

Environment dominates over host genetics in shaping human gut microbiota

Daphna Rothschild^{1,2*}, Omer Weissbrod^{1,2*}, Elad Barkan^{1,2*}, Alexander Kurilshikov³, Tal Korem^{1,2}, David Zeevi^{1,2}, Paul I. Costea^{1,2}, Anastasia Godneva^{1,2}, Iris N. Kalka^{1,2}, Noam Bar^{1,2}, Smadar Shilo^{1,2}, Dar Lador^{1,2}, Arnau Vich Vila^{3,4}, Niv Zmora^{5,6,7}, Meirav Pevsner–Fischer⁵, David Israeli⁸, Noa Kosower^{1,2}, Gal Malka^{1,2}, Bat Chen Wolf^{1,2}, Tali Avnit–Sagi^{1,2}, Maya Lotan–Pompan^{1,2}, Adina Weinberger^{1,2}, Zamir Halpern^{7,9}, Shai Carmi¹⁰, Jingyuan Fu^{3,11}, Cisca Wijmenga^{3,12}, Alexandra Zhernakova³, Eran Elinav^{5§} & Eran Segal^{1,2§}

Human gut microbiome composition is shaped by multiple factors but the relative contribution of host genetics remains elusive. Here we examine genotype and microbiome data from 1,046 healthy individuals with several distinct ancestral origins who share a relatively common environment, and demonstrate that the gut microbiome is not significantly associated with genetic ancestry, and that host genetics have a minor role in determining microbiome composition. We show that, by contrast, there are significant similarities in the compositions of the microbiomes of genetically unrelated individuals who share a household, and that over 20% of the inter-person microbiome variability is associated with factors related to diet, drugs and anthropometric measurements. We further demonstrate that microbiome data significantly improve the prediction accuracy for many human traits, such as glucose and obesity measures, compared to models that use only host genetic and environmental data. These results suggest that microbiome alterations aimed at improving clinical outcomes may be carried out across diverse genetic backgrounds.

The gut microbiome is increasingly recognized as having fundamental roles in human physiology and health^{1,2}. A central question is the extent to which microbiome composition is determined by host genetics. Previous studies have identified several heritable bacterial taxa^{3–7} but the combined bacterial abundance accounted for by them has not yet been quantified. Other studies have found associations between host single nucleotide polymorphisms (SNPs) and individual bacterial taxa or pathways^{5,8–11}. However, most previously reported associations are not statistically significant after multiple testing correction³. A recent study identified 42 SNPs that together explained 10% of the variance of microbiome β -diversity⁹. However, the statistical significance of this result has not yet been evaluated. Thus, the extent to which human genetics shape microbiome composition remains unclear.

Here we studied microbial–genetic associations using a cohort of 1,046 healthy Israeli individuals with metagenome-sequenced and 16S rRNA gene-sequenced gut microbiomes, genotypes, anthropometric and blood measurements, and dietary habits¹². Individuals in our cohort are of several different ancestral origins but we assume, owing to their broadly similar lifestyles, that they share a relatively homogeneous environment.

Our results demonstrate that gut microbiome composition is shaped predominantly by environmental factors. Specifically, we show that the microbiome is not significantly associated with genetic ancestry or with individual SNPs, and that previously reported associations are not replicated across different studies. We further estimate that the average heritability of gut microbiome taxa is only 1.9%, by analysing

data from 2,252 twins from the TwinsUK cohort⁶. However, further and larger-scale studies are required to accurately quantify gut microbiome heritability.

To provide direct evidence that the microbiome is shaped largely by environmental factors, we show that there is significant similarity among the microbiomes of genetically unrelated individuals who share a household, but no significant microbiome similarity among relatives who do not have a history of household sharing. We further demonstrate that over 20% of the variance in microbiome β -diversity can be inferred from environmental factors associated with diet and lifestyle, consistent with previous studies^{13,14}.

Because our findings suggest that gut microbiome and host genetics are largely independent, we compare the power of the gut microbiome and of host genetics to predict host phenotypes. We define the term ‘microbiome-association index’ (b^2) that—by analogy with genetic heritability, which is typically termed h^2 —quantifies the overall association between the microbiome and a host phenotype after accounting for host genetics. We find significant b^2 levels of 22–36% for body mass index (BMI; 25%), fasting glucose levels (22%), glycaemic status (25%), levels of high-density lipoprotein (HDL) cholesterol (36%), waist circumference (29%), hip circumference (27%), waist–hip ratio (WHR; 24%) and lactose consumption (36%). We note that b^2 should be interpreted with caution, because it is a correlative measure that does not necessarily indicate causality and it may be confounded by environmental factors. We additionally demonstrate that using microbiome data together with human genetic data substantially improves the

¹Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 7610001, Israel. ²Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 7610001, Israel. ³University of Groningen, University Medical Center Groningen, Department of Genetics, 9713 GZ Groningen, The Netherlands. ⁴University of Groningen, University Medical Center Groningen, Department of Gastroenterology and Hepatology, 9713 GZ Groningen, The Netherlands. ⁵Immunology Department, Weizmann Institute of Science, Rehovot 7610001, Israel. ⁶Internal Medicine Department, Tel Aviv Sourasky Medical Center, Tel Aviv 6423906, Israel. ⁷Research Center for Digestive Tract and Liver Diseases, Tel Aviv Sourasky Medical Center, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 6423906, Israel. ⁸Day Care Unit and the Laboratory of Imaging and Brain Stimulation, Kfar Shaul Hospital, Jerusalem Center for Mental Health, Jerusalem 9106000, Israel. ⁹Digestive Center, Tel Aviv Sourasky Medical Center, Tel Aviv 6423906, Israel. ¹⁰Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem 9112001, Israel. ¹¹University of Groningen, University Medical Center Groningen, Department of Pediatrics, 9713 GZ Groningen, The Netherlands. ¹²Department of Immunology, K.G. Jebsen Coeliac Disease Research Centre, University of Oslo, 0424 Oslo, Norway.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

accuracy with which human phenotypes can be predicted, consistent with a previous smaller-scale study¹⁵.

Finally, we successfully replicate our results in 836 Dutch individuals, with genotypes and metagenomic data, from the LifeLines DEEP (LLD) cohort⁶. Taken together, our results demonstrate that the gut microbiome is predominantly shaped by environmental factors, and is strongly correlated with many human phenotypes after accounting for host genetics.

Results

We studied a cohort of 1,046 healthy Israeli adults from whom we collected blood for genotyping and phenotyping, stools for metagenome sequencing and 16S rRNA gene sequencing, anthropometric measurements, and answers to food frequency and lifestyle questionnaires¹² (Extended Data Table 1 and Supplementary Table 1). We performed genotyping at 712,540 SNPs and imputed them to 5,567,647 SNPs (Methods). Stool samples were profiled using both metagenome sequencing and 16S rRNA gene sequencing, and then analysed at multiple taxonomic levels; the results presented here are based on metagenome species analysis (results at metagenome phylum, class, order, family, genus or bacterial gene levels, and for 16S genus and operational taxonomic unit levels, are provided in Supplementary Tables where appropriate). We included covariates for age, gender, stool collection method, and self-reported daily median caloric, fat, protein and carbohydrate consumption (Methods).

Limited evidence of microbiome–genetic associations

Our sample consists of self-reported Ashkenazi ($n = 508$), North African ($n = 64$), Middle Eastern ($n = 34$), Sephardi ($n = 19$), Yemenite ($n = 13$) and ‘admixed/other’ ($n = 408$) ancestries¹⁶ (Supplementary Table 2). We first successfully verified that the top two host genetic principal components are strongly associated with self-reported ancestry ($P < 10^{-32}$ for both principal component 1 and principal component 2, Kruskal–Wallis test; Fig. 1a, Extended Data Table 2 and Supplementary Table 3). By contrast, we found no significant association between ancestry and microbiome composition. Specifically, there was no significant correlation between any of the top five host genetic principal components and any of the top five microbiome β -diversity principal coordinates (PCOs), computed using Bray–Curtis dissimilarity; $P > 0.49$ for all pairwise associations, Spearman correlation; Supplementary Table 4).

We also found no significant differences between ancestries in terms of microbiome composition (quantified by PCOs of Bray–Curtis dissimilarities), α -diversity (quantified by the Shannon diversity index) or abundance of specific taxa (Kruskal–Wallis test for non-admixed individuals; Fig. 1b–d, Extended Data Table 2 and Supplementary Table 3). We obtained similar results when testing whether individuals who are more ancestrally similar, quantified by the fraction of grandparents from the same ancestry, have more similar microbiomes (quantified by Bray–Curtis dissimilarity), α -diversity or abundance of specific taxa (Mantel test¹⁷; Methods, Fig. 1e, Extended Data Table 2 and Supplementary Table 5). As a control, we verified that ancestrally similar individuals are significantly similar in terms of their genetics ($P < 10^{-5}$, Mantel test; Methods, Extended Data Table 2 and Supplementary Table 5). We also found no significant association between microbiome composition and genetic kinship (Fig. 2a, Extended Data Table 2 and Supplementary Table 6), though we note that SNP-based kinship tests are less powerful than ancestry-based tests (Supplementary Tables 7–10 and Supplementary Information).

