whether the cumulative relative abundance of food SGBs in gut microbiomes differs according to diet pattern or food group, we fit another zero-inflated linear mixed model estimated by maximum likelihood with the same model structure as above. The model's intercept corresponded to omnivore microbiomes and meat SGBs at −0.41, $P < 0.001$, with conditional $R^2 = 0.05$ and marginal $R^2 = 0.05$. Again, vegans were moved to the intercept in a following model with the same parameters. In addition and using a slightly different approach, we tested for differences in the number of food SGBs and in their cumulative relative abundance between the diet patterns and within each food group and within each cohort using Dunn's tests coupled with a BH correction for multiple testing.

To establish whether there were any significant differences in the prevalence of the 20 most common food SGBs between the three diet patterns, we ran a chi-squared test on the number of omnivore, vegetarian and vegan microbiomes in which each of the SGBs was present versus absent with the option to compute *P* values by Monte Carlo simulation (99,999 replicates). The tests were run for the larger cohorts separately, namely, for P1, P3 UK22A and P3 US22A.

Gut microbial functional potential

To generate gut microbial functional potential, we ran HUMAnN (v.3.6)[57] with default parameters (Supplementary Code 1). We focused on the pathway abundance output and removed any unmapped or unintegrated pathways, as well as pathways with a prevalence of <0.05 across samples of at least one diet pattern and a coverage of <0.2. This left us with 87 pathways in P1, 85 pathways in P3 UK22A and 87 pathways in P3 US22A. We then measured the statistical association between the relative abundances of these pathways and each diet pattern pair, which we first computed on each of the three PREDICT cohorts separately and then meta-analysed as described in detail above (Supplementary Code 1).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The publicly available datasets used in this work are available from their respective publications in refs. 12,13. Raw metagenomic samples are provided for all participants of the ZOE PREDICT studies. Specifically, PREDICT 1 has already been made publicly available as reported previously[10] under the NCBI-SRA bioproject ID PRJEB39223, whereas PREDICT 2 is deposited in EBI under accession number PRJEB75460, and PREDICT 3 cohorts under EBI accession numbers PRJEB75463 and PRJEB75464. Sex, age, BMI, country and the quantitative taxonomic profiles are available for each sample within the curatedMetagenomicData package[58]. The ZOE Microbiome Rankings for the full list of species are made available (and kept up-to-date) at https://zoe.com/our-science/microbiome-ranking. ZOE is the owner of the pseudonymized data and metadata and researchers interested in follow-up studies requiring additional specific metadata information should fill out a research request proposal at https://zoe.com/our-science/collaborate that will be evaluated by a subpanel of the ZOE Scientific Advisory Board once per month for their priority, relevance and in compliance with privacy and data protection regulations.

Code availability

The code for the analyses conducted here is provided in Supplementary Code 1. The pooled estimate of effect sizes from linear models was computed on the basis of the pipeline in GitHub at https://github.com/waldronlab/curatedMetagenomicDataAnalyses/blob/main/python_tools/metaanalyze.py. An importable meta-analysis Python library is also freely available in GitHub at https://github.com/SegataLab/inverse_var_weight/blob/main/meta_analyses.py.

References

1. Afshin, A. et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **393**, 1958–1972 (2019).
2. Willett, W. et al. Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *Lancet* **393**, 447–492 (2019).
3. Scarborough, P. et al. Vegans, vegetarians, fish-eaters and meat-eaters in the UK show discrepant environmental impacts. *Nat. Food* **4**, 565–574 (2023).
4. Clark, M. A., Springmann, M., Hill, J. & Tilman, D. Multiple health and environmental impacts of foods. *Proc. Natl Acad. Sci. USA* **116**, 23357–23362 (2019).
5. Valles-Colomer, M. et al. Cardiometabolic health, diet and the gut microbiome: a meta-omics perspective. *Nat. Med.* **29**, 551–561 (2023).
6. de Vos, W. M., Tilg, H., Van Hul, M. & Cani, P. D. Gut microbiome and health: mechanistic insights. *Gut* **71**, 1020–1032 (2022).
7. Corrêa, T. A. F., Rogero, M. M., Hassimotto, N. M. A. & Lajolo, F. M. The two-way polyphenols–microbiota interactions and their effects on obesity and related metabolic diseases. *Front. Nutr.* **6**, 188 (2019).
8. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).
9. Thomas, A. M. et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat. Med.* **25**, 667–678 (2019).
10. Asnicar, F. et al. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nat. Med.* **27**, 321–332 (2021).
11. Berry, S. E. et al. Human postprandial responses to food and potential for precision nutrition. *Nat. Med.* **26**, 964–973 (2020).
12. Tarallo, S. et al. Stool microRNA profiles reflect different dietary and gut microbiome patterns in healthy individuals. *Gut* **71**, 1302–1314 (2022).
13. De Filippis, F. et al. Distinct genetic and functional traits of human intestinal *Prevotella copri* strains are associated with different habitual diets. *Cell Host Microbe* **25**, 444–453.e3 (2019).
14. Diener, C. & Gibbons, S. M. Metagenomic estimation of dietary intake from human stool. Preprint at *bioRxiv* https://doi.org/10.1101/2024.02.02.578701 (2024).
15. Satija, A. et al. Plant-based dietary patterns and incidence of Type 2 diabetes in US men and women: results from three prospective cohort studies. *PLoS Med.* **13**, e1002039 (2016).
16. Rothschild, D. et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
17. Pasolli, E., Truong, D. T., Malik, F., Waldron, L. & Segata, N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput. Biol.* **12**, e1004977 (2016).
18. Asnicar, F., Thomas, A. M., Passerini, A., Waldron, L. & Segata, N. Machine learning for microbiologists. *Nat. Rev. Microbiol.* **22**, 191–205 (2024).
19. Natividad, J. M. et al. *Bilophila wadsworthia* aggravates high fat diet induced metabolic dysfunctions in mice. *Nat. Commun.* **9**, 2802 (2018).
20. Qiu, P. et al. The gut microbiota in inflammatory bowel disease. *Front. Cell. Infect. Microbiol.* **12**, 733992 (2022).
21. Wiredu Ocansey, D. K. et al. The diagnostic and prognostic potential of gut bacteria in inflammatory bowel disease. *Gut Microbes* **15**, 2176118 (2023).
22. Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**, 209–218 (2012).

