nature methods

BRIEF COMMUNICATION

https://doi.org/10.1038/s41592-018-0301-y

Metagenomic engineering of the mammalian gut microbiome in situ

Carlotta Ronda[1,5], Sway P. Chen[1,2,5], Vitor Cabral[1,5], Stephanie J. Yaung[3] and Harris H. Wang[1,4*]

Engineering of microbial communities in open environments remains challenging. Here we describe a platform used to identify and modify genetically tractable mammalian microbiota by engineering community-wide horizontal gene transfer events in situ. With this approach, we demonstrate that diverse taxa in the mouse gut microbiome can be modified directly with a desired genetic payload. In situ microbiome engineering in living animals allows novel capabilities to be introduced into established communities in their native milieu.

In nature, microbes live in open, dynamic, and complex habitats that are difficult to recapitulate in a laboratory setting. Although recent advances in deep sequencing have shed light on the vast microbial diversity in nature, the ability to genetically alter these microbiomes remains limited, despite advances in culturomics and synthetic biology[1–4]. Genetic intractability is often attributed to host immunity, such as restriction methylation[5] or CRISPR–Cas processes[6], although myriad other factors (e.g., DNA transformation, growth state, fitness burden) can also influence gene transfer potential[7]. Here we devised an approach, metagenomic alteration of gut microbiome by in situ conjugation (MAGIC), to genetically modify gut microbiota in their native habitat by engineering the mobilome—the repertoire of mobile genetic elements in the gut microbiome.

We applied MAGIC to the mammalian gut because it harbors a diverse microbial community with key functional roles in host physiology[8]. We constructed an *Escherichia coli* donor strain that can deliver a genetic payload into target recipients by broad-host-range bacterial conjugation (Fig. 1). We integrated the IncPα-family RP4 conjugation system[9], which can efficiently conjugate into both Gram-positive and Gram-negative cells, into the EcGT1 donor genome, along with a constitutively expressing mCherry-specR cassette (Δ*galK::mCherry-specR*). To strengthen biocontainment of the donor and to facilitate in vitro selection of recipients, we generated an alternative strain, EcGT2 (Δ*asd::mCherry-specR*), to be auxotrophic for the essential cell-wall component diaminopimelic acid (DAP), thus requiring DAP supplementation in the growth media[10].

We developed a modular suite of mobile plasmids (pGT) that featured replicative origins with narrow to broad host ranges, an RP4 transfer origin, a selectable marker, and the desired genetic payload (Supplementary Tables 1–3, Supplementary Fig. 1). We also used a broad-host-range Himar transposon system for delivering integrative payloads. As a demonstration of the system, we used a dual-reporter payload harboring a green fluorescent protein (GFP) and an antibiotic-resistance gene (AbR). The use of fluorescence-activated cell sorting (FACS) combined with 16S metagenomic analysis enabled us to identify successfully modified recipients or transconjugants, which could then be readily isolated on antibiotic selective plates. This multi-pronged strategy can increase the diversity of genetically tractable microbiota that can be captured. We first validated and optimized MAGIC protocols in vitro by assessing the gating stringency of FACS with control spike-ins of GFP-tagged bacteria into a complex sample community (Supplementary Fig. 2). Subsequently, in vitro conjugations with defined recipient species (Supplementary Fig. 3) and live bacterial communities extracted from mouse feces (Supplementary Fig. 4) demonstrated the transfer of the payload from donors to recipients to yield GFP$^+$ transconjugants that could be enriched by FACS (Supplementary Fig. 5), which we confirmed by fluorescence microscopy (Supplementary Fig. 6). 16S rRNA sequencing of FACS-enriched transconjugant populations revealed a diverse range of recipient bacteria (Supplementary Fig. 7).

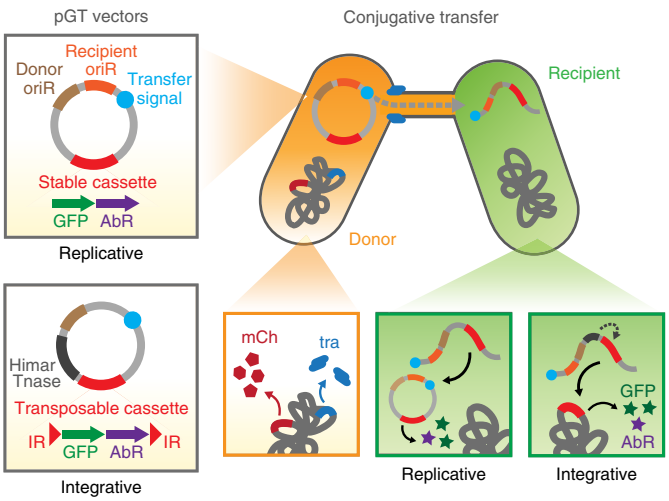

Fig. 1 | Overview of metagenomic alteration of gut microbiome by in situ conjugation (MAGIC). MAGIC implementation to transfer replicative or integrative pGT vectors from an engineered donor strain into amenable recipients in a complex microbiome. Replicative vectors feature a broad-host-range origin of replication (oriR), whereas integrative vectors contain a transposable Himar cassette and transposase (Tnase). The donor *E. coli* strain contains genomically integrated conjugative transfer genes (tra) and an mCherry gene (mCh). Transconjugant bacteria are detectable on the basis of expression of an engineered payload that includes GFP and an antibiotic-resistance gene (AbR).