One caveat of our study is the presence of imbalanced per-population sample sizes. Although an ideal study should have equal per-population sample sizes, we verified that our study is well-powered to detect microbiome–ancestry associations. Specifically, we found that the probability of finding a statistically significant microbiome–ancestry association is 70% if only 10% of the microbiome variance is explained

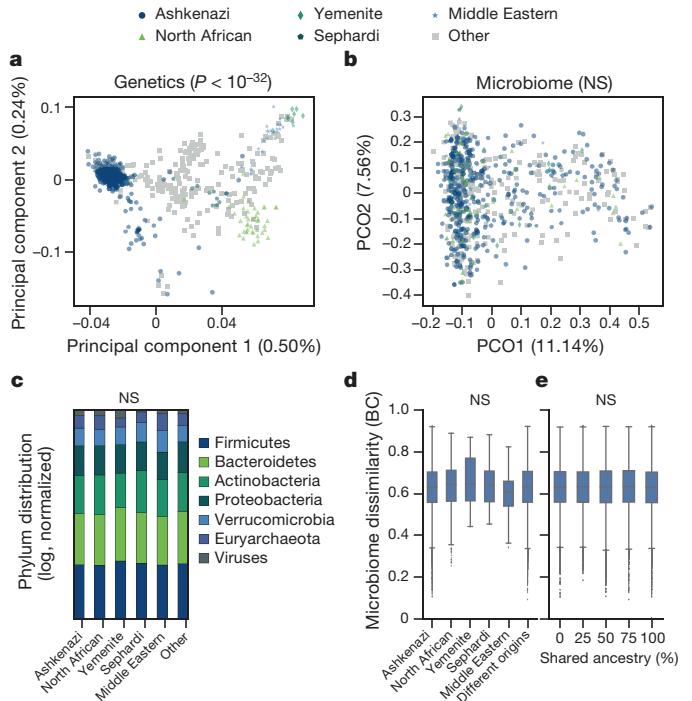


Figure 1 | Genetic ancestry is not significantly associated with microbiome composition. **a**, Genetic principal components are strongly associated with self-reported ancestry, with Ashkenazi ($n = 345$), North African ($n = 42$), Middle Eastern ($n = 24$), Sephardi ($n = 10$), Yemenite ($n = 8$) and admixed/other (other) ($n = 286$) ancestries ($P < 10^{-32}$; Kruskal–Wallis). **b**, As in **a**, but for microbiome principal coordinate analysis ($P > 0.08$; Kruskal–Wallis). **c**, The distribution of average phylum abundance among 582 non-admixed individuals (in log scale, normalized to sum to 1.0) is not associated with ancestry ($P > 0.05$; Kruskal–Wallis). NS, not significant. **d**, Box plots of Bray–Curtis (BC) dissimilarities across all pairs of 737 individuals for whom the ancestries of all grandparents are known, demonstrating that microbiome composition is not associated with ancestry ($P > 0.06$; Kruskal–Wallis test for the top five Bray–Curtis PCOs). $n = 105,570$ (Ashkenazi), 1,711 (North African), 528 (Middle Eastern), 136 (Sephardi) and 78 (Yemenite) same ancestry pairs; $n = 61,048$ different ancestry pairs. The lower and upper limits of the boxes represent the 25% and 75% percentiles, respectively, and the top and bottom whiskers represent the 5% and 95% percentiles, respectively. **e**, Box plots of Bray–Curtis dissimilarities across pairs of 946 individuals (including admixed individuals), organized according to shared ancestry fraction (the fraction of grandparents of the same ancestry), for pairs with 0% ($n = 167,618$), 25% ($n = 33,119$), 50% ($n = 100,163$), 75% ($n = 34,187$) and 100% ($n = 111,898$) shared ancestry fractions. The lower and upper limits of the boxes represent the 25% and 75% percentiles, respectively, and the top and bottom whiskers represent the 5% and 95% percentiles, respectively. The figure demonstrates that microbiome similarity is not associated with ancestral similarity ($P = 0.73$; Mantel test).

by ancestry, and is greater than 90% if over 30% of the microbiome variance is explained by ancestry (Supplementary Information).

The lack of association between microbiome composition and genetic ancestry suggests that the microbiome is not strongly associated with host genetics. Because twin studies are ideal for heritability estimation¹⁸, we analysed a previously studied⁶ dataset of 2,252 twins to directly quantify microbiome heritability. First, we found that the sum of the relative abundances of all 33 taxa reported as significantly heritable in the previous study⁶ accounted for only 5.6% of total microbiome composition (Methods). Next, we estimated the overall microbiome heritability using the formula $H^2 = \sum_{t \in S} r_t h_t^2$, in which r_t and h_t^2 are the relative abundance and estimated heritability of taxon t , respectively, and S is the set of significantly heritable taxa (making sure not to count the same taxon multiple times; see Methods). The resulting heritability estimate was only 1.9% or, at most, 8.1% when performing

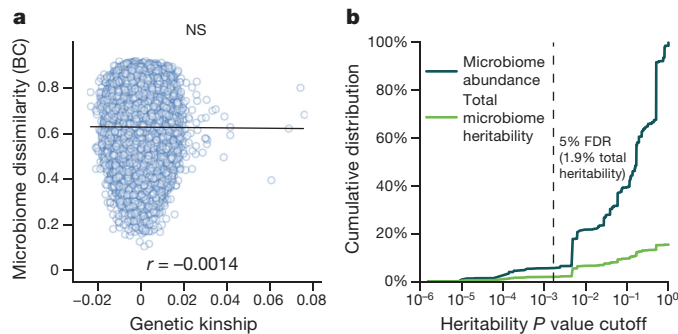


Figure 2 | Genetic kinship is weakly associated with microbiome composition. **a**, Scatter plot of genetic kinship of pairs of individuals (x axis) and their microbiome dissimilarity (y axis), among all pairs of $n = 715$ unrelated genotyped individuals, demonstrating that genetic kinship and microbiome similarity are uncorrelated ($P = 0.59$; Mantel test). NS, not significant. **b**, The overall heritability of significantly heritable taxa in a cohort of 2,252 twins⁶ (light green) and their cumulative relative abundance (dark green). The x axis indicates the P -value cutoff required to declare a taxon as significantly heritable (using P values computed in a previous study⁶). The figure demonstrates that the overall microbiome heritability is small regardless of the cutoff. Under a cutoff corresponding to a 5% FDR, the overall microbiome heritability is 1.9%. FDR, false discovery rate.

no correction for multiple testing in the definition of S (Fig. 2b and Supplementary Table 11). These numbers serve as estimates of the lower and upper bound of the true overall microbiome heritability.

In addition, we applied several machine-learning algorithms to predict ancestry proportions from microbiomes, but none were successful (prediction $R^2 < 0.01$ for all ancestries; Methods). We also tried to predict top microbiome PCOs from ancestral or genetic data, and again found no significant associations ($P > 0.1$, permutation testing; Methods and Supplementary Table 12).

Finally, we verified that similar results are obtained when repeating the above experiments using any of the following: other metagenome-derived taxonomic and functional levels (phylum, class, order, family, genus and bacterial genes; see Methods); 16S rRNA gene sequencing; Unifrac- and Jaccard-based dissimilarity measures; non-metric multidimensional scaling¹⁷ instead of principal coordinate analysis; dichotomization of relative abundance into presence/absence patterns; and when omitting covariates (Extended Data Fig. 1 and Supplementary Tables 3–6, 12).

We next investigated associations between individual SNPs and microbiome β -diversity, using a distance-based F test¹⁹ (Extended Data Fig. 2 and Methods). This analysis found two loci with marginal genome-wide significance (rs6563994, $P = 3 \times 10^{-8}$; rs13149273, $P = 4.2 \times 10^{-8}$). However, as we show later, we could not replicate these findings in an additional cohort. In addition, our data did not replicate any of the 42 SNPs previously reported⁹ as being significantly associated with microbiome β -diversity, either when using an F test or the previously applied method⁹ ($P > 0.05$ for all previously reported SNPs; Methods).

The previous study⁹ reported that these 42 SNPs could be used to infer 10% of the β -diversity variance, but did not report the statistical significance of this result. We were able to explain 12.1% of β -diversity variance using the 42 SNPs that were most closely associated with β -diversity in our own data, but this result was not statistically significant under permutations ($P = 0.74$; Methods). We conclude that inferring $> 10\%$ of β -diversity variance using top-ranked SNPs may be an inherent property of the method used rather than a biologically meaningful result. Thus, we find very limited evidence in our data for the association of any individual SNP with microbiome β -diversity.

We next tested for associations between individual SNPs and specific taxa, using a linear mixed model (LMM) and dichotomization of zero-inflated taxa (Methods; Supplementary Table 13). This analysis identified 43 loci with $P < 5 \times 10^{-8}$, but none remained

statistically significant at a false discovery rate (FDR) of 5% (Fig. 3a and Supplementary Table 14).

We also investigated the association of 225 SNPs in 211 loci reported as significantly associated with specific taxa or with β -diversity in any of five previous studies^{5,6,8–10} (Methods). To maximize replication power, we used the minimal P value obtained for each SNP across all taxa belonging to the same phylum. Only 7 of the 211 loci (3.3%) replicated at $P < 0.05/211$ (Fig. 3b and Supplementary Table 15; Methods). Two of these seven loci are found close to the *LCT* gene, which encodes the lactase enzyme that enables lactose consumption, and were found by previous studies^{6–10} to be associated with *Bifidobacterium*, possibly owing to its association with lactose consumption.

Notably, the *LCT* gene is the only case in which there was an overlap between the SNPs reported in any pair of the five previous studies^{5,6,8–10}. Moreover, no pair of previously reported SNPs from any two studies were within 100 kb of one another, or within 1 Mb of one another and associated with at least one bacterial taxa of the same phylum (Supplementary Table 15).

Microbiome–environment associations

We next investigated whether 24 pairs of related individuals—using second-to-fifth-degree relatives—with no history of household sharing had a similar microbiome composition, when compared to non-related pairs with no household sharing (using Bray–Curtis dissimilarity). We found no such evidence of similar microbiomes ($P > 0.4$, permutation testing; Methods; Extended Data Fig. 3, Supplementary Table 16). By contrast, when investigating 55 first-degree-relative pairs, who are likely to have a history of household sharing, we found significant similarities in their microbiomes at the genus and species taxonomic levels, and at the level of bacterial genes ($P < 5 \times 10^{-3}$, Methods; Extended Data Fig. 3 and Supplementary Table 16).

To test the effect of recent household sharing, we repeated the above analysis for 32 pairs of genetically unrelated individuals who reported sharing a household, and again found significant microbiome similarities at the level of species and of bacterial genes ($P < 2 \times 10^{-3}$; Extended Data Fig. 3 and Supplementary Table 16).

