nature research

Corresponding author(s): Martin F Polz

Last updated by author(s): Dec 12, 2019

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.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
☐	☒ A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All publicly available microbial community profiling data analyzed in this study was downloaded from National Center for Biotechnology Information, European Nucleotide Archive, or MG-RAST databases. All microbial community profiling data generated in this study was generated using an Illumina MiSeq 2x250 paired-end at the MIT BioMicroCenter and was processed using custom code available on GitHub at https://github.com/polzlab/VanInsberghe_2019_Cdiff_colonization. The genome of C. difficile BAA1801 was sequenced on a Illumina NextSeq paired-end 2x150 bp to 60x coverage and assembled using CLC genomics workbench version 8 (CLC Bio, Denmark).
Data analysis	Sequencing reads for microbial community profiling analyses were processed using Vxtractor V2.1, MOTHUR V1.35.1, NCBI blast V2.2.30, and dada2. All custom code generated during this study are accessible on GitHub at https://github.com/polzlab/VanInsberghe_2019_Cdiff_colonization. Image processing was performed using the R package EBImage v4.24.0. Primer design was performed using Primer3 v4.1.0. Gene alignment was performed using MUSCLE v3.8.31 Whole genome alignments were performed using MUGSY v1.2.3.

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. 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

The publicly available data that we re-analyzed here were generated by David et al accessible on the European Nucleotide Archive (ENA) under the accession number ERP006059, Hsaio et al on the NCBI Short Read Archive (SRA) under the accession number PRJEB6358, by Caporaso et al on MG-RAST under the accession number MG-RAST:4457768.3-4459735.3, Parker et al on the SRA under accession number ERP022953, and The NIH Human Microbiome Project on NCBI under the

accession number PRJNA43017. The amplicon sequencing data from the mouse model of varied disturbance intensity are available on the NCBI SRA under the accession number PRJNA507320. The genome of C. difficile BAA1801 is available on NCBI under the accession number PRJNA562328.All custom computer code necessary to reproduce our results are available on GitHub (https://github.com/polzlab/VanInsberghe_2019_Cdiff_colonization), and all source data necessary to reproduce the figures presented in this manuscript are available online at figshare (https://figshare.com/projects/Source_Data_for_Diarrheal_events_can_trigger_long-term_Clostridium_difficile_colonization_with_recurrent_blooms/72824).

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	For our animal study, four groups of 10 mice were included. Groups of 10 mice were chosen to increase the chance of detecting blooms in low PEG3350 treatment groups. This group size was sufficient for the purposes of this study because it provided statistically significant differences between treatment groups.
Data exclusions	No data were excluded from our analyses.
Replication	All experimental findings related to our model of varied disturbance intensity in mice were internally consistent with one another since the measured effect of treatment scaled with the gradient of treatment intensity. To validate the reproducibility of our observations in this experimental model, we performed re-analysis of five additional data sets derived from different human patient cohorts, including data from over 500 people and over 3000 individual samples. All quantitative PCRs were performed in triplicate, all automated cell counts were performed on a minimum of 20 microscope images per sample, all colony forming unit counts were performed in duplicate, and tests of DNA extraction efficiency were performed in triplicate.
Randomization	All mice used in this study were randomly assigned to the treatment groups.
Blinding	The investigators performing the manipulations of the model of varied disturbance intensity in mice were not blinded to the treatment groups because the same person needed to collect samples and perform the experimental manipulations. The histopathological scoring of tissues from this animal model were performed blind, however.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

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

Methods

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

Animals and other organisms

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

Laboratory animals	Our animal study used 40 male 6 week old C57BL/6J mice from Jackson Laboratories.
Wild animals	This study did not involve wild animals.
Field-collected samples	No samples were collected from the field for this study.
Ethics oversight	All experiments utilizing animals were approved by the Massachusetts Institute of Technology's Committee on Animal Care.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Data from human research participants was obtained from publicly available sources online. Demographics from these studies can be found through their original publications.
Recruitment	Information on the recruitment methods of each study can be found through their original publications (see above).
Ethics oversight	Information on the ethical oversight of each study can be found through their original publications (see above).

Note that full information on the approval of the study protocol must also be provided in the manuscript.