Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/17/eaaz2299/DC1

Supplementary Materials for

Metagenomic growth rate inferences of strains in situ

Akintunde Emiola, Wei Zhou, Julia Oh*

*Corresponding author. Email: julia.oh@jax.org

Published 22 April 2020, *Sci. Adv.* **6**, eaaz2299 (2020)
DOI: 10.1126/sciadv.aaz2299

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/17/eaaz2299/DC1)

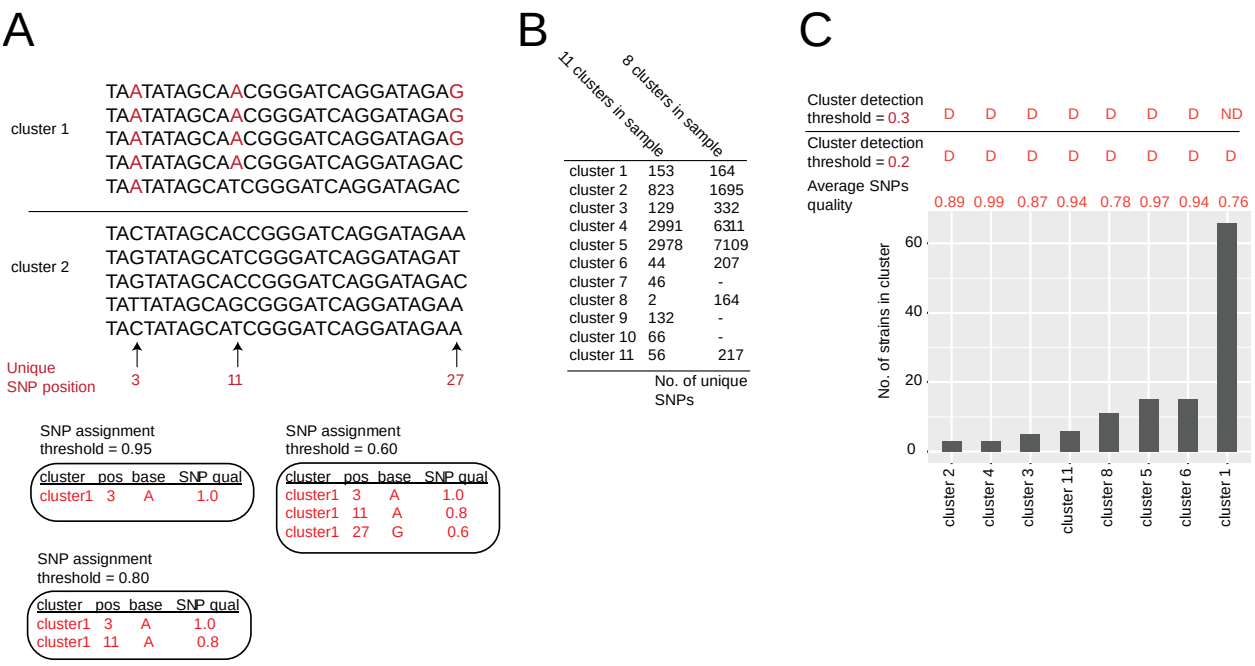

Fig. S1: Description and evaluation of SMEG parameters

(A) A hypothetical grouping of 10 strains into 2 clusters. Unique SNPs for cluster 1 are located in positions 3, 11, and 27 in the alignment depending on the *SNP assignment threshold*. If the threshold is set at 0.95 in this example, only one SNP is output in the cluster-specific SNPs profile. The SNP quality is the proportion of strains of a cluster harboring a unique SNP at a given position (0.8 ~ 80% of strains have that SNP at that position). **(B)** SNP counts for clusters of *C. acnes,* when all clusters are present in a sample in comparison to samples with 8 clusters after re-generation of a sample-specific SNPs. **(C)** Effect of SNP quality and *cluster detection thresholds* to accurately detect clusters of *C. acnes.* D = detected; ND = non-detected.