[1]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA. [2]Integrated Program in Cellular, Molecular and Biomedical Studies, Columbia University Irving Medical Center, New York, NY, USA. [3]Program in Medical Engineering and Medical Physics, Harvard–MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY, USA. [5]These authors contributed equally: Carlotta Ronda, Sway P. Chen, Vitor Cabral. *e-mail: hw2429@columbia.edu

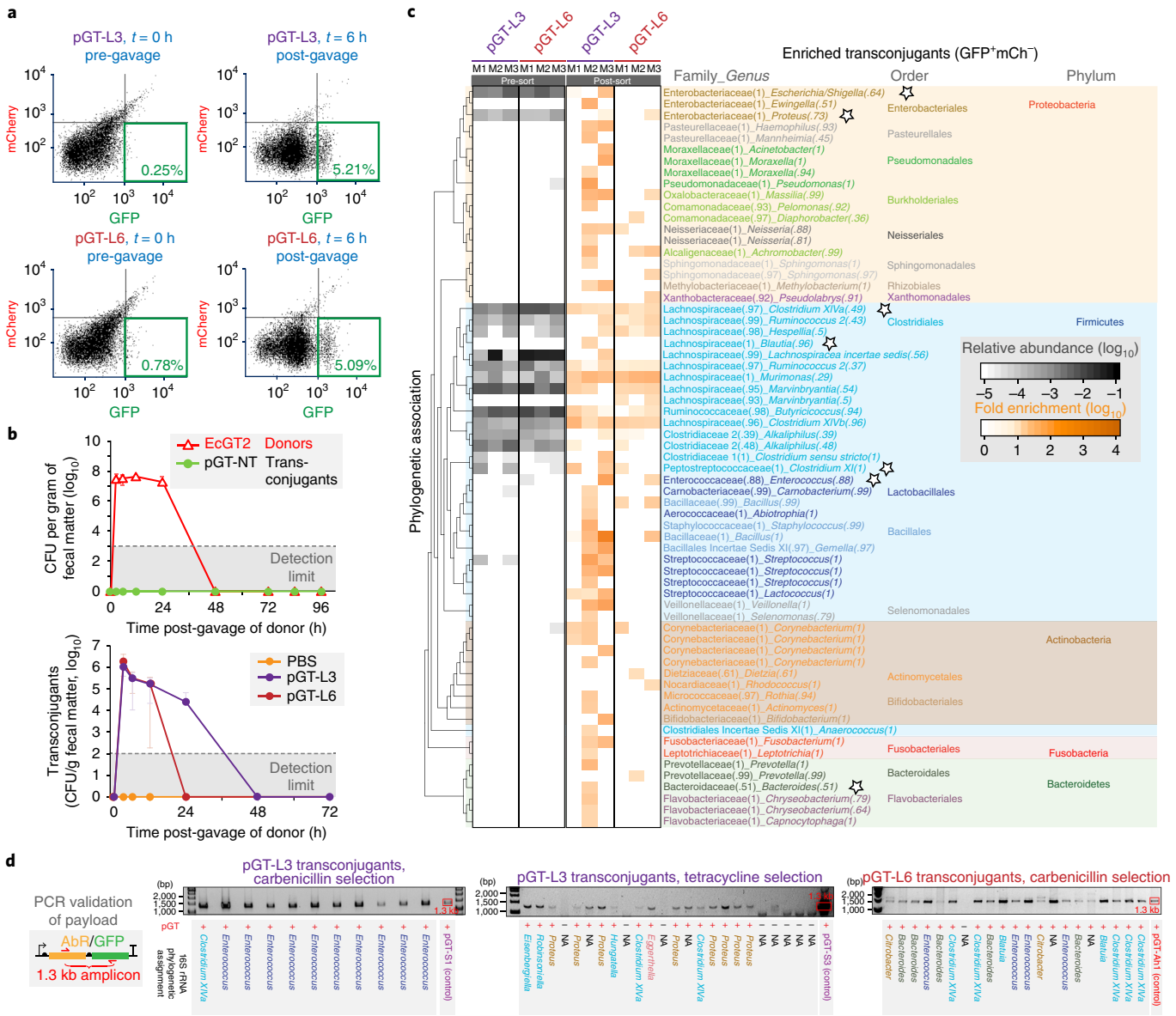

Fig. 2 | Identification and isolation of genetically tractable bacteria from the mouse gut with MAGIC. a, Fecal bacterial analysis by FACS, antibiotic selection, and sequencing after implementation of MAGIC in a mouse model. The dot plots represent FACS analysis of fecal bacteria from EcGT2 donors, pre- and post-gavage with pGT-L3 or pGT-L6 vector libraries. Green boxes define the sorted GFP$^+$mCherry$^-$ transconjugant populations. For each vector library, fecal samples from three cohoused mice were independently evaluated by flow cytometry, with similar results. **b**, Longitudinal analysis of fecal microbiome by flow cytometry for the presence of EcGT2 pGT-NT donor cells ($n = 4$ mice) and of transconjugants of vector libraries pGT-L3 ($n = 3$ mice), pGT-L6 ($n = 3$ mice), pGT-NT control ($n = 4$ mice), or PBS (no donor) control ($n = 2$ mice). Donor cells and transconjugants were lost within 48 h. The dashed line indicates the detection limit. **c**, 16S taxonomic classification of transconjugants (GFP$^+$mCherry$^-$) enriched by FACS of pGT-L3 and pGT-L6 recipient groups at 6 h post-gavage. Each heat map column represents transconjugants from one mouse. The relative abundance of each operational taxonomic unit (OTU) in the total bacterial population is shown in the grayscale heat map, and each OTU's fold enrichment among transconjugants is shown in the orange heat map. In the table on the right, numbers in parentheses indicate the confidence of taxonomic assignment by RDP Classifier. Genera with successfully cultivated isolates are denoted by white stars. **d**, PCR confirmed the presence of the antibiotic resistance–GFP payload cassette from pGT-L3 and pGT-L6 vectors in diverse isolates that were engineered in the mouse gut and isolated by selective plating with carbenicillin or tetracycline. "NA" indicates 16S sequences that were not available.