23. Geirnaert, A. et al. Butyrate-producing bacteria supplemented in vitro to Crohn's disease patient microbiota increased butyrate production and enhanced intestinal epithelial barrier integrity. *Sci. Rep.* **7**, 11450 (2017).
24. Chang, S.-C. et al. A gut butyrate-producing bacterium *Butyricicoccus pullicaecorum* regulates short-chain fatty acid transporter and receptor to reduce the progression of 1,2-dimethylhydrazine-associate d colorectal cancer. *Oncol. Lett.* **20**, 327 (2020).
25. Patterson, A. M. et al. Human gut symbiont *Roseburia hominis* promotes and regulates innate immunity. *Front. Immunol.* **8**, 1166 (2017).
26. Biddle, A., Stewart, L., Blanchard, J. & Leschine, S. Untangling the genetic basis of fibrolytic specialization by Lachnospiraceae and Ruminococcaceae in diverse gut communities. *Diversity* **5**, 627–640 (2013).
27. Iyer, R., Tomar, S. K., Uma Maheswari, T. & Singh, R. *Streptococcus thermophilus* strains: multifunctional lactic acid bacteria. *Int. Dairy J.* **20**, 133–141 (2010).
28. Meng, L. et al. The nutrient requirements of *Lactobacillus acidophilus* LA-5 and their application to fermented milk. *J. Dairy Sci.* **104**, 138–150 (2021).
29. Glendinning, L., Stewart, R. D., Pallen, M. J., Watson, K. A. & Watson, M. Assembly of hundreds of novel bacterial genomes from the chicken caecum. *Genome Biol.* **21**, 34 (2020).
30. Asnicar, F. et al. Gut microbiome species indicative of cardiometabolic health are modulated by diet in large and interventional cohorts of over 34,000 individuals. Preprint at *bioRxiv* (in the press).
31. Manghi, P. et al. Coffee consumption is associated with intestinal *Lawsonibacter asaccharolyticus* abundance and prevalence across multiple cohorts. *Nat. Microbiol.* **9**, 3120–3134 (2024).
32. Carlino, N. et al. Unexplored microbial diversity from 2,500 food metagenomes and links with the human microbiome. *Cell* **187**, 5775–5795.e15 (2024).
33. Hols, P. et al. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS Microbiol. Rev.* **29**, 435–463 (2005).
34. Ranawat, B., Bachani, P., Singh, A. & Mishra, S. *Enterobacter hormaechei* as plant growth-promoting bacteria for improvement in *Lycopersicum esculentum*. *Curr. Microbiol.* **78**, 1208–1217 (2021).
35. Rehr, B. & Klemme, J.-H. Formate dependent nitrate and nitrite reduction to ammonia by *Citrobacter freundii* and competition with denitrifying bacteria. *Antonie Van Leeuwenhoek* **56**, 311–321 (1989).
36. Hajjar, R., Ambaraghassi, G., Sebajang, H., Schwenter, F. & Su, S.-H. *Raoultella ornithinolytica*: emergence and resistance. *Infect. Drug Resist.* **13**, 1091–1104 (2020).
37. Iniguez, A. L., Dong, Y. & Triplett, E. W. Nitrogen fixation in wheat provided by *Klebsiella pneumoniae* 342. *Mol. Plant Microbe Interact.* **17**, 1078–1085 (2004).
38. Liu, D. et al. *Klebsiella pneumoniae* SnebYK mediates resistance against *Heterodera glycines* and promotes soybean growth. *Front. Microbiol.* **9**, 1134 (2018).
39. Mamphogoro, T. P., Kamutando, C. N., Maboko, M. M., Babalola, O. O. & Aiyegoro, O. A. Genome assembly of a putative plant growth-stimulating bacterial sweet pepper fruit isolate, *Enterobacter hormaechei* SRU4.4. *Microbiol. Resour. Announc.* **12**, e0123722 (2023).
40. Wicaksono, W. A. et al. The edible plant microbiome: evidence for the occurrence of fruit and vegetable bacteria in the human gut. *Gut Microbes* **15**, 2258565 (2023).
41. Kamathewatta, K. et al. Colonization of a hand washing sink in a veterinary hospital by an *Enterobacter hormaechei* strain carrying multiple resistances to high importance antimicrobials. *Antimicrob. Resist. Infect. Control* **9**, 163 (2020).
42. Anderson, M. T., Mitchell, L. A., Zhao, L. & Mobley, H. L. T. *Citrobacter freundii* fitness during bloodstream infection. *Sci. Rep.* **8**, 11792 (2018).
43. Wyres, K. L., Lam, M. M. C. & Holt, K. E. Population genomics of *Klebsiella pneumoniae*. *Nat. Rev. Microbiol.* **18**, 344–359 (2020).
44. Weber, M. & Fuchs, T. M. Metabolism in the niche: a large-scale genome-based survey reveals inositol utilization to be widespread among soil, commensal, and pathogenic bacteria. *Microbiol. Spectr.* **10**, e0201322 (2022).
45. Hoch, J. A. & Nester, E. W. Gene–enzyme relationships of aromatic acid biosynthesis in *Bacillus subtilis*. *J. Bacteriol.* **116**, 59–66 (1973).
46. Dosselaere, F. & Vanderleyden, J. A metabolic node in action: chorismate-utilizing enzymes in microorganisms. *Crit. Rev. Microbiol.* **27**, 75–131 (2001).
47. Chassagnole, C. et al. An integrated study of threonine-pathway enzyme kinetics in *Escherichia coli*. *Biochem. J.* **356**, 415–423 (2001).
48. Handzlik, M. K. & Metallo, C. M. Sources and sinks of serine in nutrition, health, and disease. *Annu. Rev. Nutr.* **43**, 123–151 (2023).
49. Watanabe, F., Yabuta, Y., Tanioka, Y. & Bito, T. Biologically active vitamin B12 compounds in foods for preventing deficiency among vegetarians and elderly subjects. *J. Agric. Food Chem.* **61**, 6769–6775 (2013).
50. de Crécy-Lagard, V., El Yacoubi, B., de la Garza, R. D., Noiriel, A. & Hanson, A. D. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *BMC Genomics* **8**, 245 (2007).
51. Nagy, P. L., Marolewski, A., Benkovic, S. J. & Zalkin, H. Formyltetrahydrofolate hydrolase, a regulatory enzyme that functions to balance pools of tetrahydrofolate and one-carbon tetrahydrofolate adducts in *Escherichia coli*. *J. Bacteriol.* **177**, 1292–1298 (1995).
52. Green, R. et al. Vitamin B12 deficiency. *Nat. Rev. Dis. Prim.* **3**, 17040 (2017).
53. Blanco-Míguez, A. et al. Extension of the *Segatella copri* complex to 13 species with distinct large extrachromosomal elements and associations with host conditions. *Cell Host Microbe* **31**, 1804–1819.e9 (2023).
54. Satija, A. et al. Healthful and unhealthful plant-based diets and the risk of coronary heart disease in U.S. adults. *J. Am. Coll. Cardiol.* **70**, 411–422 (2017).
55. Blanco-Miguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
56. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662 (2019).
57. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* **10**, e65088 (2021).
58. Pasolli, E. et al. Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **14**, 1023–1024 (2017).