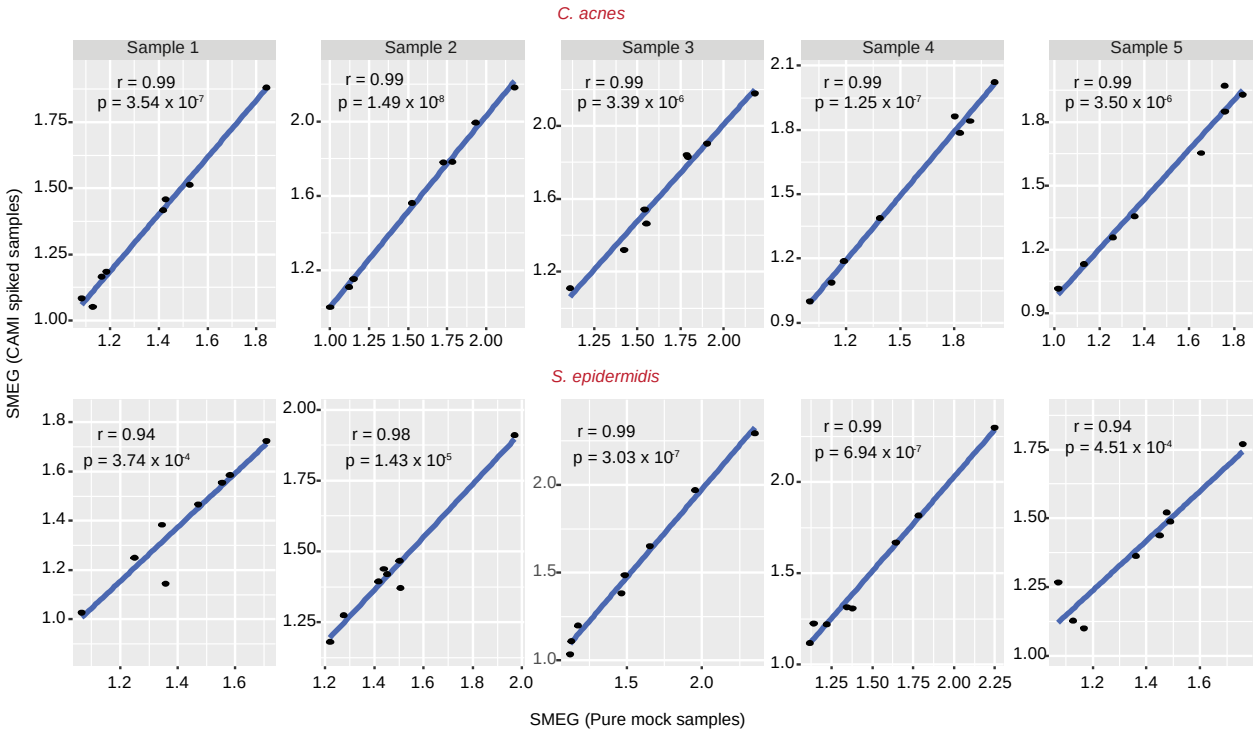

Fig. S2: Evaluation of SMEG in high complexity dataset. Pearson correlation between SMEG output in pure mock samples and spiked-in high complexity CAMI samples.