Next, we explored the possibility of implementing MAGIC in vivo, directly in the native gut microbiome of an animal. We hypothesized that different groups of microbiota could be modified through the use of a library of pGT vectors with a range of gene expression levels and plasmid replication elements suitable for different gut bacteria. We generated libraries of pGT vectors (pGT-L1 to pGT-L6) by modularly permuting pGT parts, including regulatory sequences of varying activity, payload-selectable genes (*bla*, *catP*, *tetQ*), transposon elements (Himar), and plasmid origins (RSF1010, pBBR1, p15A) (Supplementary Tables 1 and 2). We carried out four separate in vivo studies in which EcGT2 donors containing pGT libraries were orally gavaged into conventionally raised C57BL/6J mice obtained from commercial vendors (Supplementary Fig. 8a). To assess the transfer capacity of individual pGT replicative or integrative designs (pBBR1, p15A-Himar, and RSF1010), we introduced the pGT libraries pGT-L1, pGT-L2, and pGT-L3

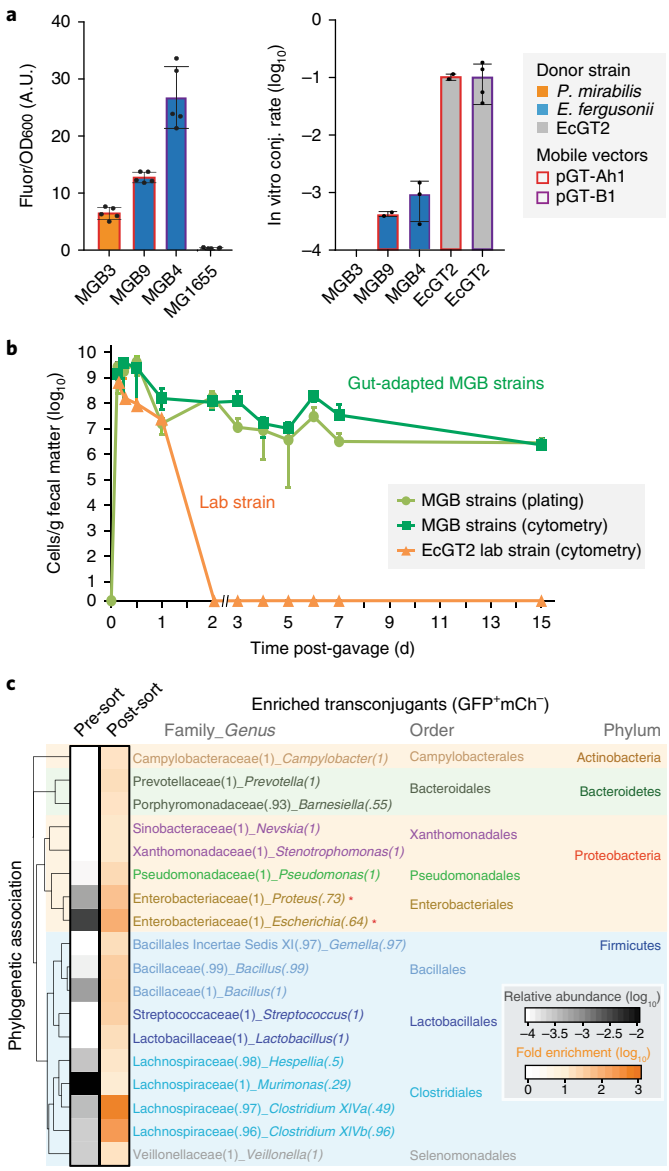

Fig. 3 | Transconjugant native gut bacteria recolonize the gut and mediate secondary transfer of engineered genetic payloads. a, Left, GFP expression profiles of three isolates (MGB3, MGB4, and MGB9; $n = 5$ for each) versus the control strain (*E. coli* MG1655; $n = 5$). MGB isolates were *P. mirabilis* (orange bar) and *E. fergusonii* (blue bars) containing either vector pGT-Ah1 (red border) or vector pGT-B1 (purple border). *E. fergusonii* strains were genetically identical, but received two different vectors. Right, efficiency of in vitro conjugation (conj.) of pGT vectors from MGB strains to *E. coli* MG1655 recipients. EcGT2 donors were used as positive controls (gray bars). Sample sizes: $n = 2$–4. Data shown as mean ± s.d. **b**, Colonization of MGB strains and the EcGT2 lab strain in mice ($n = 6$ and 4, respectively) over time, after initial oral gavage. Cell densities were determined by both plating (light green) and flow cytometry (dark green) of fecal bacteria, and by flow cytometry only for *E. coli* (orange). Data shown as mean ± s.d. **c**, FACS enrichment and 16S taxonomic classification of the top in vivo transconjugants at 6 h post-gavage with MGB strains. Fecal samples from 6 mice were combined for analysis. The relative abundance of each operational taxonomic unit (OTU) in the total bacterial population is shown in the grayscale heat map, and each OTU's fold enrichment among transconjugants is shown in the orange heat map. In the table on the right, numbers in parentheses indicate the confidence of taxonomic assignment by RDP classifier. Red asterisks denote OTUs that share the same genus as MGB donors.

separately into a cohort of mice from Taconic (Supplementary Fig. 8b–d). We tested larger combinatorial libraries (pGT-L3 to pGT-L6) in two independent mouse cohorts to assess variability across cohorts (Fig. 2, Supplementary Fig. 9). To compare in situ transfer in different gut communities, we tested the pGT-L6 library in mice from a different source (Charles River) (Supplementary Fig. 10).

We carried out FACS enrichment and 16S metagenomic analysis on fecal material from all mice studied, collected over time after oral gavage with pGT libraries. Across in situ studies, up to 5% of resulting bacteria seemed to be successful transconjugants (i.e., $GFP^+mCherry^-$) 6 h post-gavage, compared with those in samples from control groups (mice gavaged with PBS or EcGT2 carrying a nontransferable vector, pGT-NT) (Fig. 2a and Supplementary Figs. 8b, 9a, and 10a). These $GFP^+mCherry^-$ transconjugants persisted for up to 72 h post-gavage (Fig. 2b, Supplementary Fig. 9b). 16S metagenomic sequencing of these transconjugant populations revealed a wide phylogenetic breadth (Fig. 2c and Supplementary Figs. 8c, 9c, and 10b). We observed substantial reproducible enrichment of Proteobacteria and Firmicutes, especially Clostridiales and Bacillales, among successful transconjugants across multiple independent experiments. Use of the same pGT-L6 library in mice from different vendors, which harbored distinct microbiomes (Supplementary Fig. 10c), yielded shared and distinct transconjugants (Supplementary Fig. 10d). In parallel to FACS metagenomic studies, we isolated individual transconjugants from these fecal samples by selective plating for the payload AbR, and confirmed the presence of the GFP–AbR payload by PCR (Fig. 2d). Across all experiments, we isolated and validated more than 297 transconjugants belonging to 19 genera across 4 phyla (Supplementary Fig. 11, Supplementary Table 4), thus validating the capacity of MAGIC to broadly transfer genetic material in situ to diverse recipients in the mammalian gut. In contrast, we could isolate only seven genera from in vitro conjugation experiments using the same pGT vectors, despite a similar diversity of transconjugants detected by FACS metagenomics (Supplementary Fig. 7). This difference may be due to in vitro conditions that suboptimally support the growth of diverse species during conjugation reactions, which underscores the value of implementing MAGIC in situ in an established complex microbiome.

