Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria

Fabian Imdahl\textsuperscript{1,3}, Ehsan Vafadarnejad\textsuperscript{1,3}, Christina Homberger\textsuperscript{1}, Antoine-Emmanuel Saliba\textsuperscript{1,5} and Jörg Vogel\textsuperscript{1,2,5}

Bacteria respond to changes in their environment with specific transcriptional programmes, but even within genetically identical populations these programmes are not homogeneously expressed\textsuperscript{1}. Such transcriptional heterogeneity between individual bacteria allows genetically clonal communities to develop a complex array of phenotypes\textsuperscript{2}, examples of which include persisters that resist antibiotic treatment and metabolically specialized cells that emerge under nutrient-limiting conditions\textsuperscript{3}. Fluorescent reporter constructs have played a pivotal role in deciphering heterogeneous gene expression within bacterial populations\textsuperscript{4} but have been limited to recording the activity of single genes in a few genetically tractable model species, whereas the vast majority of bacteria remain difficult to engineer and/or even to cultivate. Single-cell transcriptomics is revolutionizing the analysis of phenotypic cell-to-cell variation in eukaryotes, but technical hurdles have prevented its robust application to prokaryotes. Here, using an improved poly(A)-independent single-cell RNA-sequencing protocol, we report the faithful capture of growth-dependent gene expression patterns in individual Salmonella and Pseudomonas bacteria across all RNA classes and genomic regions. These transcriptomes provide important reference points for single-cell RNA-sequencing of other bacterial species, mixed microbial communities and host-pathogen interactions.

Single-cell RNA-sequencing (scRNA-seq) is becoming routine practice for examination of eukaryotic cells, including small-sized species such as yeasts\textsuperscript{5} and protozoa\textsuperscript{6}. This technique has catalysed the discovery of new cell types and provided an unprecedented view of tissue anatomy and cellular transitions\textsuperscript{6}. Pioneering attempts notwithstanding, however, the application of scRNA-seq to single\textsuperscript{7,8} or small numbers of bacteria\textsuperscript{9,10} remains unproven in regard to its ability to infer the physiological state of individual cells within a population\textsuperscript{11,12}.

Several important technical barriers exist regarding bacterial scRNA-seq. Because bacterial cells contain only femtograms of RNA\textsuperscript{13}—that is, >100-fold less than typical eukaryotic cells—a sensitive complementary DNA synthesis and amplification protocol is required. In addition, bacterial messenger RNAs are intrinsically labile (with half-lives of several minutes, as compared to hours for eukaryotes), meaning that perforation of the thick bacterial envelope, cell lysis and subsequent RNA stabilization must all occur rapidly. Most importantly, the absence of a poly(A) tail in functional bacterial transcripts precludes the straightforward reverse transcription (RT) strategy that is commonly used for eukaryotes to selectively enrich mRNAs and concomitantly deplete ribosomal RNAs. Lastly, whereas most current eukaryotic scRNA-seq protocols have a lower detection limit of between five and ten copies per transcript per cell\textsuperscript{14}, bacterial scRNA-seq must deal with a much lower average mRNA copy number (0.4 copies per cell\textsuperscript{14}).

To overcome these barriers in bacterial single-cell transcriptomics, here we applied a highly sensitive random-hexamer, priming-based scRNA-seq protocol\textsuperscript{15} and successfully benchmarked it with the model bacterial pathogen Salmonella enterica serovar Typhimurium (henceforth, Salmonella), the transcriptome of which is extensively and robustly annotated\textsuperscript{16–19}. This enabled us to obtain high-resolution global transcriptomes of individual Salmonella bacteria that faithfully reflect physiologically relevant gene expression profiles established by previous work on bulk populations.

To achieve highly sensitive bacterial scRNA-seq, we developed a generic workflow (Fig. 1a) that starts with the isolation of single cells from culture, followed by one-tube cell lysis and cDNA synthesis and amplification. To obtain sufficient cDNA from single bacteria, we used the poly(A)-independent multiple annealing and dC-tailing-based quantitative scRNA-seq (MATQ-seq)\textsuperscript{15} protocol, in which RT primers are hybridized to internal transcript regions at low temperature, thus enabling the theoretical detection of all transcripts including those of low abundance\textsuperscript{15}. We analysed Salmonella raised under three common growth conditions: (1) ‘late stationary phase’, reflecting mainly resting cells; (2) ‘anaerobic shock’, mimicking the intestinal environment; and (3) ‘NaCl shock’, caused by increased NaCl concentration in the medium (Fig. 1b and Methods). Samples of either ten (10-pooled) or single bacteria were systematically isolated using fluorescence-activated flow cytometry (FACS). The accuracy of single-bacterium sorting was >97% according to the formation of single colonies (Extended Data Fig. 1), whereas control sorting of 2, 10 and 100 bacteria resulted in multiple colonies.