These results suggest that past or present household sharing may partly determine gut microbiome composition, whereas we find very little supporting evidence for microbiome similarities among relatives with no past household sharing. Our results corroborate previous studies²⁰, including a recent twin study that showed that the gut microbiomes of twins become more genetically dissimilar over time when living apart¹¹, and another study that showed that microbiome similarity among monozygotic twins compared to dizygotic twins is only marginally significant⁴ ($P = 0.032$ under an unweighted UniFrac dissimilarity, $P > 0.05$ under Bray–Curtis and weighted UniFrac dissimilarity).

We next directly quantified the fraction of microbiome β -diversity variance that can be inferred from environmental factors (based on Bray–Curtis dissimilarities) using answers to food frequency, lifestyle and drug use questions, as well as blood measurements, self-reported median daily intake of calories, protein, fat and carbohydrates, age, gender, glycaemic status, BMI, fasting glucose levels and the top 5 host genetic principal components (Methods; Supplementary Table 17).

We used a feature-selection algorithm that selected 95 environmental features that together could be used to infer 20.03% of the variance of microbiome β -diversity via PERMANOVA²¹ (Methods, $P < 0.002$; Extended Data Fig. 4 and Supplementary Table 18), consistent with previous studies^{13,14}. By contrast, host SNPs could not be used to infer a statistically significant fraction of β -diversity variance ($P = 0.11$, Methods).

Microbiome–phenotype associations

We next investigated how well host phenotypes can be inferred on the basis of the microbiome as compared to host genetics. The fraction of phenotypic variance that can be inferred from the microbiome after accounting for other explanatory variables including host genetics (that

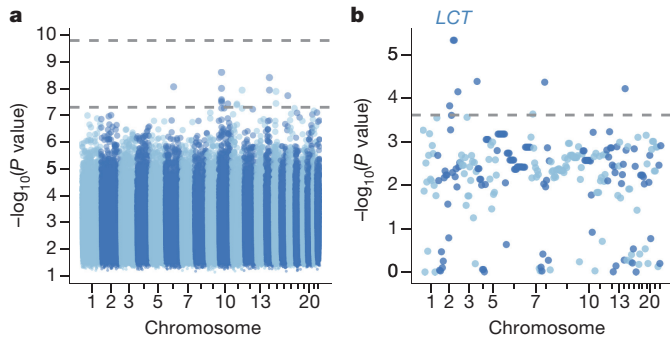


Figure 3 | Limited evidence for microbiome associations with specific SNPs. **a**, Manhattan plot showing the lowest P value obtained at every SNP tested for association with 313 taxa (computed by FaST-LMM³⁸ using $n = 814$ individuals) and with microbiome β -diversity (computed with a distance-based F test using $n = 715$ non-related individuals). The dashed lines represent a genome-wide significant P value corrected (top line) and not corrected (bottom line) for testing 313 taxa. **b**, The lowest P value obtained across 313 taxon association tests for each of 225 SNPs in 211 loci previously reported to be significantly associated with the microbiome^{5,6,8–10} (computed by FaST-LMM using $n = 814$ individuals). Seven SNPs are successfully replicated at $P < 0.05/211$ (dashed line; rs4988235, rs6730157, rs7656342, rs10112815, rs11626933, rs56006724 and rs7782745), two of which reside near the *LCT* gene.

is, b^2), represents a formal measure of predictability: larger b^2 values indicate that the microbiome is more informative with respect to a phenotype of interest. We estimate the value of b^2 in a narrow sense that cannot capture gene–gene or gene–environment interactions, as is common in heritability estimation²² (Supplementary Information).

Heritability estimation is typically performed in an LMM framework²³ and requires a kinship matrix, which is typically estimated from SNPs²⁴. We define the analogous bacterial kinship matrix on the basis of bacterial genes (Methods). We note, however, that the results may be confounded by unmeasured environmental factors, and that b^2 cannot be used to determine causality because microbiome composition can both affect and be affected by host phenotypes.

We used FIESTA²⁵ to verify that b^2 is consistently estimated in a more reliable fashion than h^2 for a given sample size, and that samples as large as 4,000 individuals are required for h^2 estimates to be as accurate as the b^2 estimate obtained in our cohort of 715 unrelated genotyped individuals (based on previous genetic data²⁶, Methods; Fig. 4a and Supplementary Table 19). We conclude that b^2 estimation is more accurate than h^2 estimation for a given sample size, and can be carried out with hundreds rather than thousands of individuals.

We next estimated b^2 for several phenotypes of interest (Extended Data Table 1), and used polygenic risk scores (PRS) based on summary statistics as an additional covariate to account for host genetic factors (Methods; Supplementary Table 20). We found 8 of the 12 traits we investigated to be significantly associated with the microbiome, after accounting for age, gender, diet and host genetics, with estimated b^2 levels of 36% for non-fasting HDL cholesterol levels, 36% for lactose consumption, 29% for waist circumference, 27% for hip circumference, 25% for BMI, 25% for glycaemic status, 24% for WHR and 22% for fasting glucose (Methods; Fig. 4b, c, Extended Data Fig. 5a, d, g, j and Supplementary Tables 21, 22). These b^2 estimates are comparable to previous SNP heritability estimates^{27–34} (Fig. 4b and Supplementary Table 21), which indicates that the microbiome is strongly associated with these traits.

To provide another comparison between host genetics and microbiome, we evaluated the ability of a linear prediction model (Methods) to predict human phenotypes from bacterial gene abundances, PRS, age, sex, as well as daily median caloric, carbohydrate, fat and protein consumption. The contribution of a specific data source to the phenotype can be assessed by the reduction in prediction power when

excluding this data source. We found that the prediction accuracy for 10 of the 12 traits we investigated—including BMI, HDL cholesterol and fasting glucose levels—is substantially improved when microbiome data is added to PRS (Fig. 4d, Extended Data Fig. 5b, c, e, f, h, i, k, l and Supplementary Table 23). Moreover, the contribution of both data sources is largely additive, consistent with our finding that microbiome and host genetics are largely independent of one another (Fig. 4e).

Taken together, these results demonstrate that host genetics and microbiome are complementary for predicting host phenotypes, and that phenotype prediction can be substantially improved by using both host genetics and microbiome data.

Independent validation on the LLD cohort

To verify that our results are broadly applicable across different study designs and populations, we repeated our analyses on a cohort of 836 Dutch individuals from the LLD cohort⁸, with metagenome-sequenced gut microbiomes, genotypes and the same covariates that we used in our analysis of the Israeli cohort.

In the LLD cohort, as in the Israeli cohort, there were no statistically significant associations between top host genetic principal components and microbiome PCOs (Supplementary Table 24), or between genetic kinship and microbiome β -diversity or bacterial taxa (Supplementary Table 25). A meta-analysis of both studies did not yield significant associations (Methods). Notably, the combined dataset is the largest cohort (1,551 individuals) of genotypes and metagenomic sequenced gut microbiomes analysed to date.

Next, we searched for significant associations between SNPs and microbiome β -diversity in the LLD cohort via the distance-based F test¹⁹ but could not replicate the two marginally-significant loci found in the Israeli cohort. Two loci had a genome-wide significance ($P < 5 \times 10^{-8}$) in the LLD cohort (Supplementary Table 26). One of these SNPs (rs4988235) is associated with lactase persistence, was replicated in the Israeli cohort ($P = 0.018$) and has previously been reported in microbiome–genetic association studies^{6–8,10}. This was also the only genome-wide significant SNP in a meta-analysis of the Israeli and LLD cohorts ($P = 2.2 \times 10^{-9}$), and is the only SNP reported by multiple studies as being associated with the gut microbiome. Thus, multiple lines of evidence suggest that rs4988235 is the SNP most strongly associated with gut microbiome composition.

Next, we found that all phenotypes that were significantly associated with the microbiome in the Israeli cohort, according to the b^2 measure, were also significantly associated in the LLD cohort, with the exception of WHR (Extended Data Fig. 5m and Supplementary Table 27). We performed phenotype predictions using the LLD cohort, and observed results highly similar to those obtained when using data from the Israeli cohort (Extended Data Fig. 5n, o and Supplementary Table 28).

Overall, we conclude that there is considerable congruency between the results obtained in the two cohorts, despite the substantial differences in the study populations, data collection method and processing pipeline. This demonstrates that our results are robust to diverse populations and experimental settings.

Discussion

In this study we used a range of statistical analyses across multiple cohorts, all of which led to the conclusion that the environment has a substantially greater role than host genetics in shaping the human gut microbiome. Several recent studies have reported that the microbiome is not only stable over time^{35,36} but also—to some extent—resilient to perturbations such as antibiotics and pathogens³⁷; the extent and determinants of such stability remain unresolved. As a small minority of heritable microbes are unlikely to generate this stability, it will be interesting to discover which mechanisms underlie microbiome stability and which perturbations cause the dysbiosis that can lead to disease susceptibility.

