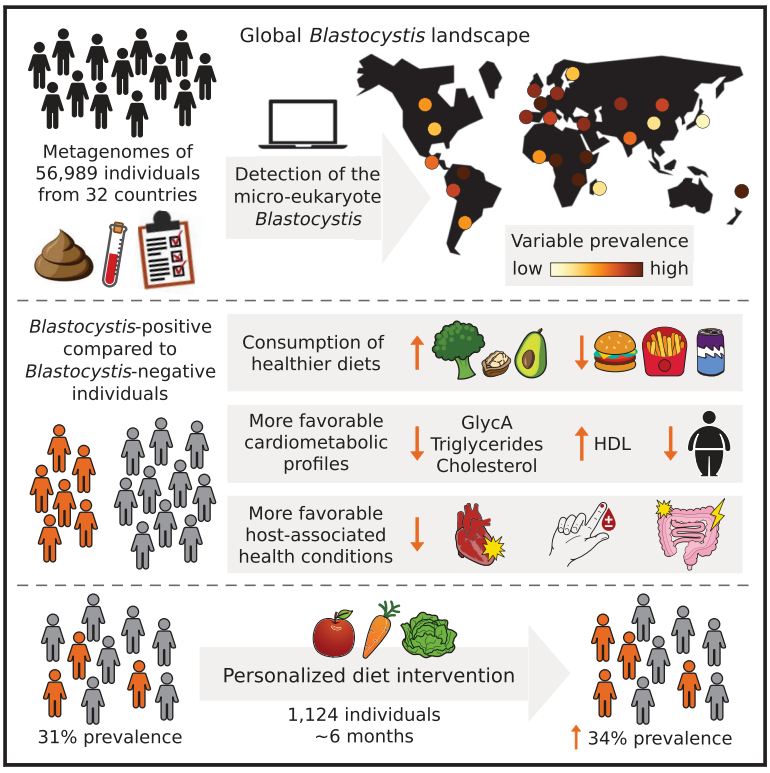
Cell

Article

Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries

Graphical abstract



Highlights

- Prevalence of *Blastocystis* in the gut microbiome varies across geography and lifestyle
- Healthier plant-based diets are linked with higher gut *Blastocystis* carriage
- *Blastocystis*-positive subjects tend to have healthier cardiometabolic profiles
- *Blastocystis* presence increases after a diet-improving intervention program

Authors

Elisa Piperni, Long H. Nguyen, Paolo Manghi, ..., Tim D. Spector, Nicola Segata, Francesco Asnicar

Correspondence

nicola.segata@unitn.it (N.S.), f.asnicar@unitn.it (F.A.)

In brief

A global-scale metagenomic exploration of *Blastocystis* was performed from 56,989 individuals across 32 countries. *Blastocystis* was seen to be widely present in healthy adults, with prevalence and abundance patterns linked to geography, dietary habits, and cardiometabolic profiles.

Piperni et al., 2024, Cell *187*, 1–17
August 22, 2024
https://doi.org/10.1016/j.cell.2024.06.018

CellPress

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

Cell

CellPress
OPEN ACCESS

Article

Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries

Elisa Piperni,[1,2,16] Long H. Nguyen,[3,4,16] Paolo Manghi,[1] Hanseul Kim,[3] Edoardo Pasolli,[5] Sergio Andreu-Sánchez,[6,7] Alberto Arrè,[1,8] Kate M. Bermingham,[8,9] Aitor Blanco-Míguez,[1] Serena Manara,[1] Mireia Valles-Colomer,[1] Elco Bakker,[8] Fabio Busonero,[10] Richard Davies,[8] Edoardo Fiorillo,[10] Francesca Giordano,[8] George Hadjigeorgiou,[8] Emily R. Leeming,[11] Monia Lobina,[10] Marco Masala,[10] Andrea Maschio,[10] Lauren J. McIver,[12] Mauro Pala,[10] Maristella Pitzalis,[10] Jonathan Wolf,[8] Jingyuan Fu,[6,7] Alexandra Zhernakova,[6] Simone M. Cacciò,[13] Francesco Cucca,[10,14] Sarah E. Berry,[9] Danilo Ercolini,[5] Andrew T. Chan,[3,4] Curtis Huttenhower,[12,15] Tim D. Spector,[11,17] Nicola Segata,[1,2,11,17,*] and Francesco Asnicar[1,17,18,*]

[1]Department CIBIO, University of Trento, Trento, Italy
[2]IEO, Istituto Europeo di Oncologia IRCSS, Milan, Italy
[3]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA, USA
[4]Harvard Chan Microbiome in Public Health Center, Boston, MA, USA
[5]Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy
[6]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
[7]Department of Pediatrics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
[8]Zoe Ltd, London, UK
[9]Department of Nutritional Sciences, King's College London, London, UK
[10]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy
[11]Department of Twins Research and Genetic Epidemiology, King's College London, London, UK
[12]Harvard T.H. Chan School of Public Health, Boston, MA, USA
[13]Department of Infectious Diseases, Istituto Superiore Di Sanità, Rome, Italy
[14]Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy
[15]The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[16]These authors contributed equally
[17]These authors contributed equally
[18]Lead contact
*Correspondence: nicola.segata@unitn.it (N.S.), f.asnicar@unitn.it (F.A.)
https://doi.org/10.1016/j.cell.2024.06.018

SUMMARY

Diet impacts human health, influencing body adiposity and the risk of developing cardiometabolic diseases. The gut microbiome is a key player in the diet-health axis, but while its bacterial fraction is widely studied, the role of micro-eukaryotes, including *Blastocystis*, is underexplored. We performed a global-scale analysis on 56,989 metagenomes and showed that human *Blastocystis* exhibits distinct prevalence patterns linked to geography, lifestyle, and dietary habits. *Blastocystis* presence defined a specific bacterial signature and was positively associated with more favorable cardiometabolic profiles and negatively with obesity ($p < 1e{-}16$) and disorders linked to altered gut ecology ($p < 1e{-}8$). In a diet intervention study involving 1,124 individuals, improvements in dietary quality were linked to weight loss and increases in *Blastocystis* prevalence ($p = 0.003$) and abundance ($p < 1e{-}7$). Our findings suggest a potentially beneficial role for *Blastocystis*, which may help explain personalized host responses to diet and downstream disease etiopathogenesis.

INTRODUCTION

Decades of evidence have linked dietary intake with chronic disease development.[1,2] While specific recommendations vary, leading authorities[3–5] endorse a high-quality diet for the amelioration of long-term disease risk, despite great variability in cardiometabolic benefits from person to person.[6] We recently showed in a single-cohort study that human gut microbial communities harboring *Blastocystis* were associated with improved postprandial glucose responses, measures of body adiposity, and other favorable biomarker profiles.[7] *Blastocystis* is a prevalent unicellular eukaryote whose role in host health is still unclear.[8] *Blastocystis* is remarkably diverse at the genetic level[9] and is currently categorized into at least 28 distinct species, termed subtypes (STs).[10,11] At least eight of these STs have been identified in humans, with ST1, ST2, and ST3 being the

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

most common.[12,13] Despite earlier metagenomic work on the prevalence of *Blastocystis* (up to 1,689 participants[9,14]), larger and more diverse cohorts with precise dietary information are needed to establish the relationship between gut *Blastocystis*, nutrition, and cardiometabolic health outcomes that refer here to the development of obesity, insulin resistance culminating in type 2 diabetes (T2D), or cardiovascular disease such as stroke or coronary artery disease.

We integrated and harmonized data from multiple cohorts to characterize the global distribution of *Blastocystis* and investigate whether its presence could help explain the heterogeneous and individualized contributions of food intake on cardiometabolic health. We found highly variable prevalences of *Blastocystis* and its STs influenced by host species (i.e., human vs. non-human), age, geography, and lifestyle. In the ZOE PREDICT studies (Personalised REsponses to DIetary Composition Trial; PREDICT 1, PREDICT 2, and PREDICT 3 US21, US22A, and UK22A) for which dietary and health risk variables are available, we observed a tight coupling between diet quality and *Blastocystis* carriage, as well as its presence among participants with more favorable cardiometabolic profiles. We externally validated these findings to demonstrate a negative association between *Blastocystis* carriage and body adiposity, as well as disorders for which a link with the gut microbiome was previously established, including inflammatory bowel diseases (IBD),[15] colorectal cancer (CRC),[16] and diabetes.[17] Finally, we observed a significant increase in *Blastocystis* prevalence and abundance among the participants of a 6-month personalized diet intervention study that improved diet quality and lost weight. Together, our findings may implicate *Blastocystis* as a marker of a healthier lifestyle with a potentially beneficial role in personalized host responses to food and downstream disease susceptibility.

RESULTS

A large, diverse, and harmonized metagenomic survey of the global *Blastocystis* landscape

To link *Blastocystis* presence with diet, health outcomes, and geography, we leveraged our access to five large multinational studies with granular information on microbiome, food intake, and cardiometabolic biomarkers: the ZOE PREDICT studies. These include the previously published PREDICT 1[7] and four other cohorts, namely PREDICT 2 and three PREDICT 3 cohorts (US21, US22A, and UK22A) (STAR Methods).[7,18,19] Together, these cohorts comprise 13,354 individuals from diverse areas of the United Kingdom (UK) (PREDICT 1 n = 1,001, PREDICT 3 UK22A n = 12,353) and 21,333 from the United States (US) (PREDICT 1 n = 97, PREDICT 2 n = 968, PREDICT 3 US21 n = 11,798, and US22A n = 8,470), for a total of 34,687 individuals, of which 25,548 were considered as healthy adults (STAR Methods). For each participant in the PREDICT studies, demographic information and cardiometabolic blood markers were collected at baseline. Long-term habitual *ad libitum* diet was assessed using established food frequency questionnaires (FFQs), while daily diet (i.e., short-term) was documented using digital food logs (STAR Methods).

We also included 2 cohorts (SardiNIA with 2,596 metagenomes sequenced as part of this study and 6,047 samples from the Dutch Microbiome Project [DMP]; STAR Methods) with information on age, sex, body mass index (BMI), and health status of the host; and five cohorts with additional dietary preference information, three of which enroll participants from Italy[20–22] (DeFilippisF_2019, MeslierV_2020, and TaralloS_2021; STAR Methods) and two from the US[23,24] (Mind-Body Study [MBS] and Men's Lifestyle Validation Study [MLVS]; STAR Methods). To characterize *Blastocystis* presence at the global level, we additionally identified and collected 61 publicly available cohorts with shotgun metagenomic samples of the human gut microbiome from curatedMetagenomicData ver. 3 (cMD3),[25] with hosts' information including age, sex, country of origin, and BMI (STAR Methods).

Together, these human data add up to a total of 73 datasets and stool metagenomes from 56,989 distinct individuals (Table S1), of which n = 41,428 were from healthy/control individuals and n = 15,561 were considered "not healthy" (i.e., generally cases from various case/control microbiome studies; STAR Methods). For 1,034 of these individuals, additional longitudinal gut microbiome samples were available (STAR Methods; Table S4). We further included 1,124 healthy individuals from the ZOE PREDICT personalized diet intervention study (PREDICT 3 UK 23A RT) with microbiome samples collected before and after the intervention (STAR Methods). Moreover, we included samples from other human body sites, namely the oral cavity (n = 243) and skin (n = 237; STAR Methods; Table S1), as well as 28 ancient human metagenomes from paleofeces (STAR Methods).[26–31]

As *Blastocystis* was previously reported in non-human hosts,[32] we considered 4,590 gut metagenomes from 214 species (spanning mammals, reptiles, birds, amphibians, insects, crustaceans, mollusks, nematodes, and metazoans) from 49 public datasets (STAR Methods; Table S2). In total, the current work compiled 139 distinct human and non-human datasets totaling 65,809 metagenomes, each profiled using the same validated computational workflow[7,9] to determine the presence and abundance of the eight genetically distinct *Blastocystis* STs previously described in humans[12] (STAR Methods; Figure S1A).

Global *Blastocystis* prevalence is variable and confined to gut microbial communities

Blastocystis was found in 8,190 human stool samples (out of a total of n = 56,989), and as expected, it was not detected in other human body sites (Table S1). Among healthy adults from 32 countries, we found significant geographical heterogeneity in gut *Blastocystis* prevalence. The lowest prevalence was observed in North America (6.64%, all Fisher-test q values < 1e−3), while Europe showed comparatively higher prevalence (22.14%) (Figure 1A; Tables 1 and S1). At the country level, Fiji[33] had the highest *Blastocystis* prevalence among well-represented regions with >50 samples (56.29%, n = 167), while Japan showed the lowest (2.46%, n = 244), a finding supported by prior investigations.[34,35] We observed high *Blastocystis* prevalences in Tanzania, Ethiopia, and Cameroon (>35%; Table S1). European countries showed moderately high prevalences (typically greater than 15%), while Asian countries ranged from 2.46% in Japan to 29.18% in Israel. Consistent with prior studies,[9,36,37] we found low *Blastocystis* prevalence in the US, with 951 positive samples out of 14,692

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

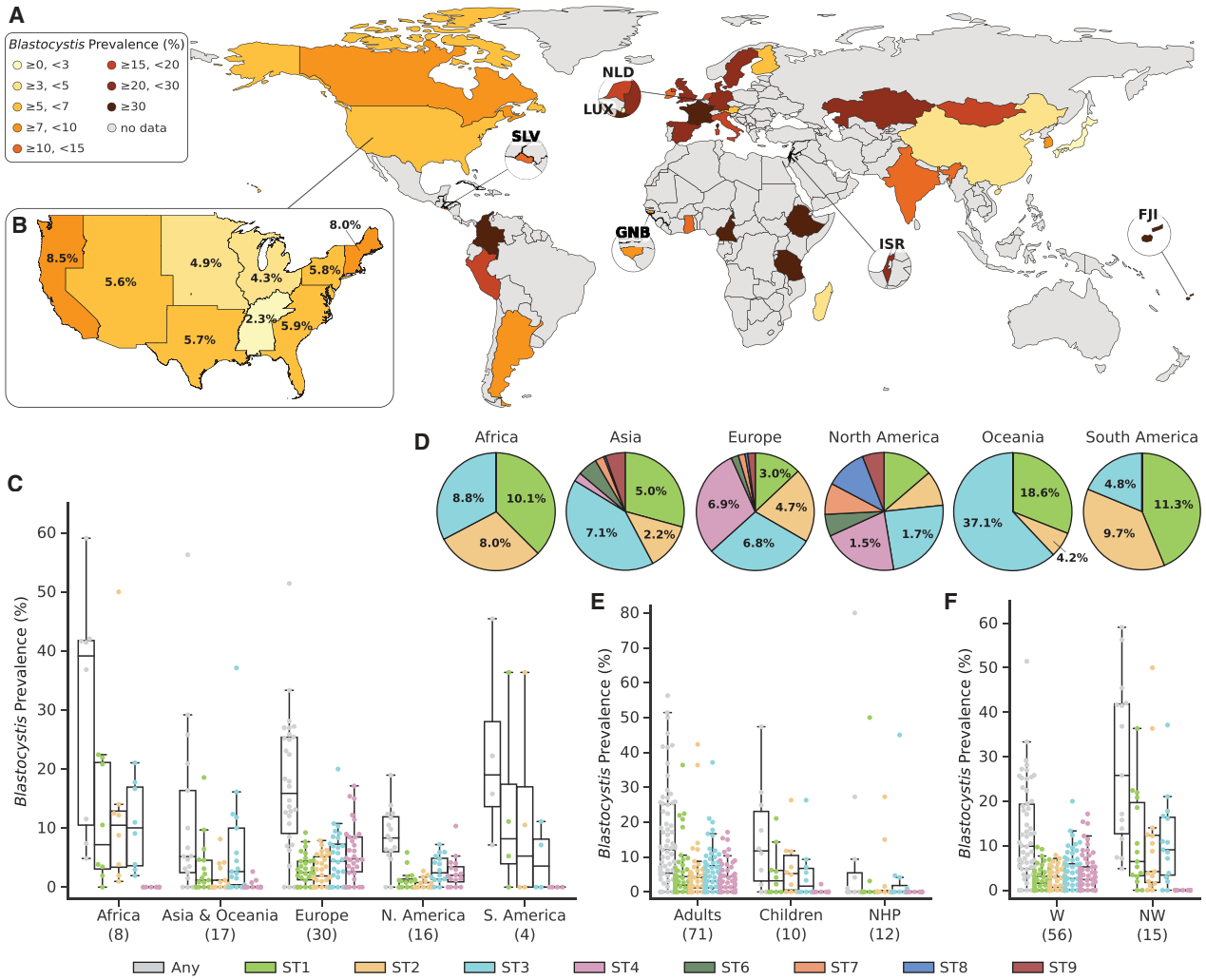


Figure 1. Worldwide *Blastocystis* prevalence suggests a non-pathogenic role in the human gut microbiome and reveals distinct ethno-geographical patterns

(A) Overall *Blastocystis* prevalence varies substantially depending on the geographical region. Prevalence estimates were calculated among samples from healthy individuals 12 years of age or older (only countries with >10 samples were colored, 32 in total; Table S1).

(B) US coastal regions have a higher prevalence of *Blastocystis* than landlocked regions (p value = 5.04e−10, Table S1). The PREDICT 1 (US), PREDICT 2, and PREDICT 3 (US21 and US22A) cohorts (n = 14,692 healthy individuals with state information) were used to assess prevalence in each US Census Bureau Division. Only the continental US is depicted.

(C) *Blastocystis* subtypes (STs) are differentially distributed by continent. ST4 was not detected in Africa and South America, while it was highly prevalent in Europe. The overall prevalence ("Any") comprises all detected STs, and single STs are plotted when prevalent above 1% in at least one continent (i.e., only ST1, ST2, ST3, and ST4). In total, we considered 71 datasets with at least 10 samples each, and dots represent the prevalence of *Blastocystis* computed on healthy individuals (age > 12 years) of that continent. Oceania was grouped with Asia as only one dataset was available, the number of datasets for each continent is reported in parentheses. Five datasets contained samples from more than one continent and were split (Table S1).