Acknowledgements

We thank all the participants of the ZOE PREDICT studies, as well as members of the Computational Metagenomics Lab for their continued input and support, and members of ZOE Ltd. that made this research possible. This work was supported by Zoe Ltd. and co-funded by the European Union under Horizon Europe Programme CoDiet [101084642] to N.S. with UK activities supported by UK Research and Innovation under the UK government's Horizon Europe funding guarantee. More information on the CoDiet project can be found at https://www.codiet.eu/. This work was also partially supported by the European Research Council (ERC-STG project MetaPG-716575 and ERC-CoG microTOUCH-101045015) to N.S., by the European Union's Horizon 2020 programme (ONCOBIOME-825410 project, MASTER-818368 project and IHMCSA-964590 project) to N.S., by

the MUR PNRR project INEST-Interconnected Nord-Est Innovation Ecosystem (ECS00000043) funded by the NextGenerationEU to N.S., by the National Cancer Institute of the National Institutes of Health (1U01CA230551 to N.S.) and by the Premio Internazionale Lombardia e Ricerca 2019 to N.S. G.F. was funded by the European Union under the Marie Sklodowska-Curie grant agreement no. 101152592–plasticOME. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Author contributions

G.F., P.M., J.W., F.A., T.D.S. and N.S. conceived and designed the study. Metadata processing was performed by F.A., A.A., E.B., A.C.C., L.F., J.C.P., R.D., K.M.B. and S.E.B. Metagenomic data were pre-processed and quality controlled by F.A. and P.M. G.F. and P.M. performed the analyses. N.C., V.H., E.P., G.P. and R.D. contributed to the analyses. Data were visualized by G.F. and P.M. L.R. and V.H. contributed to the interpretation of the results. G.F., P.M. and N.S. wrote the paper with contribution and editing from all authors.

Competing interests

T.D.S. and J.W. are co-founders of ZOE Ltd. F.A., S.E.B., T.D.S. and N.S. are consultants to ZOE Ltd. A.A., E.B., A.C.C, L.F., J.C.P., R.D., J.W. and K.M.B. are or have been employees of Zoe Ltd. A.A., J.C.P., R.D., J.W., A.C.C, K.M.B., S.E.B., T.D.S., F.A. and N.S. receive options in ZOE Ltd. All other authors declare no competing interests.

Additional information

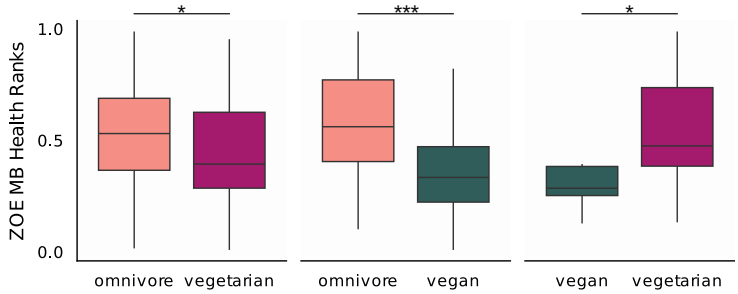
Extended data is available for this paper at https://doi.org/10.1038/s41564-024-01870-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-024-01870-z.

Correspondence and requests for materials should be addressed to Nicola Segata.