We proposed b^2 as a means of quantifying microbiome association with host phenotypes, and showed that b^2 can be reliably estimated

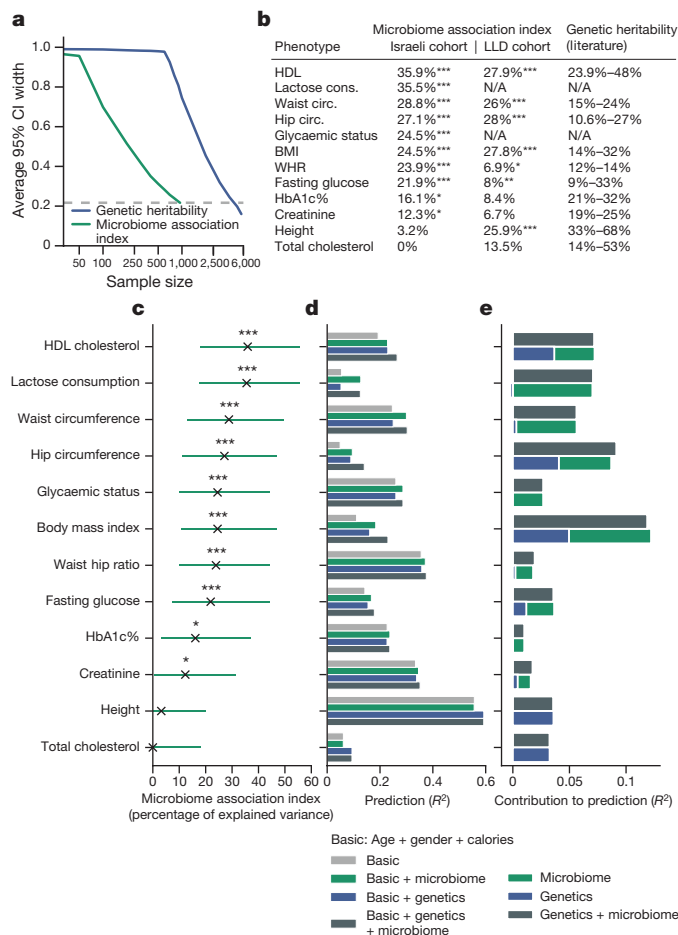


Figure 4 | The gut microbiome can be used to infer a significant fraction of the variance of several human phenotypes. **a**, For a given sample size, b^2 can be estimated more accurately than genetic heritability, as evaluated by using up to 946 individuals with gut microbiomes and up to 5,652 genotyped individuals from the Wellcome Trust Case Control Consortium 2 (ref. 26). Smaller 95% confidence interval (CI) widths indicate a greater confidence in the estimation. **b**, b^2 estimates from the analysis of 715 individuals with measured genotyped and gut microbiomes from the Israeli cohort (left column) and of 836 individuals from the LLD cohort (middle column) are comparable to previous genetic heritability estimates^{27–34} (right column). *FDR < 0.05, **FDR < 0.01 and ***FDR < 0.001. Cons., consumption, circ., circumference. **c**, b^2 estimates of several human phenotypes and their 95% confidence intervals, evaluated using 715 individuals. *FDR < 0.05, **FDR < 0.01 and ***FDR < 0.001. **d**, Phenotype prediction accuracy for 715 individuals, evaluated using a LMM under different sets of predictive features (measured using coefficient of determination (R^2)), using four different models for each phenotype: (i) ‘Basic’, age, gender and diet features; (ii) ‘Basic + microbiome’, basic features and relative abundances of bacterial genes; (iii) ‘Basic + genetics’, basic features and host genotypes; and (iv) ‘Basic + genetics + microbiome’: basic features, relative abundances of bacterial genes and host genotypes. **e**, The additive contribution of microbiome and genetics to prediction performance evaluated using a LMM across 715 individuals, over a model that includes only basic features. The joint contribution of microbiome and genetics is similar to the sum of the individual contributions, suggesting these are independent contributions.

using metagenomic cohorts of only hundreds of individuals; we then found that several phenotypes exhibit substantial b^2 levels, in the range of 22–36%. Finally, we showed that adding microbiome data to host genetics data improves prediction accuracy for several host phenotypes, and that the two data sources contribute additively. We note that b^2 should be interpreted with caution as it is a correlative measure and may be confounded by environmental factors.

Previous studies have identified heritable bacteria by observing co-occurrence among family members^{4–6,11}, or by reporting associations between specific SNPs and bacterial taxa^{5,6,8–11}. Our results are consistent with these published data, and collectively suggest that only a small number of bacterial taxa are likely to be strongly heritable, and that most SNP–bacteria associations are either weak or population-dependent. Our re-analysis of a recent study of twins⁶ estimates that the overall microbiome heritability lies between 1.9% and 8.1%. Future studies with larger sample sizes will probably identify additional heritable taxa and SNP associations, but are unlikely to change the overall conclusion that microbiome composition is predominantly shaped by non-genetic factors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 June 2017; accepted 16 January 2018.

Published online 28 February 2018.

- Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
- Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The impact of the gut microbiota on human health: an integrative view. *Cell* **148**, 1258–1270 (2012).
- Kurilshikov, A., Wijnenga, C., Fu, J. & Zhernakova, A. Host genetics and gut microbiome: challenges and perspectives. *Trends Immunol.* **38**, 633–647 (2017).
- Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).
- Turpin, W. *et al.* Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* **48**, 1413–1417 (2016).
- Goodrich, J. K. *et al.* Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* **19**, 731–743 (2016).
- Goodrich, J. K., Davenport, E. R., Clark, A. G. & Ley, R. E. The relationship between the human genome and microbiome comes into view. *Annu. Rev. Genet.* **51**, 413–433 (2017).
- Bonder, M. J. *et al.* The effect of host genetics on the gut microbiome. *Nat. Genet.* **48**, 1407–1412 (2016).
- Wang, J. *et al.* Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat. Genet.* **48**, 1396–1406 (2016).
- Blekhman, R. *et al.* Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* **16**, 191 (2015).
- Xie, H. *et al.* Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst.* **3**, 572–584 (2016).
- Zeevi, D. *et al.* Personalized nutrition by prediction of glycemic responses. *Cell* **163**, 1079–1094 (2015).
- Zhernakova, A. *et al.* Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
- Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
- Fu, J. *et al.* The gut microbiome contributes to a substantial proportion of the variation in blood lipids. *Circ. Res.* **117**, 817–824 (2015).
- Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466**, 238–242 (2010).
- Legendre P. & Legendre. L. *Numerical Ecology* Vol. 24, 3rd edn (Elsevier, 2012).
- Visscher, P. M. & Goddard, M. E. A general unified framework to assess the sampling variance of heritability estimates using pedigree or marker-based relationships. *Genetics* **199**, 223–232 (2015).
- Rühlemann, M. C. *et al.* Application of the distance-based F test in an mGWAS investigating β diversity of intestinal microbiota identifies variants in SLC9A8 (NHE8) and 3 other loci. *Gut Microbes* <https://doi.org/10.1080/19490976.2017.1356979> (2017).
- Song, S. J. *et al.* Cohabiting family members share microbiota with one another and with their dogs. *eLife* **2**, e00458 (2013).
- McArdle, B. H. & Anderson, M. J. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82**, 290–297 (2001).
- Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008).
- Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
- Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- Schweiger, R. *et al.* in *RECOMB 2017: Research in Computational Molecular Biology* (ed. Sahinalp, S.) 241–256 (Springer, 2017).
- Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.* **42**, 985–990 (2010).

27. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* **99**, 139–153 (2016).
28. Speed, D., Cai, N., Johnson, M. R., Nejentsev, S. & Balding, D. J. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).
29. Ge, T., Chen, C. Y., Neale, B. M., Sabuncu, M. R. & Smoller, J. W. Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet.* **13**, e1006711 (2017).
30. Zaitlen, N. *et al.* Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet.* **9**, e1003520 (2013).
31. Vattikuti, S., Guo, J. & Chow, C. C. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet.* **8**, e1002637 (2012).
32. Arpegård, J. *et al.* Comparison of heritability of cystatin C- and creatinine-based estimates of kidney function and their relation to heritability of cardiovascular disease. *J. Am. Heart Assoc.* **4**, e001467 (2015).
33. Xia, C. *et al.* Pedigree- and SNP-associated genetics and recent environment are the major contributors to anthropometric and cardiometabolic trait variation. *PLoS Genet.* **12**, e1005804 (2016).
34. Heckerman, D. *et al.* Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl Acad. Sci. USA* **113**, 7377–7382 (2016).
35. Antonopoulos, D. A. *et al.* Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. *Infect. Immun.* **77**, 2367–2375 (2009).
36. Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biol.* **12**, R50 (2011).
37. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
38. Widmer, C. *et al.* Further improvements to linear mixed models for genome-wide association studies. *Sci. Rep.* **4**, 6874 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Segal and Elinav group members for discussions; J. Goodrich for sharing the processed twins microbiome data with us; and participants and staff of the LifeLines DEEP cohort for their collaboration. S.C. thanks the Abisch–Frenkel Foundation. This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under awards 076113 and 085475. E.S. is supported by the Crown Human

Genome Center; the Else Kroener Fresenius Foundation; D. L. Schwarz; J. N. Halpern; L. Steinberg; and grants funded by the European Research Council and the Israel Science Foundation. E.E. is supported by Y. and R. Ungar, the Gurwin Family Fund for Scientific Research, the Leona M. and Harry B. Helmsley Charitable Trust, the Israel Science Foundation and the Helmholtz Foundation. E.E. holds the Sir Marc and Lady Tania Feldmann Professorial Chair in Immunology, is a senior fellow of the Canadian Institute for Advanced Research, and is an international scholar at the Bill and Melinda Gates Foundation and Howard Hughes Medical Institute. D.R. received a Levi Eshkol PhD Scholarship for Personalized Medicine by the Israeli Ministry of Science. LLD was made possible by grants from the Top Institute Food and Nutrition (GHO01) to C.W. C.W. is funded by a European Research Council (ERC) advanced grant (FP/2007-2013/ERC grant 2012-322698), a Netherlands Organization for Scientific Research (NWO) Spinoza prize (NWO SPI 92-266) and the Stiftelsen Kristian Gerhard Jebsen foundation (Norway). A.Z. holds a Rosalind Franklin Fellowship (University of Groningen), ERC starting grant (715772) and NWO Vidi grant (178.056). J.F. is funded by an NWO Vidi grant (NWO-VIDI 864.13.013). A.Z. and J.F. are also funded by CardioVasculair Onderzoek Nederland (CVON 2012-03).