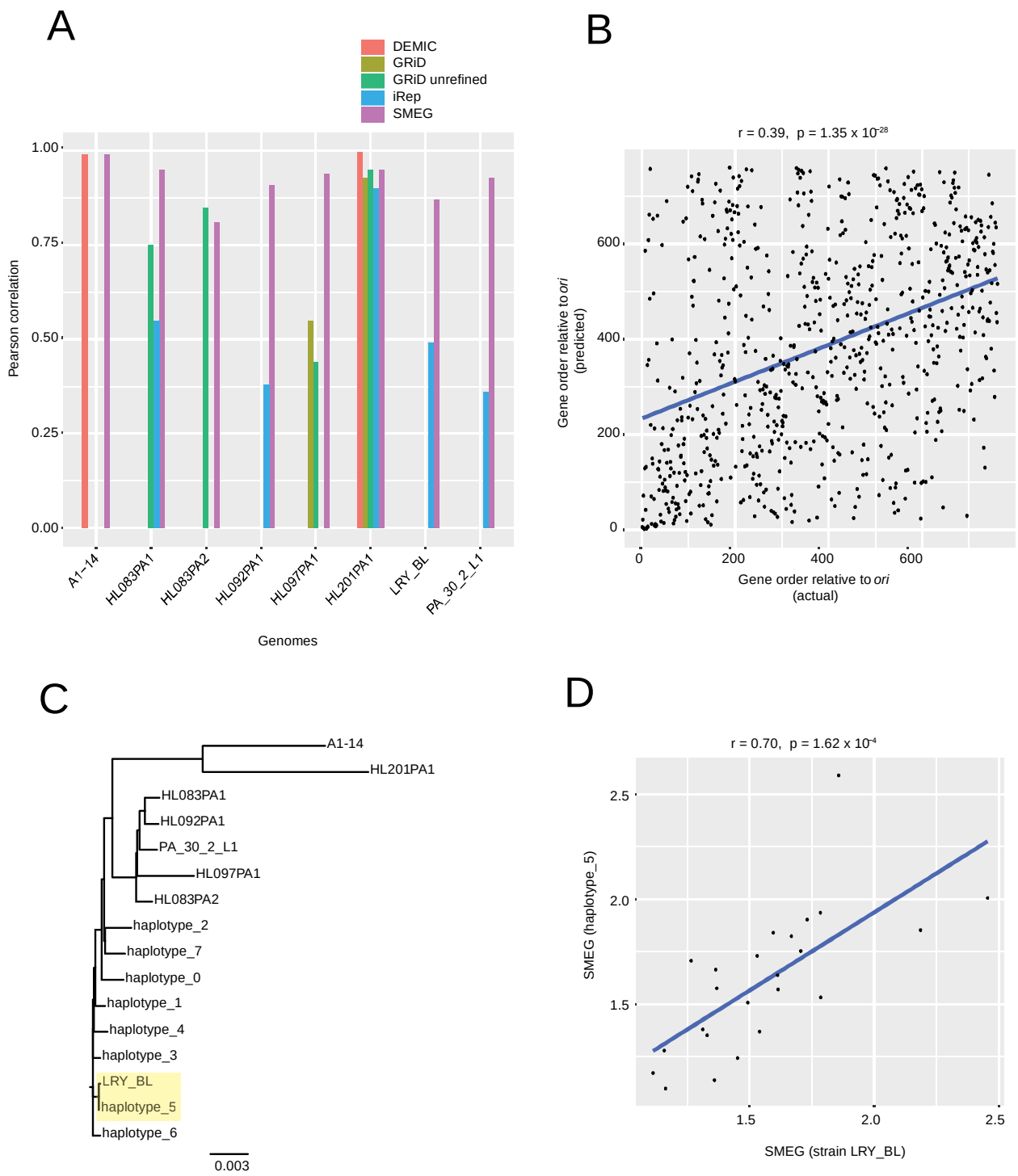

Fig. S3: Evaluation of other tools and DESMAN analyses. (A) Comparison of SMEG with other species-level growth estimation tools using a 30-sample synthetic *C. acnes* mock consortium. The barplots show the Pearson correlation between growth predictions and expected results. The reference-based approach was used in SMEG at a coverage cutoff of 5x since other tools utilize reference-based approaches and iRep requires coverage of 5x. GRiD was run in the multiplex mode (-p option) using a custom