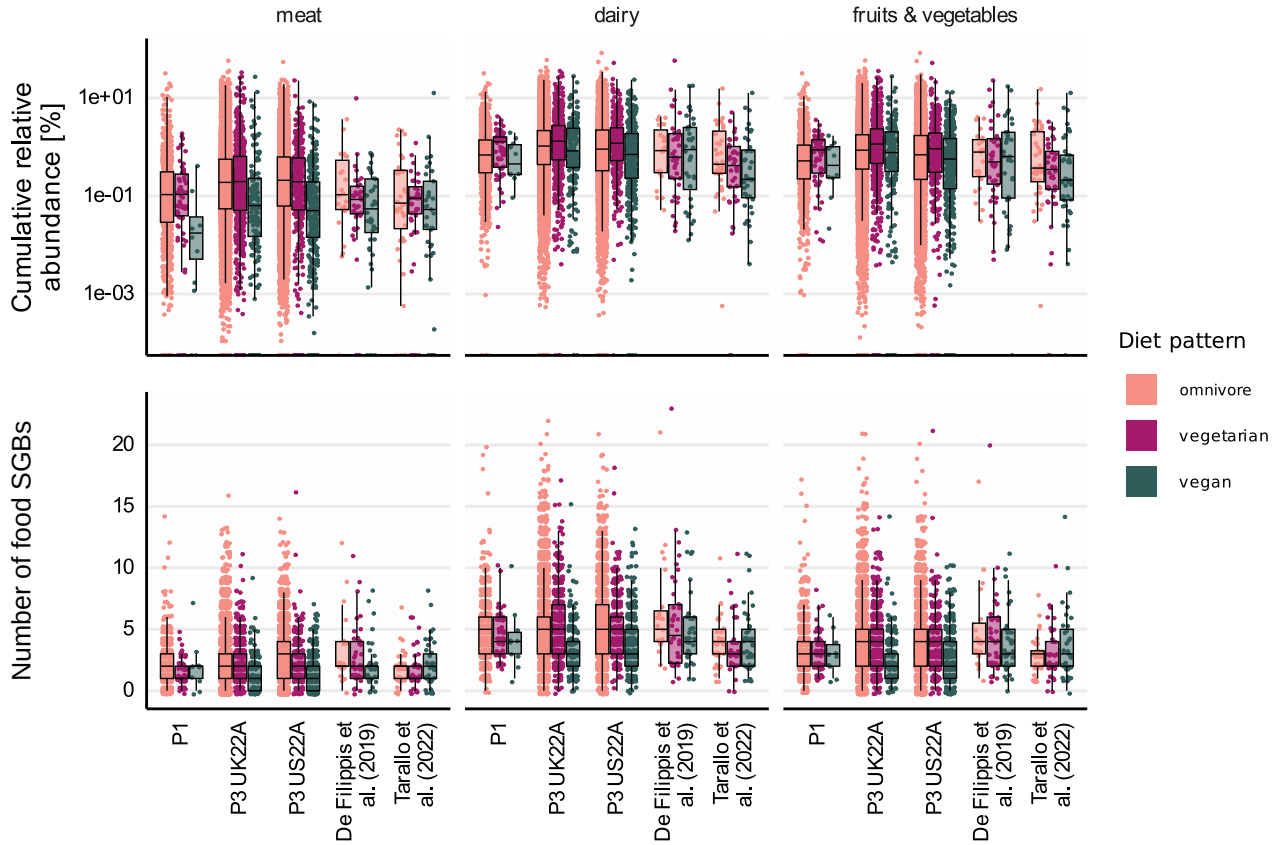
Peer review information *Nature Microbiology* thanks Sean Gibbons and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41564-024-01870-z



