nature portfolio

Corresponding author(s): Bernd Schnabl and Libby Hohmann

Last updated by author(s): 11/04/2025

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	Microsoft Excel was used to collect all patient data.
Data analysis	Data analysis was performed using R statistical software (R version 4.2.2; R Foundation for Statistical Computing). Metagenomic sequencing data was pre-processed with KneadData (KneadData version 0.12.0, Trimmomatic version 0.33, Bowtie version 2.5.3), taxonomic and functional profiling was performed using the HUMAnN 3 (version 3.8) pipeline with MetaPhlAn 4 (version 4.0.6, database version mpa_vOct22_CHOCOPhlAnSGB_202212) and Diamond (version 2.1.9, full UniRef90 database). Mapping to MetaCyc reactions, KEGG orthogroups (KOs), Pfam domains, and Gene Ontology (GO) databases was performed and MaAsLin2 was used to identify associations between host phenotype and gut microbial taxonomic and metagenomic features. A minimum abundance threshold of 0.00001 and a minimum prevalence of 20% was used. Strains of Escherichia coli were identified in the metagenomes and compared across samples using StrainPhlan6,7, a program within MetaPhlan (v4.1.1), and PanPhlan (v3.1)5,8, respectively. StrainPhlan was used to align metagenome reads to E. coli consensus marker gene sequences from the marker database ChocoPhlan version mpa vJune23CHOCOPhlAnSGB_202403 (provided by MetaPhlan) and a multiple sequence alignment (MSA) was generated. The MSA was used tc create a phylogenetic tree of samples containing the E. coli marker genes based on their phylogenetic similarity to the marker genes. The MSA and the phylogenetic tree were visualized using the ggtree (v3.14.0) package in R (v4.4.3). The length of the MSA was 2,996 nucleotides. PanPhlan was used to determine the depth of coverage for metagenome reads mapped to 10,834 E. coli reference genomes representing the E. coli pangenome from the UniProt Reference Clusters (UniRef) database. A presence/absence table was also generated based on the depth of coverage of the samples' alignment to the E. coli reference genomes.

SameStr (v1.2024.08)9 was used to determine the presence of bacterial strains in our samples, using the marker database ChocoPhlan version mpa vJune23CHOCOPhlAnSGB_202403 from MetaPhlan. Specifically, SameStr compared the samples in a pair-wise fashion to determine if two samples shared specific clades and/or strains with one another. SameStr was used to examine engraftment in our FMT patient. Clades at the species genome bin (i.e., SGB) level were compared between pre-FMT samples, post-FMT samples and the donor. The fraction of clades found to be shared between the pre-FMT and post-FMT samples (i.e., the number of shared clades found in the post-FMT sample divided by the total number of clades in the pre-FMT sample) and the fraction of clades shared between the donor and the post-FMT samples (i.e., the number of shared clades found in the post-FMT sample divided by total the number of clades in the donor) were determined. Percent engraftment was calculated by considering the number of shared strain events between post-FMT samples and the donor, such that each shared strain event was attributed to a clade not found in the pre-FMT sample.

ITS2 amplicon sequences were quality-assessed using FastQC, then trimmed and filtered with BBduk After trimming, the Divisive Amplicon Denoising Algorithm 2 (DADA2 version 1.26) pipeline was used to assign reads to amplicon sequence variants using the UNITE fungal ITS database (version 10, published on 04-04-2024) as the reference. The compositional relative abundance across samples was calculated and fungal genera associations were compared across the patient groups using Maaslin2. The arcsine square root transformation was used by Maaslin2 to transform the fungal genera relative abundances before running a linear mixed effects model. A minimum abundance threshold of 0.00001 and a minimum prevalence of 20% was used.

To analyze beta diversity, Aitchison distances were calculated by transforming amplicon sequence variant counts with a center-log ratio transformation, then calculating the Euclidean distance. Aitchison distance was input for principal coordinates analysis (PCoA). Permutational multivariate analyses of variance (PERMANOVA) were performed to determine if there were significant differences in microbial composition across groups.

For metabolomics analysis, metabolites detected in less than 95% of samples were excluded. A centered log-ratio transformation was performed on metabolomics data prior to calculating spearman correlations between metabolites and differentially abundant microbial metabolic pathways. P-values were adjusted by Benjamini-Hochberg procedure to control false discovery rate.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Shotgun metagenomics and ITS2 sequencing datasets have been uploaded to the Sequence Read Archive (SRA) under BioProject accession number PRJNA1257465. Metabolomics data are available upon request.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Findings apply to both sexes. Gender data was not elicited.
Population characteristics	Participants' sex, age, body mass index, dietary habits, and use of tobacco, alcohol, antibacterial medications, antifungal medications, and probiotics were collected.
Recruitment	Patients were recruited through the Auto-brewery Syndrome support group Facebook page and via patient advocates through the Auto-Brewery Syndrome Advocacy and Research program.
Ethics oversight	The Auto-brewery Syndrome Registry was approved by the UCSD Institutional Review Board, and written informed consent was obtained from each participant. IRB and FDA approval (Single Patient Expanded Access IND 28390/PI Hohmann/ Massachusetts General Hospital) was obtained for single donor, capsule-based fecal transplant.

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

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 power calculations, we used an estimated effect size of 1.5 based on preliminary culture data and determined that a sample size of 14 patients and 14 controls would achieve 95% power using Wilcoxon-Mann-Whitney test with α=0.05.
Data exclusions	Patients with Auto-brewery Syndrome who were screened for our study but did not meet the criteria of documented serial rise in blood alcohol concentration while in a supervised clinical setting in the documented absence of alcohol consumption were excluded from this study.
Replication	Ethanol concentrations were measured in duplicate per culture sample by HPLC or by ELISA assay and the results were averaged per sample.
Randomization	Randomization is not relevant to this study.
Blinding	Investigators were blinded to group allocation during HPLC analysis of ethanol concentration from bioreactor culture samples and during the nucleic acid extraction, DNA quantification, and sequencing steps of the metagenomic sequencing analysis.

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 and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	FMT trial was performed with FDA approval (Single Patient Expanded Access IND 28390/PI Hohmann/Massachusetts General Hospital)
Study protocol	FMT trial was performed under IND for one patient.
Data collection	After subjects were consented, we confirmed that patients with Auto-brewery Syndrome included in our study all had documentation of serial rise in blood alcohol concentration while in a supervised clinical setting in the documented absence of alcohol consumption. Baseline demographic data and laboratory test results were collected and a comprehensive clinical history was obtained from each participant. At least 10 grams of stool was collected from each patient during remissions and flares of clinical intoxication. Samples were collected from controls within 24 hours of patient samples. Concurrent blood alcohol concentrations were measured by an FDA-approved portable breathalyzer (BACtrack, San Francisco, CA), and results were recorded by the patient.
Outcomes	"Flare" was defined as periods when patients had symptoms of intoxication or had a detectable blood alcohol level. "Remission" was defined as having no symptoms nor detectable blood ethanol levels for at least one week.