As transconjugants were no longer detected by 72 h in situ (Fig. 2b, Supplementary Fig. 9b), we speculated that the genetic payload on pGT vectors might be unstable or toxic, thus causing its negative selection in transconjugants. We tested this hypothesis in vitro by carrying out 20–30 serial passages of two transconjugant isolates of *Escherichia fergusonii* that contained the GFP–carbenicillin resistance (carbR) payload on either a pGT-B1 (replicative pBBR1 origin) or a pGT-Ah1 (integrative Himar transposon) plasmid (Supplementary Fig. 12). For the pGT-B1 population, we observed a considerable decrease in the fraction of GFP^+ cells (Supplementary Fig. 12a–c). PCR assay of the origin of replication indicated that the pGT-B1 plasmid was no longer present in the GFP^- cells (Supplementary Fig. 12d). In contrast, cells in the pGT-Ah1 population remained GFP^+ despite a detectable loss of the plasmid in parts of the population over time (Supplementary Fig. 12e–g), which suggests a more stable maintenance of the GFP–carbR payload as an integrative transposon within the host genome. Together, these results highlight the challenges of maintaining the long-term in vivo stability of engineered genetic constructs in complex microbial communities, and suggest design considerations for more precise tuning of payload life span and improved payload biocontainment.

Whole-genome sequencing of three transconjugant strains of *Proteus mirabilis* and *E. fergusonii* from our studies (designated as modifiable gut bacteria MGB3, MGB4, and MBG9) revealed the presence of putative endogenous DNA mobilization systems (Supplementary Fig. 13a–c). We wondered whether these native

mobilization systems could interface with our engineered pGT vectors, and thus carried out in vitro conjugations of the MGB strains with laboratory *E. coli* recipients. We discovered that MGB4 and MGB9 (both *E. fergusonii*) were able to mobilize pGT vectors into recipients, although less efficiently than our engineered EcGT2 donor (Fig. 3a, Supplementary Fig. 13d). These results suggest that some native gut bacteria can promote secondary transfer of engineered payloads by using their endogenous conjugation machinery, which may improve payload transfer in situ.

In general, non-gut-adapted bacteria (e.g., probiotics) do not colonize an established gut microbiome. Infiltration of foreign species usually requires drastic perturbations, such as the use of broad-spectrum antibiotics to suppress the natural flora. Even then, exogenous species do not persist after discontinuation of antibiotic suppression[11]. As our donor strains did not readily colonize the mouse gut and transconjugants were lost soon after (Fig. 2b, Supplementary Figs. 9b and 14a), we reasoned that using a colonizing donor strain might extend the persistence of payload constructs in situ. To explore this possibility, we tested whether a mixed population of MGB strains (MGB3, MGB4, and MGB9) could stably recolonize the native mouse gut after a single oral dose without any antibiotic coadministration (Supplementary Fig. 15a). In contrast to the rapid disappearance of a non-gut-adapted strain (EcGT1) within 48 h, MGB strains (especially MGB4) recolonized the mouse gut and stably persisted for at least 15 d (Fig. 3b, Supplementary Fig. 15b), populating the entire gastrointestinal tract (Supplementary Fig. 15c). FACS enrichment and 16S sequencing of GFP-expressing bacteria in feces from these mice revealed transconjugants resulting from in situ transfer of the pGT payload from MGB strains to the native microbiome 6 h (Fig. 3c) and 11 d post-gavage (Supplementary Fig. 15d). These transconjugant populations had similar phylogeny but less diversity than those from prior in situ experiments using the noncolonizing EcGT2 donor (Fig. 2c, Supplementary Fig. 9c). These results highlight the utility of MAGIC for the isolation of host-derived engineerable strains that can be modified and then used to stably recolonize the native community and mediate further transfer of engineered functions in situ.

In summary, MAGIC enables metagenomic infiltration of genetic payloads into a native microbiome, and isolation of genetically modifiable strains from diverse communities. These strains can be reintroduced into their original community to maintain engineered functions via sustained vertical and horizontal transmission in situ. Future improvements to the system, such as optimization of vector stability and donor-strain dosage (Supplementary Fig. 14b), could allow for better quantitative and temporal control of retention of genetic payloads in situ, which might be useful in applications requiring short-term or long-term actuation of engineered functions[12–14]. Designing genetic programs based on recipient-specific properties should enhance the targeted execution of desired functions in a defined subset of species in a community[15,16]. MAGIC and complementary strategies to engineer the horizontal gene pool can facilitate programmable execution of genetic circuits in other microbial communities[17–20]. The isolation of genetically tractable representatives from diverse microbiomes will expand the repertoire of new microbial chassis for emerging applications in synthetic biology and microbial ecology.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41592-018-0301-y.

Received: 14 September 2018; Accepted: 12 December 2018;
Published online: 14 January 2019

