O. Cunrath, K.R. Foster. Code for: Microbiome diversity protects against pathogens by nutrient blocking, GitHub (https://github.com/MartinTJahn/Nutrient_blocking) (2023).
72. B. Periaswamy *et al.*, Live attenuated S. Typhimurium vaccine with improved safety in immuno-compromised mice. *PLoS One* **7**, e45433 (2012).
73. J. S. Johnson *et al.*, Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* **10**, 5029 (2019).
74. T. Matsuki *et al.*, Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. *Appl Environ Microbiol* **68**, 5445-5451 (2002).

Acknowledgments

We are indebted to members of the Foster lab for discussion, and to Emma Slack, Katharine Coyte, Anna Weiss and Wolf-Dietrich Hardt for feedback on the manuscript. We thank Fiona Powrie and the Oxford Centre for Microbiome Studies for germ-free mice and gnotobiotic mouse work. The Heidelberg strains were a gift from Nassos Typas, EMBL Heidelberg. The AMR *E. coli* strain 19Y000018 was from Nottingham University Hospitals Pathogen Bank (https://www.nuh.nhs.uk/pathogen-industry/). **Funding:** FS was supported by a BBSRC Studentship. EB was supported by SNSF postdoc mobility fellowships (P2EZP3_199916 and P500PB_210941). MTJ was supported by the Human Frontier Science Program (LT000798/2020). This work was supported by Wellcome Trust Investigator award 209397/Z/17/Z and by European Research Council Grant 787932 to KRF. **Author contributions:** Conceptualisation: KRF. Methodology: FS, EB, MTJ, EBNA, CFP, XW, LP, OC. Investigation: FS, EB, MTJ, OC. Visualisation: FS, EB, MTJ. Funding acquisition: KRF. Supervision: EB, OC, KRF. Writing – original draft: FS, EB, KRF. Writing – review and editing: FS, EB, MTJ, CFP, XW, LP, OC and KRF. **Competing interests:** KRF holds equity in Postbiotics Plus Research. **Data and materials availability:** All data used to generate the plots is available via DRYAD (*70*). All code is available via GitHub (*71*). Metagenomic sequencing and whole-genome sequencing of *E. coli* 0018 (NCBI accession number JAVXZX000000000) has been deposited in SRA under the BioProject PRJNA1021490. There is no restriction on use of the data, materials, or code, with the exception of *E. coli* strain 19Y000018 that is protected by an MTA which requires permission from the Nottingham University Hospitals Pathogen Bank (https://www.nuh.nhs.uk/pathogen-industry/). A Materials Design Analysis Reporting checklist is supplied with the publication.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by a Wellcome Trust Investigator award 209397/Z/17/Z, a cOAlition S organization.

The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

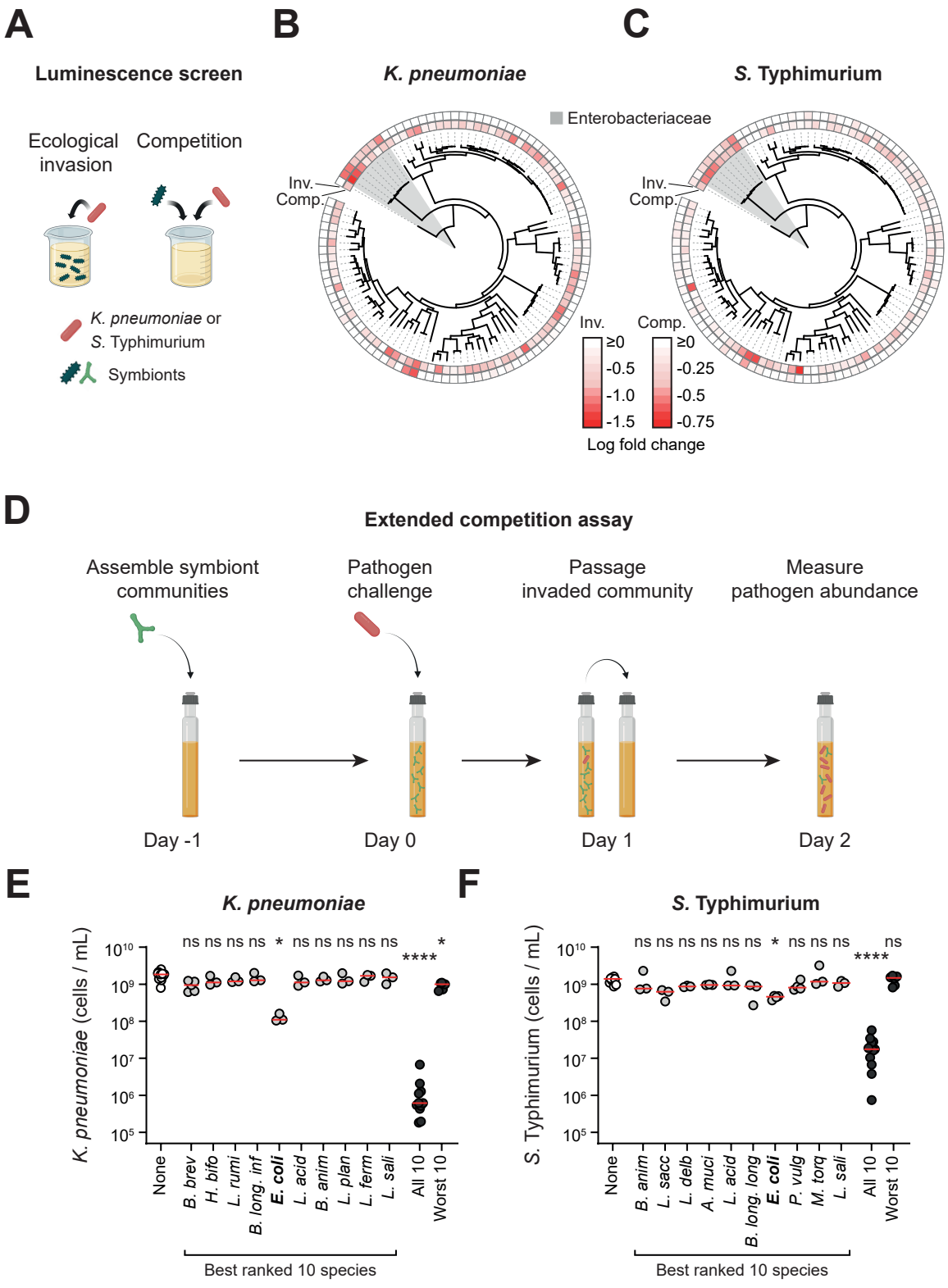
Supplementary Materials

Figs. S1 to S13

Tables S1 to S7

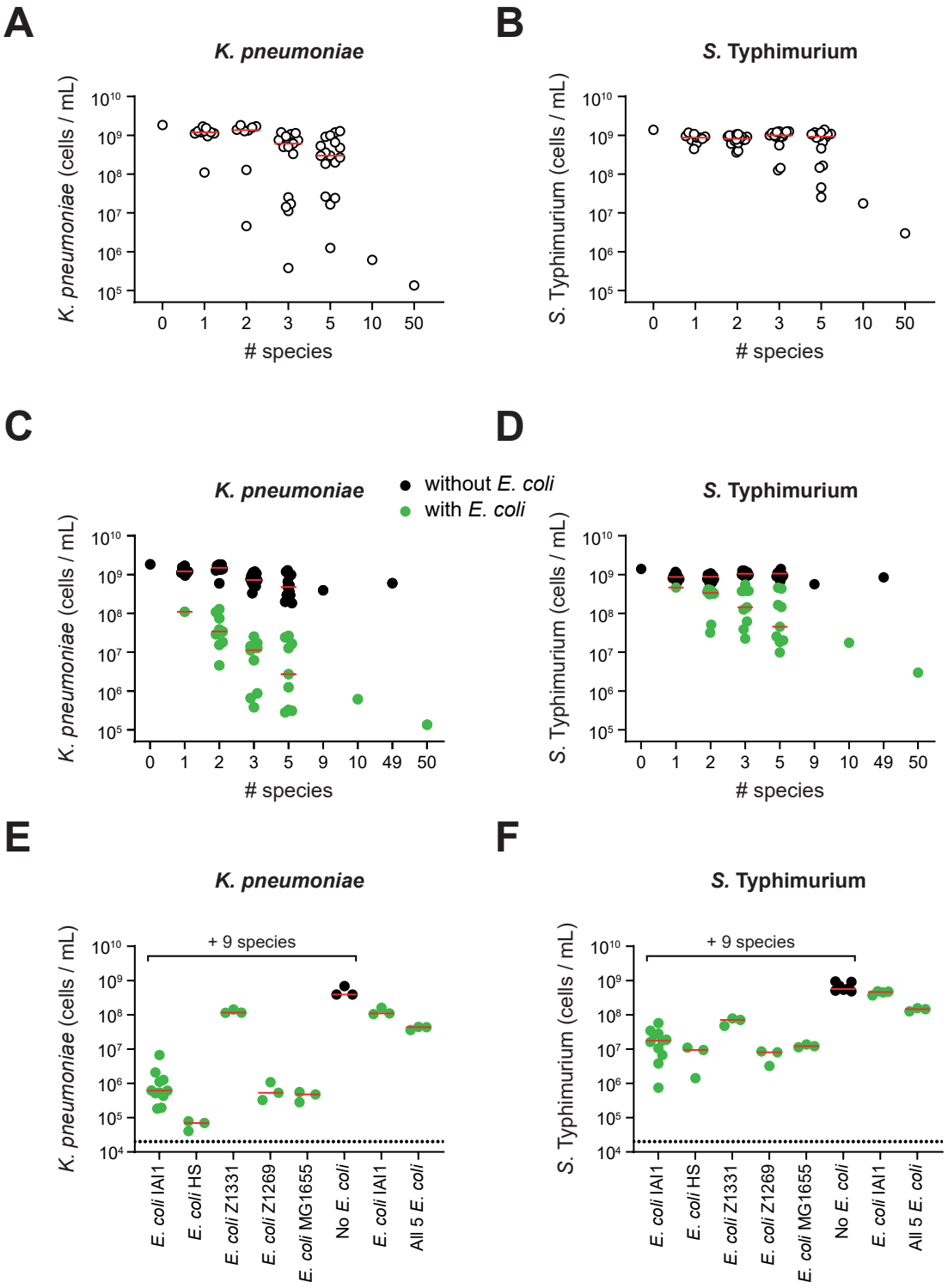
Supplementary references *(72) – (74)*

Main figures and captions


A
Luminescence screen
Ecological invasion
Competition
K. pneumoniae or S. Typhimurium
Symbionts
B
K. pneumoniae
Enterobacteriaceae
Inv.
Comp.
C
S. Typhimurium
Inv.
Comp.
Inv.
≥0
-0.5
-1.0
-1.5
Comp.
≥0
-0.25
-0.5
-0.75
Log fold change
D
Extended competition assay
Assemble symbiont communities
Pathogen challenge
Passage invaded community
Measure pathogen abundance
Day -1
Day 0
Day 1
Day 2
E
K. pneumoniae
K. pneumoniae (cells / mL)
ns ns ns ns * ns ns ns ns ns **** *
None
B. brev
H. bifo
L. rumi
B. long. inf
E. coli
L. acid
B. anim
L. plan
L. ferm
L. sali
All 10
Worst 10
Best ranked 10 species
F
S. Typhimurium
S. Typhimurium (cells / mL)
ns ns ns ns ns ns * ns ns ns **** ns
None
B. anim
L. sacc
L. delb
A. muci
L. acid
B. long. long
E. coli
P. vulg
M. torq
L. sali
All 10
Worst 10
Best ranked 10 species

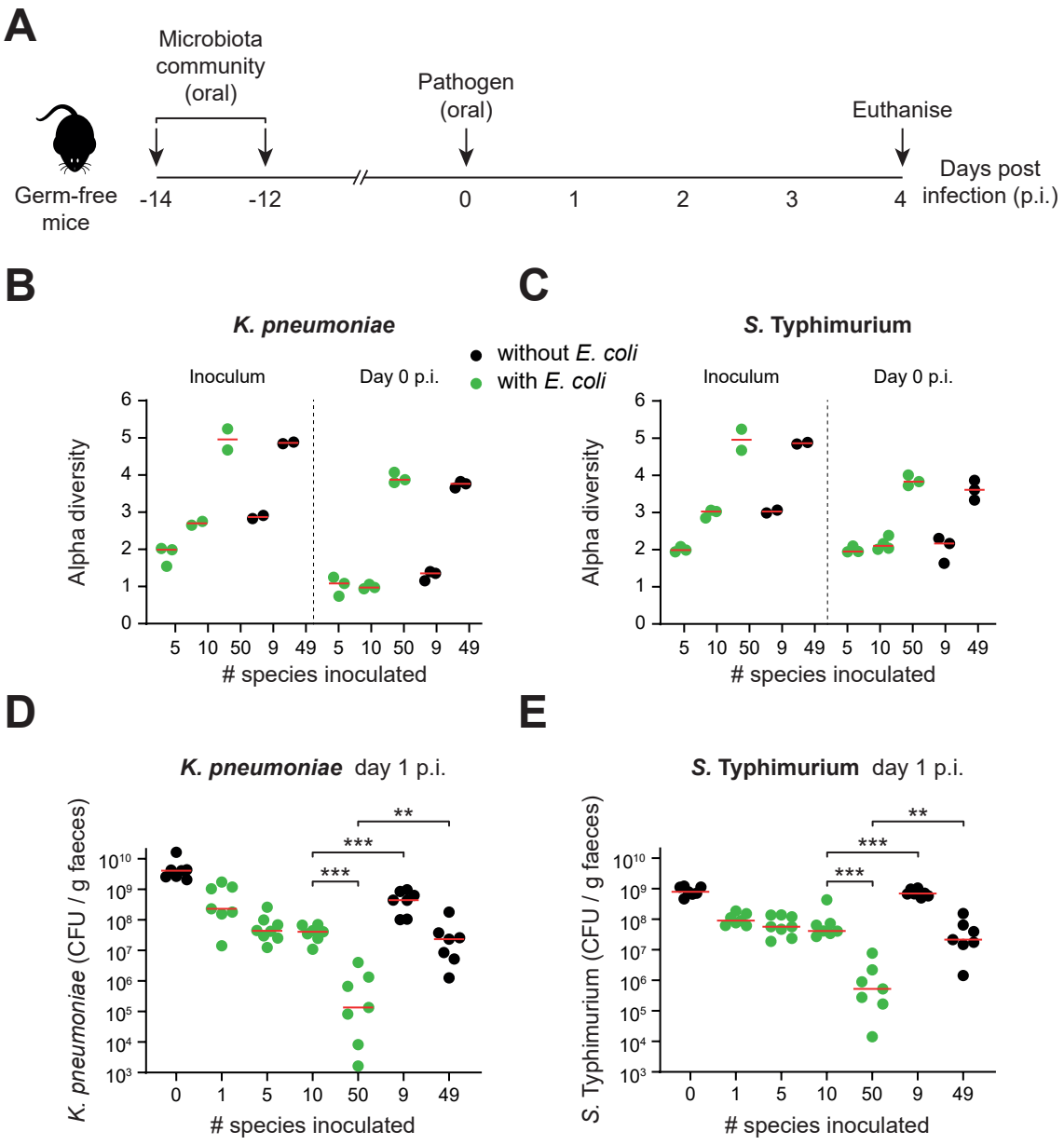
849

Figure 1. Single strains do not provide robust colonization resistance, but a diverse community can, depending on its composition. A) Overview of the luminescence co-culture assays. In the ecological invasion assay, *K. pneumoniae* or *S.* Typhimurium (red) was inoculated in co-culture with individual symbionts (different green symbols are used to represent the diversity of symbiont species screened; 19:1 ratio of symbiont to pathogen). In the competition assay, the symbionts were inoculated at an equal ratio to the pathogen to recapitulate competition between strains once the pathogen is established. In both assays, luminescence produced by the pathogen was used as a proxy for pathogen growth. Created with BioRender.com. **B-C)** Comparison of phylogenetic relatedness between symbionts, and the ability of each symbiont to compete with the pathogen (inv=ecological invasion assay; comp=competition assay). Data for *K. pneumoniae* shown in **(B)** and *S.* Typhimurium shown in **(C)**. The family Enterobacteriaceae, which includes both *K. pneumoniae* and *S.* Typhimurium, is shaded in grey. Luminescence fold change values are presented in **Fig. S1**. Data presented as the median luminescence log fold change of *N=3-10* independent experiments (biological replicates). Strains with the most negative (most red) values inhibited growth of the pathogen most strongly. **D)** Overview of the extended competition assay. Communities (or individual strains; green) of symbionts are pre-grown in anaerobic rich media before addition of the pathogen (red). The community is passaged after 24h of growth, followed by another 24h of growth before quantification with flow cytometry. Created with BioRender.com. **E-F)** The extended competition assay was performed for each individual species identified in the best ranked 10 species, as well as for combinations of 10 species (both the best- and worst-ranked 10 species; **Fig. S1**). Individual biological replicates from *N=3-15* independent experiments are shown. Red lines indicate the median. A Kruskal-Wallis test with Dunn's multiple test correction compares each group to the no-symbionts control ($p>0.05$=ns; $p<0.05$=*; $p<0.0001$=****). Data for *K. pneumoniae* shown in **(E)** and *S.* Typhimurium shown in **(F)**. See **Table S1** for species name abbreviations.