(D) Relative prevalence of *Blastocystis* subtypes shows that ST6, ST7, ST8, and ST9 were found only in Europe, North America, and Asia, always with a prevalence <1%. The inset numbers correspond to the absolute prevalence of each ST when ≥1%.

(E) Children (aged between 1 and 11 years) show a generally lower prevalence of *Blastocystis* than adults and were positive for only ST1, ST2, ST3, and ST4. *Blastocystis* profiling of non-human primates (NHPs) gut microbiome samples detected sporadically only ST1, ST2, and ST3 (Table S2).

(F) Overall *Blastocystis* prevalence was higher in non-Westernized (NW) samples, despite ST4 being detected only in Westernized (W) populations.

The boxplots show the first and third quartiles (boxes) and the median (middle line); whiskers extend to 1.5× the interquartile range (IQR).

See also Figure S1 and Tables S1 and S2.

(6.47%) from PREDICT 1 (US), PREDICT 2, and PREDICT 3 US21 and US22A participants (Figure 1B; Table S1). Our large US population allowed further investigation by US Census Bureau Divisions, demonstrating a comparative bicoastal increase vs. the Midwest and the South (Figure 1B). Generally, the proportion of *Blastocystis*-positive samples increased when considering

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

Table 1. Worldwide *Blastocystis* prevalence in the gut microbiome of healthy adults

Dataset	ST1 (%)	ST2 (%)	ST3 (%)	ST4 (%)	ST6 (%)	ST7 (%)	ST8 (%)	ST9 (%)	Prevalence (%)	*n* samples	
PREDICT 1	3.01	5.10	8.93	7.92	0.64	0.36	0.09	0.64	25.68	1,098	
PREDICT 2	1.25	1.02	2.84	1.93	0.34	0.91	0.46	0.34	8.65	879	
PREDICT 3 (UK22A)	3.52	6.12	7.35	8.98	0.52	0.54	0.14	0.53	26.99	9,855	
PREDICT 3 (US21)	0.78	0.69	1.70	1.44	0.50	0.50	0.77	0.50	6.37	7,939	
PREDICT 3 (US22A)	1.16	0.64	1.23	1.19	0.33	0.73	0.92	0.31	6.18	5,777	
									total healthy adults	25,548	
Continent	ST1 (%)	ST2 (%)	ST3 (%)	ST4 (%)	ST6 (%)	ST7 (%)	ST8 (%)	ST9 (%)	Prevalence (%)	*n* samples	*n* datasets
Africa	10.10	8.01	8.81	0	0	0	0	0	23.40	624	8
Asia	5.01	2.23	7.10	0.42	0.93	0.42	0.09	0.93	15.92	2,154	16
Europe	2.95	4.67	6.84	6.86	0.46	0.42	0.17	0.47	22.14	21,511	31
North America	0.97	0.68	1.71	1.47	0.43	0.61	0.80	0.42	6.64	15,480	16
Oceania	18.56	4.19	37.13	0	0	0	0	0	56.29	167	1
South America	11.29	9.68	4.84	0	0	0	0	0	20.97	62	4
Overall	2.49	3.06	5.02	4.28	0.47	0.48	0.41	0.47	15.97	39,998	71

The table shows *Blastocystis* prevalence for the five ZOE PREDICT cohorts and globally for the six continents represented by the recruited datasets. Subtypes (STs) distribution varied substantially based on geographical region. Prevalence estimates were calculated among samples from healthy individuals 12 years of age or older. See also Table S1.

only individuals with "healthy-weight" BMI (i.e., 18.5 $\leq$ BMI < 25 kg/m^2; Table S1).

To validate these biological and ecological differences and to rule out the contribution of variable read depth between samples, all samples were rarefied to 10 million (M) random reads (STAR Methods) and re-assessed for *Blastocystis* presence. At the country level, prevalences did not change substantially, reproducing what was reported at full read depth (Table 1; Figure S1C). Moreover, *Blastocystis* prevalences in the two UK-based PREDICT cohorts (PREDICT 1 and PREDICT 3 UK22A) were alike (26.97% and 26.99%, respectively), despite samples from the latter being sequenced to a target depth of half the former (mean number of reads: 58.3 M vs. 31.2 M, respectively; Table S1). Finally, we observed that prevalence was not saturated despite increasing the number of reads considered, even at high sequencing depths (Figure S1D). Thus, to maximize workflow sensitivity, we performed downstream analyses exploiting the complete set of reads for each sample. Importantly, a potential limit of detection does not influence intra-cohort comparisons, as samples coming from the same dataset generally had similar sequencing depths.

Distribution of *Blastocystis* STs varies by geography, age, host, and lifestyle

We explored the relationships of *Blastocystis* ST carriage with geographic and lifestyle differences (broadly defined as Westernized vs. non-Westernized, a definition that encompasses multiple aspects; STAR Methods). Both ST1 and ST2 were more prevalent in non-Westernized populations, while ST4 was detected only in samples from Westernized individuals (Figures 1C and 1D). Consistent with prior studies, we commonly found ST4 in Europe (6.86%), while it was only sparingly present in some non-European countries (China, Kazakhstan, Israel, US, and Canada, each below 3%; Table S1).[12] We did not detect ST6, ST7, ST8, and ST9 in non-Westernized populations from Africa, Asia, Oceania, and South America. Conversely, in North America where *Blastocystis* prevalence was generally low, ST6-9 were detected at comparable levels to ST1 and ST2 (prevalence < 1%). Consistent with prior findings, we did not detect *Blastocystis* in newborns (only ST1 was detected in 1 out of 831 healthy newborn samples; Table S1), suggesting it is likely acquired later in life and not subject to mother-to-infant transmission in the first months of life.[38–42] Conversely, in children (aged between 1 and 11 years), we detected only *Blastocystis* STs ST1, ST2, ST3, and ST4 (Figure 1E; Table S1). In a recent study analyzing a non-Westernized population from Ethiopia,[42] we also showed that *Blastocystis* STs were not only differentially prevalent across lifestyles and age categories but also differentially shared between mother and infant (e.g., ST2 was shared at the strain level when present, whereas other STs were rarely shared).

We then found that the eight human *Blastocystis* STs were not detected in most of the animal species tested (Table S2), and only ST1, ST2, and ST3 were rarely found in non-human primates (NHPs), almost exclusively when kept in captivity (Fisher's exact p value = 3.18e−7), suggesting potential transmission from humans.[43] Notably, we found *Blastocystis* ST7 in a sample from the wild bird species *Anas crecca*, a finding previously described among the Anatidae family.[44] Such rare detection of *Blastocystis* in animals is expected as we focused on the eight STs previously reported in humans. Other *Blastocystis* STs exclusively found in non-human hosts typically lack complete sequenced genomes and hence could not be profiled by our pipeline.[11] Taken together, the worldwide prevalence of human *Blastocystis* and its STs differed based on a variety of factors, including host type, age, and geography. Also, its somewhat common presence among healthy individuals (~16%) suggests a potential non-pathogenic role in human gut microbial communities.

To investigate whether *Blastocystis* carriage is specific to our modern times,[45,46] we profiled its presence among 28 ancient metagenomes from paleofeces found in several distinct

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

locations (STAR Methods).[26–31] We detected ST1 in three samples (two from Mexico dated ~AD 650 and one from the US dated ~AD 595), while ST3 was detected in one sample (from Mexico, ~AD 650; Table S1). The detection of *Blastocystis* in ancient human gut metagenomes suggests that it was relatively prevalent even in remote human history and hence does not serve as a hallmark of a modern microbiome configuration.

Blastocystis colonization is persistent and not linked with host genetics

Individuals with longitudinal sampling at multiple time points further allowed us to evaluate the persistence of *Blastocystis* colonization (STAR Methods). In most cases, the same individual remained either positive or negative over time (95.2%, 836 out of 878 subjects with at least two longitudinal samples, median time between samples of 180 days; Table S4). When checking the 92 and 18 individuals *Blastocystis*-positive in at least two and three time points, respectively, we found the same ST maintained in 97.8% and 94.4% of the cases (Table S4), indicating that *Blastocystis*, when present, is likely a stable colonizer of the human gut microbiome.

We next assessed the potential role of host genetics on *Blastocystis* colonization by performing subgroup analyses among monozygotic (identical, MZ) and dizygotic (fraternal, DZ) twins from three distinct datasets[7,47,48] (STAR Methods). Among 1,424 twin participants, *Blastocystis* carriage rates largely mirrored that of other non-twin participants from their respective studies (Table S4). We then calculated pairwise concordance rates (i.e., the probability that both members of a given twin pair were positive if one was), finding comparable agreement between MZ (19.8%, [95% confidence interval (CI), 13.0%–26.7%]) and DZ (21.6%, [95% CI, 13.6%–29.6%]) twins, suggesting that *Blastocystis* carriage may be determined mostly by non-genetic (i.e., environmental) factors. This conclusion was further supported by the tendency for concordant pairs of *Blastocystis*-positive siblings to harbor different STs (in 65.4% of the cases for MZ and 72.7% for DZ) and the lack of a trend in concordance from randomly paired participants vs. DZ vs. MZ twins (i.e., from non-related to closely related to identical germ lines; Table S4). This lack of association is noteworthy when considering the possibility that siblings and twins—even ones that no longer cohabitate—are likely to lead more similar lifestyles than completely unrelated persons.

Blastocystis carriage is linked to the consumption of healthier diets

Since diet can substantially affect the gut microbiome[49,50] and is influenced by host geography and lifestyle,[51–53] we sought to determine whether *Blastocystis* presence could be related to different dietary regimes and quality. We considered healthy adults from seven datasets (four PREDICT cohorts, two datasets from Italy, and the MBS/MLVS datasets from the US; STAR Methods) and stratified them into well-established diet regimes, such as omnivores, vegetarians, and vegans, showing a higher median *Blastocystis* prevalence for vegetarians (Figure 2A; Table S3).

Moreover, *Blastocystis* prevalence showed a significant dose-dependent coupling to a more prudent-style diet as assessed by FFQs (STAR Methods) in the PREDICT studies. In these US and UK-based cohorts, increased adherence to a healthful plant-based dietary index (hPDI) was associated with increased carriage of *Blastocystis*, despite country-specific differences in overall prevalence when comparing the UK-based PREDICT 1 and PREDICT 3 UK22A studies vs. the US-based PREDICT 2 and PREDICT 3 US22A studies (all Mann-Kendall test p values <0.05 but PREDICT 3 UK22A; Figure 2B; STAR Methods). Moreover, median hPDI values in each US division showed a significant correlation with *Blastocystis* prevalence (Pearson's r = 0.73, p value = 0.02; Figure S2A). The hPDI is an established dietary pattern emphasizing the intake of whole grains, fruits, and vegetables that has also been linked epidemiologically to favorable cardiometabolic outcomes.[54] The same pattern of increasing *Blastocystis* prevalence was replicated for the hPDI estimated from free-living logged dietary data (Figure 2B; STAR Methods), highlighting that the trend is robust both to the dietary assessment tool used (FFQ vs. digital logging) and the time scale of dietary intake (long- vs. short-term). The opposite trend in *Blastocystis* prevalence was observed when participants adhered to an "unhealthful" plant-based diet index (uPDI), which—while also avoiding animal proteins—is higher in refined carbohydrates and sugars (Figure S2B).

We next examined single food components and identified consistent and dose-dependent relationships with *Blastocystis* presence (Mann-Whitney U test, q value < 0.1; Figure 2C). Generally, individuals who consumed higher quantities of unprocessed plant-based foods, such as avocados, dried fruits, nuts, seeds, legumes, and cruciferous vegetables, were more likely to be *Blastocystis*-positive compared with individuals with lower intake of such foods (Figure 2C). Conversely, *Blastocystis*-negative individuals tended to more frequently consume less health-promoting plant- and animal-based foods with higher levels of food processing according to the NOVA categorization,[55] such as fried potatoes, candies, and artificially sweetened beverages (Figure 2C). Despite population-specific differences, most of these trends were directionally consistent in all PREDICT cohorts. Nevertheless, some foods were not uniformly concordant, which could reflect chance findings, differences in inter-individual responses, country-specific preparation standards or consumption patterns, or imprecise categorization of certain foods. For example, higher consumption of fish was strongly associated with higher *Blastocystis* carriage in both PREDICT 3 UK22A and US22A cohorts, but no clear trends were observed in the smaller PREDICT 1 or PREDICT 2 cohorts (Figure 2C). This discrepancy may be due to the difficulty in disentangling wild-caught from farmed fish, as well as its ultra-processed vs. unprocessed forms. Similarly, cereal intake may not adequately capture its many healthful and less-healthy versions (Figure 2C). Overall, *Blastocystis* presence was associated with higher diet quality and the intake of less-processed and fiber-enriched foods (Figure S2C).

Gut *Blastocystis* is associated with more favorable short-term cardiometabolic biomarkers

We then evaluated whether *Blastocystis* presence was associated with biomarkers of cardiometabolic health. To expand on our prior findings linking gut microbial composition to fasting and postprandial responses,[7,18,19] we tested a set of biomarkers previously reported to be representative of host health status,[7] and, to limit the effect of potentially extreme values in diseased

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

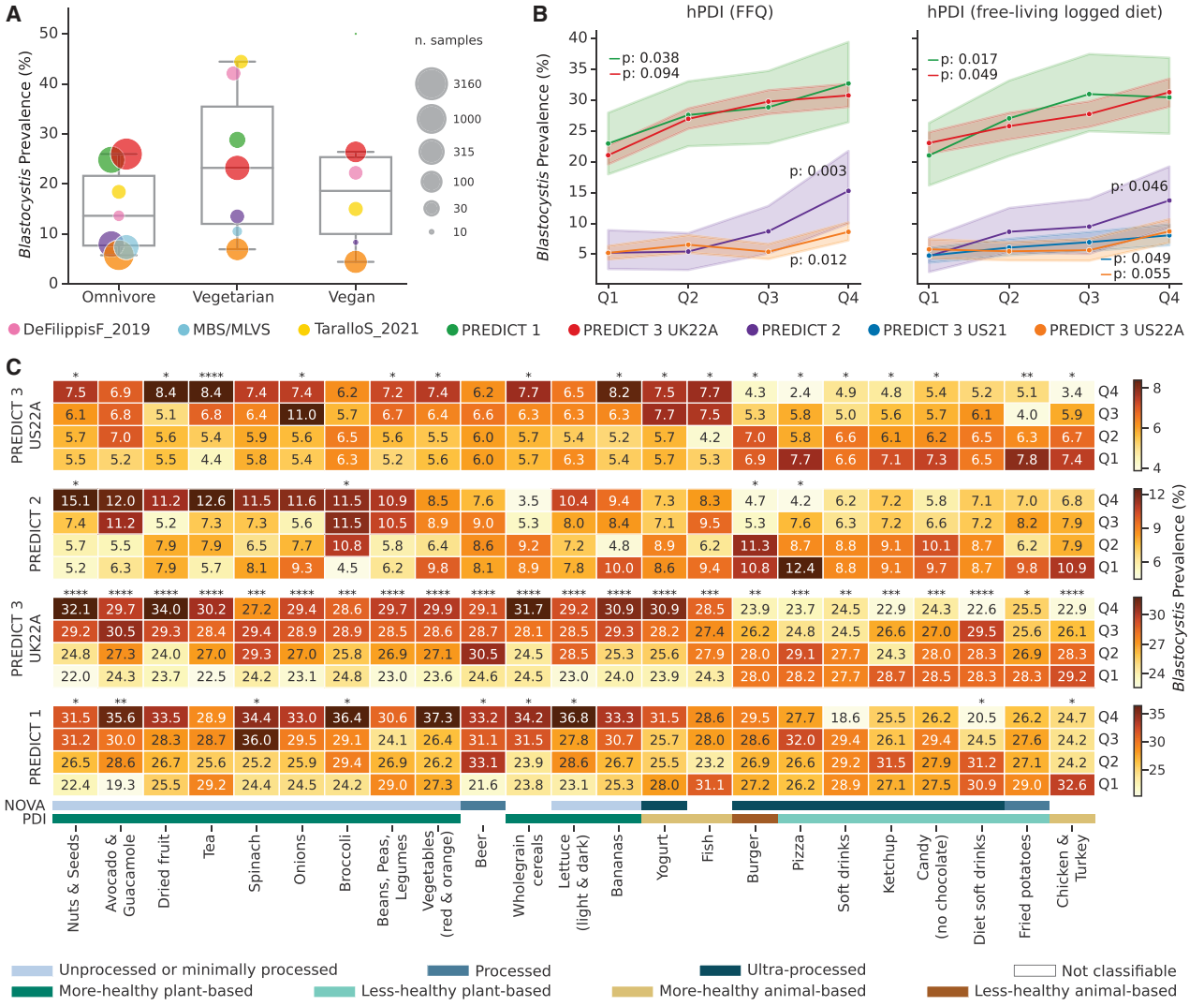


Figure 2. Higher *Blastocystis* prevalence is consistently associated with healthier diets rich in plant-based foods

(A) Individuals stratified into diet regimes (i.e., omnivores, vegetarians, and vegans; STAR Methods) showed higher median *Blastocystis* prevalences for vegetarians. The boxplots show the first and third quartiles (boxes) and the median (middle line); whiskers extend to 1.5× the interquartile range (IQR). The size of the circles reflects the number of samples ($\log_{10}$ scale, Table S3).

(B) Greater adherence to the healthy plant-based dietary index (hPDI) was significantly associated with higher *Blastocystis* prevalence across all five PREDICT cohorts. Participants were divided into quartiles according to hPDI, available from food frequency questionnaires (FFQs) and short-term dietary information (free-living logged diet). *Blastocystis* prevalence was computed in each hPDI quartile, and shadings represent the 95% confidence interval. *p* values were calculated using the Mann-Kendall test for monotonic trends (STAR Methods).