Author Contributions D.R., O.W. and E.B. conceived the project, designed and conducted all analyses, interpreted the results, wrote the manuscript and are listed in random order. A.K., A.V.V., J.F., C.W. and A.Z. performed the analyses of the Dutch cohort and interpreted the results. T.K., D.Z. and A.W. designed protocols and supervised data collection. T.K., D.Z., P.I.C., A.G., I.N.K. and N.B. conducted microbiome analyses. S.S. and D.L. designed nutritional and drug databases. N.Z., M.P.-F., D.I. and Z.H. coordinated and supervised clinical aspects of data collection. N.K., G.M. and B.C.W. coordinated and designed data collection. T.A.-S., M.L.-P. and A.W. developed protocols and performed genotyping and microbiome sequencing. S.C. designed the genetic analyses. E.E. and E.S. conceived and directed the project and analyses, designed the analyses, interpreted the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.E. (eran.elinav@weizmann.ac.il) or E.S. (eran.segal@weizmann.ac.il).

Reviewer Information *Nature* thanks M. Georges and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Description of cohorts. This study used a previously described¹² cohort of individuals collected in Israel. Study participants were healthy individuals aged between 18 and 70 (for full inclusion and exclusion criteria, see previous description¹²). Before the study, participants answered medical, lifestyle and nutritional questionnaires. All participants were monitored by a continuous glucometer (CGM) for seven days. During this period, participants were instructed to record all daily activities, including standardized and real-life meals, in real-time using their smartphones. All participants were genotyped using Illumina OMNI-EXPRESS array and provided stool samples, which were collected using a swab (88% of the individuals) or an OMNIGENE-GUT (12% of the individuals; OMR-200; DNA Genotek) stool collection kit. The stool samples were metagenome-sequenced using Illumina NextSeq and HiSeq, and 16S rRNA gene sequenced as previously described¹². We validated that SNPs extracted from human reads in pre-filtered metagenomic sequences matched SNPs extracted from the blood of their human host. We further verified that the stool collection method did not confound our results by repeating all analyses using only stool samples collected via swab, which yielded results nearly identical to those obtained under the full dataset (results not shown).

The replication LifeLines DEEP cohort includes 1,539 individuals (636 males and 903 females, age range 18–84 years) from the north Netherlands. We included 836 participants in the analysis, after excluding related individuals, unhealthy individuals and individuals without genotype data or metagenomics sequencing data.

Genome-wide genotyping for the LLD participants was performed using Illumina HumanCytoSNP-12 and Immunochip arrays, and then imputed using Haplotype Reference Consortium server with HRC 1.0 panel. Metagenomics sequencing was performed using the Illumina HiSeq platform. Within 2 weeks of participants giving a blood sample, they collected faecal samples at home and stored them immediately at -20°C . After transport to the research laboratory on dry ice, faecal samples were stored at -80°C . Aliquots were made, and DNA was isolated with the AllPrep DNA/RNA Mini kit (Qiagen, 80204) with the addition of mechanical lysis. Reads were quality-filtered, and adapter removal was performed using Trimmomatic³⁹ (v.0.32). An average of 3.0 Gb of data (around 32.3 million high-quality reads) was obtained per sample. Reads belonging to the human genome were removed by mapping the data to the human reference genome (version NCBI37) with Bowtie2⁴⁰ (v.2.1.0). The profile of microbial composition was determined using MetaPhlan⁴¹ 2.2.

Genotypes preprocessing and imputation. We performed stringent quality control in our initial set of 862 genotyped individuals and 712,540 SNPs. We excluded SNPs with a missingness rate $>5\%$, Hardy-Weinberg $P < 10^{-9}$, minor allele frequency $<5\%$, $P < 0.01$ for differential missingness between two batches of individuals, or a logistic regression $P < 10^{-6}$ for separation of the two batches, yielding 545,325 SNPs for subsequent analyses. We additionally excluded individuals with $>10\%$ missing SNPs, leaving 833 individuals.

Genotypes were pre-phased using EAGLE2⁴² without a reference panel, and imputed using IMPUTE2⁴³ using the 1000 genomes dataset⁴⁴ and 128 Ashkenazi Jewish individuals⁴⁵ as reference panels. We retained only SNPs with imputation probability $>90\%$, and applied the filtering stages above to the imputed data, yielding 5,567,647 imputed SNPs.

Microbial preprocessing. Preprocessing of 16S rRNA gene sequencing was performed as previously described¹², with the addition of rarefaction to 10,000 reads. Weighted and unweighted UniFrac matrices for 16S rRNA gene sequencing samples were computed using QIIME via the beta_diversity script⁴⁶.

For metagenome analysis, we filtered metagenomic reads containing Illumina adapters, filtered low-quality reads and trimmed low-quality read edges. We detected host DNA by mapping with GEM⁴⁷ to the human genome (hg19) with inclusive parameters, and removed human reads. We subsampled all samples to have at most 10 million reads. Relative abundances from metagenomic sequencing were computed using MetaPhlan2⁴¹ with default parameters. MetaPhlan relative abundances were capped at a level of 10^{-4} . We removed individuals with <15 observed species from the analysis. After all genotyping and metagenomics quality-control steps, 946 individuals with metagenomics data remained, 814 of whom were genotyped. Unless stated otherwise, we additionally excluded individuals self-reported to share a household and related individuals (using up to three degrees of relationship) from the analysis, yielding $n = 715$ genotyped individuals (Supplementary Table 1).

When testing for associations between specific taxa and specific SNPs, we log-transformed the data and used only taxa present in at least 5% of individuals in our cohort, which left 7/18 (remaining/total) phyla, 13/28 classes, 17/43 orders, 35/96 families, 80/221 genera and 184/652 species.

Gene mapping. We performed gene mapping for the gene-based analyses by computing the length-normalized relative abundances of genes, obtained by similarity mapping with GEM to the gene reference catalogue⁴⁸ followed by abundance correction using an iterative algorithm based on Pathoscope⁴⁹, and normalization to sum to 1.0, using single-end reads.

Fasting glucose phenotyping. In the b^2 and phenotype prediction analyses, the fasting glucose phenotype was taken from data recorded by CGMs over the course of a week, as previously described¹². The median glucose measurement over a period of 30 min from self-reported wake-up time was used as a surrogate measure for fasting glucose.

Glycaemic status. For each patient we computed a quantity which we term 'glycaemic status' that can serve as an indicator of hyperglycaemia, based on HbA1c, fasting glucose, response to standardized meals¹², and top glucose percentiles and glucose noise as obtained from the CGM over the course of one week. Each individual was first ranked according to each feature. The glycaemic status of each individual was defined as the median of the ranks of (i) HbA1c; (ii) fasting glucose; (iii) median response to standardized meals; (iv) median of 90%, 95% and 98% glucose percentiles; and (v) glucose noise. We used fasting glucose summary statistics as a surrogate measure for the PRS of this measure.

Lactose consumption computations. We computed an estimate of average monthly lactose consumption (in grams), using a questionnaire of consumption frequency of 23 dairy products. As lactose consumption was exponentially distributed in our data, we log-transformed it to induce normality for the b^2 and phenotype prediction analyses.

Genetic kinship, principal components and relatedness estimation. We used PC-Relate⁵⁰ for estimating genetic kinship and PC-AiR⁵¹ for genetic principal components computation, as these tools are robust to the presence of relatedness and admixture. We used a filtered dataset of 75,384 SNPs in approximate linkage equilibrium ($r^2 < 0.15$), and ran an iterative estimation procedure (with the initial kinship estimates provided by KING-Robust⁵²) until the principal components computation converged, as previously described⁵³. We estimated the degree of relatedness between individuals using their kinship coefficient and previously proposed cutoffs⁵². In the analysis of the LLD cohort and when testing kinship-ancestry associations, we used the kinship matrix estimated by GCTA²⁴, as the kinship matrix of PC-Relate is by definition not associated with ancestry.

Mantel tests. Mantel tests used were performed with 100,000 permutations. When associating a matrix with a vector, we constructed a distance matrix for the vector using Euclidean distances. When not using covariates, we used a Spearman correlation-based Mantel test. In the presence of covariates, we performed a Pearson correlation-based partial Mantel test (performed by first regressing the matrix of covariate differences out of the two compared matrices, and then performing a Mantel test on the resulting residualized matrices), according to previous recommendations⁵⁴. When testing association with specific taxa, we excluded taxa present in $<5\%$ of individuals.

Ancestry proportions prediction. We attempted to predict ancestry proportions from microbiome composition using a variety of different techniques: Ridge regression⁵⁵, lasso regression⁵⁵ and extreme gradient boosting⁵⁶. We used as features either the top 100 PCOs of the Bray-Curtis dissimilarity matrix, the raw bacterial abundances (under various taxonomic levels) transformed to a logarithmic scale or the principal components of presence/absence of genes. Prediction accuracy was measured via a tenfold cross validation. The hyperparameters of the methods were determined in each fold via cross validation, using only the training set of each fold.

Microbiome principal coordinates prediction. We attempted to predict top microbiome PCOs from ancestry proportions or host genotypes via ridge regression, which is robust to high-dimensional data, with covariates used as additional explanatory variables. We computed P values via permutation testing with 10,000 permutations; in each permutation we assigned to each individual the microbiome PCOs of a random individual. The P value was defined as the number of permutations in which the sum of the coefficients of determination across the top 2, 5 or 10 PCOs was greater than that obtained under the non-permuted data.

Analysis of data from twins study. We estimated the overall microbiome heritability and the abundance of heritable taxa, using a previously published⁶ dataset of 2,252 twins. Our analysis is based on two principles: First, we define the 'overall microbiome heritability' as a weighted average of taxa-specific heritabilities. The weight of each taxa is determined by its relative abundance in the TwinsUK data, and its heritability estimate is taken from the previous analysis⁶. Second, we assume that only a subset of bacterial taxa is heritable. Therefore, we include only a subset of taxa in the weighted average computation, corresponding to taxa with heritability P values (as previously computed⁶) smaller than a given cutoff.