A
K. pneumoniae
K. pneumoniae (cells / mL)
species
B
S. Typhimurium
S. Typhimurium (cells / mL)
species
C
K. pneumoniae
without E. coli
with E. coli
K. pneumoniae (cells / mL)
species
D
S. Typhimurium
S. Typhimurium (cells / mL)
species
E
K. pneumoniae
+ 9 species
K. pneumoniae (cells / mL)
E. coli IAI1
E. coli HS
E. coli Z1331
E. coli Z1269
E. coli MG1655
No E. coli
E. coli IAI1
All 5 E. coli
F
S. Typhimurium
+ 9 species
S. Typhimurium (cells / mL)
E. coli IAI1
E. coli HS
E. coli Z1331
E. coli Z1269
E. coli MG1655
No E. coli
E. coli IAI1
All 5 E. coli

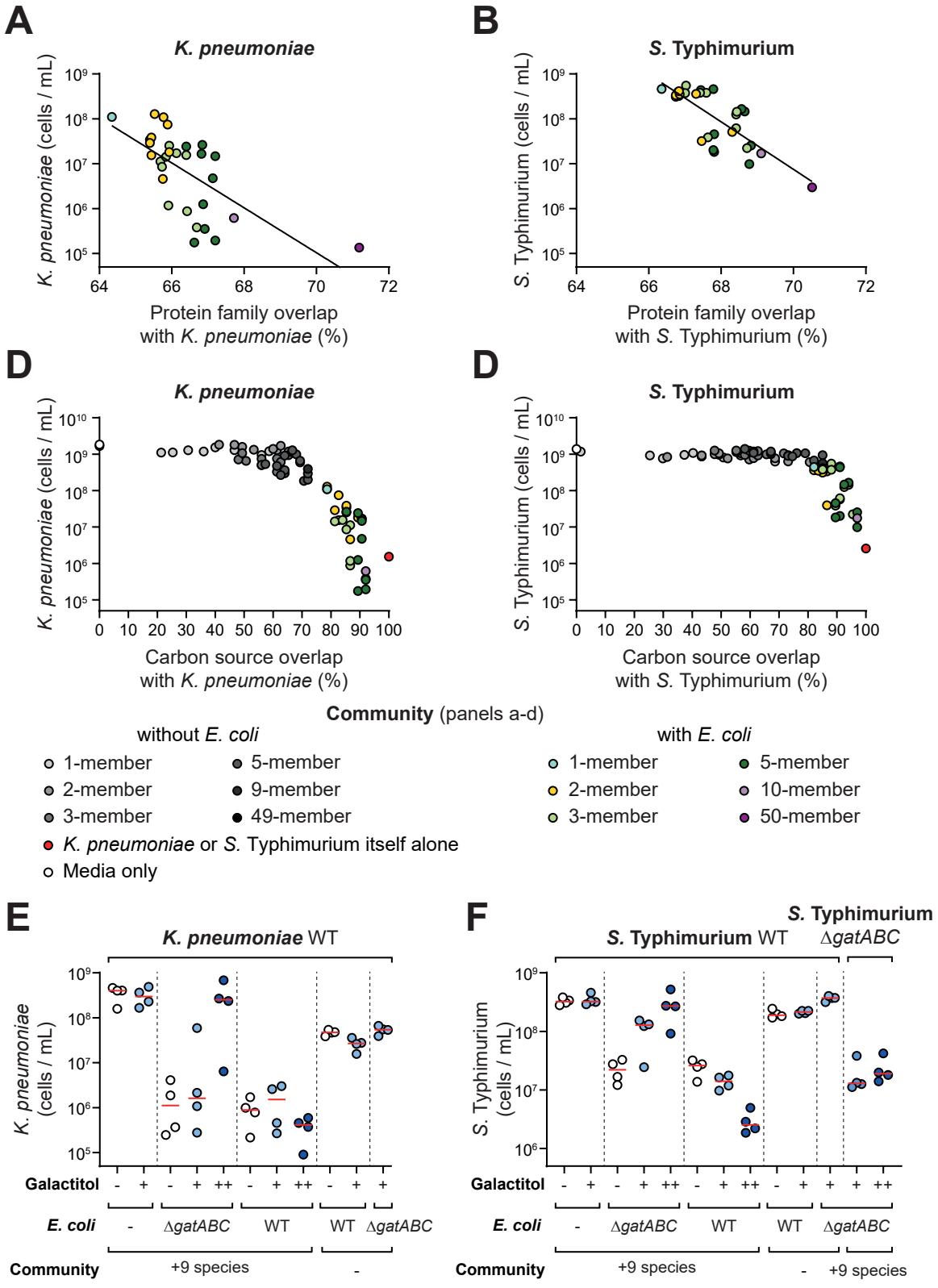
874

Figure 2. Ecological diversity and key members are needed for efficient colonization resistance *in vitro*. A-D) Extended competition assay on communities made up of an increasing number of species. Each data point represents the median pathogen cells/mL value on day 2 of the extended competition for a community (n=3-15 biological replicates from independent experiments for each community; up to 17 communities for each group). Communities with size <= 10 species were randomly selected from the 10 best ranked species for each pathogen. Community identities are shown in **Table S4-5**. Data for *K. pneumoniae* shown in **(A)** and **(C),** and for *S.* Typhimurium shown in **(B)** and **(D)**. Red lines indicate the median value of communities at a given diversity level. In **C-D)**, data from **A-B)** are replotted along with additional communities that always contained *E. coli* but were otherwise randomly selected. Communities without *E. coli* are depicted in black; communities with *E. coli* in green. Separate red median lines shown for communities with and without *E. coli*. A linear regression is performed on log-log transformed data in **Fig. S2a-b**, which shows that the association between diversity and colonization observed is statistically significant, and that this effect is greater for communities with *E. coli* than those without (F tests, $p \leq 0.0001$). **E-F)** Extended competition assay testing *E. coli* strains substituted into the best ranked 10 species community. Data for *K. pneumoniae* shown **(E)** and *S.* Typhimurium in **(F)**. Red lines indicate median values. Each data point represents a biological replicate each from independent experiments (*N=3-11*).


A
Microbiota community (oral)
Pathogen (oral)
Euthanise
Germ-free mice
-14
-12
0
1
2
3
4
Days post infection (p.i.)
B
K. pneumoniae
without E. coli
with E. coli
Inoculum
Day 0 p.i.
Alpha diversity
species inoculated
C
S. Typhimurium
D
K. pneumoniae day 1 p.i.
K. pneumoniae (CFU / g faeces)
E
S. Typhimurium day 1 p.i.
S. Typhimurium (CFU / g faeces)

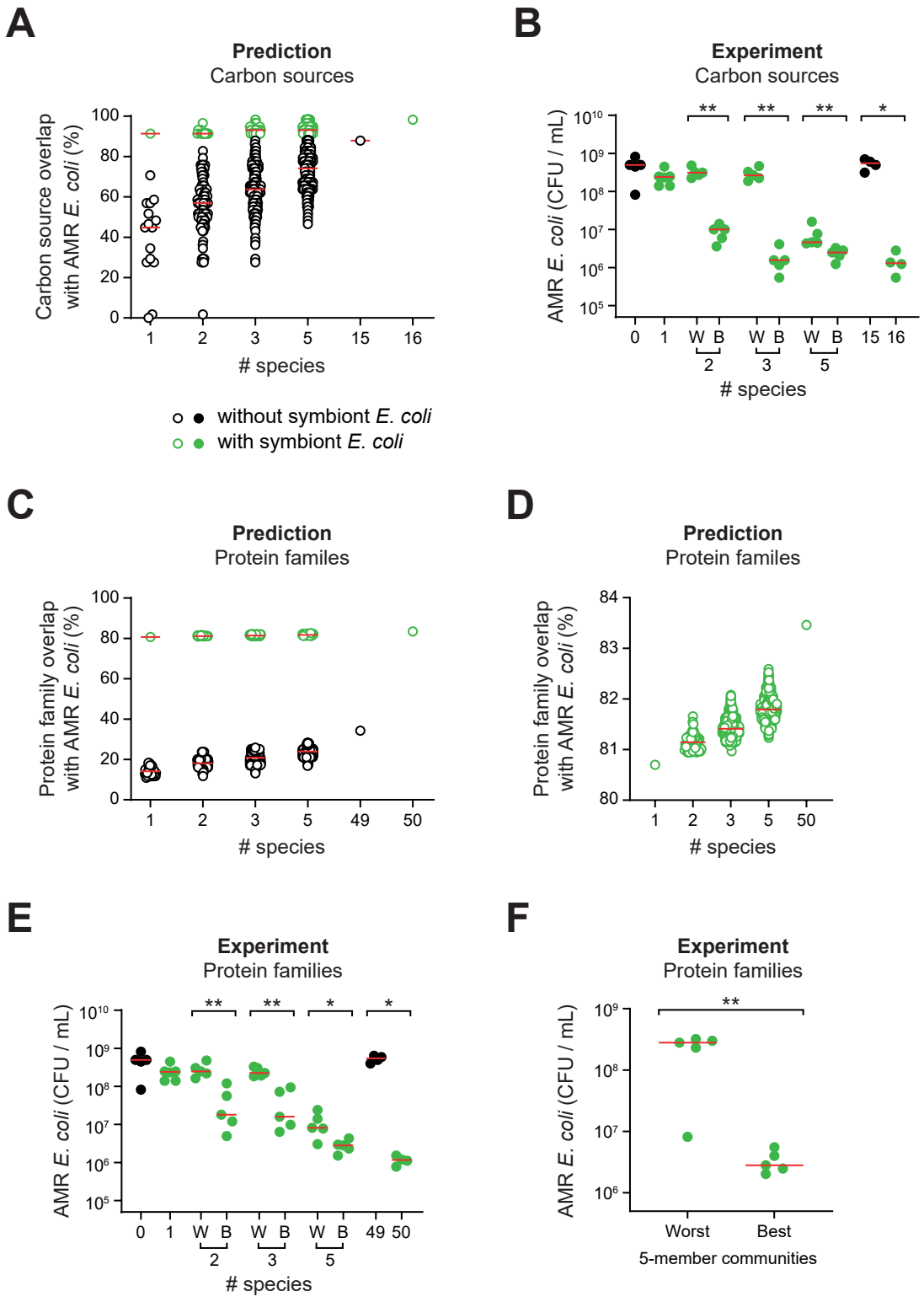
892

Figure 3. Ecological diversity and key members are needed for efficient colonization resistance *in vivo.* A) Overview of gnotobiotic mouse experiments. Symbiont communities (or *E. coli* alone) were given to germ-free mice by oral gavage twice (two days apart). 12 days later, the mice were challenged with *K. pneumoniae* or *S.* Typhimurium by oral gavage. Feces were collected from mice daily before being euthanised on day 4 post infection (p.i.). **B-C)** Alpha diversity measured by Shannon index of symbiont communities. Metagenomic sequencing was performed on the inoculum and fecal samples at day 0 (when the pathogen is introduced) and used to calculate diversity. Data for *K. pneumoniae* shown in (**B**) and *S.* Typhimurium in (**C**). Biological replicates from a representative mouse from each cage are shown (*N=2-4*; at least two independent experiments). **D-E)** Pathogen abundances in the feces of gnotobiotic mice colonized with communities of increasing diversity (mice containing communities with *E. coli* shown in green; mice containing communities without *E. coli* shown in black; *N=7-8* biological replicates of mice per group in cages of 2-3 mice; 2-3 independent experiments). Red lines indicate the medians. Two-tailed Mann-Whitney tests are used to compare the indicated groups ($p<0.01$=**; $p<0.001$=***). Data for *K. pneumoniae* shown in **(D)** and *S.* Typhimurium shown in **(E)**. Metagenomic analysis of strain diversity and relative abundance is shown in **Fig. S5**. Pathogen abundance data from days 1-4 p.i. is shown in **Fig. S6.** Community compositions are shown in **Table S6**.


A
K. pneumoniae
K. pneumoniae (cells / mL)
Protein family overlap with K. pneumoniae (%)
B
S. Typhimurium
S. Typhimurium (cells / mL)
Protein family overlap with S. Typhimurium (%)
D
K. pneumoniae
K. pneumoniae (cells / mL)
Carbon source overlap with K. pneumoniae (%)
D
S. Typhimurium
S. Typhimurium (cells / mL)
Carbon source overlap with S. Typhimurium (%)
Community (panels a-d)
without E. coli
1-member
2-member
3-member
5-member
9-member
49-member
with E. coli
1-member
2-member
3-member
5-member
10-member
50-member
K. pneumoniae or S. Typhimurium itself alone
Media only
E
K. pneumoniae WT
K. pneumoniae (cells / mL)
Galactitol
E. coli
ΔgatABC
WT
Community
+9 species
F
S. Typhimurium
S. Typhimurium WT
ΔgatABC
S. Typhimurium (cells / mL)
Galactitol
E. coli
ΔgatABC
WT
Community
+9 species
+9 species

910

Figure 4. Nutrient overlap can explain the role of ecological diversity and the effect of *E. coli* in colonization resistance. A-B) Protein family overlap is compared to the median pathogen abundance values for each community containing *E. coli* from **Fig. 2C-D**. Diversity is visualised by a color gradient**.** Data for *K. pneumoniae* shown in **(A)** and *Salmonella* shown in **(B)**. A line of best fit is shown from a linear regression on log transformed data: $R^2 = 0.4255$ for *K. pneumoniae*; $R^2 = 0.603$ for *S.* Typhimurium; both slopes are significantly different from 0 using an F-test ($p<0.0001$). Data for communities without *E. coli* is presented in **Fig. S9E-F**. **C-D**. Overlap in carbon source utilisation plotted against the median pathogen abundance measurements from experimental communities in **Fig. 2C-D**. Community carbon source overlap is calculated using an additive approach from carbon source overlap of individual strains by measurement on AN Biolog Microplates (**Fig. S10**). Diversity is visualised by a gradient of color (for *E. coli*-containing communities) or greyscale (for communities without *E. coli*). A control with the isogenic pathogen itself (100% overlap) is plotted in red. Data for *K. pneumoniae* in **(C)** and *S.* Typhimurium shown in **(D)**. **E-F)** A private nutrient, galactitol, that could only be used by the WT *E. coli* strain and the pathogens but not by the other symbionts nor an *E. coli* Δ*gatABC* mutant, was supplemented to the media and the extended competition assay performed as before. In all treatments, pathogen abundance was measured by flow cytometry after 48h of growth post-passage instead of the usual 24h. This change did not influence the control experiments without galactitol, but proved informative because we found the growth impacts of galactitol were relatively slow. Results for *K. pneumoniae* shown in **(E)** and *S.* Typhimurium in **(F).** *N=3-4* biological replicates from independent experiments per treatment. Horizontal red lines show the median of the replicates. Light blue circles show results with 0.1% galactitol supplementation (+ symbol), dark blue circles show results with 1% galactitol supplementation (++ symbol). White circles (control) show results with no nutrient supplementation (- symbol). 9 species refers to the 9 additional species in the 10 best-performing species for each respective pathogen (- symbol refers to when *E. coli* is added alone). In (**D**), a Δ*gatABC* mutant of *S.* Typhimurium was used in addition to the WT pathogen to verify the dependency of colonization on a private nutrient.