(C) The heatmap shows trends between *Blastocystis* prevalences with single food items intake grouped into quartiles. Foods were included in the heatmap if significantly associated with *Blastocystis* in at least two PREDICT cohorts (Mann-Whitney U test, *q* value < 0.1; STAR Methods). *Blastocystis*-positive individuals tended to consume less-processed and healthier plant-based foods. More- or less-healthy, animal- or plant-based foods and the food processing levels were assigned using established criteria (STAR Methods). White color represents food items too broad to classify. Foods were sorted according to their difference in mean consumption (servings/day) among *Blastocystis*-positive and *Blastocystis*-negative participants, considering all PREDICT cohorts (STAR Methods).

See also Figure S2 and Table S3.

participants, we considered only healthy adults (Figure S2E; STAR Methods).

Among available clinical risk measures, no clear relationship was found between *Blastocystis* presence and the atherosclerotic cardiovascular disease 10-year risk score (ASCVD; STAR Methods).[56] However, *Blastocystis* was associated with lower diastolic blood pressure (Figure 3A), while systolic pressure showed lower levels but no significant differences (Figure S2D). Glucose concentration from clinical measurements (STAR Methods) showed no clear relationships, while C-peptide levels were significantly lower in *Blastocystis*-positive subjects (PREDICT 1 *q* value = 3.76e−4 and PREDICT 2 *q* value = 9.96e−3; Figure 3B). Lower

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

CellPress
OPEN ACCESS

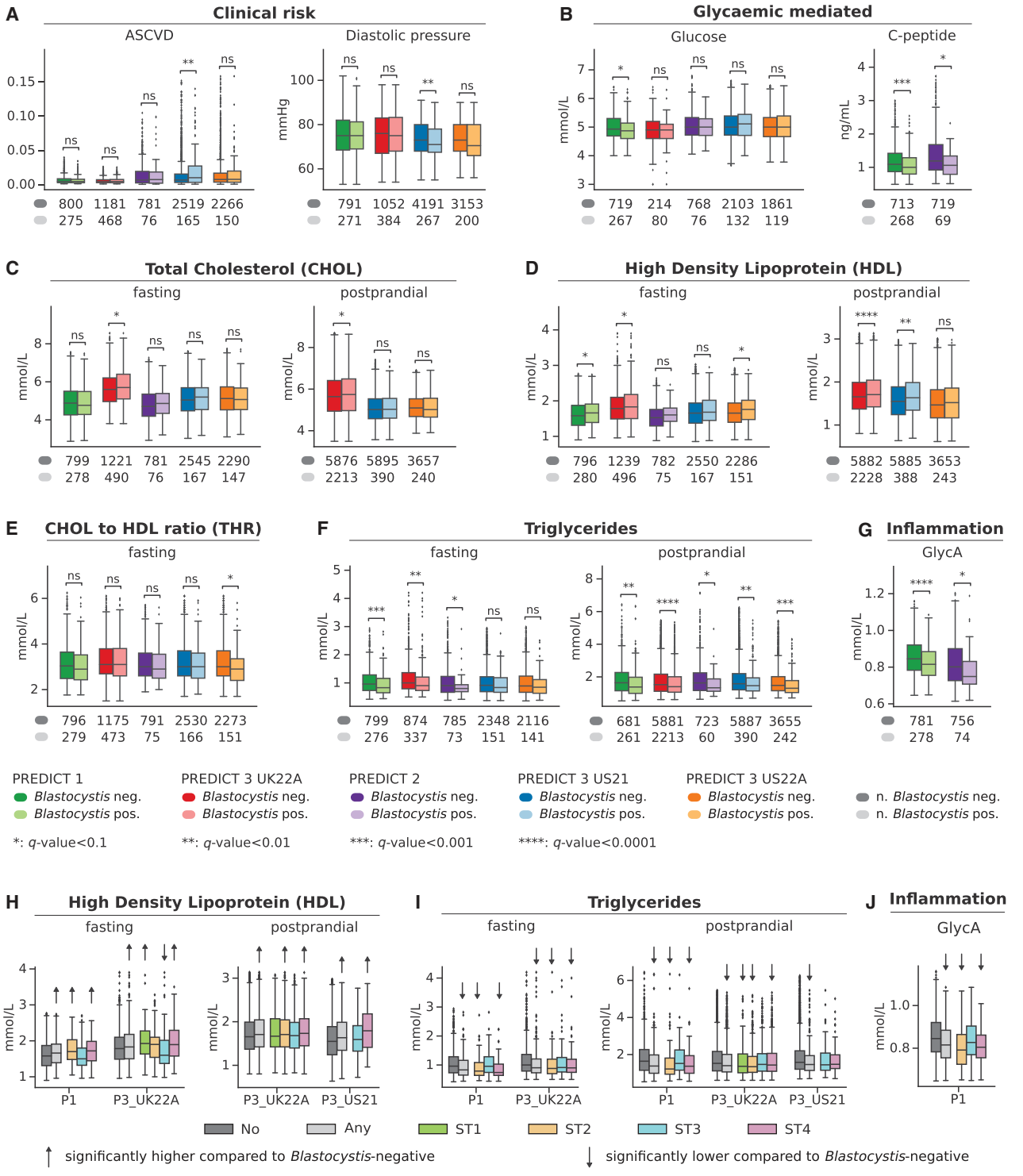


Figure 3. Gut *Blastocystis* was linked to more favorable short-term cardiometabolic profiles

(A) Clinical risk measures of atherosclerotic cardiovascular disease, specifically the ASCVD 10-year risk score[66] and diastolic blood pressure; although not always statistically significant (Mann-Whitney U test), blood pressure was lower in *Blastocystis* carriers.

(B) Fasting glucose levels showed no clear trend with *Blastocystis*, while C-peptide, a marker of endogenous insulin production, was decreased in *Blastocystis* carriers, possibly indicating improved glucose tolerance in normoglycemic populations.

(C) Total cholesterol (CHOL) levels showed no clear trend with *Blastocysis* carriage.

(legend continued on next page)

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

C-peptide levels among *Blastocystis* carriers could be interpreted as a surrogate for improved glucose tolerance or higher insulin sensitivity, previously shown to be protective against cardiovascular mortality in non-diabetic individuals.[57,58] We also found consistent relationships between *Blastocystis* carriage and more favorable lipemic profiles (Figures 3C–3F). Participants harboring *Blastocystis* generally had significantly higher fasting and postprandial high-density lipoprotein (HDL; Figure 3D) and lower triglyceride levels (Figure 3F). Both low HDL and hypertriglyceridemia are established risk factors for subsequent cardiovascular disease risk.[59,60] GlycA, an inflammatory biomarker linked to risk of T2D[61–63] and cardiovascular disease,[64,65] was uniformly depressed among *Blastocystis* carriers in both PREDICT 1 and PREDICT 2 (q value = 2.97e−5 and 9.96e−3, respectively; Figure 3G). Similar associations were observed when assessing *Blastocystis* relative abundance (Figure S9G; STAR Methods).

Among healthy adults of the PREDICT cohorts, few *Blastocystis*-positive participants harbored more than one ST (n = 130/3,555, or 3.66%; Figure S1B). Multi-ST carriage was not significantly associated with BMI or cardiometabolic markers, beyond the relationships observed for mono-colonized participants (Figure S5D). Because prior studies hinted at possible subclade-level heterogeneity linking *Blastocystis* and host outcomes,[67–69] we stratified the above findings by ST, which were generally concordant with the non-stratified results in both magnitude and direction of effect (Figures 3H–3J and S3A). To rule out the possibility that *Blastocystis* carriage is a mere indicator of diet's impact on cardiometabolic health, we divided participants into quartiles based on their hPDI values, from less (Q1) to more (Q4) healthful diet. As expected, individuals in the upper quartile (Q4) had generally better cardiometabolic profiles. Nonetheless, within each quartile of dietary quality, *Blastocystis*-positive individuals showed lower triglycerides, C-peptide, and GlycA values, and higher HDL values than *Blastocystis*-negative individuals (Figures S4A–S4F). Moreover, linear regression models showed that *Blastocystis* was a significant predictor of cardiometabolic marker values independently from hPDI (Figure S4G). These results indicate that *Blastocystis* presence may have an additive positive effect on improved cardiometabolic health.

Blastocystis is linked to decreased body mass and is enriched in healthy controls

Given the association between *Blastocystis* with a more healthful diet and more favorable cardiometabolic profiles, we explored potential links with body adiposity. We previously reported that individuals harboring *Blastocystis* showed reduced visceral fat[7] (Figures S5A and S5B). Visceral fat is an important appraisal of fat distribution, an independent risk factor for cardiometabolic disease,[70] and a distinguishing feature between patients with obesity and normal vs. high metabolic risk profiles.[71] However, its measurement requires complex techniques,[72] which are rarely available for large populations. Thus, using BMI, a more widely available proxy for body adiposity, we categorized individuals from all PREDICT cohorts into three established BMI groups[73] and computed *Blastocystis* prevalence in each (healthy-weight, 18.5 $\leq$ BMI < 25 kg/m^2; "overweight," 25 $\leq$ BMI < 30 kg/m^2; "obese," BMI $\geq$ 30 kg/m^2, excluding individuals with BMI < 18.5 kg/m^2 due to small sample size; Figure S5C; STAR Methods). All PREDICT cohorts showed higher *Blastocystis* prevalence for healthy-weight persons compared with those with overweight or obesity, showing a monotonic decrease in *Blastocystis* prevalence with increasing BMI values (Figure 4A). Similar trends were observed using *Blastocystis* relative abundance, with lower BMI values being associated with higher abundance values, although not statistically significant, likely due to sample size, as the analysis focused only on *Blastocystis*-positive individuals (Figures 4B and S5F; STAR Methods). Moreover, we found that even among individuals with comparable diet quality, *Blastocystis*-positive individuals tended to have lower BMI values, suggesting that, besides diet, *Blastocystis* carriage may have an independent impact on body adiposity (Figures S4H–S4K). Other lifestyle indices, including sleep, have been linked with diet and weight,[74,75] and we observed a positive association between *Blastocystis* presence and increased hours of sleep per night (Figure S5E).

We leveraged the cMD3 package[25] and identified 36,065 healthy adults from 24 different studies spanning 12 countries, with available BMI information and at least 10 *Blastocystis*-positive individuals within each study (Table S5). Using random-effects meta-analysis, we confirmed that *Blastocystis* was consistently and significantly associated with lower BMI values (1-point increase in BMI was linked to a 6% decrease in odds of *Blastocystis* carriage, random effect size = −0.06; 95% CI = −0.07 to −0.05, p value < 1e−16; Figure 4C). Then, to investigate potential links between chronic diseases and differences in *Blastocystis* prevalence,[76] we identified 13,548 cases (9,091 from the PREDICT cohorts) and 31,358 healthy controls from 19 studies spanning eight disease categories (Figure 4D; STAR Methods). Regardless of the disease type, *Blastocystis* was robustly enriched among the healthy controls in all studies but two, with a 25% decrease in

(D) *Blastocystis*-positive individuals exhibited higher fasting and postprandial HDL levels.
(E) *Blastocystis* presence was associated with lower levels of CHOL to HDL ratio (THR).
(F) *Blastocystis*-positive individuals showed lower fasting and postprandial triglycerides levels, which along with higher HDL levels is a favorable surrogate of long-term cardiovascular risk.
(G) GlycA, a marker of systemic inflammation, was reduced in participants harboring *Blastocystis*. Additional markers are reported in Figure S2D.
(H) We identified pairs of ST markers with at least 50 participants positive for a given ST and with available biomarker values. Arrows denote significant tests between a given ST and *Blastocystis*-negative participants (Mann-Whitney U test, q value < 0.1). The direction of the arrows indicates whether the median marker values are higher (upward) or lower (downward) in the ST-positive compared with *Blastocystis*-negative individuals. Concordantly with the non-stratified results, individuals carrying ST1, 2, or 4 showed higher HDL levels compared to *Blastocystis*-negative ones.
(I) Participants positive for ST1, 2, or 4 exhibited siginficatly lower fasting and postprandial triglycerides levels than *Blastocystis*-negative individuals.
(J) ST2- and ST4-positive participants had significantly significantly lower GlycA values.
Full panel of markers in Figure S3A. Boxplots show the IQR (box) and the median (middle line); whiskers extend to 1.5× the IQR.
See also Figures S2–S4.

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

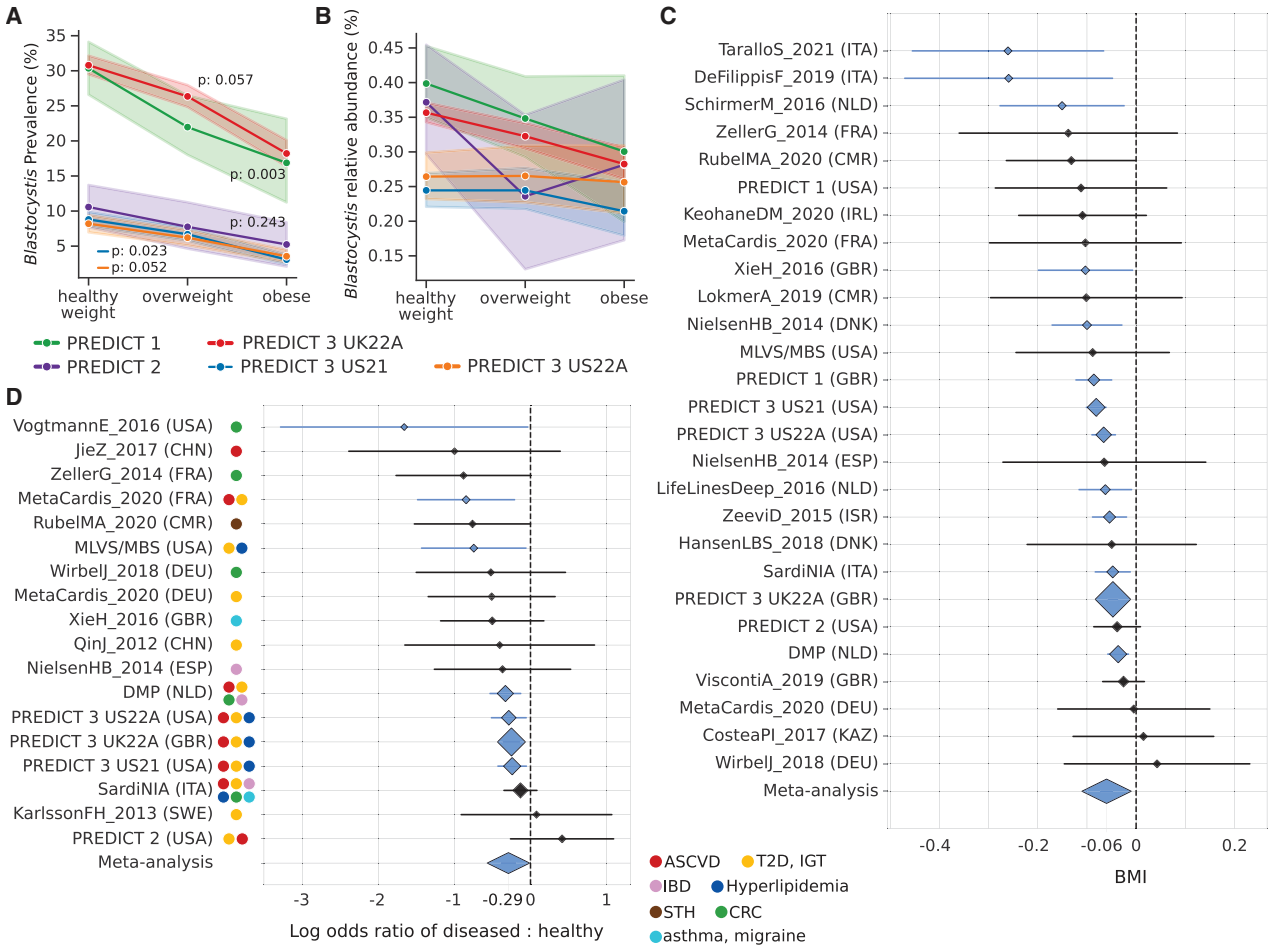


Figure 4. Meta-analysis on *Blastocystis* presence showed consistent and reproducible associations with markers of body adiposity and dysbiotic gut disorders worldwide

(A) In all PREDICT cohorts, individuals with "healthy-weight" BMI had a significantly higher *Blastocystis* prevalence compared with individuals with overweight and obesity. Reported *p* values from the rank-based Mann-Kendall test using BMI deciles (Figure S5C).

(B) Similarly, a monotonic trend toward decreased *Blastocystis* abundance among *Blastocystis*-positive participants living with overweight or obesity (Figure S57).

(C) Meta-analysis of 36,065 healthy adults from 12 countries revealed a consistent and reproducible association of lower BMI with *Blastocystis*.

(D) Meta-analysis of 13,548 individuals with various chronic diseases and their 31,358 controls revealed an inverse association between *Blastocystis* presence and disorders linked to gut microbial disturbances (Figure S5H). ASCVD, atherosclerotic cardiovascular disease; T2D, type 2 diabetes; IGT, impaired glucose tolerance; IBDs, inflammatory bowel diseases; STHs, soil-transmitted helminths; CRC, colorectal cancer (STAR Methods). Single-study *p* values were computed as coefficient extremeness of the *t*-statistics; *p* values for meta-analysis coefficients were computed as *Z* scores extreme probability, both under the null hypothesis that the average coefficient is zero. Markers' size reflect the inverse of the variance absorbed by the study without accounting for between-study heterogeneity, blue markers indicate significant *p* values.

See also Figure S5 and Table S5.

the odds of harboring *Blastocystis* among unhealthy individuals (random effect size = −0.29, 95% CI = −0.38 to −0.20, *p* value = 1.45e−9). This finding was further supported when using *Blastocystis* relative abundance as the primary exposure (random effect size = −0.13, 95% CI = −0.20 to −0.05, *p* value = 1.28e−3; Figure S5H). Overall, despite size heterogeneity among the studies of the meta-analyses, the directions of the summary estimates were consistent (Figures 4C, 4D, and S5H) and support the role of *Blastocystis* as a common member of a healthier gut microbiome configuration.