References

1. Lagier, J. C. et al. *Nat. Microbiol.* **1**, 16203 (2016).
2. Yaung, S. J., Church, G. M. & Wang, H. H. *Methods Mol. Biol.* **1151**, 3–25 (2014).
3. Cuív, P. O. et al. *Sci. Rep.* **5**, 13282 (2015).
4. Mimee, M., Tucker, A. C., Voigt, C. A. & Lu, T. K. *Cell Syst.* **1**, 62–71 (2015).
5. Tock, M. R. & Dryden, D. T. *Curr. Opin. Microbiol.* **8**, 466–472 (2005).
6. Marraffini, L. A. *Nature* **526**, 55–61 (2015).
7. Thomas, C. M. & Nielsen, K. M. *Nat. Rev. Microbiol.* **3**, 711–721 (2005).
8. Human Microbiome Project Consortium. *Nature* **486**, 207–214 (2012).
9. Pansegrau, W. et al. *J. Mol. Biol.* **239**, 623–663 (1994).
10. Hapfelmeier, S. et al. *Science* **328**, 1705–1709 (2010).
11. Myhal, M. L., Laux, D. C. & Cohen, P. S. *Eur. J. Clin. Microbiol.* **1**, 186–192 (1982).
12. Kommineni, S. et al. *Nature* **526**, 719–722 (2015).
13. Saeidi, N. et al. *Mol. Syst. Biol.* **7**, 521 (2011).
14. Steidler, L. et al. *Science* **289**, 1352–1355 (2000).
15. Wegmann, U., Horn, N. & Carding, S. R. *Appl. Environ. Microbiol.* **79**, 1980–1989 (2013).
16. Sheth, R. U., Cabral, V., Chen, S. P. & Wang, H. H. *Trends Genet.* **32**, 189–200 (2016).
17. Klümper, U. et al. *ISME. J.* **9**, 934–945 (2015).
18. Dahlberg, C., Bergström, M. & Hermansson, M. *Appl. Environ. Microbiol.* **64**, 2670–2675 (1998).
19. Bikard, D. et al. *Nat. Biotechnol.* **32**, 1146–1150 (2014).
20. Brophy, J. A. N. et al. *Nat. Microbiol.* **3**, 1043–1053 (2018).

Acknowledgements

We thank members of the Wang lab for helpful scientific discussions and feedback, especially R. Sheth for helpful insights on NGS library preparation. H.H.W. acknowledges funding from DARPA (W911NF-15-2-0065), NIH (1DP5OD009172, 5R01AI132403, and 1R01DK118044), NSF (MCB-1453219), ONR (N00014-15-1-2704), and Burroughs Wellcome PATH (1016691). C.R. is supported by a Junior Fellowship of the Simons Society of Fellows (#527896). S.P.C. is supported by an NIDDK F30 fellowship (F30 DK111145-01A1) and an NIH MSTP training grant (NIH T32GM007367). We also thank A. Figueroa at the Columbia University Medical Center Flow Cytometry Core for assistance with FACS studies.

Author contributions

C.R., V.C., S.P.C., S.J.Y., and H.H.W. designed the study. C.R., S.P.C., and V.C. performed the experiments. C.R., S.P.C., V.C., and H.H.W. analyzed the data and wrote the manuscript, with input from all other authors.

Competing interests

A provisional patent application has been filed by the Trustees of Columbia University in the City of New York based on this work.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0301-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to H.H.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Media, chemicals, and reagents. *E. coli*, *Salmonella enterica*, *Vibrio cholerae*, and *Pseudomonas aeruginosa* strains were grown in rich LB-Lennox media (BD) buffered to pH 7.45 with NaOH in aerobic conditions at 37 °C. *Lactobacillus reuteri* was grown in MRS media (BD). *Bacteroides thetaiotaomicron* and *Enterococcus faecalis* were grown anaerobically at 37 °C in Gifu anaerobic modified medium (GAM) (Nissui Pharmaceutical) or BHI media (BD) supplemented with cysteine (1 g/liter), hemin (5 mg/liter), resazurin (1 mg/liter), and vitamin K (1 µl/liter). All gut bacteria used in the study were grown in LB-Lennox media or GAM. Antibiotics were used at the following concentrations to select for *E. coli*: chloramphenicol at 20 µg/ml, carbenicillin (carb) at 50 µg/ml, spectinomycin (spec) at 250 µg/ml, kanamycin at 50 µg/ml, tetracycline at 25 µg/ml, and erythromycin at 25 µg/ml. Antibiotics were used at the following ranges of concentrations to select for transconjugant gut bacteria: chloramphenicol at 5–20 µg/ml, carb at 10–50 µg/ml, tetracycline at 5–25 µg/ml. DAP was supplemented at 50 µM as needed.

Animal ethics statement. All animal experiments were performed in compliance with Columbia University Medical Center IACUC protocols AC-AAAU6464 and AC-AAAL2503.

Isolation of live mouse gut bacteria. Fresh fecal pellets were collected from mice, and live gut bacteria were isolated by mechanical homogenization. Briefly, 250 µl of PBS was added to previously weighed pellets in a microcentrifuge tube. Pellets were thoroughly mechanically disrupted with a motorized pellet pestle, and then 750 µl of PBS was added. The disrupted pellets in PBS were then subjected to four iterations of vortex mixing for 15 s at medium speed, centrifugation at 1,000 r.p.m. for 30 s at room temperature, recovery of 750 µl of supernatant in a new tube, and replacement of that volume of PBS before the next iteration. The resulting 3 ml of isolated cells were pelleted by centrifugation at 4,000*g* for 5 min at room temperature, the supernatant was discarded, and cells were resuspended in 0.5–1.0 ml of PBS. All gut bacteria isolations were performed in an anaerobic chamber (Coy Labs).

Donor strain construction. We derived donor strains EcGT1 and EcGT2 from the S17 λpir *E. coli* strain[21] by generating modifications Δ*galK::mCherry-specR* and Δ*asd::mCherry-specR*, respectively, with λ-red recombineering using the pKD46 system[22]. Synthetic cassettes containing constitutively active mCherry and spec-resistance genes were constructed with ~40 bp of homology on both ends to *galK* or *asd* flanking regions on the *E. coli* genome. 100 ng of mCherry-specR cassette DNA were electroporated into recombineering-competent S17-pKD46 cells. Cells were allowed to recover in 3 ml of LB media plus carb at 30 °C for 3 h before being plated on LB media plus spec. Spec-resistant colonies were genotyped by PCR for validation of mutations. The pKD46 recombineering plasmid was cured out of validated recombinants by growth at 37 °C in the absence of carb to yield the EcGT1 and EcGT2 strains used throughout the study. When generating the EcGT2 strain, we supplemented the growth media with DAP at all stages of the protocol.