To minimize transcriptional changes during sorting, bacteria were treated with RNAAlater, a stabilization solution that maintains RNA integrity\textsuperscript{20,21}. In addition, cells were sorted directly onto PCR plates preloaded with lysis buffer (Fig. 1a and Methods). The cell wall was digested with 5 U of lysozyme targeting the peptidoglycan layer, using an amount of the enzyme that does not inhibit the subsequent RT step; we found that >20 U abolished cDNA synthesis. After cell lysis, RT was performed using multiple rounds of annealing (Fig. 1a). Libraries were amplified with a limited number (23)

\textsuperscript{1}Helmholtz Institute for RNA-based Infection Research, Helmholtz-Center for Infection Research, Würzburg, Germany. \textsuperscript{2}Institute of Molecular Infection Biology, University of Würzburg, Würzburg, Germany. \textsuperscript{3}These authors contributed equally: Fabian Imdahl, Ehsan Vafadarnejad. \textsuperscript{4}e-mail: emmanuel.saliba@helmholtz-hiri.de; joerg.vogel@uni-wuerzburg.de
of PCR cycles (comparable to eukaryotic scRNA-seq), indexed and pooled before each was sequenced to a depth of 62.4 ± 20.9 million reads per 10-pooled and single bacterial cells (Methods). Analysing 60 samples of 10-pooled bacteria and 71 of single bacteria (Supplementary Table 1), we consistently captured all classes of bacterial RNA (Fig. 2a). As expected for priming of total RNA,
the cDNA libraries were dominated by rRNA and transfer RNA which, on average, accounted for ~93% of all mapped reads (Fig. 2a and Supplementary Tables 2 and 3). Transcripts from mRNA and small non-coding RNA genes were represented by ~5 and ~1.2%, respectively, of total aligned reads, without major differences between pooled and single-bacterium samples (Supplementary Tables 2 and 3). The relative proportions of RNA classes are in good agreement with RNA-seq data from Salmonella populations with no previous ribosomal depletion (Fig. 2a). After removal of both unmapped and rRNA and tRNA reads, we proceeded with averages of ~1,500,000 (for 10-pooled) and ~800,000 (for single bacteria) uniquely mapped reads per cell (Extended Data Fig. 2a); in eukaryotic scRNA-seq, such read numbers are usually sufficient to assign physiological states.

After removal of rRNA and tRNA reads, MATQ-seq detected 413 ± 237 and 170 ± 81 (average ± standard, with five reads) genes in the 10-pooled and single-bacterium libraries, respectively (Fig. 2b). Interestingly, we obtained ~2.5-fold more reads for Salmonella grown under anaerobic or NaCl shock as compared to cells in late stationary phase (Fig. 2b and Extended Data Fig. 2b). Saturation analysis showed that the number of genes detected marginally increased after 500,000 uniquely aligned reads per bacterium (Fig. 2c).

The read distributions along detected genes further supported the premise that we were, indeed, sequencing cDNA derived from RNA and not genomic DNA: as shown for ssrA (transfer-mRNA), flic (flagellin) and several ribosomal protein genes, the reads obtained covered only their transcribed regions (Fig. 2d and Extended Data Fig. 3a–c). Similar to patterns in RNA-seq of a Salmonella population, we observed uneven mapping across the transcript (Fig. 2d and Extended Data Fig. 3a–c). Finally, for each gene we plotted the coefficient of variation for average read counts and obtained a profile similar to that for conventional single-cell eukaryotic transcriptomes (Extended Data Fig. 4a,b). Ten-pooled and single bacteria showed Spearman’s correlation ρ = 0.5 between anaerobic and NaCl shock and ρ = 0.42 for late stationary phase; similar values were obtained when compared with bulk-derived RNA-seq (Extended Data Fig. 5a,b). This correlation analysis shows a large proportion of values close to zero that might have arisen from dropouts and/or an averaging effect given the fact that, on average, bacterial mRNA is present in fewer than one copy per cell.