Transkingdom ecology implicates the whole microbiome in *Blastocystis* carriage

To assess potential cross-kingdom ecological interactions between *Blastocystis* and other members of the gut microbiome, we taxonomically profiled bacteria and archaea using MetaPhlAn 4 (STAR Methods). We found that alpha diversity was significantly higher in *Blastocystis*-positive participants across all datasets (Figure S6), as previously observed.[14]

To investigate whether specific microbiome configurations were linked with *Blastocystis* carriage, we trained random forest

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

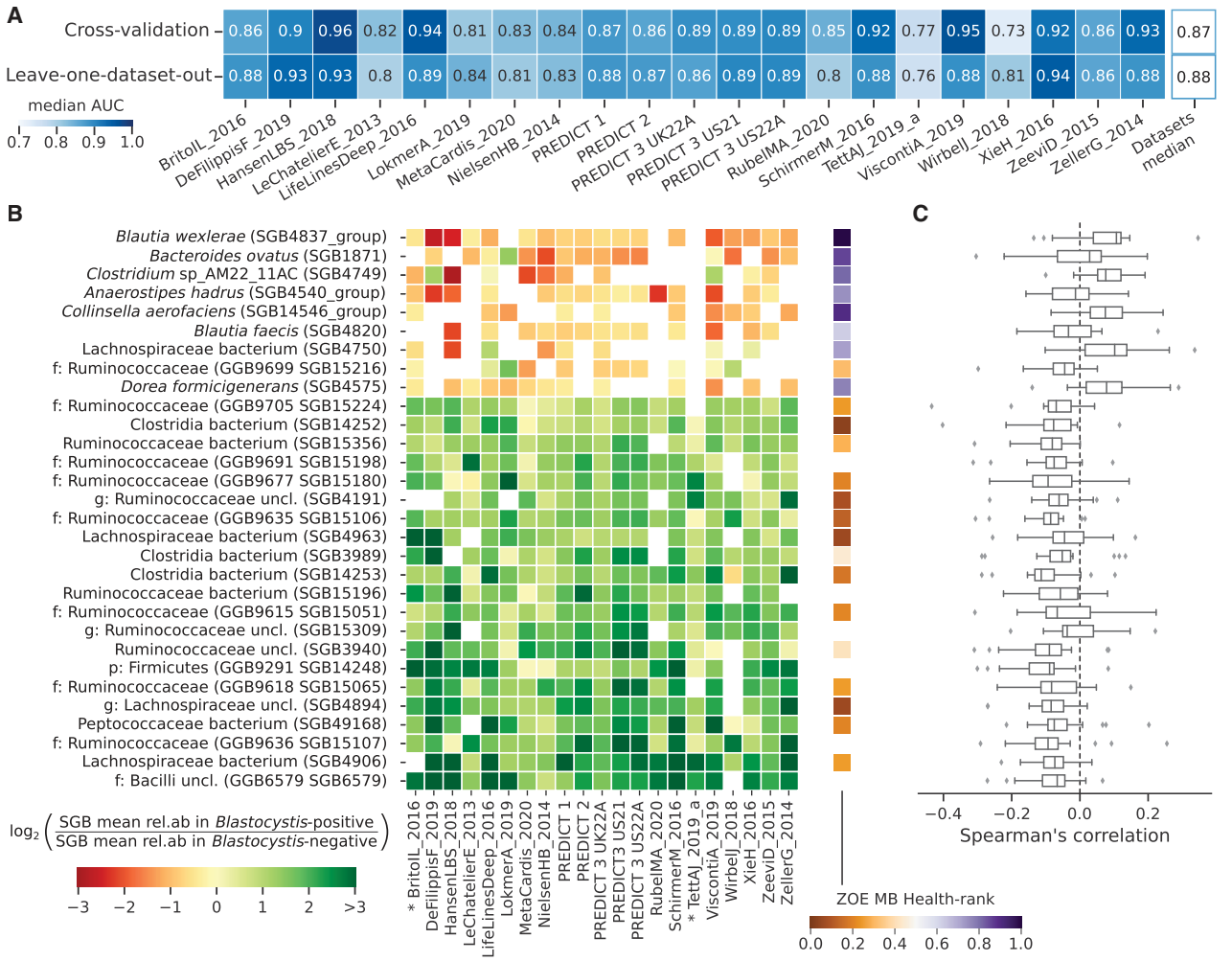


Figure 5. *Blastocystis* presence impacts gut microbiome composition and is associated with species linked with BMI and cardiometabolic health

(A) Random forest-based machine learning models trained on bacterial and archaeal species relative abundance accurately predict *Blastocystis* presence, both with a cross-validation (CV) and a leave-one-dataset-out (LODO) framework (STAR Methods).

(B) Log_2 fold change of species mean relative abundance in *Blastocystis*-positive and negative individuals for the top-ranked SGBs from CV models. The heatmap reports the top 30 species whose relative abundances were significantly higher in *Blastocystis*-positive individuals (green colors) and those significantly lower in *Blastocystis*-negative individuals (red colors). White cells indicate no statistically significant difference (Mann-Whitney U q value > 0.2; STAR Methods). Complete list of the 100 top-ranked SGBs in Table S6. Right-side column heatmap indicates the species' "ZOE MB Health-rank,"[7] with values from zero to one representing species associated with a lower and higher cardiometabolic risk, respectively (STAR Methods). Datasets without BMI information are denoted with an asterisk.

(C) Spearman's correlation between the SGB relative abundance and participants BMI in each dataset. Species enriched in *Blastocystis*-positive individuals and associated with lower cardiometabolic risk are systematically negatively correlated with BMI. Boxplots show the IQR (box) and the median (middle line); whiskers extend to 1.5× the IQR.

See also Figure S6 and Table S6.

machine learning (ML) classification models on relative abundance values of bacterial and archaeal species to predict *Blastocystis* presence (STAR Methods). Extending what was previously suggested,[9] we demonstrated high prediction accuracy using both a cross-validation (CV, median area under the receiver operating characteristic curve AUC = 0.87) and a leave-one-dataset-out (LODO, median AUC = 0.88) framework,[77,78] indicating a strong and reproducible association between the bacterial and archaeal population in the microbiome and *Blastocystis* (Figure 5A).

From our CV ML models, we identified the top-ranked species by feature importance and assessed their link with *Blastocystis* (Figure 5B; complete species' list in Table S6). *Blastocystis*-positive individuals were strongly associated with higher relative abundances of unknown species-level genome bins (uSGBs) assigned only at the family taxonomic level and mostly to Ruminococcaceae (Figure 5B). Conversely, *Blastocystis*-negative samples were enriched for two *Blautia* species and for *Bacteroides ovatus*, which was also previously reported to be associated

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

with *Blastocystis* absence.[9] In addition, species consistently and significantly more abundant in *Blastocystis*-positive individuals were linked to a more favorable cardiometabolic profile and correlated negatively with BMI, and vice versa for species more abundant in *Blastocystis*-negative individuals (Figures 5B and 5C; STAR Methods). These findings suggest strong ecological links between microbiome configuration and the presence of *Blastocystis* and that both components are associated with healthier diets and more favorable cardiometabolic profiles.

Blastocystis carriage and abundance increase following a PDP

Finally, we enrolled 1,124 healthy participants into a 6-month personalized dietary program (PDP) bookended by pre- and post-intervention stool sampling (PREDICT 3 UK 23A RT cohort, n = 2,248 samples; STAR Methods). During the study, participants were subject to individualized dietary recommendations to inform their *ad libitum* diet (as opposed to a more controlled feeding trial). Adherence to recommendations and their change vs. usual habits were ascertained through pre- and post-intervention FFQs (STAR Methods). In general, participants significantly improved their diet quality (i.e., increases in hPDI and healthy eating index [HEI], Wilcoxon p values of 1.02e−134 and 1.19e−27, respectively) and lost weight (mean 0.5 kg/month; Figures 6A and S7A).

To corroborate the links between healthy dietary intake and *Blastocystis* carriage identified from our observational (and cross-sectional) cohorts, we analyzed *Blastocystis* prevalence and relative abundance before and after the PDP intervention. Overall, *Blastocystis* prevalence significantly increased from 31.32% to 33.81% (McNemar p value = 0.003, Table S7). Among the 1,124 participants, 715 were *Blastocystis*-negative at both time points (63.61%, "No-No"), while 323 remained *Blastocystis*-positive (28.74%, "Yes-Yes"), with only three individuals switching to a different ST. Participants that became negative were 29 (2.58%, "Yes-No"), while almost double (57 individuals) acquired *Blastocystis* during the intervention (5.07%, "No-Yes"). Consistent with these differences, we observed a significant increase in *Blastocystis* relative abundance among those who were persistently positive ("Yes-Yes") throughout the study period (Wilcoxon p value = 2.38e−8; Figures 6B and S7B).

To explore specific dietary factors that may underlie this association, we noted that our aforementioned cross-sectional analyses showed a link between increased fiber intake and higher *Blastocystis* prevalence (Figure S2C). Similarly, we noticed a statistically significant increase in fiber and fat consumption (Wilcoxon p value < 0.01) among the 57 individuals who acquired *Blastocystis*, but not for proteins and carbohydrates (Wilcoxon p value > 0.05). Moreover, the observed effect size of fiber intake among individuals who acquired *Blastocystis* was greater than those who lost or did not gain *Blastocystis* (*Cohen's d* effect sizes of 0.38 for "No-Yes," 0.25 for "Yes-No," and 0.28 for "No-No"; Figure S7C), hinting at a potential role for increased fiber consumption in *Blastocystis* colonization and acquisition. Moreover, among individuals with *Blastocystis* either stably absent or present ("No-No" and "Yes-Yes") and with a comparable improvement in diet quality, *Blastocystis*-positive individuals generally had a greater reduction in BMI, suggesting a potential additive effect of gut *Blastocystis* carriage on weight loss (Figure S7D). Finally, we evaluated whether the microbiome species associated with *Blastocystis* in the cross-sectional analysis (Figure 5B) were also affected by its acquisition or loss during the PDP. We observed that the prevalences of species enriched in *Blastocystis*-positive subjects increased among those who acquired *Blastocystis* during the intervention ("No-Yes"; Figure S7E) and decreased in those who lost it ("Yes-No"), while remained constant among other individuals ("No-No" and "Yes-Yes"; Figure S7E).

DISCUSSION

We comprehensively profiled *Blastocystis* and its eight human-associated STs from 56,989 human microbiome samples, expanding prior studies by an order of magnitude. We found *Blastocystis* to be present worldwide in the gut of healthy adults with variable prevalence according to geography, lifestyle, and diet and with no apparent coupling to germline determinants. Interestingly, ST4 was prevalent in Europe and North America, only rarely found in Asia, and never in South America and Africa, suggesting a potential link with a more Westernized lifestyle. *Blastocystis* was hardly ever found in newborns, suggesting that it is likely acquired later in life and not vertically transmitted. *Blastocystis* ST1, ST2, ST3, and ST4 were the most commonly detected in human microbiome samples.[79] Other STs were less frequently found in humans, and although previously reported in wild and captive animals,[13] which could serve as potential reservoirs of zoonotic transmission, we only detected ST7 in one of the 4,590 animal metagenomes analyzed.[32,80] *Blastocystis* appears to be the most abundant eukaryotic microbe in the human gut and could be responsible for a substantial amount of reads frequently overlooked by current computational workflows focusing on prokaryotic or fungal organisms. Other human-colonizing protozoans such as *Giardia*, *Entamoeba*, and *Dientamoeba*[81–83] are less commonly detected, and metagenomics coupled with computational pipelines similar to the one we developed for *Blastocystis* can be promising for their future investigation.

We leveraged the ZOE PREDICT cohorts, which are part of a research program exploring the interplay between individual-level factors, diet, and gut ecology in cardiometabolic health. The analysis of long- and short-term diet information for 25,548 healthy PREDICT participants allowed us to link the presence of *Blastocystis* to a more prudent diet, corroborated at several levels, including associations with individual food items and overall dietary patterns favoring more healthful plant-based and minimally processed foods. Furthermore, the analysis of 1,124 individuals longitudinally sampled before and after a personalized dietary intervention showed that improvements in diet quality were linked with increases in *Blastocystis* prevalence and abundance, supporting our population-scale cross-sectional results. We also found more favorable glycemic and lipid profiles in *Blastocystis*-positive individuals, suggesting a potential positive impact on cardiometabolic health beyond the effect of a healthier diet alone. *Blastocystis* was previously associated with visceral fat,[7] a finding we expanded upon in this analysis using several independent cohorts to demonstrate a

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

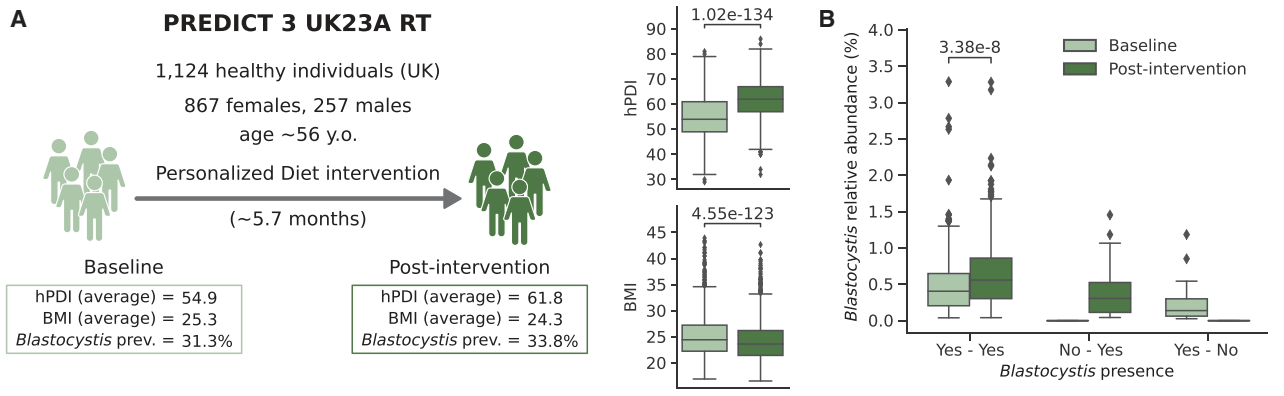


Figure 6. Improvements in dietary quality were prospectively associated with weight loss and increased *Blastocystis* prevalence and abundance

(A) The PREDICT 3 UK23A RT cohort enrolled 1,124 healthy individuals from whom 2,248 pre- and post-intervention gut metagenomic samples were collected. Participants followed personalized food-level recommendations during a 6-month personalized dietary program (PDP). After the intervention, we observed a significant (Wilcoxon test). improvement in dietary quality (measured by hPDI), a decrease in BMI, and an increase in overall *Blastocystis* prevalence.

(B) *Blastocystis* relative abundance was significantly (Wilcoxon test) higher post-intervention compared with baseline. Boxplots show the IQR (box) and the median (middle line); whiskers extend to 1.5× the IQR; *p* values from Wilcoxon tests.

See also Figure S7 and Table S7.

significant and reproducible signal between the presence and abundance of *Blastocystis* and lower BMI values. We also showed that microbiome composition accurately predicted *Blastocystis* presence, hinting at potential cross-kingdom interactions and an ecological role of (or effect on) *Blastocystis*, whose exploration could help unravel the diet-*Blastocystis*-health axis but is beyond the scope of the current work. Our metagenomic profiling approach, exploiting reference genomes and metagenome-assembled genomes clustered into SGBs[84,85] allowed us to detect and quantify both known and unknown SGBs, with the latter being among the most predictive microbes of *Blastocystis* carriage. Finally, a meta-analysis showed that *Blastocystis*-positive participants were more likely to be healthy controls in various case-control studies of diseases linked to altered gut microbial configurations (e.g., IBD, T2D, and CRC). Collectively, these results support the hypothesis that *Blastocystis* may not have detrimental host effects but, rather, is a favorable constituent of the human gut microbiome. This paradigmatic shift must be considered in future studies with analyses targeting the eukaryotic component of the human microbiome and *Blastocystis* in particular.

Microbiome-targeted interventions such as fecal microbiome transplantation (FMT) often consider *Blastocystis* carriage as an exclusion criterion for would-be donors.[86] However, our results linking *Blastocystis* to host health support its non-pathogenic, if not favorable, role. Relatedly, prior reports indicate that FMT from *Blastocystis*-positive donors was not associated with a greater risk for adverse events and yielded comparably high success rates.[87] While it remains to be clinically tested whether including *Blastocystis*-positive individuals in FMT trials could expand the candidate donor pool without risking treatment efficacy[83,88–92] or harming recipients, our data support the rationale for performing such tests. Furthermore, direct artificial administration of *Blastocystis* has never been attempted and presents technical limitations (e.g., axenizing *Blastocystis* cultures remains a significant challenge[93]) as well as potential clinical risks, especially in immunocompromised hosts, but pilot studies in animal models could start addressing the causal link with host health. Such a link could also be largely indirect as *Blastocystis* might impact the bacterial component of the microbiome,[94] further strengthening the potential of *Blastocystis*-based administration strategies.

Overall, our results offer an improved understanding of the complex interplay between the human gut microbiome, diet, and cardiometabolic diseases[40] involving the currently overlooked *Blastocystis* species. In the future, precision nutrition interventions aimed at modifying gut ecology may constitute a viable disease-prevention strategy,[95] and our work suggests that underappreciated non-bacterial taxa can have considerable roles in maintaining and promoting human health.