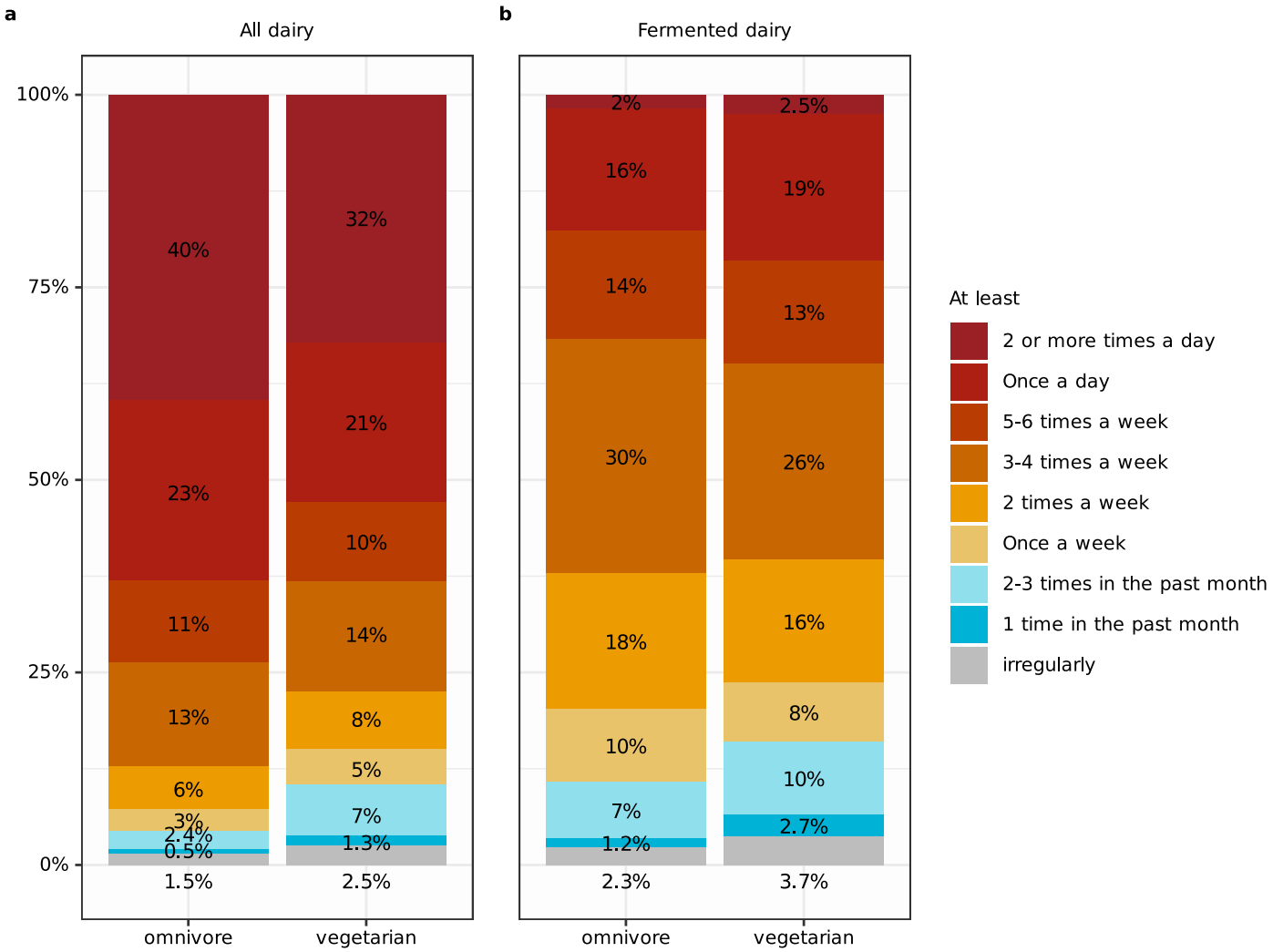
Extended Data Fig. 1 | Distribution of ZOE CMH rankings of SGBs signature of the three diet patterns. ZOE MB Health Ranks (y-axis; ranks closer to 0 indicate more favorable and those closer to 1 indicate less favorable cardiometabolic health outcomes) of all SGBs statistically significantly differentially abundant between each diet pattern pair (x-axis; $n_{omnivore\text{-}vegetarian} = 600$ SGBs, $n_{omnivore\text{-}vegan} = 724$, $n_{vegan\text{-}vegetarian} = 41$), colored by diet pattern (pink = omnivore, purple = vegetarian, green = vegan). Boxplots show the median, 25th and 75th percentiles, and whiskers extend to 1.5 times the interquartile range. Asterisks denote significance level of two sample t-tests (**Methods**), with * $0.05 > p < 0.01$ and *** $p \leq 0.001$.

https://doi.org/10.1038/s41564-024-01870-z



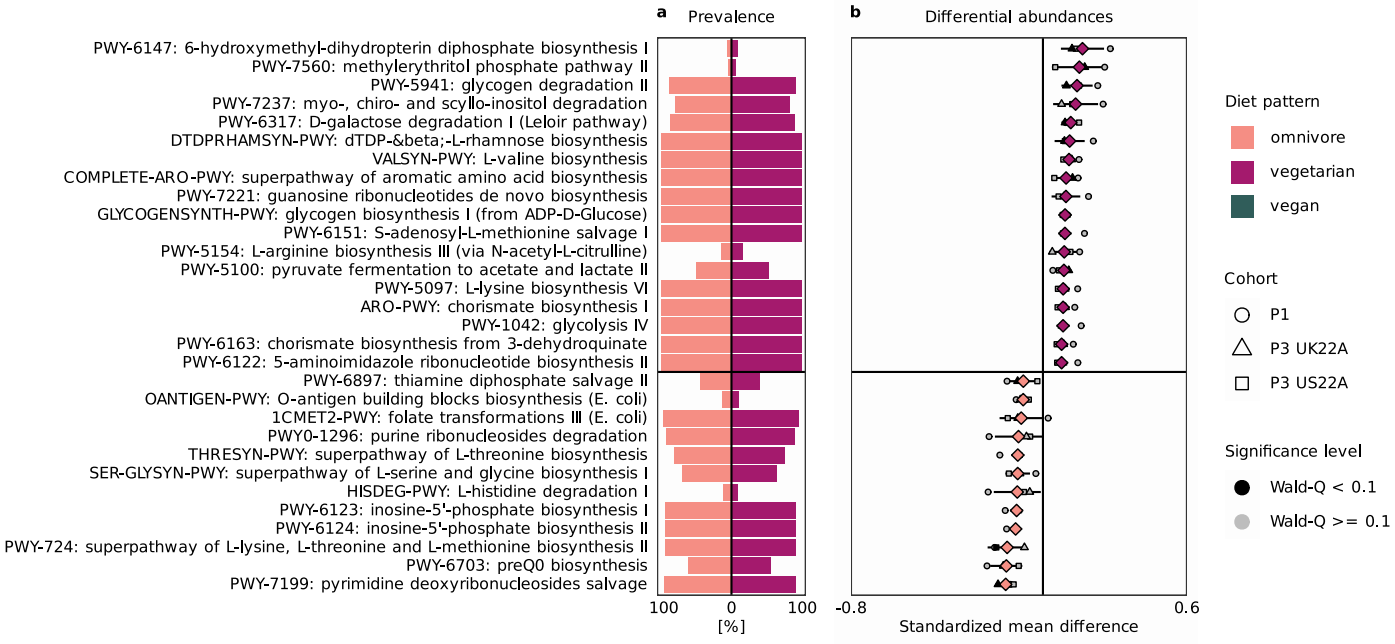
Extended Data Fig. 2 | Microbial food-to-gut transmission across the diet types and major food categories. Cumulative relative abundance (in %, $\log_{10}$ scale; upper panels) and prevalence, that is, count, (lower panels) of food SGBs (either meat, dairy, or fruits and vegetable-derived SGBs) within each individual, colored by diet pattern (pink = omnivore, purple = vegetarian, green = vegan) and grouped by cohort (P1, P3 UK22A, P3 US22A, De Filippis et al. [13], and Tarallo et al. [12]; Supplementary Tables 19, 20). Boxplots show the median, 25th and 75th percentiles, and whiskers extend to 1.5 times the interquartile range. Omnivores: $n_{P1} = 991$, $n_{P3\,UK22A} = 11{,}533$, $n_{P3\,US22A} = 7{,}228$, $n_{De\,Filippis} = 23$, $n_{Tarallo} = 40$; vegetarians: $n_{P1} = 59$, $n_{P3\,UK22A} = 623$, $n_{P3\,US22A} = 330$, $n_{De\,Filippis} = 38$, $n_{Tarallo} = 38$; vegans: $n_{P1} = 10$, $n_{P3\,UK22A} = 197$, $n_{P3\,US22A} = 373$, $n_{De\,Filippis} = 36$, $n_{Tarallo} = 40$ individuals.

