

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☐	☒ 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 <i>d</i> , Pearson's <i>r</i>), 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	Google Sheets.
Data analysis	Trim Galore (v0.6.6) and Bowtie2 (v2.3.4.3) for metagenomic reads quality check and preprocessing; MetaPhlAn v4.1 with vJun23_202307 markers db, StrainPhlAn v4.1, CroCoDeEL v1.0.6, curatedMetagenomicData v3.18; R packages ggraph (v2.2.1) and tidygraph (v1.3.1); assembly of contigs with MEGAHIT (v1.1.1), Bowtie2 (v2.2.9), SAMtools (v1.19), bedtools (v2.30), MetaBAT2 (v2.12.1) , and quality-check of bins with CheckM (v1.1.3), skani (v.0.2.1), MinCED (v0.4.2), Crass (v1.0.1); Blastocystis detection and profiling with Bowtie2 (v2.5), SAMtools (v1.19) and bedtools (v2.30); SGB9226-specific primers design with ConsensusPrime (v. 1.0) and Primer3 (v. 2.6.1); Python (v3.10.12) libraries scikit-bio (v0.5.9), scipy (v1.10.1) and statsmodels (v0.14.0). The script for strain transmission inference utilized in this study is available at https://github.com/biobakery/MetaPhlAn .

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](#)

Data availability: Shotgun metagenomic data for microTOUCH-baby are available at the NCBI-SRA under accession number PRJNA1140720, with respective sample-wise metadata information available in Supplementary Table 2. Accession numbers for sequencing data from additional longitudinal datasets can be found in the original publications. The human genome release used for host decontamination is available at NCBI RefSeq (GCF_000001405.13).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\)](#), [and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	We collected participant sex information; Participants included females (N=79) and males (N=55); we provide partial details in Suppl. Table 3.
Reporting on race, ethnicity, or other socially relevant groupings	not applicable-We did not collect data on race, ethnicity or other socially relevant groupings.
Population characteristics	Participants were N=134, including 43 infants (median age=10 months at baseline), their siblings (N=7; median age=2 y), mothers and fathers (N=39 and 30, median age parents=36 y), their pets (2 cats and 3 dogs), and 10 educators (median age=38.5). Details in Suppl. Table 3.
Recruitment	Participants were recruited and enrolled in 3 nurseries (identified as A, B, C) in the municipality of Trento (Italy) via direct oral presentation of study objectives and framework in June 2022. We acknowledge our study is susceptible to two possible selection biases: 1. Self-selection bias, where voluntary participation may skew the cohort toward more health-aware families, and 2. Socioeconomic confounding due to nursery location, as families typically attend the nursery closest to their residence or workplace, potentially linking attendance to slightly different socioeconomic groups. To mitigate the latter, we pre-selected nurseries located strategically across the town center, one outside the center, and one in a close-by village within the municipality of Trento; this, combined with our focus on relative comparisons of transmission dynamics within and between these groups, supports the robustness of our conclusions regarding the impact of social proximity.
Ethics oversight	The protocol of this study was approved by the Ethics Committee of the University of Trento (protocol nr. 2022-040) and by the Ethics Panel of the European Research Council Executive Agency upon evaluation of the project (Grant agreement ID 101045015).

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](https://www.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	We collected and utilized n=1013 fecal metagenomes from N=134 participants in this longitudinal study. As this was an observational, ecologically constrained study focusing on the first year of nursery, we did not perform a formal a priori cohort size calculation. Instead, our strategy was based on maximum achievable recruitment within the existing units in each nursery. We targeted the entire available population across three large-sized nurseries (with two classes each, with a maximum capacity of 12 babies per class). We successfully recruited in the study 43 babies out of the 71 enrolled in these specific units, representing more than 60% of the eligible population. In addition, we recruited 3–4 educators per nursery, which represented the maximum number of staff directly attending these babies (2 educators per class). This approach, together with dense longitudinal sample collection, allowed us to successfully enroll a sufficient and robust sample size for the comparative analyses of strain sharing both within and between the defined nursery groups, capturing the majority of the relevant social interactions. In addition to the 1,031 fecal samples and 342 oral samples collected in this study, 4,322 samples from 16 public longitudinal cohorts were used to build phylogenetic trees: ShaoY_2019, MehtaRS_2018, VatanenT_2016, HMP_2019_ibdmdb, BackhedF_2015, CosteaPI_2017,
-------------	---

YassourM_2018, KosticAD_2015, LouisS_2016, FerrettiP_2018, HallAB_2017, WampachL_2018, NielsenHB_2014, Heitz-BuschartA_2016, ChuDM_2017, AsnicarF_2017.

Data exclusions	We excluded metagenomes with insufficient sequencing depth (n=5) and with contaminations identified during strain sharing analysis (n = 8). Samples collected during antibiotic treatment (n = 26) were excluded from strain-sharing analyses.
Replication	not applicable-not needed for this observational metagenomic survey.
Randomization	not applicable-not needed for this observational metagenomic survey.
Blinding	not applicable-not needed for this observational metagenomic survey.

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
☒	☐ Plants

Methods

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

Plants

Seed stocks	not applicable
Novel plant genotypes	not applicable
Authentication	not applicable