A
Prediction
Carbon sources
Carbon source overlap with AMR E. coli (%)
species
without symbiont E. coli
with symbiont E. coli
B
Experiment
Carbon sources
AMR E. coli (CFU / mL)
species
C
Prediction
Protein familes
Protein family overlap with AMR E. coli (%)
species
D
Prediction
Protein familes
Protein family overlap with AMR E. coli (%)
species
E
Experiment
Protein families
AMR E. coli (CFU / mL)
species
F
Experiment
Protein families
AMR E. coli (CFU / mL)
Worst
Best
5-member communities

938

Figure 5. Nutrient blocking predicts community colonization resistance. A) *In silico* prediction of carbon source overlap with the AMR *E. coli* strain for all possible combinations of symbiont communities at the indicated diversity levels. Each circle represents a different community. Communities containing the symbiont *E. coli* IAI1 are shown in green and communities without *E. coli* IAI1 are shown in black (predictions as hollow circles; experimental data as solid circles). Predicted carbon source overlap calculated using an additive approach from carbon source use of individual strains measured using AN Biolog MicroPlates (**Fig. S10, S13**). **B)** Experimental test of *in silico* predictions in **(A).** The two *E. coli* IAI1-containing communities predicted to have the best (B) and worst (W) carbon source overlap were picked at each diversity level and competed against AMR *E. coli* in the extended competition assay. A two-tailed Mann-Whitney U test was performed on community pairs ($p>0.05$=ns; $p<0.05$=*; $p<0.01$=**) at the 2-, 3- and 5-strain diversity levels. Red horizontal bars depict the median of each community tested. *N=4-5* biological replicates from independent experiments for each community. **C-D)** *In silico* prediction of protein family overlap with the AMR *E. coli* strain for a random subset (n=59,043) of all possible symbiont communities at diversity levels 2-, 3-, and 5-strains, as well as all individual strains and the 49- and 50-species communities. Each circle represents a different community. Communities are selected from the strains comprising the 50-member community. Communities containing *E. coli* IAI1 are shown in green and communities without *E. coli* IAI1 are shown in black. **D)** Only the *E. coli*-containing communities are plotted to better visualise variation in protein family overlap. **E)** Experimental test of *in silico* predictions based on protein cluster overlap in **(C-D)**. The two *E. coli* IAI1-containing communities predicted to have the best (B) and worst (W) protein family overlap were picked at each diversity level (randomly selected, for cases where there were multiple communities with the same overlap), and competed against AMR *E. coli* in the extended competition assay. Red horizontal bars depict the median of each community tested. N=5 biological replicates from independent experiments for each community. A two-tailed Mann-Whitney U test was performed on community pairs ($p>0.05$=ns; $p<0.05$=*; $p<0.01$=**) at the 2-, 3- and 5-strain diversity levels. **F)** Experimental test of the predicted 5 best and 5 worst communities at the 5-strain diversity level, based on protein family overlap with

AMR *E. coli*. Each symbol represents the median of N=5 biological replicates from independent experiments per community. Red horizontal bars depict the median of the best and the worst predicted communities. A two-tailed Mann-Whitney U test was performed ($p<0.01$=**). Community identities for (**B, E-F**) are shown in **Table S7**.

Supplementary Materials for

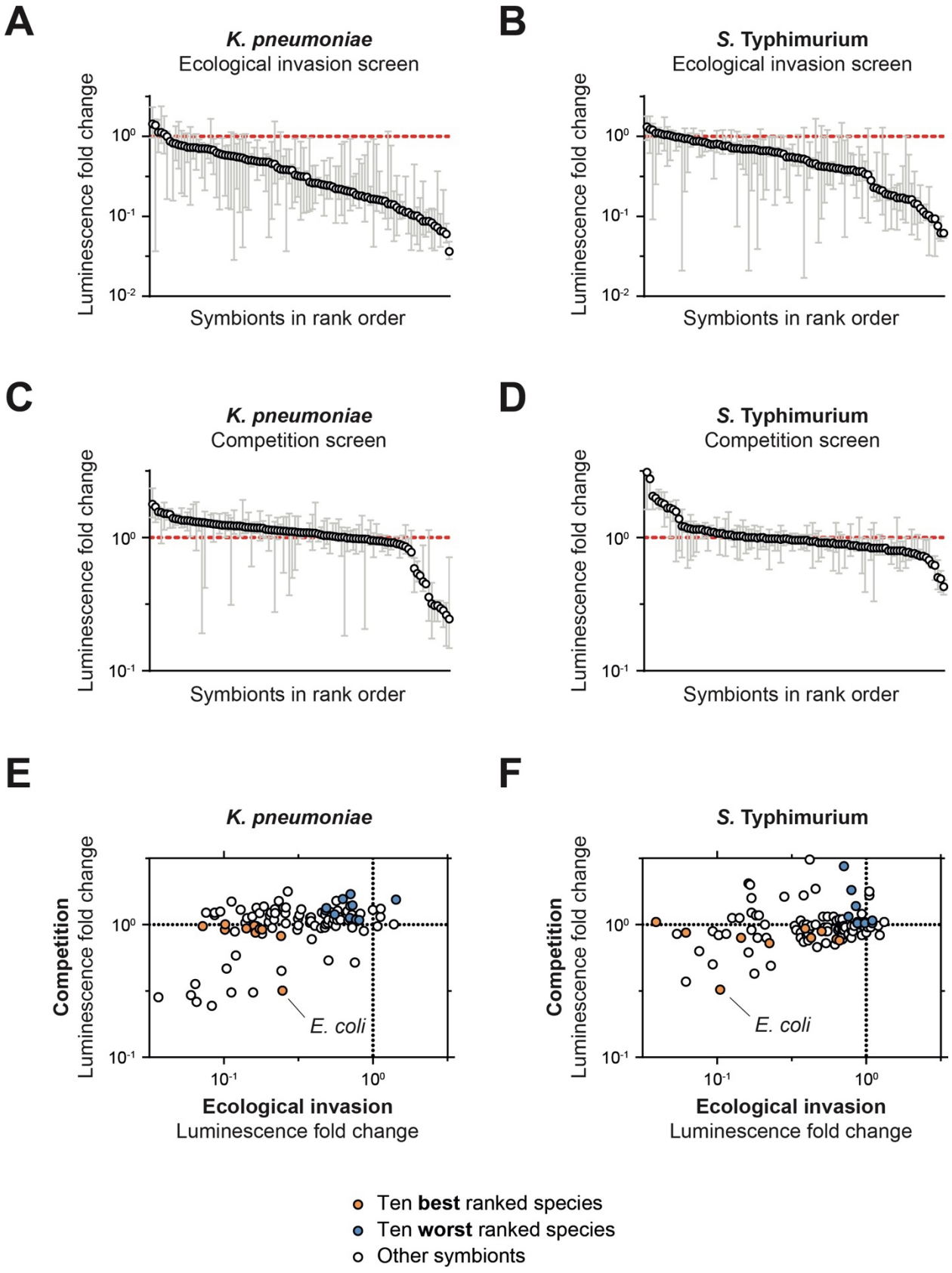
Microbiome diversity protects against pathogens by nutrient blocking

Frances Spragge+ and Erik Bakkeren+, Martin T. Jahn, Elizete B. N. Araujo, Claire F. Pearson, Xuedan Wang, Louise Pankhurst, Olivier Cunrath* and Kevin R. Foster*

*Correspondence to: kevin.foster@biology.ox.ac.uk; olivier.cunrath@unistra.fr; +These authors contributed equally to this work.

This PDF file includes:

Figs. S1 to S13
Tables S1 to S7
Supplementary references *(72) – (74)*



993

994 **Fig. S1.**

Human gut symbiont strains vary in their ability to inhibit growth of *K. pneumoniae* and *S.* Typhimurium in the luminescence screen. **A-D)** Waterfall plots show the luminescence median fold change (log ratio of the control luminescence divided by the treatment luminescence) of the pathogen in combination with each human gut symbiont tested in the luminescence screen. Results for *K. pneumoniae* are shown in **(A)** and **(C)** and for *S.* Typhimurium shown in **(B)** and **(D).** Results of the ecological invasion assay are shown in **(A)** and **(B)** and the competition assay in **(C)** and **(D).** Black circles represent the median value for each symbiont (*N*=3-10 biological replicates from independent experiments). Grey vertical lines represent range bars. **E-F)** Correlation of the results of the ecological invasion and competition assays of the luminescence screen. Each circle represents the median value for a symbiont (*N*=3-10 biological replicates from independent experiments). Results for *K. pneumoniae* shown in **(E)** and *S.* Typhimurium in **(F).** The sum of the ranks of the competition and the ecological invasion assays were used to create an overall ranking of symbionts for each pathogen. The best 10 ranked species are shown in orange and the worst 10 shown in blue, with the added criteria that they have a category 1 safety level (see Methods). Strains with the most negative competition and invasion values inhibited the growth of the pathogen most strongly compared to the media-only control.

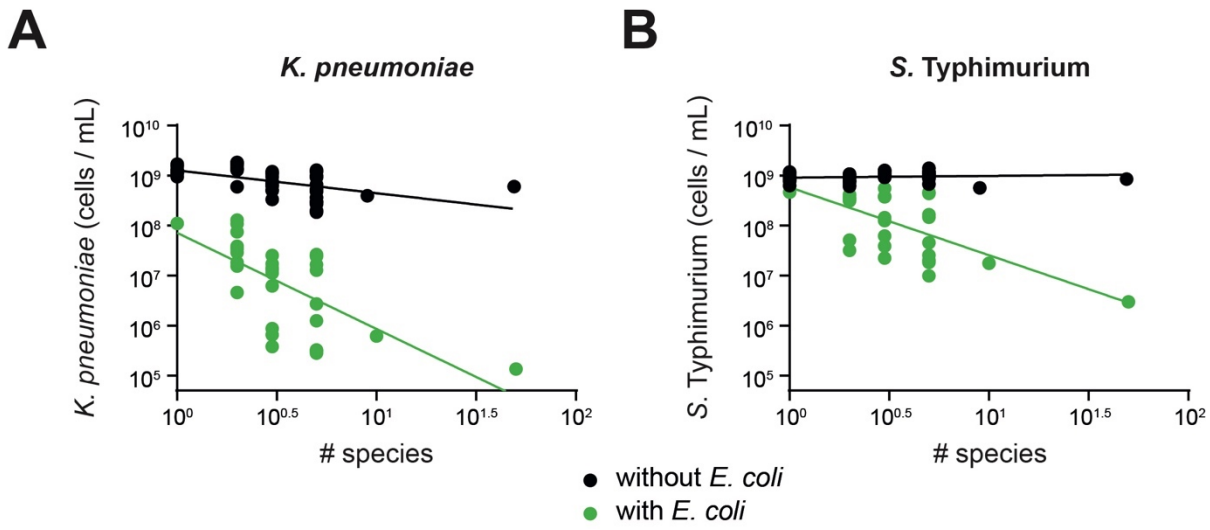


Fig. S2.

Community diversity negatively correlates with pathogen abundance. A-B) As the number of symbiont strains increases, pathogen density on day 2 of the extended competition decreases (x axis on a log scale). Data for *K. pneumoniae* shown in **(A)** and *S.* Typhimurium in **(B).** Each circle represents the median value of a community tested in the extended competition assay (data from **Fig. 2C-D**). Communities in green contain *E. coli*, communities in black do not contain *E. coli*. Linear regression of log-transformed data: **(A)** R^2=0.4296, non-zero slope for *E. coli* communities (F test, p<0.0001). R^2=0.3103, non-zero slope for communities without *E. coli* (F test, p<0.0001). Moreover, slopes of the two regressions are significantly different from each other (F test, p=0.0001). **(B)** R^2=0.4234, non-zero slope for *E. coli* communities (F test, p<0.0001). R^2=0.01378, slope not different to zero for communities without *E. coli* (F test, p=0.4218). Slopes of the two regressions are again significantly different from each other (F test, p<0.0001).

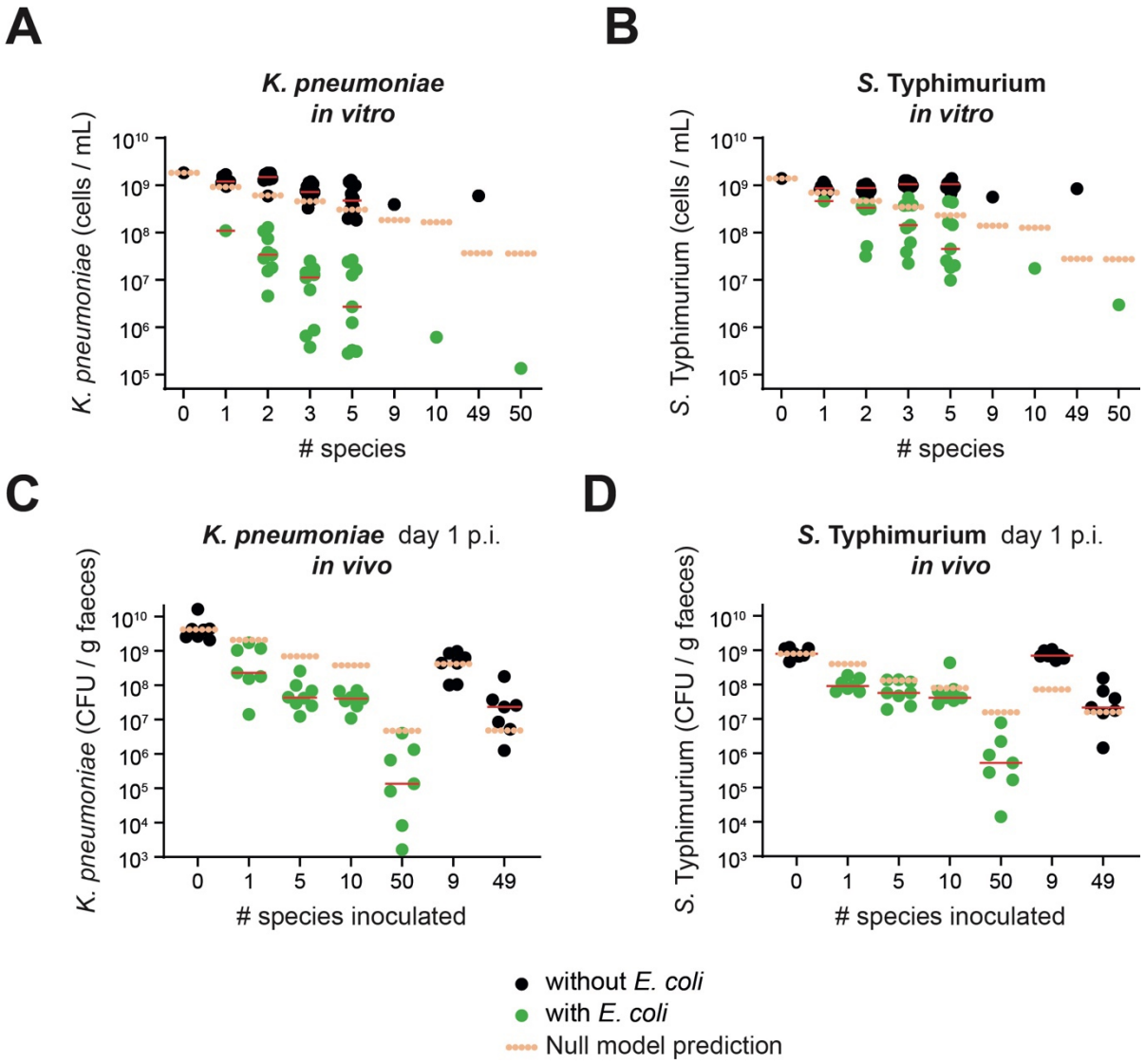


Fig. S3.

Comparison to a null model where the effect of each species proportionally restricts pathogen growth in an additive manner. We compared our experimental data to a null model where the effect of colonization resistance simply scales according to the number of species in the community. Specifically, we took the abundance of the pathogen at the end of the experiment when the pathogen is alone (ie, n=0) and multiplied it by 1/(n) where n refers to the number of species that contribute to the overall carrying capacity of the system (including the pathogen). This value is plotted in as beige dotted lines for both *in vitro* data from **Fig. 2C-D** (panels A and B; dots indicate median values of each community) and *in vivo* data from **Fig. 3D-E** (panels C and D; dots indicate individual mice). **A,C**) Data for *K. pneumoniae*, **B,D**) Data for *S.* Typhimurium. Communities containing *E. coli* are shown in green and communities without *E. coli* are shown in black. Red lines indicate medians. In all cases, the deviation between the null model and data increases at higher diversities, so long *E. coli* is present in the high diversity communities. This effect is also statistically significant in all cases: we compare the ratio of predicted to observed pathogen abundance for 1 species data to that from the 50 species case (Two-tailed Mann-Whitney U tests; p=0.0070 for **panel A**; p=0.0070 for **panel B**; p=0.0041 for **panel C**; p=0.0262 for **panel D**).