Limitations of the study

Several genetic and genomic aspects of *Blastocystis* remain to be elucidated. Comparative genetic analyses may provide a more nuanced discernment of the roles of different STs and a better characterization of relevant genes and underlying mechanisms potentially affecting host health. Genetic and functional analyses coupled with longitudinal data on disease incidence may help our understanding of whether *Blastocystis* presence has an impact on the development of cardiometabolic diseases. Our computational workflow relies on a set of nine available *Blastocystis* genomes representing eight STs previously found in the human gut. Prioritizing the discovery and validation of high-quality genomes for rarer and non-human STs may improve detection sensitivity and allow profiling of currently unreported STs. The profiling pipeline used a validated conservative 10% breadth of coverage threshold to determine *Blastocystis* presence, and future work can further optimize the trade-off between false

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

positives and false negatives. Furthermore, given evident ethnogeographical patterns of *Blastocystis* presence, as well as associations with the consumption of different foods, broader, more inclusive data-sharing practices and additional metagenomic studies, including metadata on dietary habits and also targeting underrepresented countries, will be valuable to support our results. Moreover, specific experimental works are needed to elucidate possible transmission or acquisition routes and to unravel the relative contributions of various environmental and host factors to gut *Blastocystis* colonization. Finally, although technical issues in establishing axenic *Blastocystis* cultures need to be addressed, interventional protocols could be developed to directly assess causality, and our work provides extensive data for proposing such studies with ethical support.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - The ZOE PREDICT Study cohorts
  - The PREDICT 3 UK 23A RT cohort
  - The MBS and MLVS cohort
  - DeFilippisF_2019, MeslierV_2020, TarralloS_2021
  - The SardiNIA cohort
  - The Dutch Microbiome Project cohort
  - Publicly available human metagenomic samples
  - Ancient human gut metagenomic samples
  - Publicly available animal metagenomic samples
- METHOD DETAILS
  - Sample collection, DNA extraction, sequencing, and quality control
  - Longitudinal analysis
  - Twins subgroup analysis
  - *Blastocystis* detection and ST profiling
  - Sequencing depth and rarefaction
  - Westernized vs. non-Westernized populations
  - Dietary data collection and categorization
  - Metabolic marker collection
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Prediction of *Blastocystis* carriage
  - Statistical analyses
  - Meta-analyses
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.06.018.

ACKNOWLEDGMENTS

We thank the participants of the PREDICT program. This work was supported by Zoe Ltd. and TwinsUK, which are funded by the Wellcome Trust, Medical Research Council, Versus Arthritis, European Union Horizon 2020, Chronic Disease Research Foundation, the National Institute for Health Research Clinical Research Network, and the Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. It was partially supported by the American Gastroenterological Association Research Foundation's Research Scholars Award to L.H.N., the Crohn's and Colitis Foundation Career Development Award to L.H.N., and NIH/NIDDK K23DK125838 to L.H.N. L.H.N. is the Massachusetts General Hospital/Chen Institute Transformative Scholar in Medicine. The study was 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 program (ONCOBIOME-825410 project, MASTER-818368 project, and IHMCSA-964590) to N.S., by the European Union's Horizon Europe program (DOMINO-101060218), by the European Union NextGenerationEU (INEST) to N.S., by the National Cancer Institute of the National Institutes of Health (1U01CA230551) to N.S. and R35 CA253178 to A.T.C., and by the Premio Internazionale Lombardia e Ricerca 2019 to N.S. The MBS and MLVS were supported by NIH U01CA176726 and U01CA167552. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The Lifelines Biobank initiative has been made possible by the Dutch Ministry of Health, Welfare, and Sport; the Dutch Ministry of Economic Affairs; the University Medical Center Groningen; the University of Groningen; and the Northern Provinces of the Netherlands. The authors acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants. J.F. is supported by the Dutch Heart Foundation IN-CONTROL (CVON2018-27), the ERC-CoG (No. 101001678), NWO-VICI grant VI.C.202.022, the AMMODO Science Award 2023 for Biomedical Sciences from Stichting Ammodo, and the Netherlands Organ-on-Chip Initiative, an NWO Gravitation project (024.003.001) funded by the Ministry of Education, Culture, and Science of The Netherlands. A.Z. is supported by the Dutch Heart Foundation IN-CONTROL (CVON2018-27), the ERC-StG (No. 715772), NWO-VIDI grant 016.178.056, and the NWO Gravitation grant Exposome-NL (024.004.017). A.T.C. is partially supported by an American Cancer Society Research Professor Award.

AUTHOR CONTRIBUTIONS

Conceived and supervised the study: T.D.S., N.S., and F.A.; performed the analyses: E. Piperni; contributed with data analysis, software, or sample processing: L.H.N., P.M., H.K., E. Pasolli, S.A.-S., A.A., K.M.B., A.B.-M., S.M., M.V.-C., E.B., F.B., R.D., E.F., F.G., G.H., E.R.L., M.L., M.M., A.M., L.J.M., M. Pala, M. Pitzalis, J.W., J.F., A.Z., F.C., S.E.B., D.E., A.T.C., C.H., T.D.S., N.S., and F.A; wrote and edited the manuscript: E. Piperni, L.H.N., S.M.C., C.H., T.D.S., N.S., and F.A.

DECLARATION OF INTERESTS

G.H., J.W., and T.D.S. are co-founders of ZOE Ltd. G.H., J.W., R.D., F.G., K.M.B., A.A., and E.R.L. are or have been employees of Zoe Ltd. N.S., F.A., S.E.B., and T.D.S. are consultants to ZOE Ltd. T.D.S., R.D., J.W., G.H., N.S., and S.E.B. receive options with ZOE Ltd. All other authors declare no competing interests. Zoe Ltd. holds the following patent applications on *Blastocystis*: US patent pending application 17/203,678; Europe patent pending application 21713392.5 and 24386003.8; and PCT (World) patent pending application PCT/EP2022/076241.

Received: May 12, 2023
Revised: February 23, 2024
Accepted: June 11, 2024
Published: July 8, 2024

REFERENCES

1. Willett, W.C., and Stampfer, M.J. (2013). Current evidence on healthy eating. Annu. Rev. Public Health *34*, 77–95.
2. GBD 2017 Diet Collaborators (2019). 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.
3. World Health Organization (2020). Healthy diet. https://www.who.int/en/news-room/fact-sheets/detail/healthy-diet.

4. WCRF International (2022). Cancer prevention recommendations. https://www.wcrf.org/dietandcancer/cancer-prevention-recommendations.
5. Lichtenstein, A.H., Appel, L.J., Vadiveloo, M., Hu, F.B., Kris-Etherton, P.M., Rebholz, C.M., Sacks, F.M., Thorndike, A.N., Van Horn, L., and Wylie-Rosett, J. (2021). 2021 Dietary Guidance to Improve Cardiovascular Health: A Scientific Statement From the American Heart Association. Circulation *144*, e472–e487.
6. Patnode, C.D., Evans, C.V., Senger, C.A., Redmond, N., and Lin, J.S. (2017). Behavioral Counseling to Promote a Healthful Diet and Physical Activity for Cardiovascular Disease Prevention in Adults Without Known Cardiovascular Disease Risk Factors: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA *318*, 175–193.
7. Asnicar, F., Berry, S.E., Valdes, A.M., Nguyen, L.H., Piccinno, G., Drew, D.A., Leeming, E., Gibson, R., Le Roy, C., Khatib, H.A., et al. (2021). Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nat. Med. *27*, 321–332.
8. Andersen, L.O., and Stensvold, C.R. (2016). Blastocystis in Health and Disease: Are We Moving from a Clinical to a Public Health Perspective? J. Clin. Microbiol. *54*, 524–528.
9. Beghini, F., Pasolli, E., Truong, T.D., Putignani, L., Cacciò, S.M., and Segata, N. (2017). Large-scale comparative metagenomics of Blastocystis, a common member of the human gut microbiome. ISME J. *11*, 2848–2863.
10. Stensvold, C.R., and Clark, C.G. (2020). Pre-empting Pandora's Box: Blastocystis Subtypes Revisited. Trends Parasitol. *36*, 229–232.
11. Jiménez, P., Muñoz, M., and Ramírez, J.D. (2022). An update on the distribution of Blastocystis subtypes in the Americas. Heliyon *8*, e12592. https://doi.org/10.1016/j.heliyon.2022.e12592.
12. Alfellani, M.A., Stensvold, C.R., Vidal-Lapiedra, A., Onuoha, E.S.U., Fagbenro-Beyioku, A.F., and Clark, C.G. (2013). Variable geographic distribution of Blastocystis subtypes and its potential implications. Acta Trop. *126*, 11–18.
13. Stensvold, C.R., and Clark, C.G. (2016). Current status of Blastocystis: A personal view. Parasitol. Int. *65*, 763–771.
14. Stensvold, C.R., Sørland, B.A., Berg, R.P.K.D., Andersen, L.O., van der Giezen, M., Bowtell, J.L., El-Badry, A.A., Belkessa, S., Kurt, Ö., and Nielsen, H.V. (2022). Stool Microbiota Diversity Analysis of Blastocystis-Positive and Blastocystis-Negative Individuals. Microorganisms *10*, 326. https://doi.org/10.3390/microorganisms10020326.
15. Frank, D.N., St Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc. Natl. Acad. Sci. USA *104*, 13780–13785.
16. Thomas, A.M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., Beghini, F., Manara, S., Karcher, N., Pozzi, C., et al. (2019). Author Correction: Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat. Med. *25*, 1948.
17. Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.
18. Berry, S.E., Valdes, A.M., Drew, D.A., Asnicar, F., Mazidi, M., Wolf, J., Capdevila, J., Hadjigeorgiou, G., Davies, R., Al Khatib, H., et al. (2020). Human postprandial responses to food and potential for precision nutrition. Nat. Med. *26*, 964–973.
19. Wyatt, P., Berry, S.E., Finlayson, G., O'Driscoll, R., Hadjigeorgiou, G., Drew, D.A., Khatib, H.A., Nguyen, L.H., Linenberg, I., Chan, A.T., et al. (2021). Postprandial glycaemic dips predict appetite and energy intake in healthy individuals. Nat. Metab. *3*, 523–529.
20. De Filippis, F., Pasolli, E., Tett, A., Tarallo, S., Naccarati, A., De Angelis, M., Neviani, E., Cocolin, L., Gobbetti, M., Segata, N., et al. (2019). Distinct Genetic and Functional Traits of Human Intestinal Prevotella copri Strains Are Associated with Different Habitual Diets. Cell Host Microbe *25*, 444–453.e3.
21. Meslier, V., Laiola, M., Roager, H.M., De Filippis, F., Roume, H., Quinquis, B., Giacco, R., Mennella, I., Ferracane, R., Pons, N., et al. (2020). Mediterranean diet intervention in overweight and obese subjects lowers plasma cholesterol and causes changes in the gut microbiome and metabolome independently of energy intake. Gut *69*, 1258–1268.
22. Tarallo, S., Ferrero, G., De Filippis, F., Francavilla, A., Pasolli, E., Panero, V., Cordero, F., Segata, N., Grioni, S., Pensa, R.G., et al. (2022). Stool microRNA profiles reflect different dietary and gut microbiome patterns in healthy individuals. Gut *71*, 1302–1314.
23. Al-Shaar, L., Yuan, C., Rosner, B., Dean, S.B., Ivey, K.L., Clowry, C.M., Sampson, L.A., Barnett, J.B., Rood, J., Harnack, L.J., et al. (2021). Reproducibility and Validity of a Semiquantitative Food Frequency Questionnaire in Men Assessed by Multiple Methods. Am. J. Epidemiol. *190*, 1122–1132.
24. Huang, T., Trudel-Fitzgerald, C., Poole, E.M., Sawyer, S., Kubzansky, L.D., Hankinson, S.E., Okereke, O.I., and Tworoger, S.S. (2019). The Mind–Body Study: study design and reproducibility and interrelationships of psychosocial factors in the Nurses' Health Study II. Cancer Causes Control *30*, 779–790.
25. Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D.T., Beghini, F., Malik, F., Ramos, M., Dowd, J.B., et al. (2017). Accessible, curated metagenomic data through ExperimentHub. Nat. Methods *14*, 1023–1024.
26. Maixner, F., Sarhan, M.S., Huang, K.D., Tett, A., Schoenafinger, A., Zingale, S., Blanco-Míguez, A., Manghi, P., Cemper-Kiesslich, J., Rosendahl, W., et al. (2021). Hallstatt miners consumed blue cheese and beer during the Iron Age and retained a non-Westernized gut microbiome until the Baroque period. Curr. Biol. *31*, 5149–5162.e6.
27. Tett, A., Huang, K.D., Asnicar, F., Fehlner-Peach, H., Pasolli, E., Karcher, N., Armanini, F., Manghi, P., Bonham, K., Zolfo, M., et al. (2019). The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. Cell Host Microbe *26*, 666–679.e7.
28. Wibowo, M.C., Yang, Z., Borry, M., Hübner, A., Huang, K.D., Tierney, B.T., Zimmerman, S., Barajas-Olmos, F., Contreras-Cubas, C., García-Ortiz, H., et al. (2021). Reconstruction of ancient microbial genomes from the human gut. Nature *594*, 234–239.
29. Rifkin, R.F., Vikram, S., Ramond, J.-B., Rey-Iglesia, A., Brand, T.B., Porraz, G., Val, A., Hall, G., Woodborne, S., Le Bailly, M., et al. (2020). Multiproxy analyses of a mid-15th century Middle Iron Age Bantu-speaker palaeo-faecal specimen elucidates the configuration of the "ancestral" sub-Saharan African intestinal microbiome. Microbiome *8*, 62.
30. Borry, M., Cordova, B., Perri, A., Wibowo, M., Prasad Honap, T., Ko, J., Yu, J., Britton, K., Girdland-Flink, L., Power, R.C., et al. (2020). CoprolD predicts the source of coprolites and paleofeces using microbiome composition and host DNA content. PeerJ *8*, e9001.
31. Hagan, R.W., Hofman, C.A., Hübner, A., Reinhard, K., Schnorr, S., Lewis, C.M., Jr., Sankaranarayanan, K., and Warinner, C.G. (2020). Comparison of extraction methods for recovering ancient microbial DNA from paleofeces. Am. J. Phys. Anthropol. *171*, 275–284.
32. Alfellani, M.A., Taner-Mulla, D., Jacob, A.S., Imeede, C.A., Yoshikawa, H., Stensvold, C.R., and Clark, C.G. (2013). Genetic diversity of blastocystis in livestock and zoo animals. Protist *164*, 497–509.
33. Brito, I.L., Yilmaz, S., Huang, K., Xu, L., Jupiter, S.D., Jenkins, A.P., Naisilisili, W., Tamminen, M., Smillie, C.S., Wortman, J.R., et al. (2016). Mobile genes in the human microbiome are structured from global to individual scales. Nature *535*, 435–439.
34. Yoshikawa, H., Wu, Z., Kimata, I., Iseki, M., Ali, I.K.M.D., Hossain, M.B., Zaman, V., Haque, R., and Takahashi, Y. (2004). Polymerase chain reaction-based genotype classification among human Blastocystis hominis populations isolated from different countries. Parasitol. Res. *92*, 22–29.

35. Hirata, T., Nakamura, H., Kinjo, N., Hokama, A., Kinjo, F., Yamane, N., and Fujita, J. (2007). Prevalence of Blastocystis hominis and Strongyloides stercoralis infection in Okinawa, Japan. Parasitol. Res. *101*, 1717–1719.

36. Scanlan, P.D., Knight, R., Song, S.J., Ackermann, G., and Cotter, P.D. (2016). Prevalence and genetic diversity of Blastocystis in family units living in the United States. Infect. Genet. Evol. *45*, 95–97.

37. El Safadi, D., Gaayeb, L., Meloni, D., Cian, A., Poirier, P., Wawrzyniak, I., Delbac, F., Dabboussi, F., Delhaes, L., Seck, M., et al. (2014). Children of Senegal River Basin show the highest prevalence of Blastocystis sp. ever observed worldwide. BMC Infect. Dis. *14*, 164.

38. Scanlan, P.D., Hill, C.J., Ross, R.P., Ryan, C.A., Stanton, C., and Cotter, P.D. (2018). The intestinal protist Blastocystis is not a common member of the healthy infant gut microbiota in a Westernized country (Ireland). Parasitology *145*, 1274–1278.

39. Pandey, P.K., Verma, P., Marathe, N., Shetty, S., Bavdekar, A., Patole, M.S., Stensvold, C.R., and Shouche, Y.S. (2015). Prevalence and subtype analysis of Blastocystis in healthy Indian individuals. Infect. Genet. Evol. *31*, 296–299.

40. Valles-Colomer, M., Blanco-Míguez, A., Manghi, P., Asnicar, F., Dubois, L., Golzato, D., Armanini, F., Cumbo, F., Huang, K.D., Manara, S., et al. (2023). The person-to-person transmission landscape of the gut and oral microbiomes. Nature *614*, 125–135.

41. Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D.T., Manara, S., Zolfo, M., et al. (2018). Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. Cell Host Microbe *24*, 133–145.e5.

42. Manara, S., Selma-Royo, M., Huang, K.D., Asnicar, F., Armanini, F., Blanco-Miguez, A., Cumbo, F., Golzato, D., Manghi, P., Pinto, F., et al. (2023). Maternal and food microbial sources shape the infant microbiome of a rural Ethiopian population. Curr. Biol. *33*, 1939–1950.e4.

43. Ni, Q., Dong, S., Fan, Y., Wan, W., Teng, P., Zhu, S., Liang, X., Xu, H., Yao, Y., Zhang, M., et al. (2022). Molecular Epidemiology of Blastocystis in Confined Slow Lorises, Macaques, and Gibbons. Animals (Basel) *12*, 2992. https://doi.org/10.3390/ani12212992.

44. Hublin, J.S.Y., Maloney, J.G., and Santin, M. (2021). Blastocystis in domesticated and wild mammals and birds. Res. Vet. Sci. *135*, 260–282.

45. Sonnenburg, E.D., and Sonnenburg, J.L. (2019). The ancestral and industrialized gut microbiota and implications for human health. Nat. Rev. Microbiol. *17*, 383–390.

46. Carter, M.M., Olm, M.R., Merrill, B.D., Dahan, D., Tripathi, S., Spencer, S.P., Yu, F.B., Jain, S., Neff, N., Jha, A.R., et al. (2023). Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. Cell *186*, 3111–3124.e13.

47. Visconti, A., Le Roy, C.I., Rosa, F., Rossi, N., Martin, T.C., Mohney, R.P., Li, W., de Rinaldis, E., Bell, J.T., Venter, J.C., et al. (2019). Interplay between the human gut microbiome and host metabolism. Nat. Commun. *10*, 4505.