database containing only the reference strains. iRep and DEMIC were run using default parameters. **(B)** Core genes order prediction relative to *ori* using a 30-sample mock reads of a *S. epidermidis* strain (ATCC 12228). Here, the gene closest to the *ori* will have a value of 1 in the actual gene order. The predicted gene order is based on the median coverage across all samples. The highest coverage gene had a predicted gene order value of 1. The plot displays the Pearson correlation between the actual and predicted gene order. **(C)** Phylogenetic tree constructed for eight inferred *C. acnes* strains from a 30-sample metagenomic mock. The core genes for the strains and reference genomes were aligned using mafft (*36*), indel positions excluded, and the tree was constructed using Figtree (*37*). Haplotypes accurately resolved are highlighted. **(D)** Pearson correlation between SMEG scores generated for a DESMAN-reconstituted haplotype and its phylogenetically similar reference at coverage cutoff of 5x.

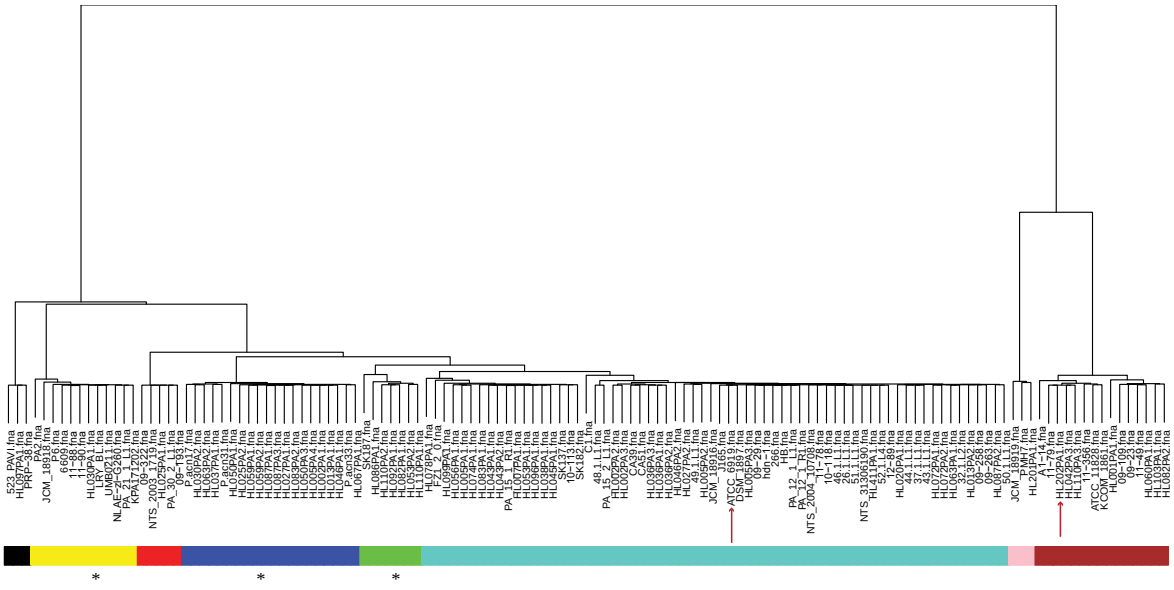

Fig. S4: Phylogenetic tree for *C. acnes* strains before iterative clustering. After iterative clustering, the asterisked clusters were further sub-grouped. Hypothetical uncharacterized strains which were excluded from the database are depicted with the pointed arrow.