1047

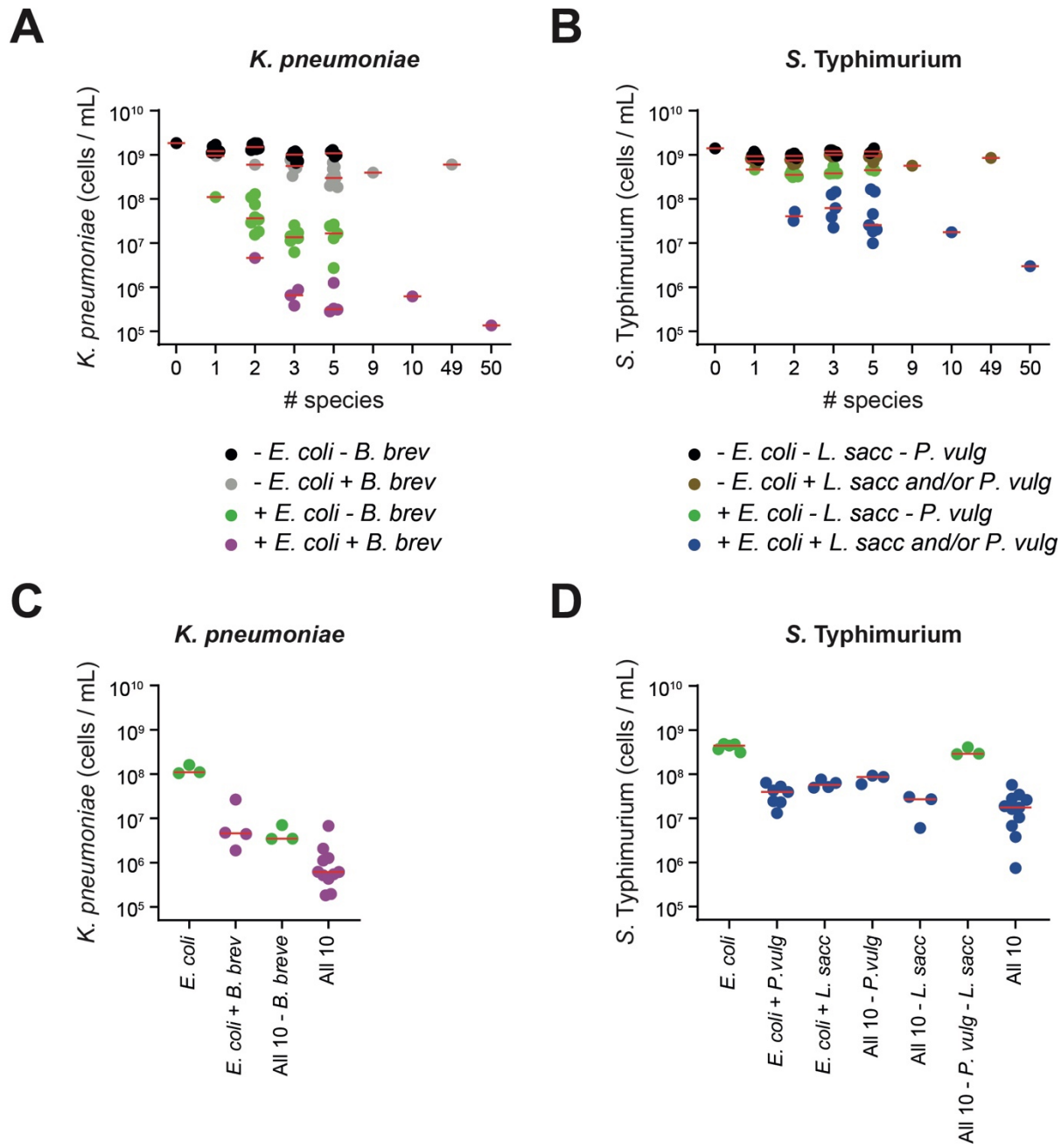
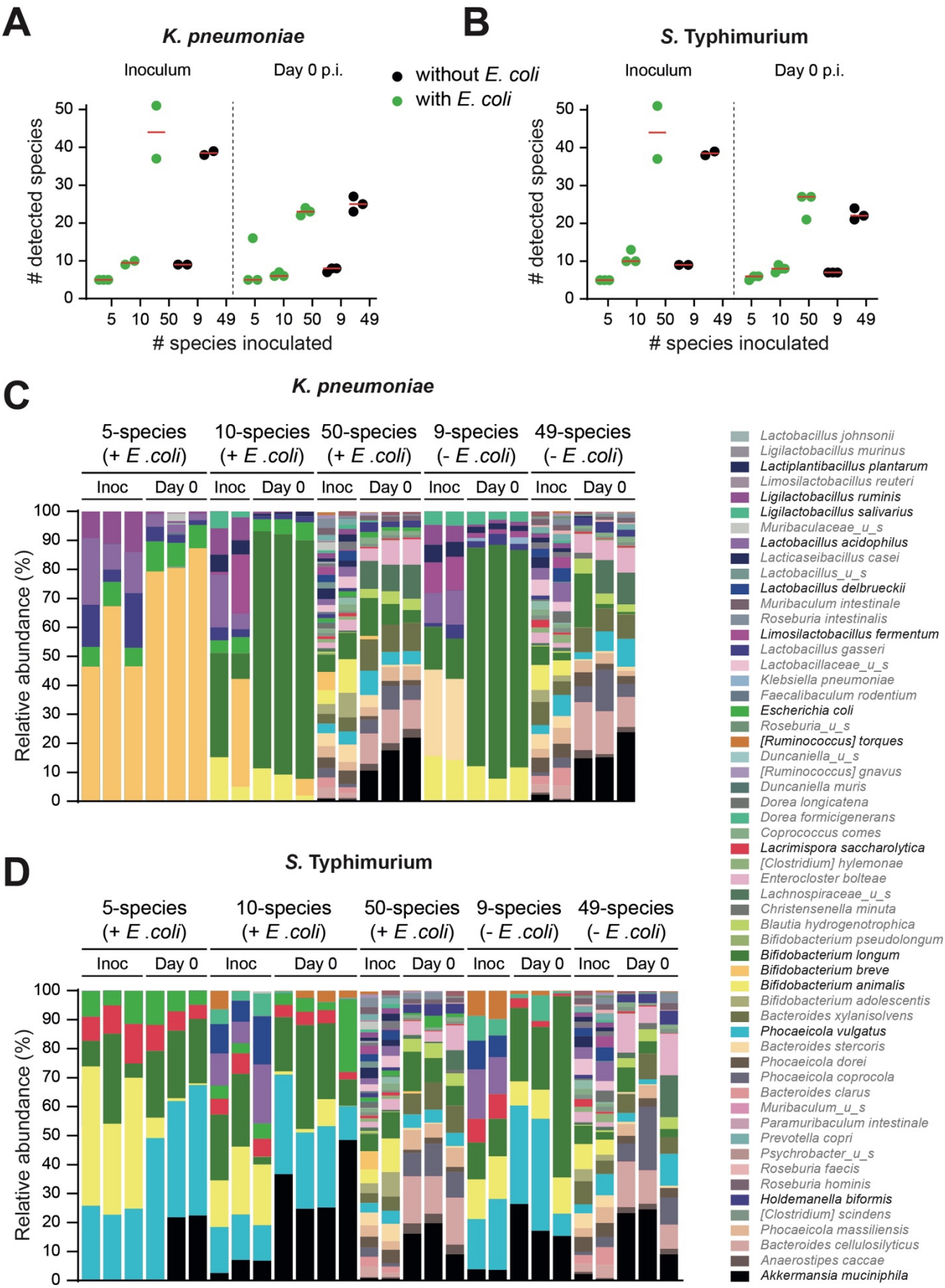


Fig. S4.

Combinations of multiple species are important for colonization resistance to each pathogen. A-B) Equivalent figures to **Fig. 2C-D**, except the presence or absence of *B. breve* within communities are shown in **(A)** (grey for communities with *B. breve* but without *E. coli* and purple for communities with *B. breve* and *E. coli*), and the presence or absence of *L. saccharolyticum* and/or *P. vulgatus* shown in **(B)** (brown for communities with *L. saccharolyticum* and/or *P. vulgatus* but not *E. coli,* and blue for communities with *L. saccharolyticum* and/or *P. vulgatus* and *E. coli*). Results for *K. pneumoniae* are shown in **(A)** and for *S.* Typhimurium shown in **(B).** In **(A-B)**, "+" and "-" refer to the presence or absence of species rather than the addition or subtraction of a species. Horizontal red lines depict the median of the communities at a particular diversity level containing the species indicated in the legend. Each circle represents the median pathogen abundance measured for a community on day 2 of the extended competition assay. **C-D)** Drop-out experiments verify the context-

dependent effect of key members on colonization resistance. *N*=3-11 biological replicates from independent experiments. In **(C-D)**, "+" and "-" refer to the addition or subtraction of a species. Results for *K. pneumoniae* are shown in **(C)** and for *S.* Typhimurium shown in **(D).** Horizontal red lines depict the median of the replicates for a particular community. See **Table S1** for species name abbreviations.


A
K. pneumoniae
Inoculum
Day 0 p.i.
without E. coli
with E. coli
detected species
species inoculated
B
S. Typhimurium
C
K. pneumoniae
5-species (+ E .coli)
10-species (+ E .coli)
50-species (+ E .coli)
9-species (- E .coli)
49-species (- E .coli)
Inoc
Day 0
Relative abundance (%)
D
S. Typhimurium
Lactobacillus johnsonii
Ligilactobacillus murinus
Lactiplantibacillus plantarum
Limosilactobacillus reuteri
Ligilactobacillus ruminis
Ligilactobacillus salivarius
Muribaculaceae_u_s
Lactobacillus acidophilus
Lacticaseibacillus casei
Lactobacillus_u_s
Lactobacillus delbrueckii
Muribaculum intestinale
Roseburia intestinalis
Limosilactobacillus fermentum
Lactobacillus gasseri
Lactobacillaceae_u_s
Klebsiella pneumoniae
Faecalibaculum rodentium
Escherichia coli
Roseburia_u_s
[Ruminococcus] torques
Duncaniella_u_s
[Ruminococcus] gnavus
Duncaniella muris
Dorea longicatena
Dorea formicigenerans
Coprococcus comes
Lacrimispora saccharolytica
[Clostridium] hylemonae
Enterocloster bolteae
Lachnospiraceae_u_s
Christensenella minuta
Blautia hydrogenotrophica
Bifidobacterium pseudolongum
Bifidobacterium longum
Bifidobacterium breve
Bifidobacterium animalis
Bifidobacterium adolescentis
Bacteroides xylanisolvens
Phocaeicola vulgatus
Bacteroides stercoris
Phocaeicola dorei
Phocaeicola coprocola
Bacteroides clarus
Muribaculum_u_s
Paramuribaculum intestinale
Prevotella copri
Psychrobacter_u_s
Roseburia faecis
Roseburia hominis
Holdemanella biformis
[Clostridium] scindens
Phocaeicola massiliensis
Bacteroides cellulosilyticus
Anaerostipes caccae
Akkermansia muciniphila

1068

Fig. S5.

Metagenomic sequencing shows that germ-free mice gavaged with more diverse communities were colonized with a higher number of bacterial strains. A-B) Number of detected bacterial species **(**above a relative abundance threshold of 0.1%) in the inoculum given to the mice and in mouse feces 14 days after the first inoculum gavage (Day 0 post infection; p.i.). Mice were given 2 identical gavages containing symbiont communities 2 days apart; the first inoculum was sequenced as a representative. Each inoculum data point depicts an independent experiment and each day 0 data point indicates a representative mouse from each cage. Horizontal red lines represent median values of the replicates at each diversity level. (*N*=2-3 for the inoculum, *N*=3-4 for fecal samples; data indicates biological replicates of a representative mouse from each cage, derived from at least two independent experiments). Green symbols represent communities that contain *E. coli*, black symbols represent communities without *E. coli*. **C-D)** Relative abundance plot of symbiont strains in the inoculum and mouse feces using metagenomic sequencing data. Data for mice challenged with *K. pneumoniae* shown in **(C)** and *S.* Typhimurium in **(D).** The 10 best ranked strains from the luminescence screen for each pathogen are shown in black writing (other detected species in grey writing).

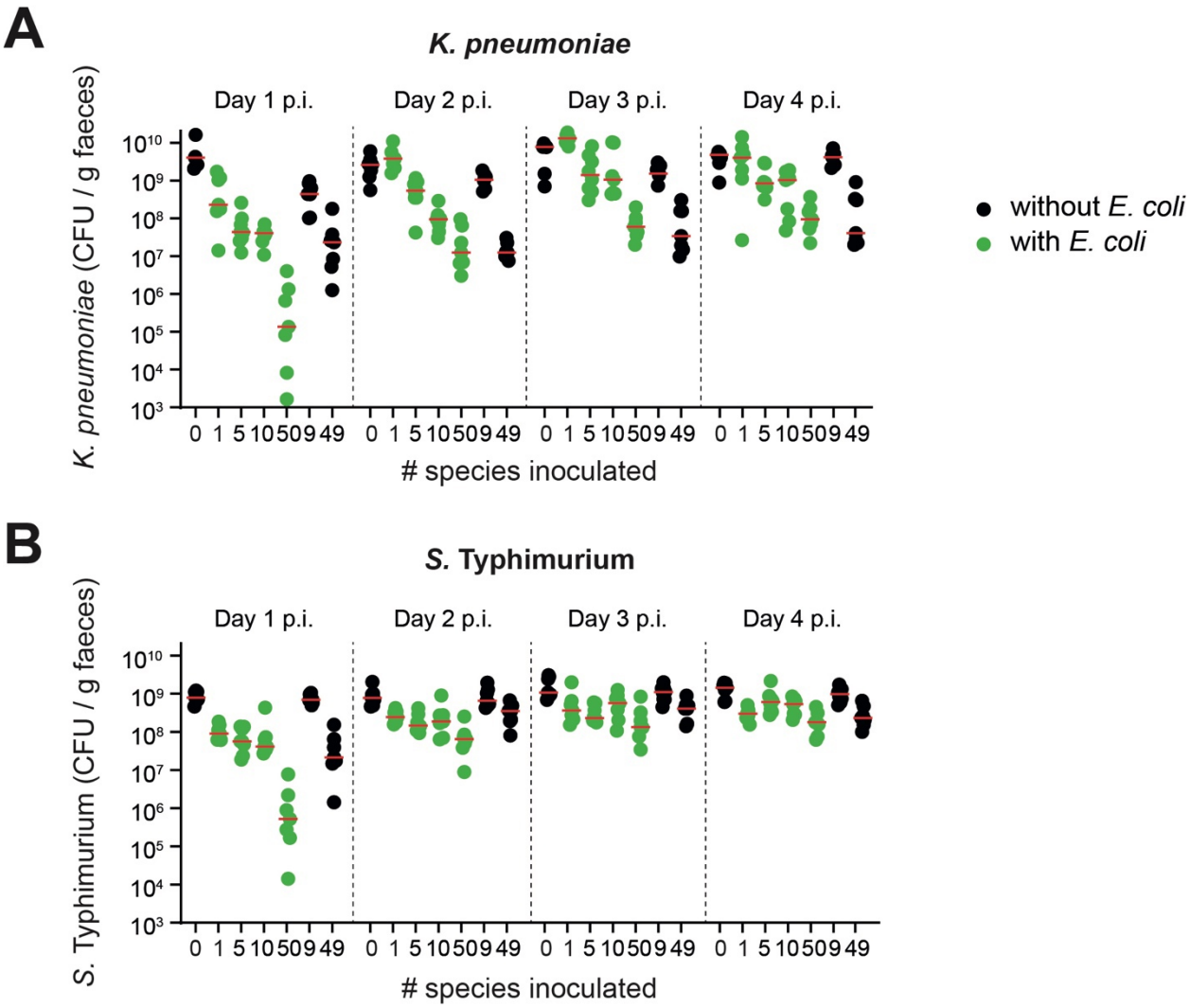


Fig. S6.