48. Xie, H., Guo, R., Zhong, H., Feng, Q., Lan, Z., Qin, B., Ward, K.J., Jackson, M.A., Xia, Y., Chen, X., et al. (2016). Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome. Cell Syst. *3*, 572–584.e3.

49. Albenberg, L.G., and Wu, G.D. (2014). Diet and the intestinal microbiome: associations, functions, and implications for health and disease. Gastroenterology *146*, 1564–1572.

50. David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563.

51. Conlon, M.A., and Bird, A.R. (2014). The impact of diet and lifestyle on gut microbiota and human health. Nutrients *7*, 17–44.

52. Vangay, P., Johnson, A.J., Ward, T.L., Al-Ghalith, G.A., Shields-Cutler, R.R., Hillmann, B.M., Lucas, S.K., Beura, L.K., Thompson, E.A., Till, L.M., et al. (2018). US Immigration Westernizes the Human Gut Microbiome. Cell *175*, 962–972.e10.

53. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227.

54. Satija, A., Bhupathiraju, S.N., Spiegelman, D., Chiuve, S.E., Manson, J.E., Willett, W., Rexrode, K.M., Rimm, E.B., and Hu, F.B. (2017). Healthful and Unhealthful Plant-Based Diets and the Risk of Coronary Heart Disease in U.S. Adults. J. Am. Coll. Cardiol. *70*, 411–422.

55. Monteiro, C.A., Cannon, G., Moubarac, J.-C., Levy, R.B., Louzada, M.L.C., and Jaime, P.C. (2018). The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. Public Health Nutr. *21*, 5–17.

56. Arnett, D.K., Blumenthal, R.S., Albert, M.A., Buroker, A.B., Goldberger, Z.D., Hahn, E.J., Himmelfarb, C.D., Khera, A., Lloyd-Jones, D., McEvoy, J.W., et al. (2019). 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation *140*, e596–e646.

57. Patel, N., Taveira, T.H., Choudhary, G., Whitlatch, H., and Wu, W.-C. (2012). Fasting serum C-peptide levels predict cardiovascular and overall death in nondiabetic adults. J. Am. Heart Assoc. *1*, e003152.

58. Min, J.Y., and Min, K.B. (2013). Serum C-peptide levels and risk of death among adults without diabetes mellitus. CMAJ *185*, E402–E408.

59. Emerging; Risk; Factors Collaboration, Di Angelantonio, E., Sarwar, N., Perry, P., Kaptoge, S., Ray, K.K., Thompson, A., Wood, A.M., Lewington, S., Sattar, N., et al. (2009). Major lipids, apolipoproteins, and risk of vascular disease. JAMA *302*, 1993–2000.

60. Triglyceride Coronary Disease Genetics Consortium and Emerging Risk Factors Collaboration, Sarwar, N., Sandhu, M.S., Ricketts, S.L., Butterworth, A.S., Di Angelantonio, E., Boekholdt, S.M., Ouwehand, W., Watkins, H., Samani, N.J., et al. (2010). Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. Lancet *375*, 1634–1639.

61. Connelly, M.A., Gruppen, E.G., Wolak-Dinsmore, J., Matyus, S.P., Riphagen, I.J., Shalaurova, I., Bakker, S.J.L., Otvos, J.D., and Dullaart, R.P.F. (2016). GlycA, a marker of acute phase glycoproteins, and the risk of incident type 2 diabetes mellitus: PREVEND study. Clin. Chim. Acta *452*, 10–17.

62. Vojinovic, D., Radjabzadeh, D., Kurilshikov, A., Amin, N., Wijmenga, C., Franke, L., Ikram, M.A., Uitterlinden, A.G., Zhernakova, A., Fu, J., et al. (2019). Relationship between gut microbiota and circulating metabolites in population-based cohorts. Nat. Commun. *10*, 5813.

63. Otvos, J.D., Shalaurova, I., Wolak-Dinsmore, J., Connelly, M.A., Mackey, R.H., Stein, J.H., and Tracy, R.P. (2015). GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation. Clin. Chem. *61*, 714–723.

64. Akinkuolie, A.O., Buring, J.E., Ridker, P.M., and Mora, S. (2014). A novel protein glycan biomarker and future cardiovascular disease events. J. Am. Heart Assoc. *3*, e001221.

65. Holmes, M.V., Millwood, I.Y., Kartsonaki, C., Hill, M.R., Bennett, D.A., Boxall, R., Guo, Y., Xu, X., Bian, Z., Hu, R., et al. (2018). Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. J. Am. Coll. Cardiol. *71*, 620–632.

66. D'Agostino, R.B., Sr., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M., and Kannel, W.B. (2008). General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation *117*, 743–753.

67. Billy, V., Lhotská, Z., Jirků, M., Kadlecová, O., Frgelecová, L., Parfrey, L.W., and Pomajbíková, K.J. (2021). Blastocystis Colonization Alters the Gut Microbiome and, in Some Cases, Promotes Faster Recovery From Induced Colitis. Front. Microbiol. *12*, 641483.

68. Deng, L., and Tan, K.S.W. (2022). Interactions between Blastocystis subtype ST4 and gut microbiota in vitro. Parasit. Vectors *15*, 80.

69. Deng, L., Lee, J.W.J., and Tan, K.S.W. (2022). Infection with pathogenic Blastocystis ST7 is associated with decreased bacterial diversity and altered gut microbiome profiles in diarrheal patients. Parasit. Vectors *15*, 312.

70. Després, J.-P. (2012). Body fat distribution and risk of cardiovascular disease: an update. Circulation *126*, 1301–1313.

71. Després, J.P., Moorjani, S., Lupien, P.J., Tremblay, A., Nadeau, A., and Bouchard, C. (1990). Regional distribution of body fat, plasma lipoproteins, and cardiovascular disease. Arteriosclerosis *10*, 497–511.

72. Shen, W., Punyanitya, M., Wang, Z., Gallagher, D., St-Onge, M.-P., Albu, J., Heymsfield, S.B., and Heshka, S. (2004). Visceral adipose tissue: relations between single-slice areas and total volume. Am. J. Clin. Nutr. *80*, 271–278.

73. Weir, C.B., and Jan, A. (2022). BMI Classification Percentile And Cut Off Points. In StatPearls (StatPearls Publishing).

74. Dashti, H.S., Scheer, F.A., Jacques, P.F., Lamon-Fava, S., and Ordovás, J.M. (2015). Short sleep duration and dietary intake: epidemiologic evidence, mechanisms, and health implications. Adv. Nutr. *6*, 648–659.

75. Hall, W.L. (2022). The emerging importance of tackling sleep-diet interactions in lifestyle interventions for weight management. Br. J. Nutr. *128*, 561–568.

76. VanEvery, H., Franzosa, E.A., Nguyen, L.H., and Huttenhower, C. (2023). Microbiome epidemiology and association studies in human health. Nat. Rev. Genet. *24*, 109–124.

77. Asnicar, F., Thomas, A.M., Passerini, A., Waldron, L., and Segata, N. (2024). Machine learning for microbiologists. Nat. Rev. Microbiol. *22*, 191–205.

78. Pasolli, E., Truong, D.T., Malik, F., Waldron, L., and Segata, N. (2016). Machine Learning Meta-analysis of Large Metagenomic Datasets: Tools and Biological Insights. PLoS Comput. Biol. *12*, e1004977.

79. Clark, C.G., van der Giezen, M., Alfellani, M.A., and Stensvold, C.R. (2013). Chapter One - Recent Developments in Blastocystis Research. In Advances in Parasitology, D. Rollinson, ed. (Academic Press), pp. 1–32.

80. Noël, C., Dufernez, F., Gerbod, D., Edgcomb, V.P., Delgado-Viscogliosi, P., Ho, L.-C., Singh, M., Wintjens, R., Sogin, M.L., Capron, M., et al. (2005). Molecular phylogenies of Blastocystis isolates from different hosts: implications for genetic diversity, identification of species, and zoonosis. J. Clin. Microbiol. *43*, 348–355.

81. Stensvold, C.R., and van der Giezen, M. (2018). Associations between Gut Microbiota and Common Luminal Intestinal Parasites. Trends Parasitol. *34*, 369–377.

82. Lukeš, J., Stensvold, C.R., Jirků-Pomajbíková, K., and Wegener Parfrey, L. (2015). Are Human Intestinal Eukaryotes Beneficial or Commensals? PLoS Pathog. *11*, e1005039.

83. Ianiro, G., Iorio, A., Porcari, S., Masucci, L., Sanguinetti, M., Perno, C.F., Gasbarrini, A., Putignani, L., and Cammarota, G. (2022). How the gut parasitome affects human health. Therap. Adv. Gastroenterol. *15*, 17562848221091524.

84. Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L.J., Thompson, K.N., Zolfo, M., Manghi, P., Dubois, L., Huang, K.D., Thomas, A.M., et al. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat. Biotechnol. *41*, 1633–1644.

85. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., et al. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell *176*, 649–662.e20.

86. Cammarota, G., Ianiro, G., Kelly, C.R., Mullish, B.H., Allegretti, J.R., Kassam, Z., Putignani, L., Fischer, M., Keller, J.J., Costello, S.P., et al. (2019). International consensus conference on stool banking for faecal microbiota transplantation in clinical practice. Gut *68*, 2111–2121.

87. Ianiro, G., Punčochář, M., Karcher, N., Porcari, S., Armanini, F., Asnicar, F., Beghini, F., Blanco-Míguez, A., Cumbo, F., Manghi, P., et al. (2022). Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. Nat. Med. *28*, 1913–1923.

88. Scanlan, P.D., Stensvold, C.R., Rajilić-Stojanović, M., Heilig, H.G.H.J., De Vos, W.M., O'Toole, P.W., and Cotter, P.D. (2014). The microbial eukaryote Blastocystis is a prevalent and diverse member of the healthy human gut microbiota. FEMS Microbiol. Ecol. *90*, 326–330.

89. Chabé, M., Lokmer, A., and Ségurel, L. (2017). Gut Protozoa: Friends or Foes of the Human Gut Microbiota? Trends Parasitol. *33*, 925–934.

90. Terveer, E.M., van Beurden, Y.H., Goorhuis, A., Seegers, J.F.M.L., Bauer, M.P., van Nood, E., Dijkgraaf, M.G.W., Mulder, C.J.J., Vandenbroucke-Grauls, C.M.J.E., Verspaget, H.W., et al. (2017). How to: Establish and run a stool bank. Clin. Microbiol. Infect. *23*, 924–930.

91. Tito, R.Y., Chaffron, S., Caenepeel, C., Lima-Mendez, G., Wang, J., Vieira-Silva, S., Falony, G., Hildebrand, F., Darzi, Y., Rymenans, L., et al. (2019). Population-level analysis of Blastocystis subtype prevalence and variation in the human gut microbiota. Gut *68*, 1180–1189.

92. Terveer, E.M., van Gool, T., Ooijevaar, R.E., Sanders, I.M.J.G., Boeije-Koppenol, E., Keller, J.J., Bart, A., and Kuijper, E.J.; Netherlands Donor Feces Bank (NDFB) Study Group (2020). Human Transmission of Blastocystis by Fecal Microbiota Transplantation Without Development of Gastrointestinal Symptoms in Recipients. Clin. Infect. Dis. *71*, 2630–2636.

93. Clark, C.G., and Stensvold, C.R. (2016). Blastocystis: Isolation, Xenic Cultivation, and Cryopreservation. Curr. Protoc. Microbiol. *43*, 20A.1.1–20A.1.8.

94. Laforest-Lapointe, I., and Arrieta, M.-C. (2018). Microbial Eukaryotes: a Missing Link in Gut Microbiome Studies. mSystems *3*, e00201-17. https://doi.org/10.1128/mSystems.00201-17.

95. Finlay, B.B.; CIFAR Humans, and Microbiome (2020). Are noncommunicable diseases communicable? Science *367*, 250–251.

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

97. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics *25*, 2078–2079.

98. Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics *47*, 11.12.1–11.1234.

99. Beghini, F., McIver, L.J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A.M., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife *10*, e65088. https://doi.org/10.7554/eLife.65088.

100. Berry, S., Drew, D., Linenberg, I., Wolf, J., Hadjigeorgiou, G., Davies, R., Khatib, H.A., Hart, D., Surdulescu, G., Yarand, D., et al. (2020). Personalised REsponses to DIetary Composition Trial (PREDICT): an intervention study to determine inter-individual differences in postprandial response to foods. Preprint at Research Square. https://doi.org/10.21203/rs.2.20798/v1.

101. Bermingham, K.M., Linenberg, I., Polidori, L., Wolf, J., Bulsiewicz, W.J., Spector, T.D., and Berry, S.E. (2023). Improved Cardiometabolic Health Using a Personalised Nutrition Approach: The ZOE METHOD Study. Proc. AMIA Annu. Fall Symp. *91*, 55.

102. Nguyen, L.H., Ma, W., Wang, D.D., Cao, Y., Mallick, H., Gerbaba, T.K., Lloyd-Price, J., Abu-Ali, G., Hall, A.B., Sikavi, D., et al. (2020). Association Between Sulfur-Metabolizing Bacterial Communities in Stool and Risk of Distal Colorectal Cancer in Men. Gastroenterology *158*, 1313–1325.

103. Wang, D.D., Nguyen, L.H., Li, Y., Yan, Y., Ma, W., Rinott, E., Ivey, K.L., Shai, I., Willett, W.C., Hu, F.B., et al. (2021). The gut microbiome

modulates the protective association between a Mediterranean diet and cardiometabolic disease risk. Nat. Med. *27*, 333–343.

104. Mehta, R.S., Lochhead, P., Wang, Y., Ma, W., Nguyen, L.H., Kochar, B., Huttenhower, C., Grodstein, F., and Chan, A.T. (2022). Association of midlife antibiotic use with subsequent cognitive function in women. PLoS One *17*, e0264649.
105. Franzosa, E.A., Morgan, X.C., Segata, N., Waldron, L., Reyes, J., Earl, A.M., Giannoukos, G., Boylan, M.R., Ciulla, D., Gevers, D., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. Proc. Natl. Acad. Sci. USA *111*, E2329–E2338.
106. Voigt, A.Y., Costea, P.I., Kultima, J.R., Li, S.S., Zeller, G., Sunagawa, S., and Bork, P. (2015). Temporal and technical variability of human gut metagenomes. Genome Biol. *16*, 73.
107. Pilia, G., Chen, W.-M., Scuteri, A., Orrú, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P., et al. (2006). Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet. *2*, e132.
108. Gacesa, R., Kurilshikov, A., Vich Vila, A., Sinha, T., Klaassen, M.A.Y., Bolte, L.A., Andreu-Sánchez, S., Chen, L., Collij, V., Hu, S., et al. (2022). Environmental factors shaping the gut microbiome in a Dutch population. Nature *604*, 732–739.
109. Sijtsma, A., Rienks, J., van der Harst, P., Navis, G., Rosmalen, J.G.M., and Dotinga, A. (2022). Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. Int. J. Epidemiol. *51*, e295–e302.
110. Levin, D., Raab, N., Pinto, Y., Rothschild, D., Zanir, G., Godneva, A., Mellul, N., Futorian, D., Gal, D., Leviatan, S., et al. (2021). Diversity and functional landscapes in the microbiota of animals in the wild. Science *372*, eabb5352. https://doi.org/10.1126/science.abb5352.
111. McGue, M. (1992). When assessing twin concordance, use the proband-wise not the pairwise rate. Schizophr. Bull. *18*, 171–176.
112. De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. USA *107*, 14691–14696.
113. Schnorr, S.L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M., et al. (2014). Gut microbiome of the Hadza hunter-gatherers. Nat. Commun. *5*, 3654.
114. Clemente, J.C., Pehrsson, E.C., Blaser, M.J., Sandhu, K., Gao, Z., Wang, B., Magris, M., Hidalgo, G., Contreras, M., Noya-Alarcón, Ó., et al. (2015). The microbiome of uncontacted Amerindians. Sci. Adv. *1*, e1500183. https://doi.org/10.1126/sciadv.1500183.
115. Martínez, I., Stegen, J.C., Maldonado-Gómez, M.X., Eren, A.M., Siba, P.M., Greenhill, A.R., and Walter, J. (2015). The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. Cell Rep. *11*, 527–538.
116. Obregon-Tito, A.J., Tito, R.Y., Metcalf, J., Sankaranarayanan, K., Clemente, J.C., Ursell, L.K., Zech Xu, Z., Van Treuren, W., Knight, R., Gaffney, P.M., et al. (2015). Subsistence strategies in traditional societies distinguish gut microbiomes. Nat. Commun. *6*, 6505.
117. Rampelli, S., Schnorr, S.L., Consolandi, C., Turroni, S., Severgnini, M., Peano, C., Brigidi, P., Crittenden, A.N., Henry, A.G., and Candela, M. (2015). Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. Curr. Biol. *25*, 1682–1693.
118. Bingham, S.A., Welch, A.A., McTaggart, A., Mulligan, A.A., Runswick, S.A., Luben, R., Oakes, S., Khaw, K.T., Wareham, N., and Day, N.E. (2001). Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. Public Health Nutr. *4*, 847–858.
119. Subar, A.F., Thompson, F.E., Kipnis, V., Midthune, D., Hurwitz, P., McNutt, S., McIntosh, A., and Rosenfeld, S. (2001). Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. Am. J. Epidemiol. *154*, 1089–1099.
120. Roe, M., Pinchen, H., Church, S., and Finglas, P. (2015). McCance and widdowson's the composition of Foods Seventh summary edition and updated composition of foods integrated dataset. Nutr. Bull. *40*, 36–39.
121. U.S. Department of Agriculture (2024). Agricultural Research Service Food Surveys Research Group Home Page. http://www.ars.usda.gov/nea/bhnrc/fsrg.
122. Guenther, P.M., Casavale, K.O., Reedy, J., Kirkpatrick, S.I., Hiza, H.A.B., Kuczynski, K.J., Kahle, L.L., and Krebs-Smith, S.M. (2013). Update of the Healthy Eating Index: HEI-2010. J. Acad. Nutr. Diet. *113*, 569–580.
123. Monteiro, C.A., Cannon, G., Lawrence, M., da Costa Louzada, M.L., and Pereira Machado, P. (2019). Ultra-Processed Foods, Diet Quality, and Health Using the NOVA Classification System*49*. FAO.
124. Goff, D.C., Jr., Lloyd-Jones, D.M., Bennett, G., Coady, S., D'Agostino, R.B., Gibbons, R., Greenland, P., Lackland, D.T., Levy, D., O'Donnell, C.J., et al. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation *129*, S49–S73.
125. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.
126. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B *57*, 289–300.
127. Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference, pp. 92–96.
128. Lyubchich, V., Gel, Y., and Vishwakarma, S. (2023). funtimes: Functions for Time Series Analysis. https://github.com/vlyubchich/funtimes.
129. Vallat, R. (2018). Pingouin: statistics in Python. J. Open Source Softw. *3*, 1026.
130. World Health Organization (2010). A healthy lifestyle – WHO recommendations. https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle—who-recommendations.
131. CDC (2022). About adult BMI (Centers for Disease Control and Prevention). https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html.