Plasmid construction. pGT vectors were designed to have modular components (e.g., selectable markers, regulatory elements, replication origins) that are interchangeable by isothermal assembly (ITA) or Golden Gate assembly. Vector selection markers for *E. coli* were constitutively expressed, whereas the deliverable cargo and transposase cassettes were expressed using different regulatory elements to enable broad-host-range or narrow-host-range gene expression. The regulatory elements used in this study exhibit a range of activity (Supplementary Table 1). Vector libraries used in this study are detailed in Supplementary Table 2. Full vector component sequences are listed in Supplementary Table 3. The nontransferable vector pGT-NT used as a negative control was a minimal p15A cloning vector with no origin of transfer, containing a constitutively expressed sfGFP gene.

All plasmids were constructed by ITA with NEBuilder HiFi DNA assembly master mix (New England Biolabs). Component parts were made by high-fidelity PCR with Q5 (NEB) or KAPA Hifi (Kapa Biosystems) polymerase, using existing vectors or gBlocks (Integrated DNA Technologies) as PCR templates. PCR products were digested with DpnI (NEB) and purified with the QIAquick PCR purification kit (Qiagen) before ITA and transformation into *E. coli*. All assembled plasmids were Sanger-sequence verified.

In vitro MAGIC studies on synthetic recipient community. Donor strains harboring pGT vectors and representative recipients (*E. coli* MG1655, *S. enterica* ATCC 700931, *V. cholerae* C9503, *P. aeruginosa* PA01, *E. faecalis* ATCC 29200, *L. reuteri* ATCC 23272, *B. thetaiotaomicron* ATCC 29148) were grown overnight in appropriate media and cultivation conditions, and a 1:1,000 dilution culture was regrown for 14 h at 37 °C before conjugation studies. To prepare cells for in vitro conjugation, we washed donor and recipient populations twice in PBS and quantified cells by OD_{600} or flow cytometry using SYTO9 staining (Thermo Fisher). 10^8 donor cells and 10^8 recipient cells were mixed together, pelleted by centrifugation, and resuspended in 10 µl of PBS. Donor and recipient mixes were spotted on an agar plate and incubated for 5 h at 30 °C or 37 °C for conjugation. In vitro conjugations were performed on LB-Lennox (*E. coli*, *S. enterica*, *V. cholerae*, *P. aeruginosa*, *E. faecalis*), MRS media (*L. reuteri*), or supplemented BHI agar (*B. thetaiotaomicron*). After conjugation, cells were scraped from the plate into 1 ml of PBS, and 100 µl was plated on appropriate antibiotics and incubated overnight at 30 °C or 37 °C so we could determine the number of colony-forming units (CFUs) of transconjugants.

In vitro MAGIC studies on natural recipient community. Donor strains harboring pGT vectors were streaked onto LB-Lennox agar plates with appropriate antibiotics and supplements, grown at 37 °C overnight, and then grown from a single colony in 2 ml of liquid media for 10 h at 37 °C before conjugation. The recipient community was isolated anaerobically from fresh mouse feces as described above, immediately before conjugation. Donor cells were washed twice in PBS and quantified by OD_{600}, whereas recipient cells were quantified by flow cytometry using SYTO9 staining. 10^8 donor cells and 10^9 recipient cells were mixed, pelleted by centrifugation at 5,000*g*, and resuspended in 25 µl of PBS. The mixes were spotted on PBS + 1.5% agar plates and incubated at 37 °C either aerobically or anaerobically overnight (9–10 h). After conjugation, cells were scraped from the plate into 1 ml of PBS and subjected to antibiotic selection on GAM, FACS enrichment, and metagenomic 16S analysis (see below).

In vitro assessment of pGT vector horizontal gene transfer mediated by natural isolates. MGB natural isolates harboring pGT vectors (MGB3, MGB9, and MBG4) were conjugated with a recipient *E. coli* strain harboring a kanamycin-resistance plasmid compatible with pGT vectors. Prior to conjugations, all strains were streaked onto GAM agar plates with appropriate antibiotics, grown at 37 °C overnight, and then grown from a single colony in 5 ml of liquid GAM for 10 h at 37 °C before conjugation. MGB donor and recipient cells were washed twice in PBS and quantified by OD_{600}. 10^9 cells each of MGB and recipient strains were mixed, pelleted by centrifugation at 5,000*g*, and resuspended in 15 µl of PBS. The mixtures were spotted on GAM agar plates and incubated at 37 °C aerobically for 6 h. After conjugation, cells were scraped from the plate into 1 ml of PBS and plated on selective and nonselective GAM. Conjugation efficiency was calculated as t/n, where t is the number of *E. coli* transconjugant CFUs and n is the total number of *E. coli* CFUs.

Measurement of GFP expression in MGB strains. MGB isolates harboring pGT vectors (MGB3, MGB9, MBG4) were streaked onto GAM agar plates with appropriate antibiotics, grown at 37 °C overnight, and then diluted to OD_{600} 0.001 in liquid GAM into a 96-well plate. The plate was incubated in a Synergy H1 (BioTek) microplate reader for 24 h at 37 °C with orbital shaking. Measurements of OD_{600} and GFP expression (excitation, 488 nm; emission, 510 nm) were acquired with Gen5 software (BioTek) at the end of 24 h.

In vivo MAGIC studies in mice. Conventionally raised C57BL/6 female mice (Taconic Biosciences or Charles River Laboratories) were used throughout the study. Two control groups of four mice each were gavaged with PBS and EcGT2 containing a nontransferable GFP vector (pGT-NT). Three to four mice were used in each group gavaged with a pGT donor mix or with MGB strains. To equilibrate the mouse gut microbiome ahead of time, we mixed mice from multiple litters, cohoused them for at least 1 week before all experiments, and randomly allocated them into groups. Mice were gavaged with 10^9 donor cells (EcGT2 or MGB strains) in 300 µl of PBS at 8–10 weeks old. Control mice were gavaged with 300 µl of PBS. Fecal matter was collected immediately before gavage and periodically after gavage for analysis of the resulting microbiome populations by FACS, metagenomic 16S sequencing, and plating. Upon completion of the study, mice were euthanized, and small and large intestinal tissues were extracted. Luminal contents were washed from each tissue sample with PBS, and bacteria were extracted by homogenization of the luminal contents for plating and final CFU determination.