Pathogen abundance *in vivo* at later timepoints. A-B) Mice were gavaged with *K. pneumoniae* **(A)** or *S.* Typhimurium **(B)** on Day 0. Each symbol represents a fecal sample from 1 mouse. Pathogen abundances were determined by selective plating aerobically on LB agar + carbenicillin (*K. pneumoniae*) or LB agar + streptomycin (*S.* Typhimurium). Horizontal red lines represent median values of the replicate mice tested at each diversity level. Communities containing *E. coli* are shown in green whereas communities without *E. coli* are in black. *N*=7-8 biological replicates of mice per group in cages of 2-3 mice; 2-3 independent experiments. The day 1 post infection (p.i.) data is the same as in **Fig. 3D-E**.

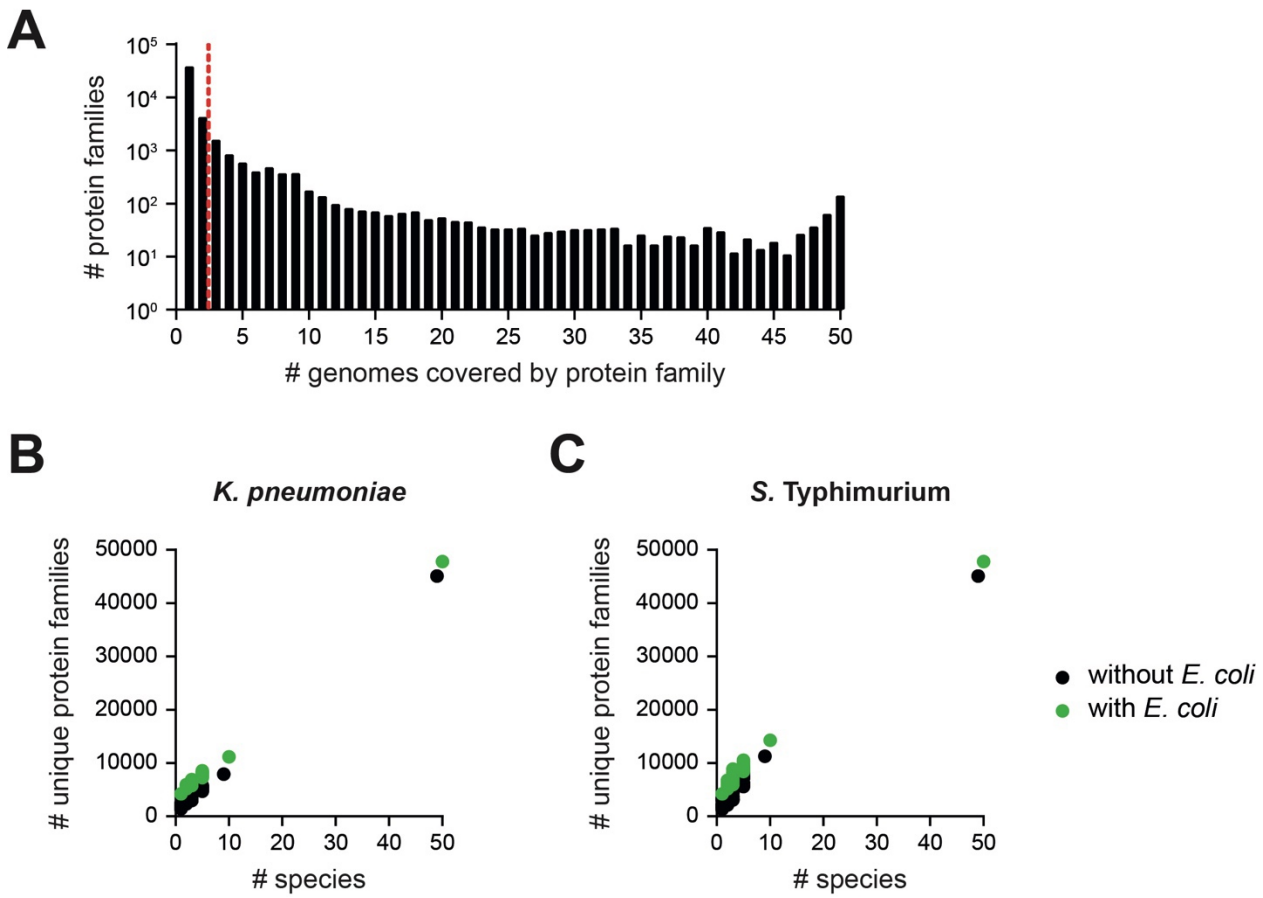


Fig. S7.

The number of protein families increase proportional to community diversity. A) Histogram showing the distribution of protein families among the 50 species subset used for protein family analysis. The vertical red dashed line represents the average number of genomes out of the 50 strains that share a particular protein family (2.44 genomes). There is an average of 3.22 proteins in each protein family. The histogram shows that many protein families are unique to a strain while others (141) are shared between all 50 strains. **B-C)** The number of protein families covered by a community increases as the number of strains in the community increases. Results for *K. pneumoniae* shown in **(B)** and for *S.* Typhimurium in **(C).** Community IDs taken from **Fig. 2C-D**. Each circle represents a different community. Green circles depict communities containing *E. coli*, while black circles are communities without *E. coli*.

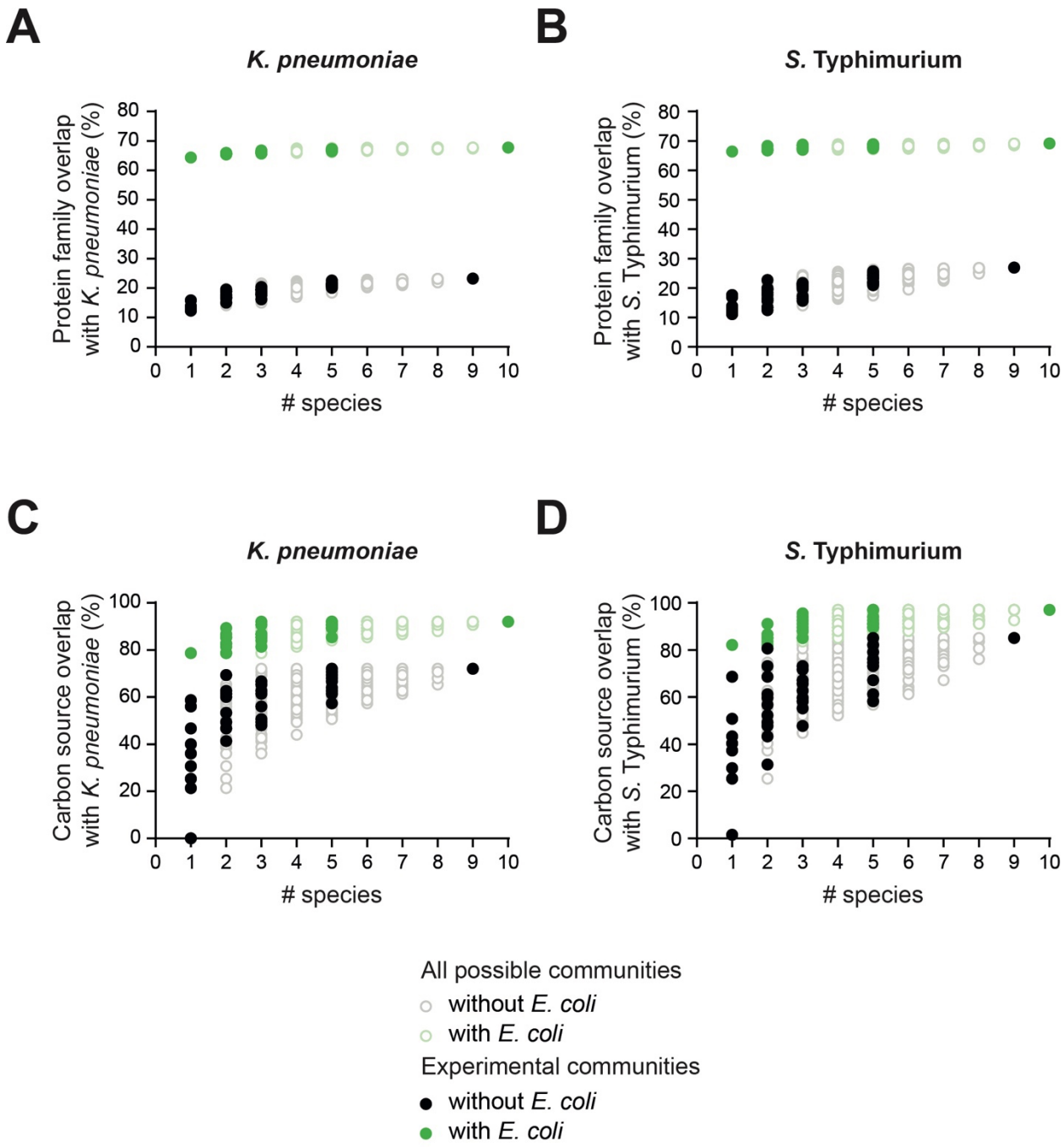
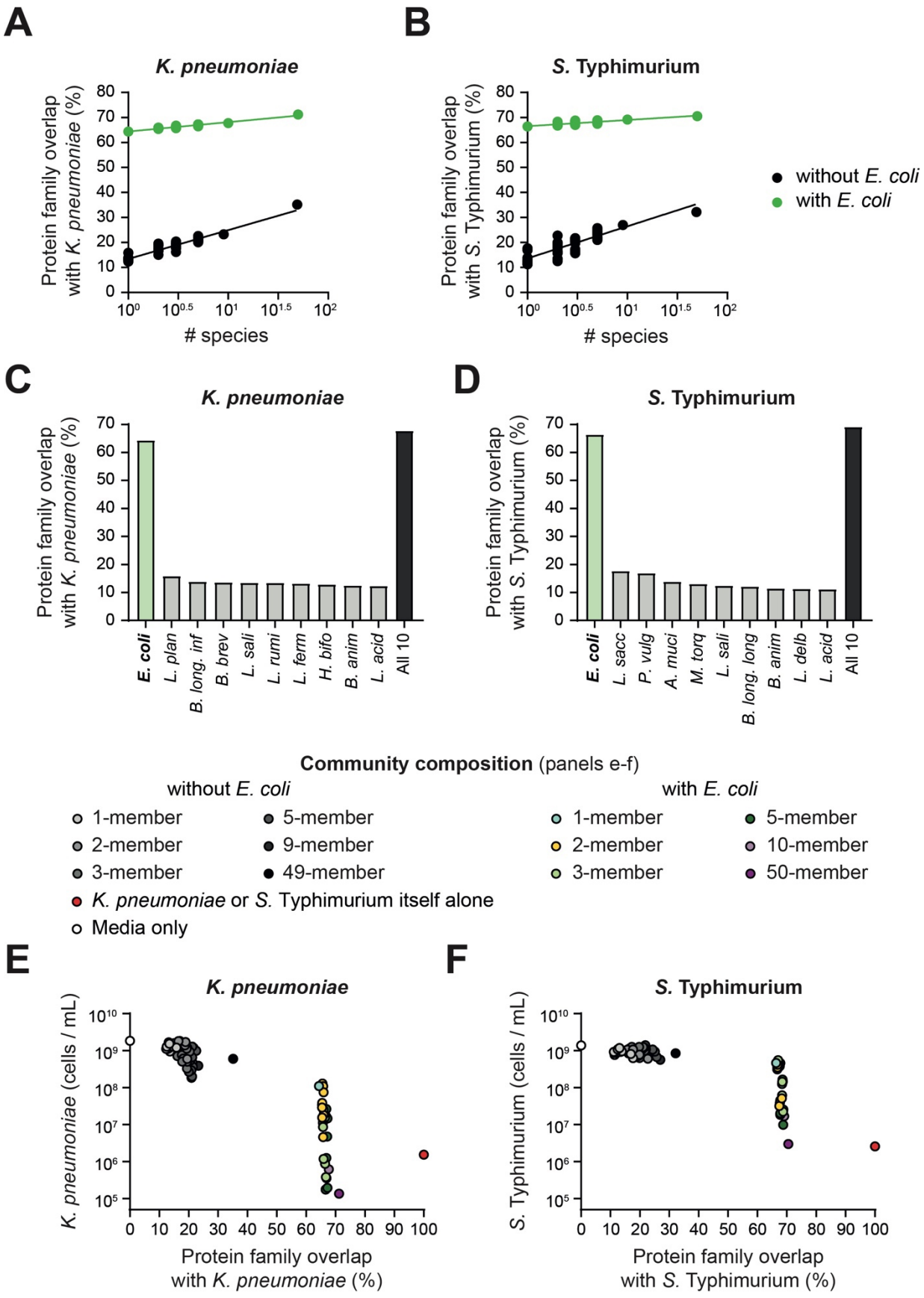


Fig. S8.

The randomly chosen communities used in the *in vitro* experiments are representative of all possible combinations of the 10 best-ranked species. A-B) Experimental communities contain representative protein family overlap to the pathogens compared to all possible combinations of the 10 best-ranked species. All possible combinations of communities are depicted by unfilled circles, experimentally tested communities are shown as filled circles. Communities in green contain *E. coli* while communities in black do not contain *E. coli*. Data for *K. pneumoniae* shown in **(A)** and for *S.* Typhimurium in **(B);** data from **Fig. 2C-D. C-D)** Experimental communities contain representative carbon source utilization overlap to the pathogens compared to all possible combinations of the 10 best-ranked species. All possible combinations of communities are depicted by unfilled circles, experimentally tested

communities are shown as filled circles. Communities in green contain *E. coli* while communities in black do not contain *E. coli*. Data for *K. pneumoniae* shown in **(C)** and for *S.* Typhimurium in **(D).**


A
K. pneumoniae
Protein family overlap with K. pneumoniae (%)
species
B
S. Typhimurium
Protein family overlap with S. Typhimurium (%)
species
without E. coli
with E. coli
C
K. pneumoniae
Protein family overlap with K. pneumoniae (%)
E. coli
L. plan
B. long. inf
B. brev
L. sali
L. rumi
L. ferm
H. bifo
B. anim
L. acid
All 10
D
S. Typhimurium
Protein family overlap with S. Typhimurium (%)
E. coli
L. sacc
P. vulg
A. muci
M. torq
L. sali
B. long. long
B. anim
L. delb
L. acid
All 10
Community composition (panels e-f)
without E. coli
1-member
2-member
3-member
5-member
9-member
49-member
with E. coli
1-member
2-member
3-member
5-member
10-member
50-member
K. pneumoniae or S. Typhimurium itself alone
Media only
E
K. pneumoniae
K. pneumoniae (cells / mL)
Protein family overlap with K. pneumoniae (%)
F
S. Typhimurium
S. Typhimurium (cells / mL)
Protein family overlap with S. Typhimurium (%)

1126

Fig. S9.