https://doi.org/10.1038/s41564-024-01870-z



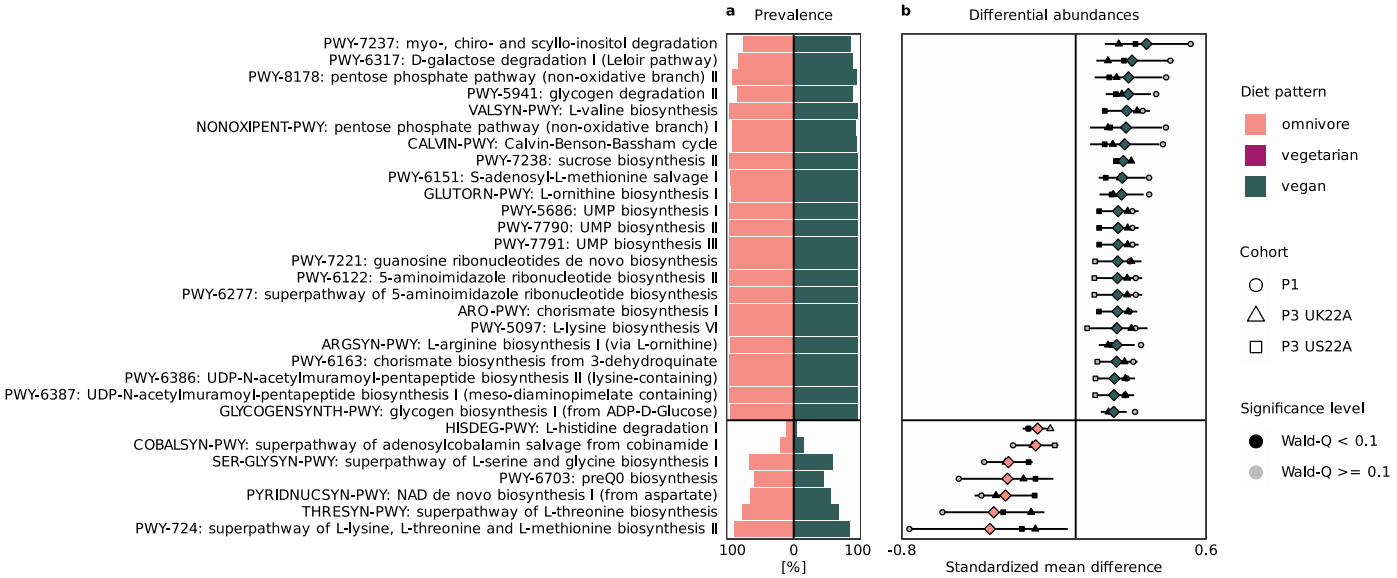
Extended Data Fig. 3 | Frequency of dairy consumption across omnivores and vegetarians in P3 UK22A and P3 US22A according to FFQs. a Percentage (y-axis) of omnivores and vegetarians (x-axis) that consume dairy products (milk, yogurt, cheese, butter, or other dairy) between 'two or more times per day' to 'irregularly'. The consumption frequency categories were given by the FFQs. Percentages within the bar plots indicate the dairy consumption prevalence of that diet pattern in that consumption category. **b** Same as in a, but considering only fermented dairy products (yogurt and cheeses).