To determine whether the obtained transcriptomes reflect cells in different states, we performed an unbiased clustering of the 10-pooled or single-bacterium libraries using principal-component analysis (PCA). Along the first two components, clustering robustly delineated cells according to the original growth condition (Fig. 3a and Extended Data Fig. 6). We also sought to identify differentially expressed genes between conditions using DESeq2 analysis. For the 10-pooled libraries, we retrieved 101 anaerobic and 274 NaCl shock-specific genes; single-bacterium libraries yielded 63 and 131 such genes, respectively (Fig. 3b and Supplementary Tables 4 and 5).
Among these genes, we found some characteristic for metabolic shift under both anaerobic (for example, gfpA and tdcC) and NaCl conditions (for example, ygiL, yadF and sodA) (Fig. 3b). More globally, Gene Ontology analysis of differentially expressed genes revealed typical signatures associated with stationary phase metabolic processes, while under anaerobic conditions we observed a global shift to anabolism and catabolism (Supplementary Table 6).

We compared our single-cell RNA-seq results with previous bulk RNA-seq results. To this end, we correlated our anaerobic and NaCl shock expression patterns with *Salmonella* RNA-seq profiles provided by the SalCom database26 (Fig. 3c, Extended Data Fig. 5b and Supplementary Tables 4 and 5). Importantly, we observed that ~75% of genes upregulated under both anaerobic and NaCl shock matched upregulated genes in the SalCom database (Fig. 3c), confirming the robustness of gene expression signatures obtained from single bacteria. However, we emphasize that even if a RNA-seq compendium such as SalCom is available, we recommend that a bulk RNA-seq sample be prepared in parallel for benchmarking of scRNA-seq data.

Finally, to establish proof of principle for the generalizability of our bacterial scRNA-seq approach, we applied MATQ-seq to another species, the clinically important pathogen *Pseudomonas aeruginosa*, whose genome has a much higher genomic GC content than that of *Salmonella* (>67% versus ~50%). With generation of libraries from 10-pooled and single *P. aeruginosa* bacteria (Extended Data Fig. 7 and Supplementary Tables 7 and 8), MATQ-seq captured on average 102 genes at the single-bacterium level, a number that compares favourably with *Salmonella* scRNA-seq data above.

In conclusion, by adopting MATQ-seq to capture transcripts of low abundance, we were able to perform scRNA-seq on minimal microbial samples down to the level of single bacteria and subsequently infer different growth conditions. Some previous reports on prokaryotic scRNA-seq detected thousands of genes expressed in a single bacterium, which is hard to reconcile with the fact that most bacterial mRNAs are predicted to be present in fewer than one copy per cell, on average1, or that scRNA-seq libraries had to be pooled back to infer expression programmes2. Here, by choosing well-characterized growth conditions and independent datasets2, we could assign robust transcriptomic signatures to single bacteria. While the number of expressed genes detected here by MATQ-seq is in the hundreds, and so allows the global study of transcription of individual bacterial cells, this number is still lower than that required to reflect a full bacterial transcriptome8. Nonetheless, our mRNA capture rates seem realistic in light of the fact that a recently published prokaryotic single-cell RNA-seq protocol achieved capture in the range of 200 mRNAs per cell for exponentially growing *Escherichia coli*22.

Our work here should pave the way for gene activity profiling in complex ecological microbial niches, such as the microbiome at the level of its many different single bacteria, and to monitor the drug susceptibility of pathogens based on RNA signatures from clinical samples23. At present, although bacterial single-cell transcriptomics using MATQ-seq technology incurs considerable cost (Supplementary Table 9), we are positive that making the protocol compatible with microfluidics devices and implementing RNA read removal before sequencing9,28 will facilitate considerable scale-up of throughput while also reducing the cost per cell analysed.

### Methods

**Bacterial growth conditions.** Wild-type *Salmonella Typhimurium* strain SL1344 was cultured in 5 ml of Lennox broth (LB) at 37 °C under constant agitation at 220 r.p.m. (New Brunswick Innova 44) to reach an optical density (OD, λ = 600 nm) of 2.0 followed by a further 6 h (late stationary phase1). A 1:100 dilution of the bacterial culture was grown at 37 °C under 220 r.p.m. agitation to reach an OD (λ = 600 nm) of 0.3 and subjected to either NaCl or anaerobic shock as described previously7 (Fig. 1b). Two millilitres of the culture was pelleted, washed twice with 1 ml of Dulbecco’s phosphate-buffered saline (DPBS, Gibco) and