Protein family overlap between single strains or communities and the pathogen shows that both diversity and key members (*E. coli*) are important in explaining predicted colonization resistance. A-B) As community diversity increases, protein family overlap with the pathogen increases. Results for *K. pneumoniae* shown in **(A)** and for *S.* Typhimurium in **(B).** Each circle represents a community (communities from **Fig. 2C-D).** Green circles depict communities containing *E. coli*, while black circles are communities without *E. coli*. Linear regression of log-transformed data: **(A)** R^2=0.9350, non-zero slope for *E. coli* communities (F test, p<0.0001). R^2=0.9182, non-zero slope for communities without *E. coli* (F test, p<0.0001). **(B)** R^2=0.6825, non-zero slope for *E. coli* communities (F test, p<0.0001). R^2=0.7484, non-zero slope for communities without *E. coli* (F test, p<0.0001). **C-D)** Bar chart showing the protein family overlap with the pathogen for the individual 10 best-ranked strains. Results for *K. pneumoniae* shown in **(C)** and for *S.* Typhimurium in **(D).** The bar for *E. coli* is shown in green and the other strains in grey. The predicted protein family overlap for all 10 strains is shown in dark grey. See **Table S1** for species name abbreviations. **E-F)** As community cluster overlap with the pathogen increases, the pathogen abundance on day 2 of the extended competition decreases. Colored circles depict communities containing *E. coli*, while black circles represent communities without *E. coli* (data from **Fig. 2C-D).** Color or greyscale gradients indicate the diversity of the community. The red circles represent the isogenic wildtype pathogens. Results for *K. pneumoniae* shown in **(E)** and for *S.* Typhimurium in **(F).**

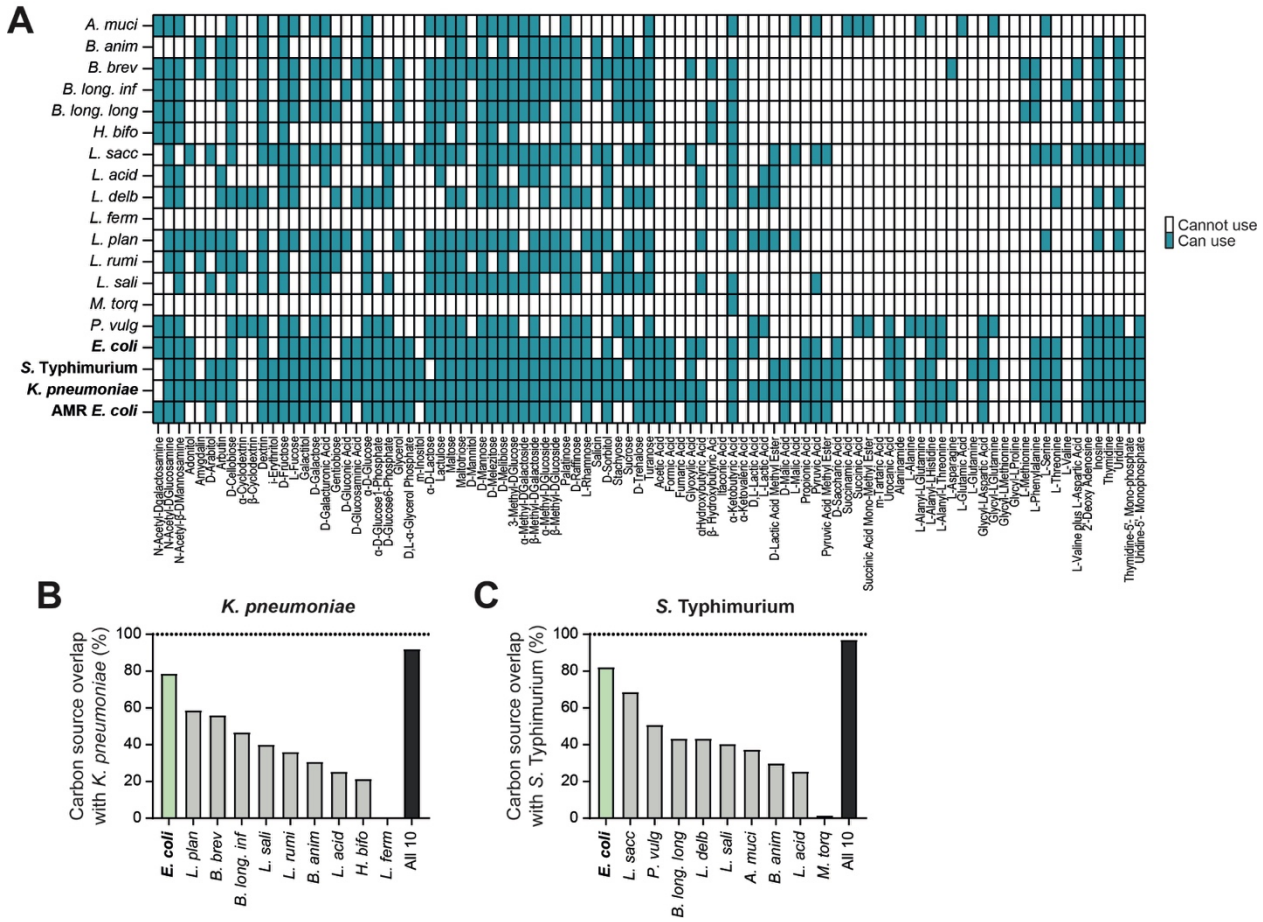


Fig. S10.

Individual carbon source utilization profiles of the 10 best-ranked symbiont strains for *K. pneumoniae* and *S.* Typhimurium (16 symbiont strains total) and their overlap with the pathogens. (A) The x-axis lists the 95 individual nutrients tested in the AN Biolog MicroPlates and the y-axis shows the 16 symbiont strains and the 3 pathogens (*K. pneumoniae*, *S.* Typhimurium, AMR *E. coli*). The pathogens are highlighted in bold on the y-axis. Nutrients shaded in blue could be used by a strain whereas those in white could not be used as defined by a threshold of background-subtracted $Abs_{590nm.}$ Data used to apply thresholding is derived from the median value of three biological replicates of AN Biolog measurements, coming from three independent experiments. **B-C)** Bar charts showing carbon source utilization overlap (%) of the 10 best ranked individual strains with the pathogens. Results for *K. pneumoniae* in **(B)** and *S.* Typhimurium in **(C).** *E. coli* is shown in green and the other individual strains in grey. The predicted utilization of all 10 best-ranked symbionts together is shown in dark grey. The percentage overlap with the pathogen is calculated as the proportion of the number of nutrients able to be used by the pathogen that can also be used by a particular symbiont strain or community. See **Table S1** for species name abbreviations.

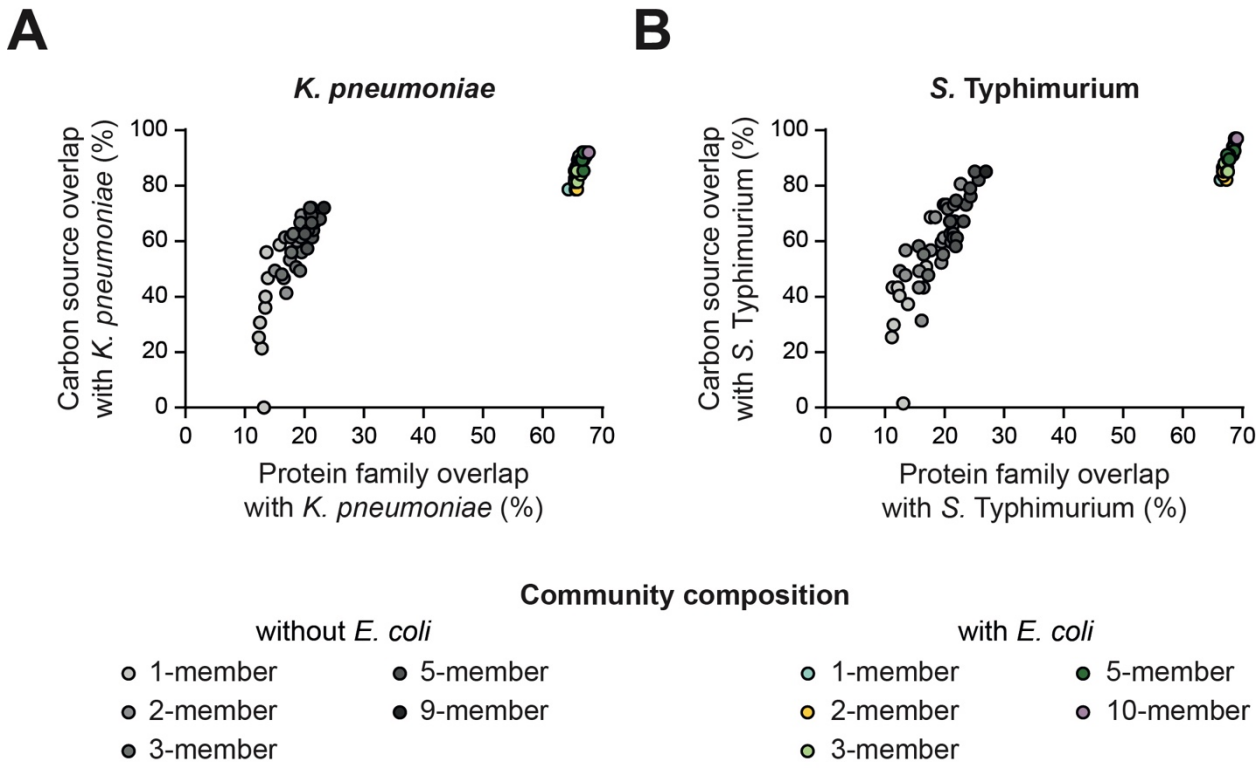
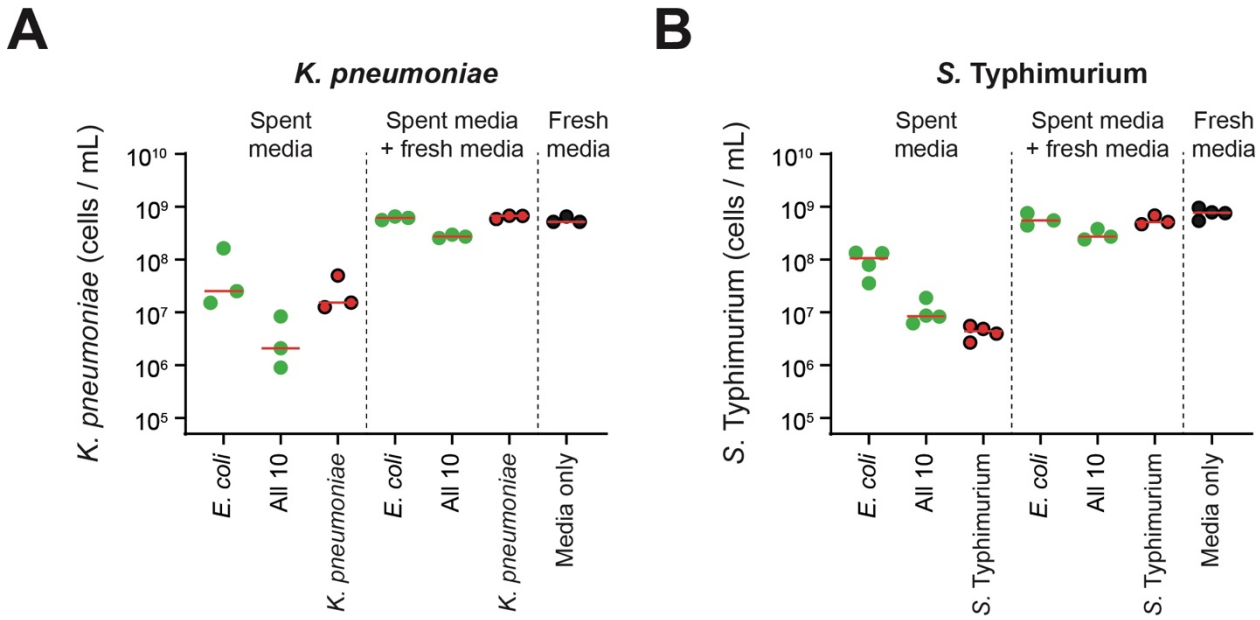


Fig. S11.

The protein family overlap and carbon source overlap prediction approaches are positively correlated. A-B) Correlation between protein cluster percentage overlap with the pathogen and carbon source utilization percentage overlap with the pathogen. Results for *K. pneumoniae* in **(A)** and *S.* Typhimurium in **(B).** Communities shown in color contain *E. coli* and those in black do not. Gradients of color or greyscale intensity show community diversity. **(A)** R^2=0.6119, slope significantly different than 0 by an F test for communities with *E. coli* ($p<0.0001$). R^2=0.6838, slope significantly different than 0 by an F test for communities without *E. coli* ($p<0.0001$). **(B)** R^2=0.8204, slope significantly different than 0 by an F test for communities with *E. coli* ($p<0.0001$). R^2=0.7027, slope significantly different than 0 by an F test for communities without *E. coli* ($p<0.0001$). The communities are the same as those in **Fig. 2C-D.** Percentage overlap calculated as the proportion of shared carbon source use or shared protein families with the pathogen. Values for communities calculated in an additive way based on the profiles of individual strains. Each data point represents a community.



1182

1183 **Fig. S12.**

1184 **Spent media experiment**. Communities were assembled and grown for 96 hours and the
1185 pathogen invaded into the spent media or re-supplemented spent media (half volume spent
1186 media and half volume nutrient media). Pathogen density measured by flow cytometry 24 hours
1187 after pathogen invasion (day 1). Results for *K. pneumoniae* shown in **(A)** and *S.* Typhimurium
1188 in **(B).** *N*=3-4 biological replicates from different independent experiments per treatment.
1189 Horizontal red lines show the median of the replicates.

1190

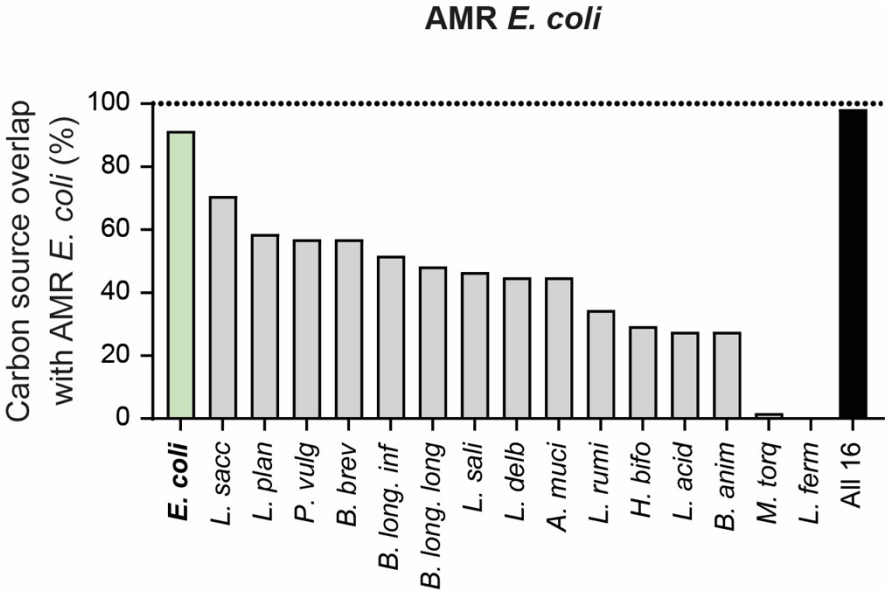


Fig. S13.

Carbon source utilization overlap with AMR *E. coli*. Bar chart showing carbon source utilization overlap (%) of the 16 best ranked individual strains (for *K. pneumoniae* and *S.* Typhimurium) with the AMR *E. coli* strain. The *E. coli* symbiont is shown in green and the other individual strains in grey. All 16 symbiont strains together are shown in black. The percentage overlap with the pathogen is calculated as the proportion of the number of nutrients that can be used by AMR *E. coli* that can also be used by a particular symbiont strain. Data used to apply thresholding is derived from the median value of 3 biological replicates of AN Biolog measurements, coming from three independent experiments.

Table S1.

Strains used in this study. *Relevant resistances only. Streptomycin= >50 ug/mL; carbenicillin=>50µg/mL; ampicillin=>100µg/mL; kanamycin=>50µg/mL; tetracycline=>50µg/mL. **According to European biosafety designation. *** Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7, 38124 Braunschwe, Science Campus Braunschweig-Süd, Germany. **** ATCC (American Type Culture Collection), 10801 University Boulevard, Manassas, Virginia 20110-2209, United States.