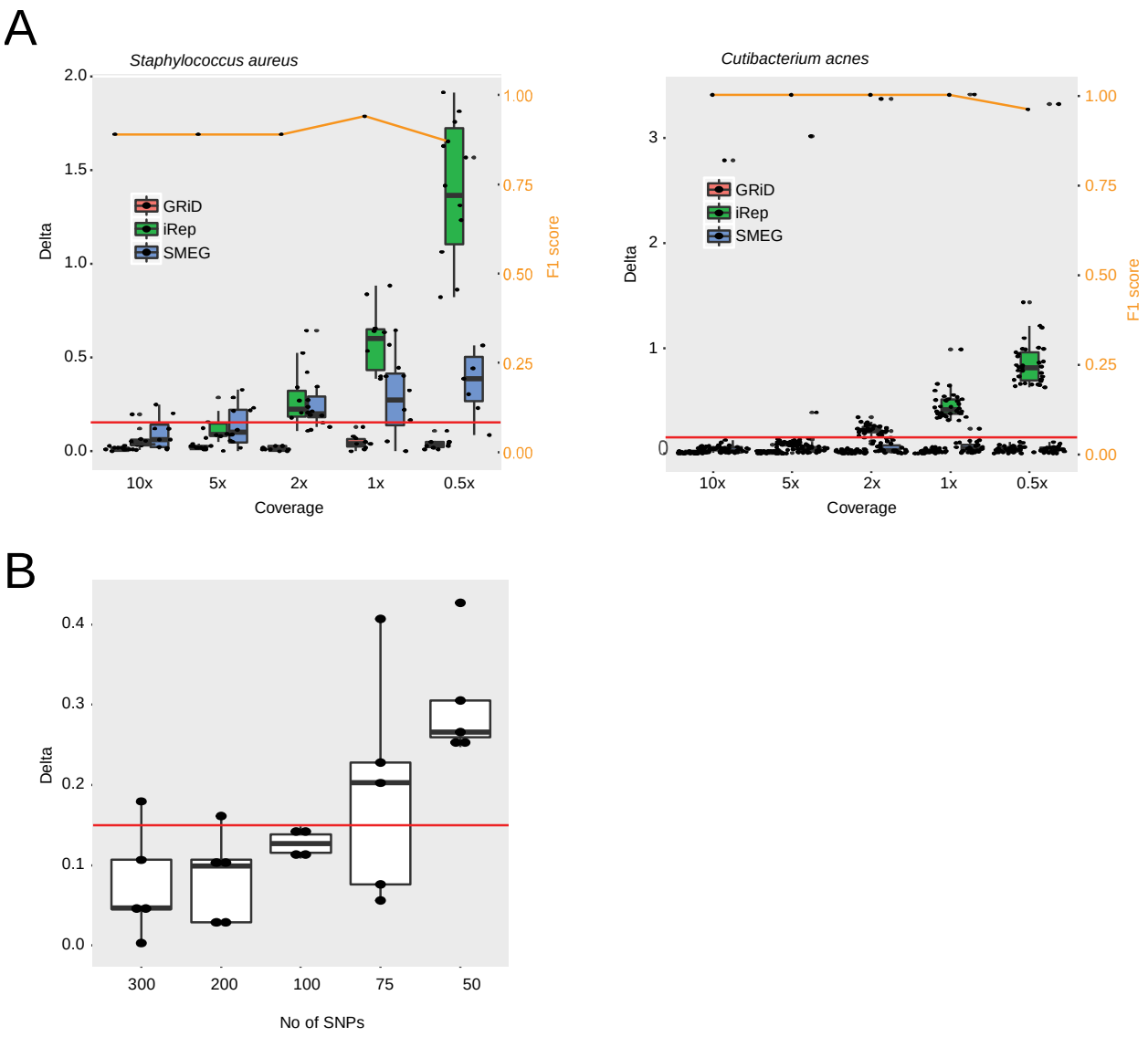

Fig. S5: Effect of coverage and unique SNP count on SMEG accuracy. (A) Growth rate reproducibility between for SMEG, GRiD, and iRep at different cluster coverages for microbes with high (*S. aureus*) low (*C. acnes*) within-species genetic diversity. The boxplot shows the difference (delta) in growth estimates before and after reads were subsampled from ≥20X to the lower coverage. The F_1 scores reflect precision and recall of clusters, as a function of coverage. The red horizontal line represents a delta cutoff of 0.15. **(B)** Growth rate reproducibility as a function of SNP count. The boxplot shows the difference (delta) in growth estimates before and after unique SNPs were subsampled. This subsampling step was conducted 5 times. The red horizontal line represents a delta cutoff of 0.15.