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
DNA/RNA-Shield Zymo buffer	Zymo	Cat#R1100 DNA/RNA Shield™ (Zymo Research)
Qiagen DNeasy PowerSoil Pro	Qiagen	Cat#47016_DNeasy PowerSoil Pro Kit (250)
Illumina DNA Prep Tagmentation kit	Illumina	Nextera DNA Flex Library Prep
Illumina NovaSeq 6000 S4 flowcell	Illumina	Cat#20012866 NovaSeq 6000 S4 Reagent Kit (300 cycles)
QIAamp PowerFecal DNA Kit	Qiagen	Cat#12830-50
Deposited data		
curatedMetagenomicData (v3)	Pasolli et al.[25]	Accession number of publicly available human metagenomic samples are available at: https://github.com/waldronlab/curatedMetagenomicData
Full table of *Blastocystis* presence/absence	This paper	Zenodo: https://doi.org/10.5281/zenodo.11202497
Metagenomic samples (PREDICT 2)	This paper	EBI: PRJEB75460
Metagenomic samples (PREDICT 3 US21)	This paper	EBI: PRJEB75462
Metagenomic samples (PREDICT 3 US22A)	This paper	EBI: PRJEB75463
Metagenomic samples (PREDICT 3 UK22A)	This paper	EBI: PRJEB75464
Metagenomic samples (PREDICT 3 UK23ART)	This paper	EBI: PRJEB75465
Software and algorithms		
Python (v3.9)	Python Software Foundation	https://www.python.org/
R (v4.3)	R Core Team	https://www.R-project.org/
MetaPhlAn4 (v4.beta)	Blanco-Míguez et al.[84]	https://github.com/biobakery/MetaPhlAn
Bowtie2 (v2.3)	Langmead and Salzberg[96]	https://github.com/BenLangmead/bowtie2
SAMtools (v1.19)	Li et al.[97]	http://github.com/samtools/
BEDtools	Quinlan[98]	https://github.com/arq5x/bedtools2
Meta-analyses	N/A	https://github.com/waldronlab/curatedMetagenomicDataAnalyses
preprocessing pipeline	N/A	https://github.com/SegataLab/preprocessing
Pyani (v0.2)	N/A	https://github.com/widdowquinn/pyani
mapsharper	N/A	www.mapshaper.org
Other		
curatedMetagenomicData (v3)	Pasolli et al.[25]	https://github.com/waldronlab/curatedMetagenomicData
2021 US Divisions Cartographic Boundary File	N/A	https://www.census.gov/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Francesco Asnicar (f.asnicar@unitn.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper analyzes existing, publicly available data as well as new data. Sequence data for all publicly available datasets and their accompanying metadata can be found at their respective public repositories as specified in their parent manuscripts (DOIs are reported in Table S1) or using the curatedMetagenomicData package,[25] where applicable. For the Dutch Microbiome Project (DMP), raw microbiome sequencing data and basic phenotypes (including age, sex and BMI) are available at the European

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

Genome-Phenome Archive under accession EGAS00001005027. These datasets can be accessed from https://forms.gle/eHeBdXJMXbVvCJRc8. Raw metagenomic samples are provided for all participants of the ZOE PREDICT Studies. Specifically, PREDICT 1 is made available as reported previously,[7] whereas PREDICT 2 and PREDICT 3 US21, US22A, UK22A, and UK23ART cohorts are deposited in EBI: PRJEB75460, PRJEB75462, PRJEB75463, PRJEB75464, PRJEB75465. These accession numbers are also listed in the key resources table. Sex, age, BMI, country, and the presence or absence of each *Blastocystis* subtype in each sample of all the human cohorts analyzed have been deposited at Zenodo: https://doi.org/10.5281/zenodo.11202497. To obtain individual-level data from the Mind Body Study and Men's Lifestyle Validation Study, including metadata not used in the current manuscript, a written application is required. As per standard controlled access procedures and consistent with prior precedent, applications to use MBS/MLVS data will be reviewed by the External Collaborators Committees for scientific merit, fit for the proposed methodology, and verification that the proposed use meets the guidelines of the Ethics and Governance Framework and consent provided by participants. All other initial inquiries may be directed nhsaccess@channing.harvard.edu (MBS), and hpfs@hsph.harvard.edu (MLVS), respectively. Table S1 reports *Blastocystis* and ST-specific prevalences for all the included datasets.

- Computational analyses were performed using the bioBakery suite of tools[84,99] and a *Blastocystis*-specific workflow previously described.[9] The code to perform meta-analyses is available at (https://github.com/waldronlab/curatedMetagenomicDataAnalyses). Other code and tools are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The ZOE PREDICT Study cohorts

The ZOE PREDICT program comprises multiple individual studies that collectively constitute one of the largest multi-omic population health initiatives, linking nutrition, individualized metabolic responses to foods, and the gut microbiome.[18,100] In brief, we jointly considered and harmonized five ZOE PREDICT Studies: PREDICT 1, PREDICT 2, PREDICT 3 US 21, PREDICT 3 US 22A, and PREDICT 3 UK 22A. The previously described PREDICT 1 cohort (Clinicaltrials.gov identifier: NCT03479866) enrolled 1,098 participants from the United Kingdom (UK) and United States (US) to a study protocol involving a monitored clinical visit to collect both blood and anthropometric information, followed by a subsequent at-home phase during which postprandial responses to both standardized test and *ad libitum* meals were recorded, along with the collection of 1,098 metagenomic stool samples.[7] The PREDICT 2 study (Clinicaltrials.gov identifier: NCT03983733) largely mirrored the collection protocol of PREDICT 1, but to offer a more representative and patient-friendly approach to precision nutrition, enrollment targeted 975 individuals from groups historically underrepresented in US clinical trials and was conducted completely remotely. Finally, the three PREDICT 3 cohorts consisted of 32,622 paying customers of a smartphone-aided personalized nutrition program (n=11,798 for P3_US21, n=8,470 for P3_US22A, and n=12,353 for P3_UK22A) for whom health questionnaires were administered and baseline stool samples collected (P3 US protocol number (IRB): Pro00044316; P3 UK ethical review reference: HR-23/24-28300; Clinicaltrials.gov identifier: NCT04735835). All the participants of the PREDICT 1 study were screened and considered only if considered healthy, while for the other cohorts, we considered as "not healthy" participants who self-reported a diagnosis or treatment for cardiovascular diseases, diabetes, pre-diabetes, hyperlipidemia, and high blood pressure. The "ZOE Microbiome Ranking 2024 (Cardiometabolic Health)" (abbreviated to "ZOE MB Health-rank") reported in Figure 5 were defined as the average summary of the five PREDICT cohorts rankings from partial correlations (corrected by sex, age, and BMI) between SGBs' relative abundance with personal, fasting, and postprandial markers representing cardiometabolic health.

The PREDICT 3 UK 23A RT cohort

The PREDICT 3 UK 23A RT (also P3_UK23ART) cohort enrolled 1,124 healthy individuals (867 females and 257 males) who followed a six-month personalized dietary program (PDP) intervention study and from whom pre and post-intervention gut microbiome samples were collected (n=2,248). The duration of the PDP intervention[101] was on average 5 months and 3 weeks (min 42 days and max 295 days; Table S7), during which participants received personalized food-level dietary advice through the ZOE mobile phone application. Recommendations were based on person-specific food quality scores computed using the ZOE 2022 Algorithm that incorporates food characteristics (e.g. macronutrients, glycemic load, fat quality, level of processing), individual's glucose control and postprandial triglyceride concentrations, gut microbial composition, baseline CVD risk, and prior health history. These person-specific ("personalized") food scores ranged from 0 to 100 with higher values indicating more healthful food items. The PDP did not specifically seek to restrict energy intake (i.e., calorie counting). Participants also received more generalized nutrition and lifestyle advice through interactive smartphone-based curriculum that covered basic nutritional and dietary health concepts (e.g., the potential benefits of increasing the intake of plant-based foods, fiber, fermented foods, or replacing refined carbohydrates with whole grains). Participants completed two validated food frequency questionnaires (FFQs), one at baseline and one post-intervention. Other than enrolling adults with no chronic health issues, participants were also required to self-report that they had not taken antibiotics in the 3 months prior to the study enrollment, nor during the two sampling times.

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

The MBS and MLVS cohort

From 2012 to 2013, longitudinal stool samples were collected from men in the Men's Lifestyle Validation Study (MLVS), nested within the long-running Health Professionals Follow-up Study (HPFS), and similarly, from women in the Mind-Body Study (MBS) nested within the Nurses' Health Study II (NHS II) from 2013 to 2014. Study design and biospecimen collection for these two prospective cohorts have been described previously,[23,24,102–104] but in brief, participants consented to provide two stool specimens from consecutive bowel movements at-home one to three days apart employing a validated self-collection protocol.[105,106] A second collection of paired samples was conducted approximately six months after the initial pair. Participants were also asked to complete questionnaires detailing the collection time, stool appearance, and information on various lifestyle factors, including diet and recent antibiotic use. As previously described, samples were shipped overnight and stored at -80°C until sequencing at the Broad Institute of MIT and Harvard on the HiSeq paired-end shotgun sequencing platform (Illumina Inc., San Diego, CA). *Blastocystis* STs were profiled using the same computational workflow described below (see *Blastocystis* detection and ST profiling). In the current study, participants with hyperlipidemia, diabetes and/or cardiovascular diseases, were considered "not healthy". For the non-longitudinal analysis, out of the four samples of each subject, we retained the one with the highest number of reads.

DeFilippisF_2019, MeslierV_2020, TaralloS_2021

The DeFilippisF_2019 cohort comprised 97 healthy Italian participants grouped into three categories based on their habitual diet: omnivores (n=23), vegetarians (n=38), and vegans (n=36).[20] The MeslierV_2020 cohort included 82 healthy Italian persons living with overweight or obesity and with a habitually low intake of fruit and vegetables and a sedentary lifestyle. Participants joined a parallel randomized controlled intervention trial of the Mediterranean diet.[21] In this case, for each subject we retained only one sample, choosing the one with the highest number of reads. Finally, the TaralloS_2021 cohort included 120 healthy individuals categorized as omnivores (n=40), vegetarians (n=40), and vegans (n=40).[22]

The SardiNIA cohort

The SardiNIA project is a longitudinal study[107] of 7,730 individuals from the general population (57% females and 43% males), ranging from 14 to 103 years, native of the central east coast of Sardinia, Italy. All participants signed informed consent to study protocols approved by the Sardinian Regional Ethics Committee (protocol no. 2171/CE). Volunteers were asked to collect stools in the morning of the visit (or the day before), and to store them at home (in the refrigerator). At the clinic site, stools were frozen at -80°. Then, all at once, were aliquoted (0,25 grams)–avoiding thawing–in safe-lock tubes and frozen at -80°C within 10 minutes, until DNA extraction. Microbial DNA extraction was performed with Bead Beating Plus Column (RBB+C), through high-speed shaking (15 Hz for 10 min with Tissuelyser II, Qiagen) coupled with QIAamp PowerFecal DNA Kit. Metagenomic sequencing was performed at the University of Trento, using the Illumina Novaseq 6000 platform to generate, on average, 8.68 Gb of 150 bp paired-end reads per sample. In this work, we considered the subset of 2,596 individuals with a sequenced metagenome available. We considered as "not healthy" those individuals who reported cardiovascular diseases (including infarction, heart failure, and stroke), diabetes (both T1D and T2D), any type of cancer, high cholesterol, Crohn's disease, peptic ulcer, ulcerative colitis, pancreatitis, and asthma.

The Dutch Microbiome Project cohort

The Lifelines Dutch Microbiome Project (DMP) cohort[108] consists of 8,208 individuals with fecal samples collected between 2015 and 2016 and is part of the Lifelines cohort study. Lifelines is a three-generation prospective population-based cohort study from the northern Netherlands.[109] The study was approved by the ethical committee of the University Medical Center Groningen (METc number: 2017/152) with additional written consent signed by participants, or parents/legal representatives for underaged (<18) participants. Sample collection and sequencing were described in the original publication.[108] In brief, fecal collection was performed by participants at home, stool samples were frozen within 15 minutes of production and shipped and stored at -80 °C until DNA extraction. Microbial DNA was isolated with the QIAamp Fast DNA Stool Mini Kit (Qiagen). Metagenomic sequencing was performed at Novogene, using the Illumina HiSeq 2000 platform to generate approximately 8 Gb of 150 bp paired-end reads per sample. In this study, we included 6,047 individuals with sex, age, BMI, and health status information and a minimum sampling depth of 10M reads. We considered an individual as "not healthy" if one of the following was reported: cancer (including breast, cervix, colon, melanoma, prostate, skin basal); cardiovascular diseases (including arrhythmia, atherosclerosis, high cholesterol, heart attack, heart failure, heart valve problems, hypertension); endocrine diseases (including T1D and T2D); and gastrointestinal diseases (including celiac disease, Crohn's disease, ulcerative colitis).

Publicly available human metagenomic samples

Publicly available metagenomic cohorts and their respective metadata were retrieved from the curatedMetagenomicData version 3 (cMD3) package,[25] including age at sample collection, sex, body mass index (BMI), country, lifestyle (i.e., Westernized or non-Westernized) and information on health status of the host. Individuals were categorized as "not healthy" as encoded in cMD3. Diseases comprise cardiovascular disease (ASCVD), diabetes, hyperlipidemia, inflammatory bowel disease, soil-transmitted helminth infection (STH), cancer; asthma; and migraine. We included only samples with more than 10M reads. Moreover, given the heterogeneity of study designs in cMD3 (e.g., cross-sectional vs. longitudinal), for each unique participant, we retained only one sample, choosing

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

the one with the highest number of reads. Then, we considered only datasets with at least 10 samples from either healthy adults or children. In total, we gathered 61 distinct datasets with 12,865 microbiome samples from as many distinct participants.

Ancient human gut metagenomic samples

We included 28 publicly-available ancient human gut metagenomic samples from seven datasets.[26–31] Reads were pre-processed as described in the corresponding works (Table S1). We only considered samples with more than 10M reads (range 14.3 M - 919.5 M), and profiled them using the same computational workflow described below (see *Blastocystis* detection and ST profiling). Overall, the samples were collected from paleofeces found in Austria, Italy, Mexico, South Africa, the United Kingdom, and the United States, with archaeological dating ranging from 3000 BC to the Post-Medieval Age (Table S1).

Publicly available animal metagenomic samples

We gathered 4,590 metagenomic samples from non-human hosts among 49 different publicly available datasets (Table S2) to assess the presence of the eight human-associated *Blastocystis* subtypes. Metadata was retrieved according to instructions offered in the data availability statements of each respective manuscript. In total, the samples encompassed 214 different host species with mice (n=3,318), non-human primates (n=560), ruminants (n=326), and birds (n=140) being the most frequently represented. Less represented animal taxa included insects (n=28), carnivores (bears, dogs, cats, seals, n=31), boned fishes (n=17), turtles (n=17), elephants (n=13), nematodes (n=10). For most species, only a few samples were available, with 179 species with less than five samples each, including crabs, mollusks, bats, cetaceous, rabbits, hedgehogs, sponges, anemones, and jellyfishes. Most datasets sampled animal hosts of a single species, a notable exception was one study that surveyed the microbiomes of 184 different animal species in the wild[110] (Table S2).

METHOD DETAILS

Sample collection, DNA extraction, sequencing, and quality control

For the PREDICT 1 cohort, details on sample collection, DNA extraction, and sequencing, were previously described.[7] The PREDICT 2 cohort followed the same procedure of PREDICT 1. For the PREDICT 3 cohorts, samples were self-collected at home into sterile tubes containing the DNA/RNA-Shield Zymo buffer. After collection, total DNA was extracted using the Qiagen DNeasy PowerSoil Pro. Sequencing libraries were prepared using the Illumina DNA Prep Tagmentation kit, following the manufacturer's guidelines. Whole-genome shotgun metagenomic sequencing was then performed targeting 3.5Gbp per sample on the Illumina NovaSeq 6000 with an S4 flowcell. All raw sequenced data were quality controlled using the preprocessing pipeline available at https://github.com/SegataLab/preprocessing, which entails three steps: *(i)* removal of reads that were low-quality (Q<20), too short (<75nt), or with more than 2 ambiguous bases, *(ii)* removal of host contaminant DNAs (Illumina's spike-in phiX 174 and human genomes, hg19), and *(iii)* synchronization of paired-end and unpaired reads. After pre-processing, the mean number of reads for each dataset was: 59.4 M in P2, 36.9 M in P3_US21, 35.0 M in P3_US22A, 31.2 M in P3_UK22A, while for the P3_UK23ART were 31.9 M at baseline and 31.1 M at post-intervention.