Flow cytometry and FACS measurements. Gut bacteria isolated from fresh fecal pellets were analyzed for evidence of successful conjugation on a flow cytometer (Guava easyCyte HT) using red (642 nm) and blue (488 nm) lasers with Red2 and Green photodiodes to detect mCherry (587/610 nm) and sfGFP (485/510 nm) fluorescence, respectively. Bacteria at 100× and 1,000× dilutions in PBS were used for optimal detection of donor material ($GFP^+mCherry^+$), gut microbes without a transferred vector ($GFP^-mCherry^-$), and transconjugants ($GFP^+mCherry^-$). Data were collected and analyzed with InCyte 3.1 software. For FACS enrichment studies, a BD FACSAria II cell sorter operated with BD FACSDiva software was used to gate for sfGFP (FITC filter 515/10 nm) and mCherry (mCherry filter 616/26 nm). Double-gating on GFP and mCherry channels was used to select for cells with $GFP^+mCherry^-$ fluorescence. In addition, we took background events into account by using the $GFP^+mCherry^-$ fluorescence detected in the fecal sample before gavage as the baseline signal. An increase over the baseline signified an enrichment of transconjugants. Population density (cells per gram of fecal matter) was calculated as the number of cells sorted over the mass of the sorted fecal sample. Additional plating and direct colony counting were used to validate flow cytometry measurements. FACS plots were formatted with FCS Express 6.

Fluorescence microscopy of fecal bacteria. We suspended bacteria in PBS and centrifuged them at 5,000*g* to concentrate them into a smaller volume, which

varied depending on the concentration of bacteria. The bacteria were resuspended by pipetting, and a volume of 15 µl was dropped onto a Superfrost Plus microscope slide (Thermo Shandon) and covered with a glass coverslip. Slides were air-dried until the PBS receded from the edges of the coverslip and then were sealed with clear nail polish. Bacteria were imaged at 40× magnification on a Nikon Eclipse Ti2 microscope on bright-field, RFP, and GFP channels using NIS-Elements-AR software.

Validation of pGT vectors in transconjugants. Transconjugant validation was done by colony PCR of the GFP–antibiotic resistance payload and/or the pGT vector backbone. PCR products with the expected size were further verified by Sanger sequencing. Taxonomy assignment of isolated colonies was based on 16S rRNA PCR amplification and Sanger sequencing. All transconjugant strains validated in the study are listed in Supplementary Table 4.

In vitro evolution of transconjugant gut bacteria. *E. fergusonii* transconjugants MGB4 and MGB9 were serially passaged in LB media for 11–15 d. Starting from a single colony, the strains were inoculated into LB and grown at 37 °C with shaking. Every 12 h the liquid culture was diluted 1:1,000 into fresh LB media. At selected time points an aliquot of the saturated culture was plated on selective (50 µg/ml carb) and nonselective plates for quantification of the percentage of cells expressing the payload antibiotic-resistance and GFP genes. MGB9 cultures were also plated on selective plates with 20 µg/ml chloramphenicol to check for maintenance of the plasmid backbone.

Metagenomic 16S sequencing. Genomic DNA was extracted from isolated bacteria populations with the MasterPure Gram-positive DNA purification kit (Epicentre). PCR amplification of the 16S rRNA V4 region and multiplexed barcoding of samples were done in accordance with previous protocols[23]. The V4 region of the 16S rRNA gene was amplified with customized primers according to the method described by Kozich et al.[23], with the following modifications: (i) alteration of 16S primers to match updated EMP 505f and 806rB primers[24–26] and (ii) use of NexteraXT indices such that each index pair was separated by a Hamming distance of >2 and Illumina low-plex pooling guidelines could be used. Sequencing was done with the Illumina MiSeq system (500V2 kit).

Analysis of 16S next-generation sequencing data. Bacteria from fecal samples taken right before gavage (T0) and 6 h post-gavage (T6) were sorted by FACS to enrich for transconjugants. The compositions of the sorted transconjugant and total populations for each sample were determined from 16S sequencing data via the UPARSE pipeline[27] (USEARCH version 10.0.240) to generate operational taxonomic unit (OTU) tables and abundances and the RDP Classifier[28] to assign the taxonomy. Phylogenetic associations were analyzed at the genus level with at least 90% confidence for 16S assignment. In all MiSeq runs, two blank controls with sterile water as input material were included to check for contaminants in the reagents and to filter out contaminant OTUs if present. Reads mapping to nonbacterial DNA (e.g., mitochondria, plastids, or other eukaryotic DNA) were also excluded from analysis. Only OTUs with more than ten reads were considered in downstream analysis.

Relative abundances of OTUs in unsorted total fecal populations were calculated as the normalized number of reads in a sample. Relative abundances of OTUs in T0 FACS-enriched populations were used to measure false positive background fluorescence, which was subtracted from the fluorescence of T6 transconjugant populations. The corrected relative abundance of each OTU in a T6 FACS-enriched population is given by the following formula:

$$\mathrm{RA}_{6,i,\mathrm{sorted}} = \frac{A_{6,i} \times N_6 - A_{0,i} \times N_0}{\sum_i (A_{6,i} \times N_6 - A_{0,i} \times N_0)}$$

where $\mathrm{RA}_{t,i,\mathrm{sorted}}$ is the corrected relative abundance of OTU i at time t, $A_{t,i}$ is the normalized number of reads of OTU i at time t in the FACS-sorted sample, and N_t is the fraction of GFP$^+$mCherry$^-$ FACS-sorted events at time t. OTUs for which $\mathrm{RA}_{6,i,\mathrm{sorted}}$ is negative are eliminated from subsequent analysis, and all remaining $\mathrm{RA}_{6,i,\mathrm{sorted}}$ values are renormalized.