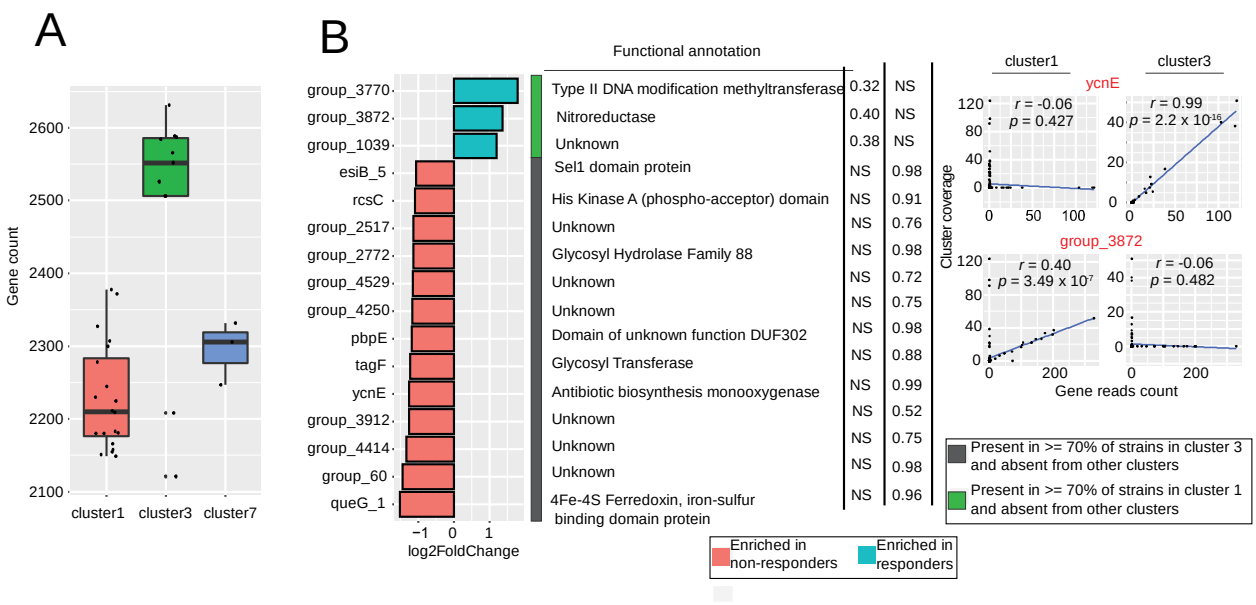

Fig. S6: SMEG analysis of *A. muciniphila* in patients treated with ICIs.
(A) Total genes count for members of clusters 1, 3 and 7. **(B)** Cluster-specific accessory genes for clusters 1 and 3 that were differentially enriched in responders and non-responders. The chart displays Pearson correlation values between cluster coverage and unambiguous gene reads count. The scatter plots on the right show the Pearson correlation between cluster coverage and unambiguous reads counts for two genes in baseline samples (NS = non-significant).

REFERENCES AND NOTES

1. T. Korem, D. Zeevi, J. Suez, A. Weinberger, T. Avnit-Sagi, M. Pompan-Lotan, E. Matot, G. Jona, A. Harmelin, N. Cohen, A. Sirota-Madi, C.A. Thaiss, M. Pevsner-Fischer, R. Sorek, R.J. Xavier, E. Elinav, E. Segal, Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101–1106 (2015).

2. A. Emiola, J. Oh, High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. *Nat. Commun.* **9,** 4956 (2018).

3. C. T. Brown, M. R. Olm, B. C. Thomas, J. F. Banfield, Measurement of bacterial replication rates in microbial communities. *Nat. Biotechnol.* **34**, 1256–1263 (2016).