https://doi.org/10.1038/s41564-024-01870-z



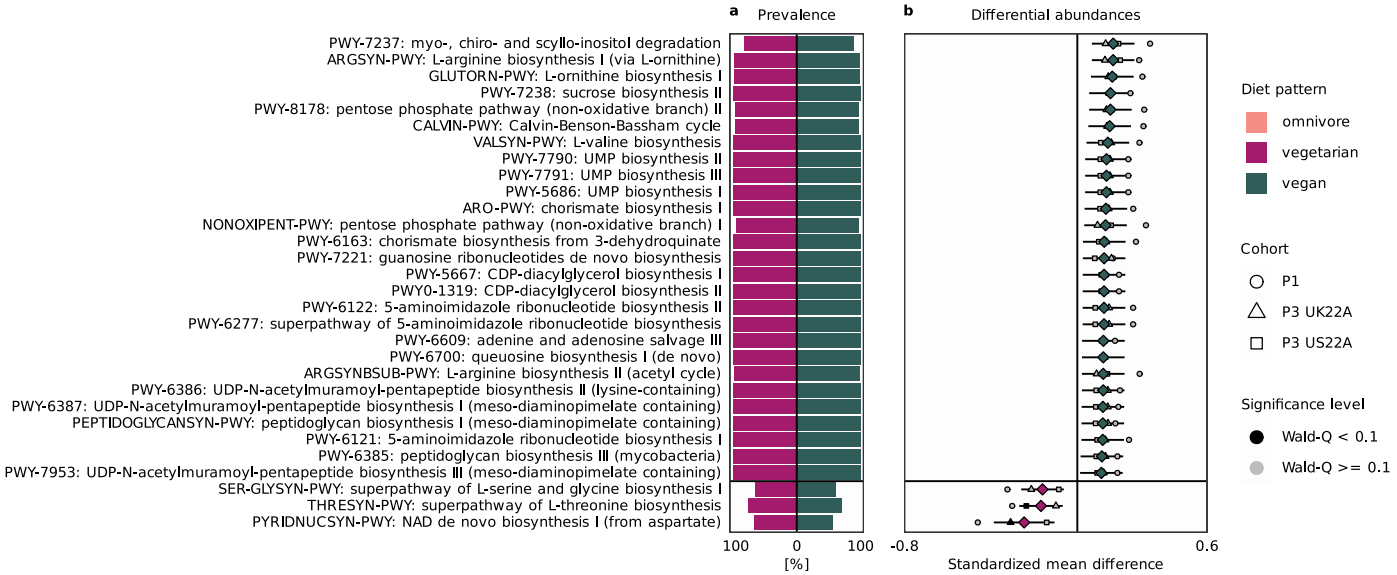
Extended Data Fig. 4 | Potential microbial functional signatures of omnivore and vegetarian gut microbiomes. a Prevalence (in %) of each functional pathway (y-axis) in omnivore (pink, left bars) and vegetarian (purple, right bars) gut microbiomes. **b** Meta-analyzed correlations between pathway relative abundance and diet pattern (omnivore vs vegetarian) for the top 30 pathways with the largest absolute standardized mean difference, upper and lower confidence intervals. Purple dots to the right indicate pathway-associations with vegetarians, while pink dots to the left indicate pathway-associations with omnivores. Also shown in smaller shapes are the per-cohort correlations, with shapes filled in black indicating a Wald q-value < 0.1 and those filled in gray indicating a Wald q-value ≥ 0.1. The black horizontal bar indicates the separation between the correlations with omnivores vs vegetarians for ease of visualization only. Shown are only the pathways with a prevalence of less than 0.05 across samples of at least one diet pattern, a coverage less than 0.2, and which were significant at q < 0.1. $n_{omnivores} = 19{,}817$, $n_{vegetarians} = 1{,}088$.

https://doi.org/10.1038/s41564-024-01870-z



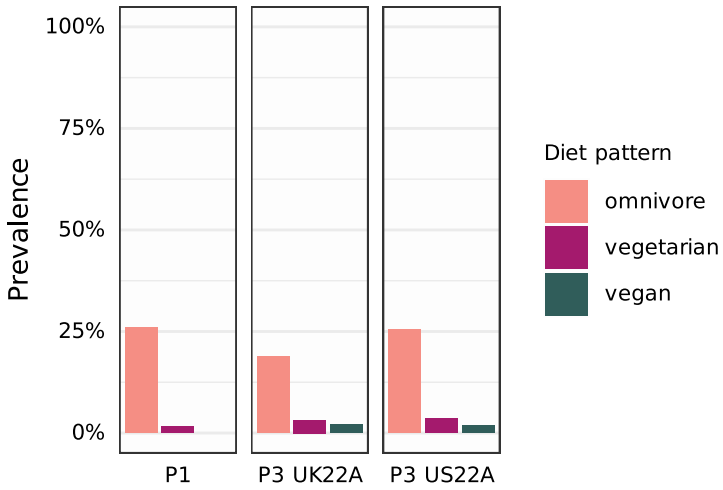
Extended Data Fig. 5 | Potential microbial functional signatures of omnivore and vegan gut microbiomes. a Prevalence (in %) of each functional pathway (y-axis) in omnivore (pink, left bars) and vegan (green, right bars) gut microbiomes. **b** Meta-analyzed correlations between pathway relative abundance and diet pattern (omnivore vs vegan) for the top 30 pathways with the largest absolute standardized mean difference, upper and lower confidence intervals. Green dots to the right indicate pathway-associations with vegans, while pink dots to the left indicate pathway-associations with omnivores. Also shown in smaller shapes are the per-cohort correlations, with shapes filled in black indicating a Wald q-value < 0.1 and those filled in gray indicating a Wald q-value ≥ 0.1. The black horizontal bar indicates the separation between the correlations with omnivores vs vegans for ease of visualization only. Shown are only the pathways with a prevalence of less than 0.05 across samples of at least one diet pattern, a coverage less than 0.2, and which were significant at q < 0.1. $n_{omnivores} = 19{,}817$, $n_{vegans} = 656$.

https://doi.org/10.1038/s41564-024-01870-z



Extended Data Fig. 6 | Potential microbial functional signatures of vegetarian and vegan gut microbiomes. a Prevalence (in %) of each functional pathway (y-axis) in vegetarian diet (purple, left bars) and vegan (green, right bars) gut microbiomes. **b** Meta-analyzed correlations between pathway relative abundance and diet pattern (vegetarian vs vegan) for the top 30 pathways with the largest absolute standardized mean difference, upper and lower confidence intervals. Green dots to the right indicate pathway-associations with vegans, while purple dots to the left indicate pathway-associations with vegetarians. Also shown in smaller shapes are the per-cohort correlations, with shapes filled in black indicating a Wald q-value < 0.1 and those filled in gray indicating a Wald q-value ≥ 0.1. The black horizontal bar indicates the separation between the correlations with vegetarians vs vegans for ease of visualization only. Shown are only the pathways with a prevalence of less than 0.05 across samples of at least one diet pattern, a coverage less than 0.2, and which were significant at q < 0.1. $n_{vegetarians} = 1{,}088$, $n_{vegans} = 656$.

https://doi.org/10.1038/s41564-024-01870-z



Extended Data Fig. 7 | Estimation of animal-based food consumption based on metagenomic reads. Prevalence (in %) of animal DNA as estimated by MEDI using the gut metagenomes of omnivores (bars in pink), vegetarians (bars in purple), and vegans (bars in green) in the P1, P3 UK22A and P3 US22A cohorts (Supplementary Table 24).