Strain	**Abbrev.**	**Relevant genotype ***	**Internal ID**	**PATRIC ID**	**Collection ID**	**Hazard group****	**100-strain screen**	**50-strain community**	***K. pneumoniae* 10 best-ranked**	***S.* Typhimurium 10 best-ranked**	**Source**
Actinomyces odontolyticus			NT5039	411466.7	DSM43331	2	+				Nassos Typas
Akkermansia muciniphila	*A.muci*		F3	349741.6	DSM22959	1	+	+		+	DSMZ***
Anaerostipes caccae			F2	105841.35	DSM14662	1	+	+			DSMZ
Bacteroides caccae			B1	411901.7	DSM19024	2	+				DSMZ
Bacteroides cellulosilyticus			B2	537012.5	DSM14838	1	+	+			DSMZ
Bacteroides clarus			NT5052	762984.10	DSM22519	1	+	+			Nassos Typas

Strain	Abbrev.	Relevant genotype *	Internal ID	PATRIC ID	Collection ID	Hazard group**	100-strain screen	50-strain community	*K. pneumoniae* 10 best-ranked	*S.* Typhimurium 10 best-ranked	Source
Bacteroides eggerthii			B13	483216.6	DSM20697	2	+				DSMZ
Bacteroides fragilis enterotoxigenic 20656-2-1			NT5033	817.95	ATCC43860	2	+				Nassos Typas
Bacteroides fragilis 3_1_12			NT5057	457424.5	HM-20	2	+				Nassos Typas
Bacteroides fragilis CL07T12C05			NT5059	997883.3	HM-710	2	+				Nassos Typas
Bacteroides fragilis CL05T00C42			NT5060	997880.3	HM-711	2	+				Nassos Typas
Bacteroides fragilis CL05T12C13			NT5062	997881.3	HM-712	2	+				Nassos Typas
Bacteroides fragilis CL03T00C08			NT5063	997878.3	HM-713	2	+				Nassos Typas

Strain	Abbrev.	Relevant genotype *	Internal ID	PATRIC ID	Collection ID	Hazard group**	100-strain screen	50-strain community	*K. pneumoniae* 10 best-ranked	*S.* Typhimurium 10 best-ranked	Source
Bacteroides fragilis CL03T12C07			NT5061	997879.7	HM-714	2	+				Nassos Typas
Bacteroides fragilis nontoxigenic			B4	272559.17	DSM2151	2	+				DSMZ
Bacteroides ovatus			B6	411476.11	DSM1896	2	+				DSMZ
Bacteroides stercoris VPI B5-21			B8	46506.156 2	DSM19555	1	+	+			DSMZ
Bacteroides stercoris CC31F			NT5055	1073351.3	HM-1036	2	+				Nassos Typas
Bacteroides thetaiotaomicron			B9	226186.12	DSM2079	2	+				DSMZ
Bacteroides uniformis			B10	820.37	DSM6597	2	+				DSMZ
Bacteroides uniformis CL03T12C37			NT5066	997890.6	HM-716	2	+				Nassos Typas

Strain	Abbrev.	Relevant genotype *	Internal ID	PATRIC ID	Collection ID	Hazard group**	100-strain screen	50-strain community	*K. pneumoniae* 10 best-ranked	*S.* Typhimurium 10 best-ranked	Source
Bacteroides xylanisolvens XB1A			B12	657309.4	DSM18836	1	+	+			DSMZ
Bacteroides xylanisolvens CL03T12C04			NT5064	997892.50	DSM2079	2	+				Nassos Typas
Bifidobacterium adolescentis			NT5022	367928.6	DSM20083	1	+	+			Nassos Typas
Bifidobacterium animalis	*B. anim*		A2	555970.3	DSM10140	1	+	+	+	+	DSMZ
Bifidobacterium animalis subsp. lactis Bi-04			NT5043	580050.3		1	+				Nassos Typas
Bifidobacterium animalis subsp. lactis Bi-07			NT5044	742729.3	DGCC2908	1	+	+			Nassos Typas
Bifidobacterium bifidum			A1	500634.6	DSM20456	1	+	+			DSMZ
Bifidobacterium breve	*B. brev*		A3	518634.19	DSM20213	1	+	+	+		DSMZ
Bifidobacterium longum subsp. infantis	*B. long.* inf		A4	391904.8	DSM20088	1	+	+	+		DSMZ

Strain	**Abbrev.**	**Relevant genotype ***	**Internal ID**	**PATRIC ID**	**Collection ID**	**Hazard group****	**100-strain screen**	**50-strain community**	***K. pneumoniae* 10 best-ranked**	***S.* Typhimurium 10 best-ranked**	**Source**
Bifidobacterium longum subsp. longum	*B. long*. long		A5	565042.3	DSM20219	1	+	+		+	DSMZ
Blautia hansenii			NT5005	537007.6	DSM20583	1	+	+			Nassos Typas
Blautia hydrogenotrophica			F4	476272.21	DSM10507	1	+	+			DSMZ
Christensenella minuta			F7	626937.8	DSM22607	1	+	+			DSMZ
Clostridium difficile 630			NT23006	272563.8	DSM27543	2	+				Nassos Typas
Clostridium leptum			F12	428125.8	DSM753	1	+	+			DSMZ
Clostridium perfringens			NT5031	195103.10	DSM756	2	+				Nassos Typas
Clostridium perfringens			NT5032	451754.18	DSM11782	2	+				Nassos Typas
Collinsella aerofaciens			A6	411903.6	DSM3979	2	+				DSMZ

Strain	**Abbrev.**	**Relevant genotype ***	**Internal ID**	**PATRIC ID**	**Collection ID**	**Hazard group****	**100-strain screen**	**50-strain community**	***K. pneumoniae* 10 best-ranked**	***S.* Typhimurium 10 best-ranked**	**Source**
Coprococcus comes			NT5048	470146.3	ATCC27758	1	+	+			Nassos Typas
Dorea formicigenerans			NT5076	411461.20	DSM3992	1	+	+			Nassos Typas
Dorea longicatena			F13	411462.6	DSM13814	1	+	+			DSMZ
Eggerthella lenta			NT5024	479437.5	DSM2243	2	+				Nassos Typas
Enterocloster bolteae			NT5026	208479.10	DSM15670	1	+	+			Nassos Typas
Erysipelatoclostridium ramosum			NT5006	445974.19	DSM1402	2	+				Nassos Typas
AMR *Escherichia coli*		Wildtype (Ampicillin)	19Y000018		-	2					Nottingham University Hospital Pathogen Bank
Escherichia coli ED1a			NT5078	585397.9		1	+				Nassos Typas
Escherichia coli HS			e-OPC-323	331112.6		1	+				(*34*)

Strain	**Abbrev.**	**Relevant genotype ***	**Internal ID**	**PATRIC ID**	**Collection ID**	**Hazard group****	**100-strain screen**	**50-strain community**	***K. pneumoniae* 10 best-ranked**	***S.* Typhimurium 10 best-ranked**	**Source**
Escherichia coli IAI1	*E. coli*		NT5077	585034.5		1	+	+	+	+	Nassos Typas
Escherichia coli IAI1	*E. coli* Δ*gatABC*	Δ*gatABC*	eOPC-364			1					This study
Escherichia coli JKE201	+pOPC-231		eOPC-362			1					This study
Escherichia coli JKe201	+pOPC-232		eOPC-363			1					This study
Escherichia coli MG1655			e-OPC-292	511145.12		1	+				(*35*)
Escherichia coli Z1269			Z1269		-	1					(*33*)
Escherichia coli Z1331			Z1331		-	1					(*33*)
Eubacterium rectale			NT5009	657318.12	DSM17629	1	+	+			Nassos Typas
Eubacterium siraeum			NT5040	428128.19	DSM15702	1	+	+			Nassos Typas
Faecalibacterium prausnitzii			F25	411483.3	DSM17677	1		+			DSMZ

Strain	Abbrev.	Relevant genotype *	Internal ID	PATRIC ID	Collection ID	Hazard group**	100-strain screen	50-strain community	*K. pneumoniae* 10 best-ranked	*S.* Typhimurium 10 best-ranked	Source
Fusobacterium nucleatum CTI-01			NT24006	1204474.3		2	+				Nassos Typas
Fusobacterium nucleatum MJR7757B			NT24015	851.8		2	+				Nassos Typas
Fusobacterium nucleatum subsp. nucleatum			NT5025	190304.8	DSM15643	2	+				Nassos Typas
Fusobacterium nucleatum subsp. vincentii			NT24005	155615.5	DSM19508	2	+				Nassos Typas
Fusobacterium nucleatum subsp. vincentii			NT5030	209882.4	DSM19507	2	+				Nassos Typas
Fusobacterium periodonticum 1_A_54/D10			NT24011	546275.3	ATCC33693	1	+	+			Nassos Typas
Fusobacterium periodonticum 2_1_31			NT24012	469599.3		2	+				Nassos Typas
Gemella morbillorum			NT24013	562982.3		2	+				Nassos Typas
Holdemanella biformis	*H. bifo*		F15	518637.5	DSM3989	1	+	+	+		DSMZ

Strain	Abbrev.	Relevant genotype *	Internal ID	PATRIC ID	Collection ID	Hazard group**	100-strain screen	50-strain community	*K. pneumoniae* 10 best-ranked	*S.* Typhimurium 10 best-ranked	Source
Intestinibacter bartletti			NT5086	261299.118	DSM16795	1	+				Nassos Typas
Klebsiella pneumoniae			K1	272620.9	ATCC 700721	2	+				Modernising Medical Microbiology, Nuffield Department of Medicine
Klebsiella pneumoniae subsp. pneumoniae		Wildtype (Carbenicillin)	bFS-26	1162296.3	DSM 30104	2	+				DSMZ
Klebsiella pneumoniae subsp. pneumoniae		+ pRSJ-p_{nptII}::ilux	bFS-29		DSM 30104	2					This study
Klebsiella pneumoniae subsp. pneumoniae		+pBC11	bFS-34		DSM 30104	2					This study
Lachnoclostridium symbiosum WAL-14163			NT24007	742740.3		2	+				Nassos Typas

Strain	Abbrev.	Relevant genotype *	Internal ID	PATRIC ID	Collection ID	Hazard group**	100-strain screen	50-strain community	*K. pneumoniae* 10 best-ranked	*S.* Typhimurium 10 best-ranked	Source
Lachnoclostridium symbiosum WAL-14673			NT24014	742741.3		2	+				Nassos Typas
Lachnoclostridium hylemonae			NT27002	553973.19	DSM15053	1	+	+			Nassos Typas
Lachnoclostridium scindens			F9	411468.41	DSM5676	1	+	+			DSMZ
Lachnoclostridium symbiosum			F8	411472.5	DSM934	2	+				DSMZ
Lacrimispora saccharolytica	*L. sacc*		NT5037	610130.3	DSM2544	1	+	+		+	Nassos Typas
Lacticaseibacillus casei			F16	219334.4	DSM20011	1	+	+			DSMZ
Lacticaseibacillus paracasei			NT5042	1226298.3	ATCCSD5275	1	+	+			Nassos Typas
Lactiplantibacillus plantarum JDM1	*L. plan*		F14	644042.3	-	1	+	+	+		Department of Food and Nutritional Sciences, University of Reading

Strain	**Abbrev.**	**Relevant genotype ***	**Internal ID**	**PATRIC ID**	**Collection ID**	**Hazard group****	**100-strain screen**	**50-strain community**	***K. pneumoniae* 10 best-ranked**	***S.* Typhimurium 10 best-ranked**	**Source**
Lactobacillus acidophilus	*L. acid*		NT5041	272621.13	ATCC700936	1	+	+	+	+	Nassos Typas
Lactobacillus delbrueckii subsp. Delbrueckii	*L. delb*		NT14075	1423823.4	DSM20074	1	+	+		+	Nassos Typas
Lactobacillus gasseri			F18	324831.13	DSM20243	1	+	+			DSMZ
Ligilactobacillus ruminis	*L. rumi*		F17	1423798.5	DSM20403	1	+	+	+		DSMZ
Ligilactobacillus salivarius	*L. sali*		NT14072	1423799.3	DSM20555	1	+	+	+	+	Nassos Typas
Limosilactobacillus fermentum	*L. ferm*		NT14076	1613.547	DSM20052	1	+	+	+		Nassos Typas
Mediterraneibacter gnavus			NT5046	411470.47	ATCC29149	1	+	+			Nassos Typas
Mediterraneibacter torques	*M. torq*		NT5047	411460.6	ATCC27756	1	+	+		+	Nassos Typas
Odoribacter splanchnicus			NT5081	709991.142	DSM20712	2	+				Nassos Typas

Strain	Abbrev.	Relevant genotype *	Internal ID	PATRIC ID	Collection ID	Hazard group**	100-strain screen	50-strain community	*K. pneumoniae* 10 best-ranked	*S.* Typhimurium 10 best-ranked	Source
Parabacteroides distasonis			NT5074	435591.48	DSM20701	2	+				Nassos Typas
Parabacteroides merdae			NT5071	411477.88	DSM19495	1	+				Nassos Typas
Peptostreptococcus stomatis			NT24002	596315.3	DSM17678	2	+				Nassos Typas
Phocaeicola coprocola			B3	470145.69	DSM17136	1	+	+			DSMZ
Phocaeicola dorei			NT5049	357276.1035	DSM17855	1	+	+			Nassos Typas
Phocaeicola massiliensis			B5	1121098.3	DSM17679	1	+	+			DSMZ
Phocaeicola vulgatus	*P. vulg*		B11	435590.9	DSM1447	1	+	+		+	DSMZ
Phocaeicola vulgatus CL09T03C04			NT5056	997891.3		1	+				Nassos Typas
Prevotella buccae			B14	873513.3	DSM19025	2	+				DSMZ

Strain	Abbrev.	Relevant genotype *	Internal ID	PATRIC ID	Collection ID	Hazard group**	100-strain screen	50-strain community	*K. pneumoniae* 10 best-ranked	*S.* Typhimurium 10 best-ranked	Source
Prevotella copri			B15	537011.439	DSM18205	1	+	+			DSMZ
Roseburia faecis			F24	301302.4	DSM16840	1	+	+			DSMZ
Roseburia hominis			F22	585394.18	DSM16839	1	+	+			DSMZ
Roseburia intestinalis			NT5011	536231.75	DSM14610	1	+	+			Nassos Typas
Roseburia inulinivorans			NT5012	622312.48	DSM16841	1	+	+			Nassos Typas
Salmonella enterica Typhimurium SL1344		Wildtype (Streptomycin)	SB300	216597.6	DSM24522	2	+				(*57*)
Salmonella enterica Typhimurium SL1344		+ pRSJ-p_{nptII}::ilux	sOPC-406		DSM24522	2					This study
Salmonella enterica Typhimurium SL1344		+pBC11	sOPC-404		DSM24522	2					This study

Strain	Abbrev.	Relevant genotype *	Internal ID	PATRIC ID	Collection ID	Hazard group**	100-strain screen	50-strain community	*K. pneumoniae* 10 best-ranked	*S.* Typhimurium 10 best-ranked	Source
Salmonella enterica Typhimurium SL1344		Avirulent	M2702		DSM24522	2					(*72*)
Salmonella enterica Typhimurium SL1344		*hisG* prototroph (Streptomycin)	EB199		DSM24522	2					This study
Salmonella enterica Typhimurium SL1344		Δ*gatABC* (Streptomycin)	sOPC-463		DSM24522	2					This study
Salmonella enterica Typhimurium SL1344		Δ*gatABC* +pBC11	EB149		DSM24522	2					This study
Staphylococcus epidermis			bOPC-105	176280.85	DSM1798	2	+				ATCC****
Streptococcus parasanguinis			NT5072	760570.3	DSM6778	2	+				Nassos Typas
Streptococcus salivarius			NT5038	1304.1829	DSM20560	2	+				Nassos Typas

Strain	**Abbrev.**	**Relevant genotype ***	**Internal ID**	**PATRIC ID**	**Collection ID**	**Hazard group****	**100-strain screen**	**50-strain community**	***K. pneumoniae* 10 best-ranked**	***S.* Typhimurium 10 best-ranked**	**Source**
Veillonella parvula			NT5017	479436.6	DSM2008	2	+				Nassos Typas

Table S2.

Plasmids used in this study. *Relevant resistances only. Streptomycin= >50 ug/mL; carbenicillin=>50µg/mL; ampicillin=>100µg/mL; kanamycin=>50µg/mL; tetracycline=>50µg/mL.

Plasmid name	**Relevant genotype**	**Resistance***	**Source**
pRSJ-p_{nptII}::ilux	*lux*CDBAE-*frp* expression	Tetracycline	(*61*)
pBC11	YPet expression	Kanamycin	(*64*)
pOPC-231	SV-*aphT-tetR-sceI-sacB-S*Tm-*Δ*gatABC	Kanamycin	This study
pOPC-232	SV-*aphT-tetR-sceI-sacB-Ec*IAI-*Δ*gatABC	Kanamycin	This study

Table S3.
Primers used in this study.