4. Y. Gao, H. Li, Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. *Nat. Methods* **15**, 1041–1044 (2018).

5. A. L. Byrd, A. L. Byrd, C. Deming, S. K. B. Cassidy, O. J. Harrison, W. I. Ng, Conlan S; NISC Comparative Sequencing Program, Y. Belkaid, J. A. Segre, H. H. Kong, *Staphylococcus aureus* and *Staphylococcus epidermidis* strain diversity underlying pediatric atopic dermatitis. *Sci. Transl. Med.* **9**, eaal4651 (2017).

6. C. Luo, R. Knight, H. Siljander, M. Knip, R. J. Xavier, D. Gevers, ConStrains identifies microbial strains in metagenomic datasets. *Nat. Biotechnol.* **33**, 1045–1052 (2015).

7. D. Albanese, C. Donati, Strain profiling and epidemiology of bacterial species from metagenomic sequencing. *Nat. Commun.* **8**, 2260 (2017).

8. P. I. Costea, R. Munch, L. P. Coelho, L. Paoli, S. Sunagawa, P. Bork, metaSNV: A tool for metagenomic strain level analysis. *PLOS ONE* **12**, e0182392 (2017).

9. C. Hong, S. Manimaran, Y. Shen, J. F. Perez-Rogers, A. L. Byrd, E. Castro-Nallar, K. A. Crandall, W. E. Johnson, PathoScope 2.0: A complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome* **2**, 33 (2014).

10. C. F. Scholz, H. Brüggemann, H. B. Lomholt, H. Tettelin, M. Kilian, Genome stability of *Propionibacterium acnes*: A comprehensive study of indels and homopolymeric tracts. *Sci. Rep.* **6**, 20662 (2016).

11. S. Conlan, L. A. Mijares; NISC Comparative Sequencing Program, J. Becker, R. W. Blakesley, G. G. Bouffard, S. Brooks, H. Coleman, J. Gupta, N. Gurson, M. Park, B. Schmidt, P. J. Thomas, M. Otto, H. H. Kong, P. R. Murray, J. A. Segre, *Staphylococcus epidermidis* pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. *Genome Biol.* **13**, R64 (2012).

12. A. K. Szafrańska, V. Junker, M. Steglich, U. Nübel, Rapid cell division of *Staphylococcus aureus* during colonization of the human nose. *BMC Genomics* **20**, 229 (2019).

13. C. Quince, T. O. Delmont, S. Raguideau, J. Alneberg, A. E. Darling, G. Collins, A. M. Eren, DESMAN: A new tool for de novo extraction of strains from metagenomes. *Genome Biol.* **18**, 181 (2017).

14. M. Scholz, D. V. Ward, E. Pasolli, T. Tolio, M. Zolfo, F. Asnicar, D. T. Truong, A. Tett, A. L. Morrow, N. Segata, Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat. Methods* **13**, 435–438 (2016).

15. N. J. Loman, C. Constantinidou, M. Christner, H. Rohde, J. Z.-M. Chan, J. Quick, J. C. Weir, C. Quince, G. P. Smith, J. R. Betley, M. Aepfelbacher, M. J. Pallen, A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104: H4. *JAMA* **309**, 1502–1510 (2013).

16. T. de Sablet, C. Chassard, A. Bernalier-Donadille, M. Vareille, A. P. Gobert, C. Martin, Human microbiota-secreted factors inhibit shiga toxin synthesis by enterohemorrhagic *Escherichia coli* O157: H7. *Infect. Immun.* **77**, 783–790 (2009).

17. B. Routy, E. Le Chatelier, L. Derosa, C. P. M. Duong, M. T. Alou, R. Daillère, A. Fluckiger, M. Messaoudene, C. Rauber, M. P. Roberti, M. Fidelle, C. Flament, V. Poirier-Colame, P. Opolon, C. Klein, K. Iribarren, L. Mondragón, N. Jacquelot, B. Qu, G. Ferrere, C. Clémenson, L. Mezquita, J.