Corresponding author(s): Nicola Segata

Last updated by author(s): Oct 9, 2024

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	n/a
Data analysis	R (version 4.2.2), MetaPhlAn 4 (version 4.beta.2, database vJan21_CHOCOPhlAnSGB_202103), HUMAnN (version 3.6), scikit-learn python library (version 0.22.2). The code for the analyses conducted here is provided in Supplementary File 1. The pooled estimate of effect sizes from linear models was computed based on the pipeline https://github.com/waldronlab/curatedMetagenomicDataAnalyses/blob/main/python_tools/metaanalyze.py. An importable meta-analysis python library is also freely available at https://github.com/SegataLab/inverse_var_weight/blob/main/meta_analyses.py.

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

The publicly available datasets used in this work are available from their respective publications in Tarallo et al. 2022 and De Filippis et al. 2019. Raw metagenomic samples are provided for all participants of the ZOE PREDICT Studies. Specifically, PREDICT 1 has already been made publicly available as reported previously (Asnicar, Berry, et al. 2021) under the NCBI-SRA bioproject ID PRJEB39223, whereas the PREDICT 2 is deposited in EBI under accession number PRJEB75460, and PREDICT 3 cohorts under EBI accession numbers PRJEB75463 and PRJEB75464. Sex, age, BMI, country, and the quantitative taxonomic profiles are available for each sample within the curatedMetagenomicData package Pasolli et al. 2017. The ZOE Microbiome Rankings for the full list of species are made available (and kept up-to-date) at https://zoe.com/our-science/microbiome-ranking and in their current version are reported in Supplementary Table 14. ZOE is the owner of the pseudonymized data and metadata and researchers interested in follow-up studies requiring additional specific metadata information should fill out a research request proposal at https://zoe.com/our-science/collaborate [to appear upon acceptance] that will be evaluated by a sub-panel of the ZOE Scientific Advisory Board once per month for their priority, relevance and in compliance with privacy and data protection regulations.

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	Data on sex (not gender) were collected with informed consent of participants (see Data Availability Statement). Sex was considered throughout the analysis as an explanatory variable, as is indicated throughout the Methods.
Reporting on race, ethnicity, or other socially relevant groupings	Data on socially relevant variables such as race or ethnicity were neither collected nor considered in this research, where the focus was on dietary patterns pertaining to veganism, vegetarianism and mixed diets.
Population characteristics	Participants were aged 52 +/- 12.5 years (mean +/- standard deviation). Genotype and diagnosis information was not collected.
Recruitment	Information pertaining to the publicly available datasets used in this work are available from their respective publications: (Tarallo et al. 2022) and (De Filippis et al. 2019). Participants of P1, P2, P3 US22A, and P3 UK22A all gave informed study consent either written or electronically. In addition, P3 US22A and P3 UK22A participants gave product research consent during the course of product purchase at ZOE Ltd. Only the US subset of P1 received modest direct financial compensation for their participation. All other participants did not receive direct financial compensation beyond reimbursement of expenses incurred.
Ethics oversight	Information pertaining to the publicly available datasets used in this work are available from their respective publications: Asnicar et al. 2021, Tarallo et al. 2022 and De Filippis et al. 2019. Both P3 plus P2 clinical trials were registered at https://www.clinicaltrials.gov (clinical trial identifier for P3: NCT04735835; P2: NCT03983733) and ethical approval was obtained (P3 US protocol number (IRB): Pro00044316; P3 UK ethical review reference: HR-23/24-28300; P2 IRB: Pro00033432).

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	This study encompassed two published, publicly available datasets (Tarallo et al. 2022 with 118 individuals and De Filippis et al. 2019 with 97 individuals) along with three ZOE PREDICT datasets: P1, a UK cohort with 1,062 individuals; P3 UK22A a UK cohort with 12,353 individuals; and P3 US22A, a US cohort with 7,931 individuals. In total, 656 vegans, 1,088 vegetarians, and 19,817 omnivores were sampled. When possible, we also included samples from the ZOE PREDICT 2 (P2) cohort, which encompassed only omnivores from the US (843 individuals), thus limiting its usability in this analysis.
Data exclusions	Data was only excluded in the alpha diversity analysis, in which we considered all observations outside the 95% CI to be outliers, which removed 22 out of the 21,561 samples.
Replication	The gut microbial signatures of the 3 diet patterns were replicable across 5 independent cohorts (P1, P3 US22A, P3 UK22A, De Filippis et al.

Replication	(2019) and Tarallo et al. (2022) using a cross-LODO machine learning approach. Linking these patterns to FFQs and individual physiological data was replicable across 4 cohorts for which FFQ data were present (P1, P2, P3 US22A, and P3 UK22A) using a meta-analytical approach.
Randomization	Participants were grouped into dietary patterns according to their reported dietary patterns. Randomization was only necessary when cross-LODO machine learning was applied, in which case a per-cohort (ten-times, ten-folds) cross-validation was performed, in which the rest of the cohorts is added to each training set as a support.
Blinding	Authors who extracted and sequenced stool samples did not conduct the microbiome 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
☒	☐ Plants

Methods

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

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com