We considered two cutoffs: a 5% FDR cutoff, and a liberal 5% false-positive rate cutoff, with no multiple testing corrections. The first cutoff probably yields a subset

of the truly heritable taxa, and the second cutoff probably yields a subset with all heritable taxa as well as many non-heritable taxa. The heritability estimates using these two subsets therefore serve as lower and upper bounds, respectively, on the overall microbiome heritability estimate.

Given a subset of taxa considered as significantly heritable, we estimated the overall heritability by using a weighted average of the estimated heritabilities of operational taxonomic units (OTUs) associated with these taxa, weighted by the relative abundances of these OTUs. The resulting quantity was then averaged across individuals. The estimated heritability of an OTU was the maximal heritability estimate among all heritable taxa with which it was associated.

Testing for SNP–microbiome associations using the vegan package. We repeated previously proposed⁹ techniques for testing SNP–microbiome associations and estimating the variance inferred by several SNPs using the `envfit` and `ordiR2step` functions, respectively, in the `vegan` package in R⁵⁴.

Permutation testing for the fraction of genus β -diversity variance that can be inferred from the top 42 SNPs (corresponding to the number used in the previous analysis⁹) was carried out as follows. We performed 10,000 permutation analyses. In each analysis we (i) randomly assigned to each individual the genotype of a randomly selected individual; (ii) ranked all SNPs according to their association with microbiome β -diversity, using the `envfit` function; and (iii) estimated the fraction of β -diversity variance that can be inferred from the combined top 42 SNPs, using the `ordiR2step` function. The resulting P value was the fraction of permutations in which the fraction of inferred variance was greater than observed under the real data.

Testing for SNP associations with individual taxa. We tested for associations between individual bacteria and individual SNPs using `FaST-LMM`³⁸. We used all 814 genotyped individuals who passed quality control, including related individuals and individuals with a shared household, and controlled for these potential confounding sources using two variance components that encode kinship (as computed via `PC-Relate`⁵⁰) and household sharing (using a binary co-sharing covariance matrix). When testing each SNP, we used the covariates described earlier, as well as the top five genetic principal components, and a genetic kinship matrix based only on SNPs from other chromosomes, to avoid proximal contamination⁵⁷.

The abundance of bacteria present in at least 95% of individuals was encoded using the log-abundance; we excluded outlier individuals who were more than five standard deviations away from the mean. Otherwise, we dichotomized bacteria into presence/absence patterns and encoded the phenotype as a binary vector to prevent zero inflation, which leads to a bimodal distribution (LMMs handle binary phenotypes properly if the data are not ascertained⁵⁸).

Comparing results of different studies. We evaluated the consistency of previous association studies^{5,6,8–10} using the number of associations that are in the same locus (<100 kb apart) and associated with taxa belonging to the same phylum. We evaluated replication power by counting the number of SNPs in our own study with $P < 0.05/211$ (corresponding to 211 previously reported loci), using the closest imputed SNP to the reported one.

Relatives and household-sharing tests. We tested for significant microbiome sharing among related individuals or individuals sharing a household, by comparing their average Bray–Curtis dissimilarity to that of pairs with no family relation or household sharing using a permutation test with 100,000 permutations. In each permutation, we randomly divided the combined set of all pairs into two disjoint sets while preserving the original set sizes, and asked whether the mean difference in Bray–Curtis dissimilarity between individuals in the two sets was greater than the difference observed in the real data. To prevent confounding effects, we considered only individuals whose stool was collected with a swab (one of the two stool collection methods).

Associating environmental factors with the microbiome. We tested for associations between 201 environmental factors and microbiome β -diversity at the species level with `PERMANOVA`²¹, using data from self-reported questionnaires¹² (Supplementary Table 17). To quantify the fraction of microbiome variance that could be inferred from environmental factors in combination we performed a greedy stepwise algorithm, in which at each iteration we added the environmental factor that contributed the greatest fraction of inferred variance to factors added in previous iterations. Before adding each factor, we permuted it 100 times and verified that its contribution was greater than in at least 55% of these permutations. If not, we stopped the algorithm. The statistical significance of the resulting estimate was evaluated using a permutation testing with 100,000 permutations, in which for each permutation we assigned all 201 environmental factors of each individual to a random individual, and then reran the entire analysis (including the feature selection procedure).

To perform the above procedure with SNPs instead of environmental factors, we first selected a set of SNPs in approximate linkage equilibrium that are maximally associated with the microbiome β -diversity, and then performed the analysis using these SNPs. Specifically, we first sorted the SNPs according to their fraction of

inferred β -diversity variance, and then iteratively selected 201 top-ranking SNPs (corresponding to the number of environmental variants) that are not within 200 kb of a previously selected SNP. We then reran the `PERMANOVA` analysis with the selected SNPs.

Computing polygenic risk scores. PRSs were computed using $\hat{y}_i = \sum_{j \in R(c)} x_i^j \hat{b}^j$, in which \hat{y}_i is the predicted phenotype, x_i^j is SNP j of individual i , \hat{b}^j is the effect of SNP j reported in summary statistics and $R(c)$ is the set of SNPs found in both the genotyping array and in the summary statistics with $P < c$ for the cutoff c . The optimal c value was selected by searching over the grid $[10^0, 3 \times 10^{-1}, 10^{-1}, \dots, 10^{-8}]$ and finding the value maximizing the Spearman correlation between the true and predicted phenotypes. To prevent overfitting, the value of c used to compute the PRS of every individual was estimated using a subset of 90% of the data that did not include this individual. Similarly, when performing phenotype prediction, we estimated c using only individuals in the training set. SNPs were normalized to have a unit variance, according to their reported allele frequency. We used the original rather than the imputed set of SNPs, as we empirically verified that using the imputed set of SNPs in conjunction with linkage disequilibrium pruning did not improve prediction results. The list of summary statistics used is provided in Supplementary Table 20.

Construction of a kinship matrix based on microbial genes. We encoded the bacterial kinship of individuals i, j using $\sum_k g_i^k g_j^k / n$, in which k iterates over all genes present in >1% of individuals, g_i^k is the presence/absence indicators of gene k in individual i (using a relative abundance cutoff of 10^{-6} , and normalized to have a zero mean and a unit variance), and n is the number of genes (1,360,337).

Microbiome-association index estimation. Microbiome-association index (b^2) was estimated using `GCTA`²⁴, a tool used in statistical genetics for estimating SNP-based genetic kinship. Instead of a matrix of host SNPs, as is commonly used in `GCTA`, we used a microbial genes-based kinship matrix. For all phenotypes (except lactose consumption), the covariates included the PRS of the investigated phenotype, the covariates described earlier and the top five genetic principal components. In the analysis of lactose consumption, we replaced the PRS with the SNPs rs4988235 and rs182549, which largely explain the genetic component of lactase persistence in European populations⁵⁹. P values were computed using `RL-SKAT`⁶⁰ and confidence intervals were computed using `FIESTA`²⁵. Outlier individuals with phenotypes more than five standard deviations away from the mean were excluded from the analysis.

We defined b^2 estimation accuracy for a given covariance matrix using the average width of 95% confidence intervals (assuming that b^2 is uniformly distributed in $[0,1]$). We estimated this quantity by invoking `FIESTA` 100 times with 100 different b^2 values evenly spaced in the interval $[0,1]$ and averaging the resulting 95% confidence interval widths.

Analysis of data from the Wellcome Trust. We computed confidence intervals for genetic heritability estimation using 5,652 previously described²⁶ control individuals from the Wellcome Trust National Blood Service and 1958 birth cohorts. SNPs with >0.5% missing data, $P < 0.01$ for allele frequency difference between the two groups, $P < 0.000005$ for deviation from Hardy–Weinberg equilibrium or minor allele frequency <1% were removed. The genetic kinship matrix was computed using `GCTA`²⁴ and confidence intervals were estimated using `FIESTA`²⁵.

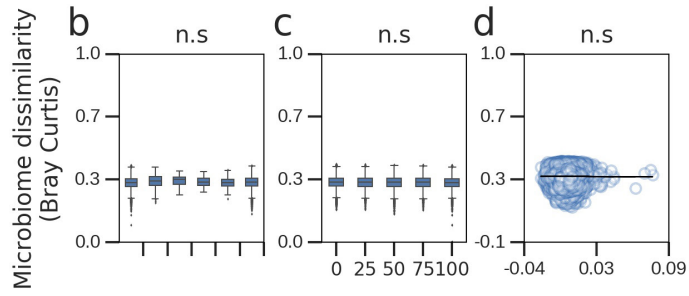
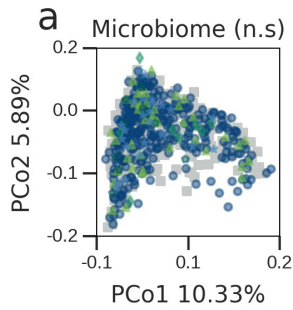
Phenotype prediction. Phenotype prediction was performed with an `LMM`⁶¹, using a kinship matrix based on presence/absence of genes, constructed as described in ‘Construction of a kinship matrix based on microbial genes’ (LMMs are mathematically equivalent to a ridge regression model that uses the principal components of the kinship matrix as covariates, and they reduce to linear regression when not using a kinship matrix). The covariates included age, sex, and daily median caloric, carbohydrate, fat and protein consumption. In some experiments we additionally included covariates for host genetic effects, represented either as PRS (for all phenotypes except lactose consumption) or as the SNPs rs4988235 and rs182549 for lactose consumption. Prediction performance was evaluated using a tenfold cross validation. Outlier individuals with phenotypes more than five standard deviations away from the mean were excluded from all analyses. We also evaluated additional types of kinship matrices: (i) a β -diversity matrix, which we transformed to a kinship matrix as previously described⁶²; and (ii) kinship matrices based on relative abundances or presence/absence of bacterial taxa instead of genes (Supplementary Table 23).

Meta-analysis. Meta-analysis of the Israeli and LLD cohorts was performed using `Stouffer's method`.