R. Masip, C. Naltet, S. Brosseau, C. Kaderbhai, C. Richard, H. Rizvi, F. Levenez, N. Galleron, B. Quinquis, N. Pons, B. Ryffel, V. Minard-Colin, P. Gonin, J.-C. Soria, E. Deutsch, Y. Loriot, F. Ghiringhelli, G. Zalcman, F. Goldwasser, B. Escudier, M. D. Hellmann, A. Eggermont, D. Raoult, L. Albiges, G. Kroemer, L. Zitvogel, Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors. *Science* **359**, 91–97 (2018).

18. A. Palleja, K. H. Mikkelsen, S. K. Forslund, A. Kashani, K. H. Allin, T. Nielsen, T. H. Hansen, S. Liang, Q. Feng, C. Zhang, P. T. Pyl, L. P. Coelho, H. Yang, J. Wang, A. Typas, M. F. Nielsen, H. B. Nielsen, P. Bork, J. Wang, T. Vilsbøll, T. Hansen, F. K. Knop, M. Arumugam, O. Pedersen, Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nat. Microbiol.* **3**, 1255–1265 (2018).

19. L. Kraal, S. Abubucker, K. Kota, M. A. Fischbach, M. Mitreva, The prevalence of species and strains in the human microbiome: A resource for experimental efforts. *PLOS ONE* **9**, e97279 (2014).

20. M. F. Laursen, M. F. Laursen, R. P. Laursen, A. Larnkjær, C. Mølgaard, K. F. Michaelsen, H. Frøkiær, M. I. Bahl, T. R. Licht, *Faecalibacterium* gut colonization is accelerated by presence of older siblings. *mSphere* **2**, e00448-17 (2017).

21. A. J. Lopatkin, J. M. Stokes, E. J. Zheng, J. H. Yang, M. K. Takahashi, L. You, J. J. Collins, Bacterial metabolic state more accurately predicts antibiotic lethality than growth rate. *Nat. Microbiol.* **4**, 2109–2117 (2019).

22. A. I. Rissman, B. Mau, B. S. Biehl, A. E. Darling, J. D. Glasner, N. T. Perna, Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* **25**, 2071–2073 (2009).

23. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010).

24. T. Seemann, Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

25. A. J. Page, C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. G. Holden, M. Fookes, D. Falush, J. A. Keane, J. Parkhill, Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).

26. P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: The dynamic tree cut package for R. *Bioinformatics* **24**, 719–720 (2007).

27. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

28. K. Kafadar, The efficiency of the biweight as a robust estimator of location. *J. Res. Natl. Bur. Stand.* **88,** 105–116 (1983).

29. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

30. J. Oh, A. L. Byrd, M. Park; NISC Comparative Sequencing Program, H. H. Kong, J. A. Segre, Temporal stability of the human skin microbiome. *Cell* **165**, 854–866 (2016).

31. W. Zhou, N. Gay, J. Oh, ReprDB and panDB: Minimalist databases with maximal microbial representation. *Microbiome* **6**, 15 (2018)

32. X. Guo, S. Li, J. Zhang, F. Wu, X. Li, D. Wu, M. Zhang, Z. Ou, Z. Jie, Q. Yan, P. Li, J. Yi, Y. Peng, Genome sequencing of 39 *Akkermansia muciniphila* isolates reveals its population structure, genomic and functional diverisity, and global distribution in mammalian gut microbiotas. *BMC Genomics* **18**, 800 (2017).

33. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2013).

34. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

35. J. Huerta-Cepas,D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M. C. Walter, T. Rattei, D. R. Mende, S. Sunagawa, M. Kuhn, L. J. Jensen, C. von Mering, P. Bork, eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2015).

36. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

37. A. Rambaut, FigTree version 1.3.1 [computer program] (2009); http://tree.bio.ed.ac.uk/software/figtree/.