...continued...
References

1. Ackermann, M. A functional perspective on phenotypic heterogeneity in microorganisms. Nat. Rev. Microbiol. 13, 497–508 (2015).
2. Gollan, B., Grabe, G., Michaux, C. & Helaine, S. Bacterial persisters and infection: past, present, and progressing. Annu. Rev. Microbiol. 73, 359–385 (2019).
3. Kreibich, S. & Hardt, W. D. Experimental approaches to phenotypic diversity in infection. Curr. Opin. Microbiol. 27, 25–36 (2015).
4. Gasch, A. P. et al. Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. PLoS Biol. 15, e2004050 (2017).
5. Muller, L. S. M. et al. Genome organization and DNA accessibility control antigenic variation in trypanosomes. Nature 563, 121–125 (2018).
6. Stubbington, M. J. T., Rozenblatt-Rosen, O., Regev, A. & Teichmann, S. A. Single-cell transcriptomics to explore the immune system in health and disease. Science 358, 58–63 (2017).
7. Fang, Y. et al. Transcript amplification from single bacterium for transcriptome analysis. Genome Res. 21, 925–935 (2011).
8. Wang, J., Chen, L., Chen, Z. & Zhang, W. RNA-seq based transcriptomic analysis of single bacterial cells. Integr. Biol. (Camb.) 7, 1466–1476 (2015).
9. Avital, G. et al. scDual-Seq: mapping the gene regulatory program of Salmonella infection by host and pathogen single-cell RNA-sequencing. Genome Biol. 18, 200 (2017).
10. Betin, V. et al. Hybridization-based capture of pathogen mRNA enables paired host–pathogen transcriptional analysis. Sci. Rep. 9, 19244 (2019).
11. Penaranda, C. & Hung, D. T. Single-cell RNA sequencing to understand host–pathogen interactions. ACS Chem. Div. 5, 336–344 (2019).
12. Milo, R. & Phillips, R. Cell Biology by the Numbers (Garland Science, 2015).
13. Bagnoli, J. W. et al. Sensitive and powerful single-cell RNA sequencing using mcSCRIB-seq. Nat. Commun. 9, 2937 (2018).
14. Taniguchi, Y. et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–538 (2010).
15. Sheng, K., Cao, W., Niu, Y., Deng, Q. & Zong, C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. Nat. Methods 14, 267–270 (2017).
16. Kroger, C. et al. An infection-relevant transcriptomic compendium for Salmonella enterica serovar Typhimurium. Cell Host Microbe 14, 683–695 (2013).
17. Smirnov, A. et al. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. Proc. Natl Acad. Sci. USA 113, 11591–11596 (2016).
18. Chao, Y. et al. In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. Mol. Cell 65, 39–51 (2017).
19. Hor, J., Matera, G., Vogel, J., Gottesman, S. & Storz, G. Trans-acting small RNAs and their effects on gene expression in Escherichia coli and Salmonella enterica. EcoSal Plus https://doi.org/10.1128/ecosalplus.ESP-0030-2019 (2020).
20. Westermann, A. J. et al. Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. Nature 529, 496–501 (2016).
21. Westermann, A. J. & Vogel, J. Host–pathogen transcriptomics by dual RNA-seq. Methods Mol. Biol. 1737, 59–75 (2018).
22. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods 10, 1096–1098 (2013).
23. Brennecke, P. et al. Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods 10, 1093–1095 (2013).
24. Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W. & Livny, J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? BMC Genomics 13, 734 (2012).
25. Blattman, S. B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. Nat. Microbiol. https://doi.org/10.1038/s41564-020-0729-6 (2020).
26. Barczak, A. K. et al. RNA signatures allow rapid identification of pathogens and antibiotic susceptibilities. Proc. Natl Acad. Sci. USA 109, 6217–6222 (2012).
27. Gu, W. et al. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol. 17, 41 (2016).
28. Premza, G. et al. Improved bacterial RNA-seq by Cas9-based depletion of ribosomal RNA reads. RNA 26, 1069–1078 (2020).
29. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344, 1396–1401 (2014).
30. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).

Acknowledgements

We thank S. Gorski for critical comments on the manuscript; K. Secener and X. Hartmann for performing preliminary experiments; C. Michaux for advice on bacterial culture; L. Barquist and F. Erhard for advice on statistical analysis; and P. Arampatzi, T. Heckel (SysMed Würzburg) and R. Geffers (Genome Analytics, HZI) for conducting the sequencing.