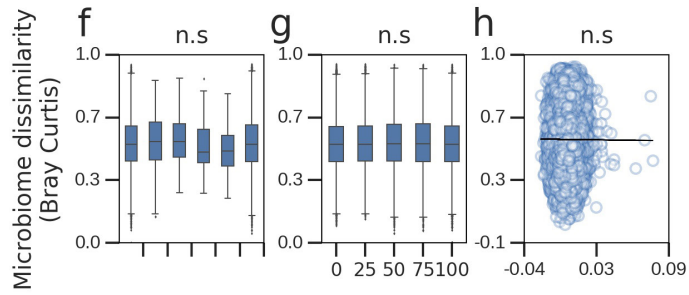
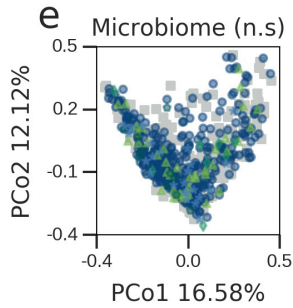
Israeli cohort. The Israeli cohort study was approved by Tel Aviv Sourasky Medical Center Institutional Review Board, approval numbers TLV-0658-12, TLV-0050-13 and TLV-0522-10; Kfar Shaul Hospital Institutional Review Board, approval number 0-73; and Weizmann Institute of Science Bioethics and Embryonic Stem Cell Research oversight committee. The study was reported to <http://clinicaltrials.gov/>, NCT number: NCT01892956.

- Data availability.** The accession number for the datasets analysed in this paper are: (i) Israeli metagenome and 16S rRNA gene sequences, European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena>): PRJEB11532; (ii) TwinsUK 16S rRNA gene sequences, ENA: ERP015317; (iii) LifeLines DEEP sequencing data, European Genome-phenome Archive (EGA; <https://www.ebi.ac.uk/ega/>): EGAS00001001704; and (iv) Wellcome Trust 2 genotypes, EGA: EGAD00000000021 and EGAD00000000023. All relevant data are available from the corresponding authors upon reasonable request. Source data for Fig. 3a and Extended Data Figs 1, 2 is available from http://genie.weizmann.ac.il/genomica_links.html.
39. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
 40. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
 41. Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814 (2012).
 42. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
 43. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
 44. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
 45. Carmi, S. *et al.* Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat. Commun.* **5**, 4835 (2014).
 46. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
 47. Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* **9**, 1185–1188 (2012).
 48. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
 49. Hong, C. *et al.* PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome* **2**, 33 (2014).
 50. Conomos, M. P., Reiner, A. P., Weir, B. S. & Thornton, T. A. Model-free estimation of recent genetic relatedness. *Am. J. Hum. Genet.* **98**, 127–148 (2016).
 51. Conomos, M. P., Miller, M. B. & Thornton, T. A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet. Epidemiol.* **39**, 276–293 (2015).
 52. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
 53. Conomos, M. P. *et al.* Genetic diversity and association studies in US Hispanic/Latino populations: applications in the Hispanic Community Health Study/Study of Latinos. *Am. J. Hum. Genet.* **98**, 165–184 (2016).
 54. Oksanen, J. *et al.* vegan: community ecology package. <https://cran.r-project.org/web/packages/vegan/index.html> (2017).
 55. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009).
 56. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016).
 57. Listgarten, J. *et al.* Improved linear mixed models for genome-wide association studies. *Nat. Methods* **9**, 525–526 (2012).
 58. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).
 59. Ingram, C. J. E., Mulcare, C. A., Itan, Y., Thomas, M. G. & Swallow, D. M. Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.* **124**, 579–591 (2009).
 60. Schweiger, R. *et al.* RL-SKAT: an exact and efficient score test for heritability and set tests. *Genetics* **207**, 1275–1283 (2017).
 61. de los Campos, G., Vazquez, A. I., Fernando, R., Klimentidis, Y. C. & Sorensen, D. Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet.* **9**, e1003608 (2013).
 62. Zhao, N. *et al.* Testing in microbiome-profiling studies with MiRKAT, the microbiome regression-based kernel association test. *Am. J. Hum. Genet.* **96**, 797–807 (2015).

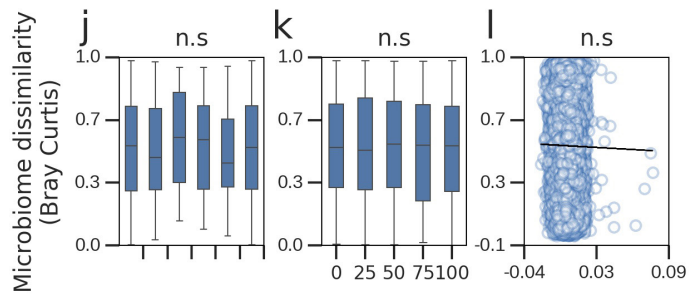
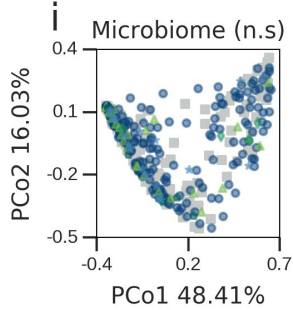
Gene level



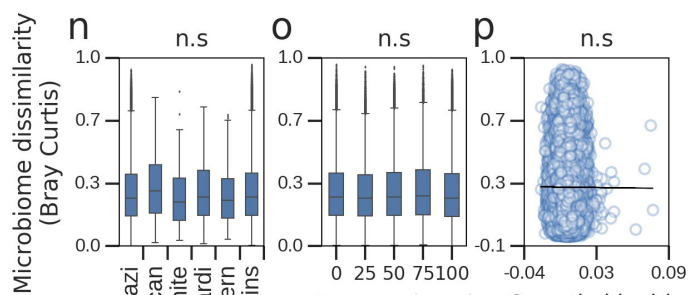
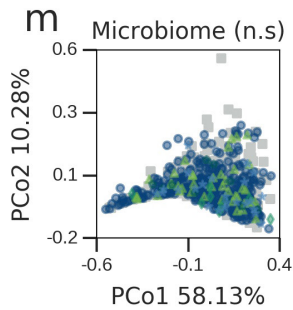
Genus level



16S Genus level



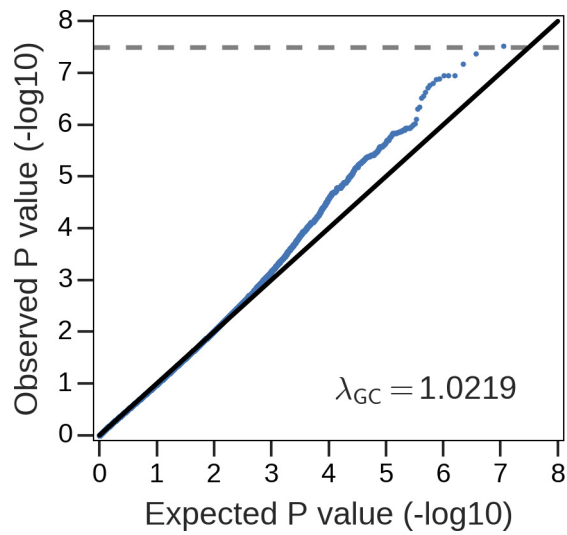
Phylum level



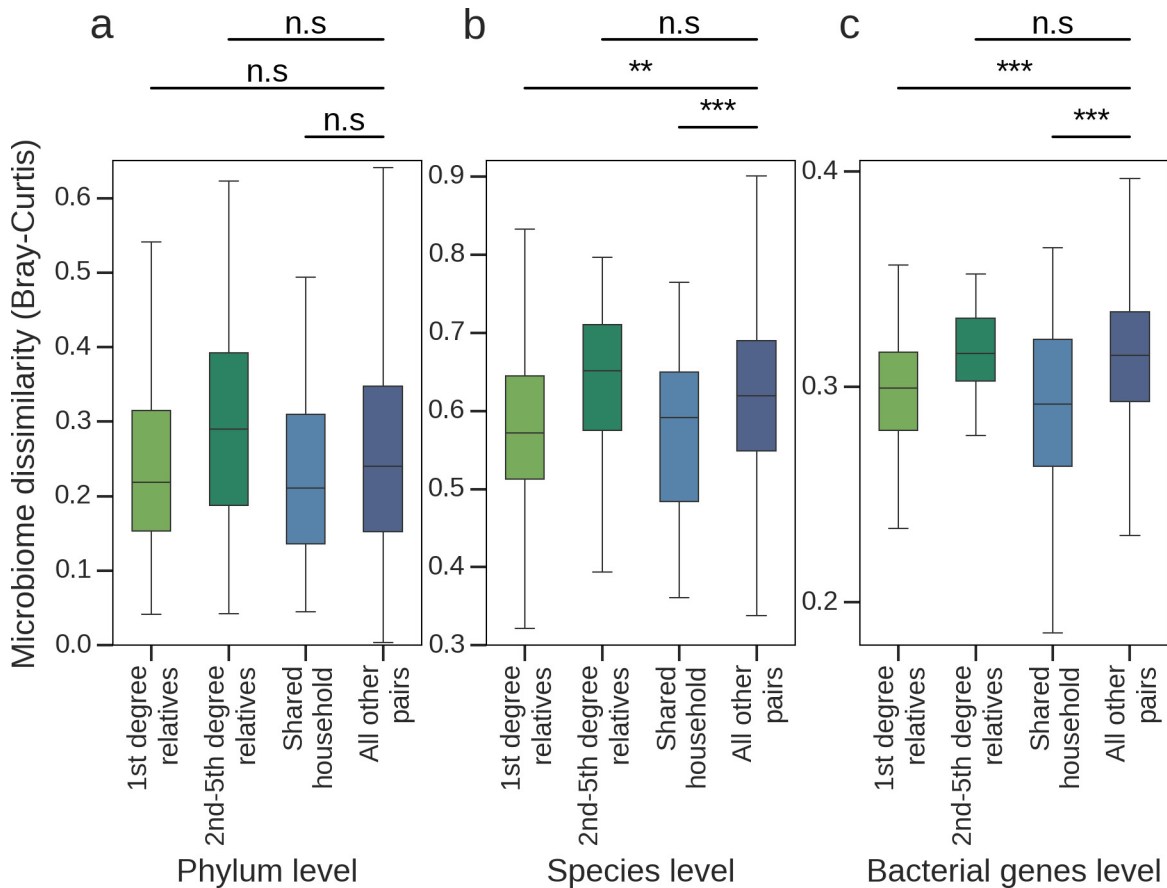
- Ashkenazi
- ▲ North African
- ◆ Yemenite
- Sephardi
- ★ Middle Eastern
- Other

Extended Data Figure 1 | Limited evidence for microbiome associations with genetic ancestry or kinship across multiple functional and taxonomic levels. a–p. Each row is similar to Figs 1b, d–e, 2b, but is based on the abundance of bacterial genes (a–d), genera (e–h), genera based on 16S rRNA gene sequencing data (i–l) or phyla (m–p). a, d, e, h, m, p, n = 715 genotyped individuals; i, l, n = 481 individuals

with 16S rRNA gene sequencing data; b, f, n, n = 737 individuals for whom the ancestries of all grandparents are known; j, n = 509 individuals with 16S rRNA gene sequencing data for whom the ancestries of all grandparents are known; c, g, o, n = 946 individuals; and k, n = 650 individuals with 16S rRNA gene sequencing data.