Primer name	Sequence	Purpose	Source
oOPC-953	AGAGTTTGATCCTGGCTCAG	16S sequencing (27F)	(*73*)
oOPC-954	TACGGYTACCTTGTTACGACTT	16S sequencing (1492R)	(*73*)
g-Bifid-F	5′-CTCCTGGAAACGGGTGG-3′	16S sequencing for Bifidobacteria	(*74*)
g-Bifid-R	5′-GGTGTTCTTCCCGATATCTACA-3′		(*74*)
oOPC-975	CCCAGTCTCGAGGTCGACGGTATCGATAAGCTTGA TATCGAATTCaaatcgcgtttcgtgaatcagg	amplify 700 upstream *gatABC* for deletion - STm SL1344	This study
oOPC-976	taaaattaagaggcgatttgaaatagttggctcataaattctccattattcagg		This study
oOPC-977	aatttatgagccaactatttcaaatcgcctcttaattttaggggag	amplify 700 downstream *gatABC* for deletion - *S*Tm	This study
oOPC-978	CTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAAC TAGTGGATCCttcccggctattacaggtatgcgttgcgc		This study
oOPC-979	agtcgatgctgcacgtacgc	check *gatABC* deletion - *S*Tm	This study
oOPC-980	atgtcggacaacgcggtctg		This study
oOPC-981	CCCAGTCTCGAGGTCGACGGTATCGATAAGCTTGA TATCGAATTCctactgttaaatgtttgcatcgcacc	amplify 700 upstream *gatABC* for deletion - *E.coli* IAI1	This study
oOPC-982	ggtatatgactaacctgtttgttctcgcagaataatttttacctgaggg		This study
oOPC-983	aaaaattattctgcgagaacaaacaggttagtcatataccgtccttattccg	amplify 700 downstream *gatABC* for deletion - *E.coli* IAI1	This study
oOPC-984	CTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAAC TAGTGGATCCtttacgtacgcatcaaaaggcctttattgcc		This study
oOPC-985	aaacgctctgcatttgcggc	check *gatABC* deletion - *E.coli* IAI1	This study
oOPC-986	cccatgttgaagatgccgcc		This study

Table S4.

Species compositions of communities used for *in vitro* experiments with *K. pneumoniae* DSM 30104. *Communities that were additionally selected to contain *E. coli* IAI1, but otherwise were selected at random.

Community ID	Community members	No. replicates	Median day 2 pathogen density (cells/mL)
No commensals	N/A	15	1.84E+09
KTS1	A3	5	9.45E+08
KTS2	F15	3	1.11E+09
KTS3	F17	3	1.19E+09
KTS4	A4	3	1.29E+09
KTS5	NT5077	3	1.10E+08
KTS6	NT5041	3	1.12E+09
KTS7	A2	3	1.28E+09
KTS8	F14	3	1.21E+09
KTS9	NT14076	3	1.68E+09
KTS10	NT14072	3	1.53E+09
KTD1	A4, NT5041	3	1.33E+09
KTD2	NT5041, F14	3	1.29E+09
KTD3	A3, NT5077	4	4.58E+06
KTD4	F15, F14	3	1.40E+09
KTD5	F15, A2	3	1.83E+09
KTD6	A2, F14	3	1.70E+09
KTD7	F17, NT14072	3	1.59E+09
KTD8	NT5077, NT14076	4	1.28E+08
KTD9	A3, F14	3	5.96E+08
KTD10*	NT5077, F14	3	1.82E+07
KTD11*	F15, NT5077	3	1.09E+08
KTD12*	F17, NT5077	3	3.42E+07

Community ID	Community members	No. replicates	Median day 2 pathogen density (cells/mL)
KTD13*	A4, NT5077	3	7.44E+07
KTD14*	NT5077, NT5041	3	1.55E+07
KTD15*	NT5077, A2	3	3.84E+07
KTD16*	NT5077, NT14072	3	2.90E+07
KTD17*	NT5041, A2	3	1.81E+09
KTT1	A3, F15, NT5041	3	3.30E+08
KTT2	A3, F15, A2	3	7.44E+08
KTT3	A2, NT14076, NT14072	3	6.56E+08
KTT4	F17, NT5077, NT14072	3	1.12E+07
KTT5	NT5041, NT14076, NT14072	3	7.11E+08
KTT6	F17, NT5077, A2	3	2.52E+07
KTT7	F17, A4, NT14076	3	1.07E+09
KTT8	A3, F15, F17	4	5.03E+08
KTT9	A3, NT5041, A2	3	7.59E+08
KTT10	A3, A4, NT5041	3	6.13E+08
KTT11	A2, F14, NT14072	3	1.19E+09
KTT12	A4, NT5041, F14	3	9.36E+08
KTT13	F15, F14, NT14072	3	1.13E+09
KTT14	A3, NT5077, F14	3	3.81E+05
KTT15	A3, F17, A2	3	5.07E+08
KTT16	NT5077, F14, NT14072	3	1.72E+07
KTT17	NT5077, NT14076, NT14072	3	1.43E+07
KTT23*	A3, F15, NT5077	2	8.73E+05
KTT24*	A4, NT5077, NT14072	3	1.29E+07
KTT25*	A3, NT5077, A2	3	6.53E+05
KTT26*	NT5077, NT5041, NT14072	3	6.20E+06
KTP1	A3, F15, F17, NT5041, NT14076	4	4.77E+08
KTP2	A3, F15, NT5041, A2, NT14072	4	2.98E+08

Community ID	Community members	No. replicates	Median day 2 pathogen density (cells/mL)
KTP3	NT5077, A2, F14, NT14076, NT14072	4	1.66E+07
KTP4	F17, NT5077, NT5041, A2, NT14072	4	2.42E+07
KTP5	A3, F15, NT5041, NT14076, NT14072	4	3.63E+08
KTP6	A3, F15, F17, A4, A2	4	5.33E+08
KTP7	F17, A4, NT5077, A2, NT14076	4	2.63E+07
KTP8	A3, F15, F17, NT5077, NT5041	4	1.25E+06
KTP9	A3, A4, NT5041, A2, F14	4	1.85E+08
KTP10	A3, A4, NT5041, F14, NT14072	4	1.99E+08
KTP11	F15, NT5041, A2, F14, NT14072	4	1.19E+09
KTP12	F17, NT5041, A2, F14, NT14076	4	9.19E+08
KTP13	F15, A4, F14, NT14076, NT14072	4	9.89E+08
KTP14	A3, A4, A2, F14, NT14076	4	6.60E+08
KTP15	F17, A4, F14, NT14076, NT14072	4	1.28E+09
KTP16	A3, F17, A4, F14, NT14072	4	2.95E+08
KTP17	A3, A4, A2, NT14076, NT14072	4	2.67E+08
KTP18*	F15, NT5077, NT5041, A2, F14	3	2.72E+06
KTP19*	F17, A4, NT5077, F14, NT14076	3	1.28E+07
KTP20*	A3, NT5077, NT5041, A2, NT14076	3	2.82E+05
KTP21*	A3, F15, NT5077, NT5041, F14	3	3.26E+05
KTP22*	A3, F17, NT5077, NT5041, F14	3	3.08E+05
KALLTOP	A3, F15, F17, A4, NT5077, NT5041, A2, F14, NT14076, NT14072	11	6.17E+05

Table S5.

Species compositions of communities used for *in vitro* experiments with *S. enterica* serovar Typhimurium SL1344. *Communities that were additionally selected to contain *E. coli* IAI1, but otherwise were selected at random.

Community ID	Community members	No. replicates	Median day 2 pathogen density (cells/mL)
No commensals	N/A	11	1.38E+09
STS1	A2	3	7.64E+08
STS2	NT5037	3	6.29E+08
STS3	NT14075	3	8.71E+08
STS4	F3	3	9.55E+08
STS5	NT5041	3	9.26E+08
STS6	A5	3	8.69E+08
STS7	NT5077	5	4.46E+08
STS8	B11	4	8.14E+08
STS9	NT5047	3	1.18E+09
STS10	NT14072	3	1.08E+09
STD1	NT5037, B11	3	6.04E+08
STD2	NT5077, NT5047	3	3.60E+08
STD3	NT5041, B11	3	7.16E+08
STD4	A5, NT5047	3	9.19E+08
STD5	NT14075, NT5077	3	3.98E+08
STD6	NT5037, NT14072	3	6.30E+08
STD7	B11, NT5047	3	7.64E+08
STD8	NT14075, NT14072	3	9.23E+08
STD9	NT5037, NT5047	3	7.74E+08
STD10	NT14075, NT5047	3	9.25E+08
STD11	NT14075, NT5041	3	9.62E+08
STD12	NT5041, NT14072	3	9.98E+08

Community ID	Community members	No. replicates	Median day 2 pathogen density (cells/mL)
STD13	A2, NT14072	3	1.06E+09
STD14	A2, NT5047	3	8.40E+08
STD15	B11, NT14072	3	7.79E+08
STD16	NT14075, F3	3	1.07E+09
STD17*	A2, NT5077	3	3.36E+08
STD18*	NT5037, NT5077	4	5.73E+07
STD19*	F3, NT5077	3	3.23E+08
STD20*	NT5041, NT5077	3	3.11E+08
STD21*	A5, NT5077	3	3.52E+08
STD22*	NT5077, B11	7	3.94E+07
STD23*	NT5077, NT14072	3	4.14E+08
STT1	A2, NT14075, A5	3	9.50E+08
STT2	NT14075, F3, B11	3	1.00E+09
STT3	F3, NT5041, NT5077	3	5.51E+08
STT4	A2, A5, NT14072	3	1.15E+09
STT5	A2, NT5037, A5	3	9.38E+08
STT6	NT5037, NT5077, NT14072	3	1.25E+08
STT7	F3, NT5041, A5	3	1.26E+09
STT8	NT5037, A5, NT5047	3	1.05E+09
STT9	NT5037, F3, NT5077	3	1.44E+08
STT10	NT5037, A5, NT14072	3	9.17E+08
STT11	NT14075, B11, NT5047	3	9.20E+08
STT12	A2, NT14075, B11	3	1.10E+09
STT13	A5, B11, NT14072	3	9.89E+08
STT14	NT5041, NT5047, NT14072	3	1.26E+09
STT15	A5, B11, NT5047	3	1.25E+09
STT16	A2, F3, B11	3	1.22E+09
STT17*	NT5037, NT5077, B11	5	2.23E+07

Community ID	Community members	No. replicates	Median day 2 pathogen density (cells/mL)
STT18*	A2, NT14075, NT5077	3	3.75E+08
STT19*	NT5041, NT5077, NT5047	3	3.80E+08
STT20*	NT5037, NT5077, NT5047	3	6.17E+07
STT21*	F3, NT5077, NT5047	3	3.79E+08
STT22*	NT14075, NT5077, B11	3	3.87E+07
STP1	NT14075, F3, NT5041, NT5047, NT14072	4	1.19E+09
STP2	A2, NT5037, NT14075, NT5077, NT14072	4	1.65E+08
STP3	NT14075, F3, A5, B11, NT14072	4	9.01E+08
STP4	A2, F3, A5, B11, NT5047	4	1.05E+09
STP5	F3, NT5041, A5, NT5077, B11	4	4.53E+07
STP6	NT5037, F3, B11, NT5047, NT14072	4	6.69E+08
STP7	NT14075, F3, NT5041, A5, NT5047	4	1.09E+09
STP8	A2, NT5037, F3, NT5041, NT5047	4	1.07E+09
STP9	NT5037, NT14075, A5, B11, NT14072	4	9.40E+08
STP10	A2, F3, NT5041, A5, NT5047	4	1.40E+09
STP11	A2, NT5037, A5, NT5077, B11	4	2.55E+07
STP12	NT5037, A5, NT5077, NT5047, NT14072	4	1.46E+08
STP13	A2, NT14075, A5, B11, NT14072	4	8.71E+08
STP14	F3, NT5041, NT5077, NT5047, NT14072	4	4.59E+08
STP15	A2, NT5037, NT14075, F3, A5	4	1.06E+09
STP16	A2, NT14075, F3, A5, NT14072	4	1.19E+09
STP17*	NT5041, A5, NT5077, B11, NT14072	3	2.01E+07
STP18*	NT5037, NT5041, A5, NT5077, B11	3	9.83E+06
STP19*	NT14075, F3, A5, NT5077, NT14072	3	4.38E+08
STP20*	A2, F3, A5, NT5077, B11	3	1.83E+07
SALLTOP	A2, NT5037, NT14075, F3, NT5041, A5, NT5077, B11, NT5047, NT14072	11	1.76E+07

Table S6.
Species compositions of communities used for gnotobiotic mouse experiments.

Pathogen gavaged	**No. symbionts in community**	**Community members**	***E. coli* IAI1 present**	**No. replicates**	**Median pathogen fecal density 24 hours p.i. (CFU/g)**
K. pneumoniae	0	N/A	No	7	4.0375e+09
K. pneumoniae	1	NT5077	Yes	7	2.28846e+08
K. pneumoniae	5	KTP8 (see Table S4)	Yes	8	4.35148e+07
K. pneumoniae	10	KALLTOP (see Table S4)	Yes	7	4.07019e+07
K. pneumoniae	50	See Table S1	Yes	7	134667
K. pneumoniae	9	KALLTOP minus NT5077	No	7	4.4e+08
K. pneumoniae	49	50 minus NT5077	No	7	2.32e+07
S. Typhimurium	0	N/A	No	7	7.89e+08
S. Typhimurium	1	NT5077	Yes	7	9.01408e+07
S. Typhimurium	5	STP11 (see Table S5)	Yes	8	5.68115e+07
S. Typhimurium	10	SALLTOP (see Table S5)	Yes	7	4.1e+07
S. Typhimurium	50	See Table S1	Yes	7	521000
S. Typhimurium	9	SALLTOP minus NT5077	No	8	6.91e+08
S. Typhimurium	49	50 minus NT5077	No	7	2.15e+07

Table S7.

Species compositions of communities used for prediction experiments with the AMR *E. coli* strain 19Y000018.

Community ID	Community members	No. replicates	Median day 2 pathogen density (CFU/mL)
No commensals	N/A	5	4.96E+08
Best 2 Biolog	NT5077, NT5037	5	1.00E+07
Worst 2 Biolog	NT5077, A2	5	3.12E+08
Best 3 Biolog	NT5077, B11, NT5037	5	1.56E+06
Worst 3 Biolog	NT5077, F17, NT5041	5	2.64E+08
Best 5 Biolog	NT5077, NT5037, NT14075, NT5041, B11	5	2.48E+06
Worst 5 Biolog	NT5077, NT14075, A5, A3, F17	5	4.64E+06
Best 2 protein family	NT5077, NT5026	5	1.80E+07
Worst 2 protein family	NT5077, F3	5	2.48E+08
Best 3 protein family	NT5077, NT5026, F2	5	1.60E+07
Worst 3 protein family	NT5077, A2, NT5044	5	2.24E+08
Best 5 protein family	NT5077, NT5026, F2, NT5049, NT24011	5	2.80E+06
Worst 5 protein family	NT5077, NT14072, NT5041, F24, NT5040	5	8.16E+06
Best 10 protein family	NT5077, F2, NT14076, NT5026, NT5049, NT5047, F25, NT24011, NT5037, NT5012	5	1.60E+06
Worst 10 protein family	NT5077, F17, NT5041, NT5076, NT5048, A1, F15, A2, NT5009, NT5044	5	3.40E+06
Best 5 protein family #2	NT5077, NT5026, NT24011, NT5037, NT5052	5	4.00E+06

Community ID	Community members	No. replicates	Median day 2 pathogen density (CFU/mL)
Best 5 protein family #3	NT5077, NT5026, NT24011, NT5037, NT5049	5	2.02E+06
Best 5 protein family #4	NT5077, F2, NT24011, NT5037, NT5052	5	5.52E+06
Best 5 protein family #5	NT5077, F2, NT5026, F16, NT5049	5	2.48E+06
Worst 5 protein family #2	NT5077, F17, F15, A2, NT5044	5	3.00E+08
Worst 5 protein family #3	NT5077, NT5048, F15, A2, NT5044	5	2.32E+08
Worst 5 protein family #4	NT5077, NT14072, NT5040, A2, NT5044	5	3.20E+08
Worst 5 protein family #5	NT5077, F17, A1, A2, NT5044	5	2.80E+08

Supplementary references

72. B. Periaswamy et al., Live attenuated S. Typhimurium vaccine with improved safety in immuno-compromised mice. PLoS One 7, e45433 (2012).

73. J. S. Johnson et al., Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun 10, 5029 (2019).

74. T. Matsuki et al., Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. Appl Environ Microbiol 68, 5445-5451 (2002).