Longitudinal analysis

To evaluate the persistence of *Blastocystis* in the gut microbiome over time, we studied datasets with available longitudinal microbiome samples for the same individual from cMD3, the MBS, MLVS, and MeslierV_2020 cohorts. In total, we identified nine datasets with at least five *Blastocystis*-positive individuals in at least one of their samples (Table S4). These comprised 878 and 156 individuals with two or three time points, respectively. In each dataset, we considered the total number of persons for which *Blastocystis* was always detected (*Ba*) or never detected (*Bn*) across all time points. They were compared with the number of individuals found *Blastocystis*-positive at least at one time point (*Bf*). Moreover, among the "*Ba*" individuals, we assessed whether the detected subtype(s) was the same across all time points (Table S4). For each dataset, we reported the minimum, median, and maximum number of days elapsed between the sample collections for the same individual.

Twins subgroup analysis

To investigate the potential contribution of host genetics to gut *Blastocystis* ecology, we focused on available twin pairs. In total, we analyzed 712 twin pairs from the TwinsUK cohort (n=376 monozygotic [MZ] and n=336 dizygotic [DZ] pairs), published in three distinct works.[7,47,48] We assessed the probability that both members of a twin pair were *Blastocystis*-positive given that at least one was *Blastocystis*-positive (i.e., pairwise concordance rate[111]). We would anticipate that if genetic factors played a major role in *Blastocystis* colonization, pairwise *Blastocystis* concordance would be greater among genetically identical MZ compared to genetically unique DZ twins. Moreover, in concordantly positive twin pairs, we analyzed whether siblings carried the same *Blastocystis* ST. Finally, we compared *Blastocystis* concordance in twin pairs (both DZ and MZ) to randomly matched pairs of participants (Table S4).

Blastocystis detection and ST profiling

The presence of *Blastocystis* was assessed using a previously described computational workflow.[9] Briefly, nine reference genomes for eight different *Blastocystis* STs (ST1 [ST1_LXWW01], ST2 [ST2_JZRJ01], ST3 [ST3_JZRK01], ST4 [GCF_000743755 & ST4_BT1_JZRL01], ST6 [ST6_JZRM01], ST8 [ST8_JZRN01], and ST9 [ST9_JZRO01]) were mapped against metagenomic reads

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

using Bowtie2.[96] Then SAMtools[97] and bedtools[98] were used to compute the breadth of coverage for each subtype ('genomecov –bg' parameter). Finally, we report a sample to be positive for *Blastocystis* if any genomes of its subtypes had a breadth of coverage of at least 10% (the minimum sequencing breadth of coverage was zero, i.e. when no reads mapped to the reference genomes). The 10% threshold was selected based on the similarity between different *Blastocystis* ST genomes (Figure S1A) and the false positive detection rate for a second *Blastocystis* ST when another was already detected.[9] Conversely, depth of coverage is not involved in *Blastocystis* calling but was used to approximate relative abundance.

To assess *Blastocystis* relative abundance, we estimated within-sample *Blastocystis* DNA concentration for each ST reference genome. In each sample and for each detected ST, we summed the number of reads that mapped to each genome's base and divided it by the sample depth (estimated as the number of reads multiplied by the median read length). When a sample was positive for more than one ST, we summed up this estimation for all detected STs.

Sequencing depth and rarefaction

Variability in sequencing depth can affect mapping results, for this reason, shallow sequenced samples may not meet the above-described criterion for the detection of *Blastocystis*. Thus, we considered only human metagenomic stool samples with at least 10M reads. Moreover, to evaluate the effect of the sequencing depth on *Blastocystis* detection, we also profiled all samples by considering only 10M random reads for each (Figure S1C). We additionally performed a step-by-step rarefaction analysis on the five datasets with the highest median sequencing depth and at least 15 *Blastocystis*-positive samples (minimum above 70M reads, namely, BengtssonPalmeJ_2015, BritoIL_2016, LeChatelierE_2013, XieH_2016, CosteaPI_2017; Figure S1D). Briefly, for each dataset we computed the median number of reads (only healthy adult samples) and retained samples with a number of reads greater than the median. For these samples, we assessed *Blastocystis* presence considering an increasing number of random reads from 10M and up to the median for each specific dataset.

Westernized vs. non-Westernized populations

Consistent with prior studies,[52,53,85,112–117] in this work, we use the terms "Westernized" and "non-Westernized" as umbrella categorizations to broadly describe different lifestyles, including diet.[85] In brief, cohorts enrolling participants from populations with comparatively limited access to pharmaceuticals, lower caloric and fat intake, higher consumption of locally produced and less-processed foods, and/or a larger exposure to wildlife and domesticated animals are herein referred to as "non-Westernized" cohorts. Conversely, those in more industrialized countries with comparatively greater exposure to xenobiotics and pollutants, higher-calorie diets enriched in fats and shelf-stable ultra-processed foods, and greater availability and use of antibiotics and other pharmaceuticals are referred to as "Westernized" cohorts. According to these definitions, admittedly somewhat simplistic categorizations unable to fully capture the unique and distinguishing factors between populations, we analyzed 14 non-Westernized and 55 Westernized datasets, and three datasets that included samples from both non-Westernized and Westernized populations (Table S1).

Dietary data collection and categorization

For PREDICT 1 (UK), we employed a modified 131-item European Prospective Investigation into Cancer and Nutrition (EPIC) food frequency questionnaire (FFQ) designed to capture the average intake of foods in the past year, an instrument that has previously been validated against established nutrient biomarkers.[118] Participants in PREDICT 2 (US) were surveyed using a similarly validated DHQ-III FFQ including 135 items about food and beverage as well as 26 questions about dietary supplements.[119] For participants in the PREDICT 3 US21, UK22A, and US22A cohorts, short-term logged dietary data were collected via the ZOE mobile phone application. The PREDICT 3 UK22A and US22A were also surveyed using a 264-item FFQ. Food items (both FFQ and logged) were assembled into food groups by mapping them onto a "Food Tree" consisting of a database of nutrient information arranged according to a hierarchical tree structure as follows; Level 1 (9 food groups); Level 2 (52 food groups); Level 3 (195 food groups). UK foods were mapped onto the Composition of Foods Integrated Dataset (CoFID)[120] for the UK using food categories or sub-group codes, while US foods were similarly mapped onto the USDA FNDDS database.[121] Level 3 foods were aggregated and harmonized by nutrition scientists to allow for comparisons across cohorts. From both short and long-term dietary data, we computed three versions of the Plant-based Dietary Index (PDI), namely, the overall PDI (oPDI), the healthy PDI (hPDI) and the unhealthy PDI (uPDI), as well as the Healthy Eating Index (HEI).[54,122] The oPDI dietary index assigns positive scores to plant foods and negative scores to animal foods. For the hPDI, healthy plant foods (whole grains, fruits/vegetables, nuts/legumes, oils, tea/coffee) receive positive scores, while less-healthy plant foods (juices/sweetened beverages, refined grains, potatoes/fries, sweets) and animal foods receive negative scores. For the uPDI, less-healthy plant foods receive positive scores and animal and healthy plant foods receive negative scores. The HEI-2015 measures to which extent the individual's diet aligns with dietary guidelines. The guidelines cover 13 food groups and nutrients of which 9 adequacy (encouraged) and 4 moderation (discouraged) components are summed to represent overall diet quality.

To assess the intake of ultra-processed foods–defined as those that are shelf-stable, hyper-palatable, energy-dense, and made from refined ingredients using industrial processes–we classified food items according to established NOVA criteria[55,123] based on the nature, purpose, and extent of food processing: unprocessed or minimally processed foods, processed culinary ingredients, processed foods, and ultra-processed foods (e.g., bacon, cola, energy bars, and ice cream). Finally, individuals were stratified into well-established dietary regimes (i.e., omnivores, vegetarians, and vegans) based on self-reported answers.

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

Metabolic marker collection

In the PREDICT studies cardiometabolic markers selected for this research included sex, age, height, weight, BMI, blood pressure, fasting HbA1c (%), total cholesterol, C-peptide, GlycA, THR, ASCVD risk, and fasting and postprandial glucose, high-density lipoprotein cholesterol (HDL-C) and triglyceride (TG) concentrations. In PREDICT 1 descriptive information including sex and age were self-reported. Height, weight, and blood pressure were measured at a clinic visit (day 0) and dual-energy X-ray absorptiometry (DEXA) scans were used to measure visceral fat mass following standard manufacturer's recommendations (DXA; Hologic QDR 4500 plus). Participants were also fitted with wearable CGM devices (Abbott Freestyle Libre Pro (FSL; Abbott, Abbott Park, IL, US)) at the clinic visit (day 0). BMI (kg/m^2) was categorized as "healthy-weight," $18.5 \leq$ BMI<25 kg/m^2; "overweight", $25 \leq$ BMI<30 kg/m^2; and "obese", BMI ≥ 30 kg/m^2. Fasting GlycA was measured using a high-throughput NMR metabolomics (Nightingale Health) 2016 panel. Fasting and postprandial venous blood samples were collected at the clinic; plasma glucose and serum total cholesterol, HDL-C, and TG were measured using Affinity 1.0 and whole blood HbA1c (%) was measured using Viapath. The 10-year Atherosclerotic cardiovascular disease (ASCVD) risk was calculated as per the 2019 ACC/AHA Guidelines.[124] T2D and hyperlipidemia were self-reported via health questionnaires. Additional data were collected over the subsequent 13-day period at home; postprandial responses to eight standardized meals (seven in duplicate) of differing macronutrient (fat, carbohydrate, protein, and fiber) content were measured using continuous glucose monitors (CGMs) and dried-blood-spot (DBS) analysis as previously described.[18] The PREDICT 2 and PREDICT 3 studies took place after the PREDICT 1 study and were fully remote. Demographic information, cardiometabolic blood biomarkers, and postprandial responses to standardized test meals were collected in PREDICT 2 and 3 but with some differences from PREDICT 1. Height, weight, and blood pressure were self-reported, while fasting and postprandial responses for total cholesterol, HDL-C, TG and HbA1c, were assessed using whole blood finger-prick samples collected at home using DBS by commercial labs (CRL, Eurofins Biomnis). CGMs were fitted at home by participants. A smaller selection of standardized meals (a metabolic challenge meal, and medium fat and carbohydrate breakfast and lunch meals) were tested in PREDICT 2 & 3. To address potential noise due to extreme outliers, we filtered out values below the 1st and above the 99th percentiles, respectively. Detailed descriptions of sample collection, processing, and analysis have been previously described.[18,100]

QUANTIFICATION AND STATISTICAL ANALYSIS

Prediction of *Blastocystis* carriage

Species-level profiling was performed on all samples using MetaPhlAn 4[84] (version 4.beta.1 and the database vJan21_CHOCOPhlAnSGB_202103) with default parameters. We then trained a random forest-based machine learning algorithm, consistent with prior investigations,[16] only on species-level bacterial and archaeal relative abundance to predict *Blastocystis* carriage. In each model, we performed a cross-validation (CV) by randomly splitting the data into training and testing sets with an 80/20 ratio, repeated 100 times. We also applied a LODO (leave-one-dataset-out) classification framework in which each training cohort is left out of training with subsequent prediction modeling. Models were evaluated using the median area under the receiver operating characteristic curve (AUC).[77] To avoid biases due to low sample size, we included in the machine learning framework only cohorts with at least 50 samples from healthy adults and *(i) Blastocystis* prevalence >20% or *(ii)* more than 50 *Blastocystis*-positive samples. Moreover, we excluded the datasets without available MetaPhlAn 4 profiles (i.e., TaralloS_2021, SardiNIA, GacesaR_2022).

Statistical analyses

The Mann-Whitney U test was performed using the function mannwhitneyu from the scipy Python library.[125] Otherwise, when explicitly stated in the text and paired samples were available, the Wilcoxon signed-rank test was applied, using the function wilcoxon from the scipy Python library.[125] Unless otherwise specified, nominal p values were corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure[126] with the function fdrcorrection from the statsmodels Python library.[127] False discovery rate (FDR) corrected p values were reported as q-values. Mann-Kendall test was used to evaluate the presence of monotonic trends (no-trend_test (test="MK") function from the R package funtimes[128]). The hPDI scores and BMI values were divided into deciles to have sufficient points for the trend test (Figures 2B, 4A, and 4B). Effect sizes were calculated using the function compute_effsize from the pingouin Python library.[129] Unless otherwise noted, q-values are represented in figures with asterisks as follows: *=q-value<0.1, **=q-value<0.01, ***=q-value<0.001, and ****=q-value<0.0001.

GeoPandas v0.11.0 was used to work with the geospatial data in Python. The world map with contours of countries was downloaded from the GeoPandas dataset "naturalearth_lowres". The 2021 US Divisions Cartographic Boundary File (shapefile, 1:500,000 resolution) was downloaded from https://www.census.gov/. The mapshaper Interactive web interface (www.mapshaper.org) was used to remove the states "Alaska", "Puerto Rico", and "Hawaii", from the Pacific division for better visualization, and to convert the shapefile in the GeoJSON format.

To create the heatmaps in Figure 2C, we computed the 25th, 50th, and 75th percentiles of each food in each cohort and divided individuals into the four quartiles based on their relative food consumption. Within each quartile, we calculated *Blastocystis* prevalence and colored cells accordingly. In a few cases, due to the discrete nature of dietary data, values for two of the three quantile cutoffs were identical, thus leaving one quartile empty. To address this, we randomly split individuals into the two adjacent quartile groups stratifying for *Blastocystis* presence/absence. If the empty quartile was either Q2 or Q3, both with two adjacent quartiles (i.e. Q1 and Q3 for Q2; Q2 and Q4 for Q3), we selected the adjacent one with more individuals to be randomly split. Foods were

Please cite this article in press as: Piperni et al., Intestinal *Blastocystis* is linked to healthier diets and more favorable cardiometabolic outcomes in 56,989 individuals from 32 countries, Cell (2024), https://doi.org/10.1016/j.cell.2024.06.018

categorized based on whether they were concordantly higher in *Blastocystis*-positive or negative individuals in all PREDICT cohorts. Then, foods were sorted by the sum, across all PREDICT cohorts, of the differences in consumption between *Blastocystis*-positive and negative individuals, normalized by the overall maximum value.

For the heatmap in Figure 5B, we selected the 100 top-ranked SGBs according to their feature importances from the CV ML models, considering their average percentile rank across the 21 datasets used in the ML analysis (see prediction of Blastocystis carriage). In each dataset and for each of the top 100 SGBs, we compared the relative abundance values with the Mann-Whitney U test and computed the Log_2 Fold Change of the mean relative abundance between *Blastocystis*-positive and negative individuals. To avoid biases due to low prevalence or low sample size, an SGB was tested only if *(i)* present in at least 5% of *Blastocystis*-positive and negative individuals and the total number of such individuals was greater than 10; or *(ii)* the SGB was present in at least 50 *Blastocystis*-positive and negative samples. SGBs were then sorted according to their average Log_2FC across datasets. For visual purposes, we selected only 30 of the 100 SGBs, choosing those tested in all 21 datasets and with an absolute average Log_2FC greater than or equal to 0.5. Only 9 SGBs with higher relative abundance in *Blastocystis*-negative samples (i.e., negative avg. Log_2FC) passed these filters. We then selected the remaining 21 as those with the highest positive average Log_2FC values. White cells indicate no statistically significant difference after FDR correction of the Mann-Whitney U test (q-value >0.1).

Meta-analyses

We performed meta-analyses using a random-effects model to assess the strength of the associations between *Blastocystis* presence with both BMI as a continuous measure and the likelihood of sampling from chronic disease participants vs. healthy/control participants in our aggregated dataset. For the BMI meta-analysis, we included healthy adults (age >12 years) with available information on sex, age, and BMI. We excluded the few individuals who reported BMI <18.5 kg/m^2, the usual cutoff for patients living with underweight according to World Health Organization criteria[130,131] (n=22 in PREDICT 1, n=14 in PREDICT 2, n=50 in PREDICT 3 US21, n=54 in PREDICT 3 US22A, and n=120 in PREDICT 3 UK22A). Studies with less than 10 *Blastocystis*-positive and 10 *Blastocystis*-negative samples were also excluded (total samples n=36,065 from 24 datasets[25]). For the healthy/disease meta-analysis, we included adults (age >12 years) with age, sex, and BMI information and excluded underweight individuals. We further restricted the analysis to the cohorts with at least 10 *Blastocystis*-positive and 10 *Blastocystis*-negative samples and with at least 10 samples from healthy/control and 10 from diseased individuals (n=44,906 [13,548 cases and 31,358 controls] from 17 studies spanning eight disease categories; Table S5). The eight disease categories are atherosclerotic cardiovascular disease (ASCVD), which includes coronary artery disease, cerebrovascular disease, and heart failure; diabetes, which includes Type 2 Diabetes (T2D), Impaired Glucose Tolerance, and Metabolic Syndrome; Hyperlipidemia; inflammatory bowel disease (IBD); soil-transmitted helminth infection (STH); colorectal cancer (CRC), which includes precursor adenomas; asthma; and migraine.

For each analysis and each study, we performed logistic regression modeling on *Blastocystis* presence/absence accounting for sex, age, and BMI (BMI only in the latter). In the BMI meta-analysis, BMI was used as a continuous predictor in the logistic regression model, and in the healthy/disease meta-analysis we used logistic regression to assess the likelihood of samples being from those with chronic illnesses compared to referent healthy/control individuals, the log odds ratio of diseased over healthy individuals coefficient was extracted with their std. errors from each study and synthesized with the program available at (https://github.com/waldronlab/curatedMetagenomicDataAnalyses), estimating between-study variance with the Paule-Mandel formula. Single-study p values were computed as coefficient extremeness of the t-statistics; p values for meta-analysis coefficients were computed as Z scores extreme probability, both under the null hypothesis that the average coefficient is zero.

ADDITIONAL RESOURCES

The PREDICT cohorts are part of the following clinical trials. PREDICT 1 ClinicalTrials.gov ID: NCT03479866 (https://clinicaltrials.gov/study/NCT03479866); PREDICT 2 ClinicalTrials.gov ID: NCT03983733 (https://clinicaltrials.gov/study/NCT03983733); PREDICT 3: P3 US protocol number (IRB): Pro00044316; P3 UK ethical review reference: HR-23/24-28300; ClinicalTrials.gov ID: NCT04735835 (https://clinicaltrials.gov/study/NCT04735835).