The fold enrichment of each OTU in the FACS-sorted population is defined as its relative abundance in the FACS-sorted population divided by its relative abundance in the unsorted total population at T6. To overcome the problem of detection limits (i.e., OTU i appears in the sorted population but is present at levels below the detection limit in the total population), we added a pseudo-count of p to all relative abundances when calculating fold enrichments. p is given by

$$p = 10^{\lfloor -\log_{10} n \rfloor}$$

where n is the total number of reads in the FACS-sorted sample, and $\lfloor -\log_{10} n \rfloor$ is the floor function of (i.e., the greatest integer less than or equal to) $-\log_{10} n$. The fold enrichment of OTU i with the pseudo-count correction is calculated as

$$F_i = \frac{\mathrm{RA}_{6,i,\mathrm{sorted}} + p}{\mathrm{RA}_{6,i,\mathrm{unsorted}} + p}$$

If the relative abundance of OTU i in the unsorted population is below the detection limit, then the fold enrichment is calculable as $(\mathrm{RA}_{6,i,\mathrm{sorted}} + p)/p$, instead of $\mathrm{RA}_{6,i,\mathrm{sorted}}/0$.

The pseudo-count-corrected fold enrichment F_i overestimates the true fold enrichment ($\mathrm{RA}_{6,i,\mathrm{sorted}}/\mathrm{RA}_{6,i,\mathrm{unsorted}}$) by at most 10%, or possibly underestimates it. Because $0 < p \leq 1/n$ and $\mathrm{RA}_{6,i,\mathrm{sorted}} \geq 10/n$,

$$F_i = \frac{\mathrm{RA}_{6,i,\mathrm{sorted}} + p}{\mathrm{RA}_{6,i,\mathrm{unsorted}} + p} \leq \frac{\mathrm{RA}_{6,i,\mathrm{sorted}} + p}{\mathrm{RA}_{6,i,\mathrm{unsorted}}} \leq \frac{1.1 \times \mathrm{RA}_{6,i,\mathrm{sorted}}}{\mathrm{RA}_{6,i,\mathrm{unsorted}}}$$

In all heat maps showing fold enrichment versus relative abundance, only OTUs with $F_i > 10$ are displayed, to show more stringent and high-confidence results. R code for this analysis is available from the corresponding author upon request.

Whole-genome sequencing of engineered mouse gut bacteria isolates. To sequence MGB isolates, we prepared a sequencing library using the Nextera kit (Illumina) and used the Illumina HiSeq 2500 platform for 100-bp single-end reads. The SPAdes single-cell assembler pipeline (version 3.9.1)[29] was used to generate whole-genome contigs. BLAST and PlasmidFinder (version 1.3)[30] were used to analyze the sequences and identify native mobilization systems. Geneious (version 7.1.5) was used to visualize contig alignments to genomes and plasmids.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Materials availability. All modular vector part sequences are listed in Supplementary Table 3. Full plasmid maps, vectors, and strains used in this study are available from the corresponding author upon request or will be available on Addgene.

Data availability

The raw data from this study are available from the corresponding author upon request.

References

21. Simon, R., Priefer, U. & Pühler, A. *Nat. Biotechnol.* **1**, 784–791 (1983).
22. Datsenko, K. A. & Wanner, B. L. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
23. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
24. Caporaso, J. G. et al. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).
25. Parada, A. E., Needham, D. M. & Fuhrman, J. A. *Environ. Microbiol.* **18**, 1403–1414 (2016).
26. Apprill, A., McNally, S., Parsons, R. & Weber, L. *Aquat. Microb. Ecol.* **75**, 129–137 (2015).
27. Edgar, R. C. *Nat. Methods* **10**, 996–998 (2013).
28. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
29. Nurk, S. et al. *J. Comput. Biol.* **20**, 714–737 (2013).
30. Carattoli, A. et al. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).

nature research

Corresponding author(s): Harris H. Wang

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Flow cytometry was performed using InCyte 3.1 software on the Guava easyCyte HT flow cytometer. BD FACSDiva software was used for FACS on the BD FACSAria II cell sorter. NIS-Elements-AR software was used for fluorescence microscopy. Gen5 software was used to operate the plate reader for measurement of GFP expression in isolate strains.
Data analysis	All 16S data were processed using the UPARSE pipeline and the RDP classifier (USEARCH v.10.0.240) and subsequently analyzed in R, using the calculations stated in the Methods section. For whole genome sequencing assembly, we used SPAdes (v.3.9.1) software to generate contigs and then performed sequence analysis using BLAST, PlasmidFinder(v.1.3), and Geneious (v.7.1.5). FCS Express 6 was used for formatting FACS plots.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data and analysis code are available from the corresponding author upon request

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes of mice were chosen to ensure that the effect of treatment with the engineered bacteria was robust and replicable. At least 3 mice were used for each treated group, and at least 2 mice were used for control groups.
Data exclusions	No data were excluded from the analysis
Replication	All attempts at replication were successful. We ran multiple iterations of the study using different cohorts of mice, with multiple mice in each treatment group (see samples size above).
Randomization	Mice were randomly allocated to different treatments. We ensured that animals shipped to the animal facility in different cages were mixed appropriately in order to avoid microbiome cage bias.
Blinding	Blinding does not apply to this study because the investigators needed to identify the cages of mice for subsequent FACS sorting and analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All strains used in the study are available upon request from the corresponding author.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	7-8 week old C57BL/6 female mice from Taconic and Charles River Laboratories were used.

Wild animals	No wild animals were used.
Field-collected samples	The study did not involve samples collected from the field.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Bacteria were extracted from murine feces as described in the Methods section by resuspension in PBS and filtration. The bacteria were run directly on the flow cytometer/cell sorter without additional treatment.
Instrument	BD FACSAria II, Guava easyCyte HT
Software	BD FACSAriaII was operated using BD FACSDiva. Guava easyCyte HT was operated using InCyte3.1. FCS Express 6 software was used to format FACS plots.
Cell population abundance	Representative population abundances pre- and post-sorting are shown in the manuscript. The purity of samples is addressed in the manuscript, as autofluorescent cells were filtered out of the post-sort population.
Gating strategy	FSC/SSC gates were determined by comparison of fecal bacterial samples and in vitro cultures of E. coli against the PBS background to gate in the signal for live bacteria and exclude noise. GFP and mCherry gates were set by comparing GFP+/mCherry+ E. coli, GFP+/mCherry- E. coli, GFP-/mCherry+ E. coli, and GFP-/mCherry- E. coli. To minimize sorting of autofluorescent fecal bacteria, we adjusted the fluorescence gates to stringently gate out the natural murine gut bacterial community.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.