Author contributions

F.I. and C.H. conducted experiments. E.V. performed data analysis. A.-E.S. designed research. J.V and A.-E.S. directed research. A.-E.S. and J.V. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-020-0774-1. Supplementary information is available for this paper at https://doi.org/10.1038/s41564-020-0774-1. Correspondence and requests for materials should be addressed to A.-E.S. or J.V. Reprints and permissions information is available at www.nature.com/reprints. Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020
Extended Data Fig. 1 | Single-bacteria sorting. Evaluation of single-bacteria sorting efficiency. On a LB agar molded in a 96 well format dish, GFP-fluorescent and non-fluorescent Salmonella 100, 10, 2 and 1 bacteria are sorted systematically in alternative manner. After overnight growth colonies are observed (upper bright field image) and absent colonies are indicated with a red circle. Observing the plate under fluorescent reader allows to count the sorting mismatches (indicated by an orange square) and define the sorting precision.
Extended Data Fig. 2 | Comparison of library size and number of detected genes between different growth conditions. a, Violin plots represent the size of libraries cultured in the different growth conditions for 10-pooled and single bacteria. b, Scatter plots depicting the relation between the number of detected genes and library size. All libraries labelled as outliers have been removed from the analysis (Methods, Supplementary Table 1).
Extended Data Fig. 3 | Coverage plot and read densities. a, Read densities of structural genes with a high expression in 10- and single-cell libraries. b, Gene coverage of selected differentially expressed genes in the respective conditions in 10-cell libraries. c, Read densities of csrA in 10-pooled libraries and respective conditions overlapping to 5’ UTR. Libraries have been automatically log scaled by Integrative Genomics Viewer.
Extended Data Fig. 4  | Technical noise assessment of 10-pooled bacteria and single-bacteria RNA-seq data.  

**a, b.** Coefficient of variation is plotted against the log₂ (average of normalized read count) for (a) 10-pooled bacteria and (b) single bacteria conditions. Color code refers to salt (NaCl) shock (green), anaerobic shock (red) and late stationary phase (blue). Such analysis are routinely conducted when analyzing mammalian single-cell RNA-seq data (see ref. 23). Note that upper bound corresponds to CV = √n with n the number of samples analyzed.
Extended Data Fig. 5 | Correlations between matching growth conditions. 

a. Scatter plots show the correlation between the matching pooled (10-pooled) and single bacteria in the respective conditions with the associated Spearman’s correlation coefficient p<2.2e10^-16 in all three growth conditions.

b. Scatter plots show the correlation between 10-pooled and single bacteria (this study) and bulk RNA-seq in the respective conditions with the associated Spearman’s correlation coefficient (ρ) with the associated p-values.
Extended Data Fig. 6 | Technical parameters associated to the Principal Component Analysis (PCA) of 10-pooled and single bacteria transcriptomes.

a, b, For single-bacteria (a) and 10-pooled bacteria transcriptomes (b) library size (left), number of detected genes (middle) and genes associated with PC loading 1 and 2 are overlaid on top of the PCA. The top 15 genes with the highest contribution to PC were selected and their related loading vectors are shown on the PCA plot. The vectors show how the original variables contribute to creating the principal component. c, Scree plots give the variance associated to each PC loading.
Extended Data Fig. 7 | Application of MATQ-seq to Pseudomonas aeruginosa. a, Scatter plots depicting the relation between the number of detected genes versus library size for single bacteria libraries (blue) and 10-pooled bacteria libraries (red). b, Violin plots representing the number genes detected for 10-pooled and single bacteria. c, Average proportions of transcript categories after removal of unmapped reads obtained for 10-pooled and single bacteria. CDS: coding sequences; IGR, intergenic region; ncRNA: non-coding RNA; sRNA: small RNA; other: all other RNA classes (Supplementary Tables 7 and 8). d, Representative reads aligning to the reference sequence of tmRNA-encoding ssrA gene (±300 bp upstream and downstream the CDS) across 10-pooled and single bacteria.
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
Salcom database was used: http://bioinf.gen.tcd.ie/cgi-bin/salcom.pl?_HL

Data analysis
FastQC (v 0.11.7); DESeq2 (1.26.0); featureCounts (2.0.0); STAR (2.5.1b); cutadapt (2.8)

All the codes are available from https://github.com/saliba-lab/single-bacterium-transcriptome-profiling

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

All RNA-seq data has been deposited in NCBI’s Gene Expression Omnibus under accession number GSE119888

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
Life sciences study design

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

**Sample size**

60 of 10-pooled bacteria and 71 single-bacterium samples were analysed (Supplementary Table 1).

**Data exclusions**

After PCA analysis of 10-pooled bacterium and single-bacterium, we removed a total of 5 libraries out of 131 (<4%) (three 10-pooled bacterium and two single bacterium (Suppl. Table 1)), as they appeared to be outliers on the analysis. The exclusion criteria for outliers were not pre-established.

**Replication**

n/a

**Randomization**

n/a

**Blinding**

The person in charge of data analysis was aware of sample identities and groups as this would not have been possible during sample preparation, data collection and analysis. Furthermore, blinding was not relevant to this study. Only PCA analysis enables us to differentiate the growth conditions.

---

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 |