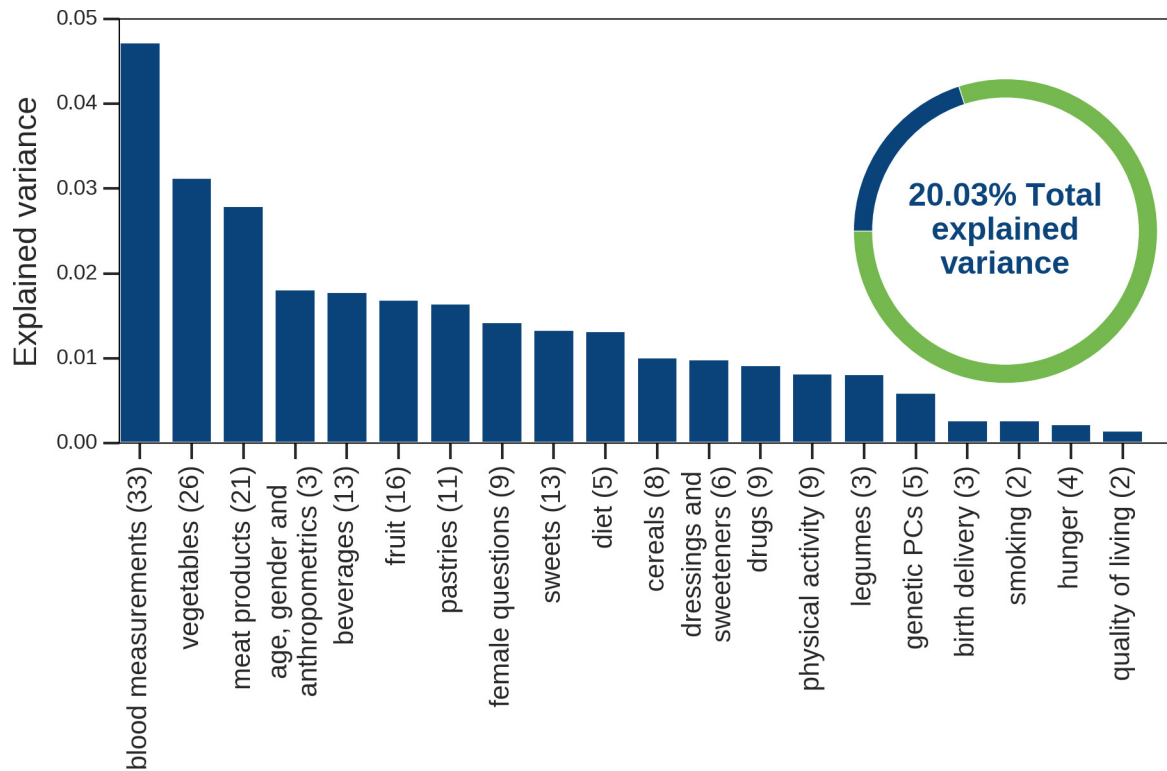


Extended Data Figure 2 | Limited evidence for associations between microbiome β -diversity and specific SNPs. The quantile–quantile plot shows that only two SNPs are significantly associated with microbiome β -diversity at $P < 5 \times 10^{-8}$, computed using a distance-based F test with $n = 715$ unrelated genotyped individuals. λ_{GC} , genomic inflation factor.



Extended Data Figure 3 | Individuals who share a household at present or have shared one in the past have significantly similar microbiomes. First-degree relatives and individuals with present household sharing have significantly similar species and bacterial gene abundances ($P < 0.01$; permutation testing). **a–c**, Box plots depict the distribution of Bray–Curtis dissimilarities across pairs of individuals at the phylum (**a**), species (**b**) and bacterial genes (**c**) level. Each panel shows the Bray–Curtis dissimilarities among all pairs of (i) first-degree relatives, who are likely to

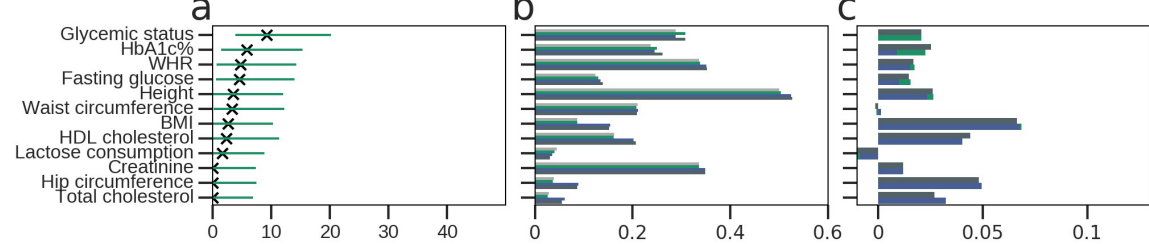
have experienced present or past household sharing ($n = 55$ pairs); (ii) second-to-fifth-degree relatives, who are unlikely to have experienced present or past household sharing ($n = 24$ pairs); (iii) unrelated individuals self-reported to currently share a household ($n = 32$ pairs); and (iv) all other individuals ($n = 255,891$ pairs). The lower and upper limits of the boxes represent the 25% and 75% percentiles, respectively, and the top and bottom whiskers represent the 5% and 95% percentiles, respectively. The P value ranges for all panels are: ** $P < 0.01$ and *** $P < 0.005$.



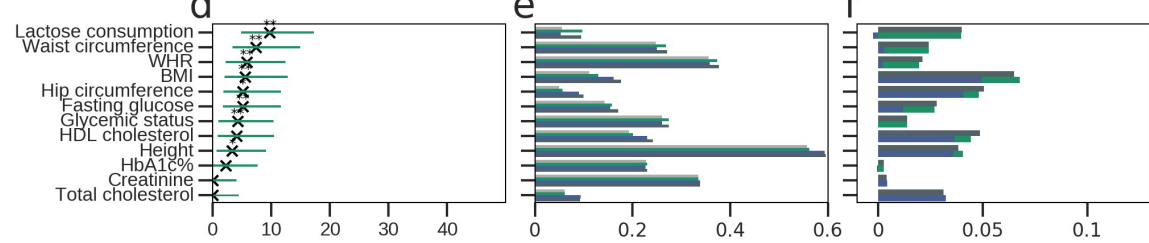
Extended Data Figure 4 | The gut microbiome is significantly associated with multiple environmental factors. The fraction of variance of the microbiome β -diversity matrix that can be inferred from different categories of environmental factors is shown. $n = 715$ individuals (Supplementary Table 17); numbers in parentheses indicate the number of

features in each category. The fraction of inferred variance can reflect both the information that the category conveys on the microbiome as well as the number of factors in the category, which depends on the questionnaire used in the study.

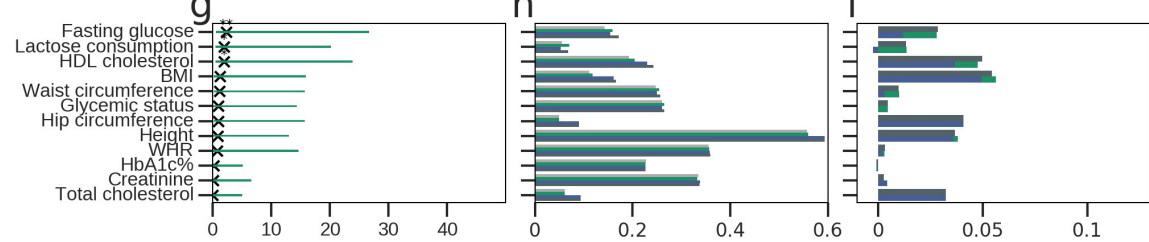
16S Genus level



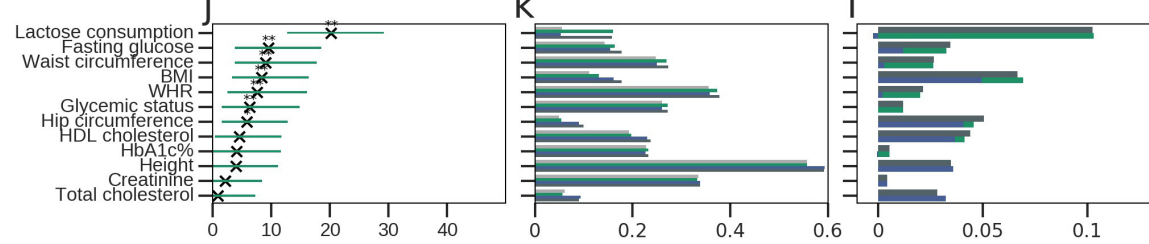
Genus level



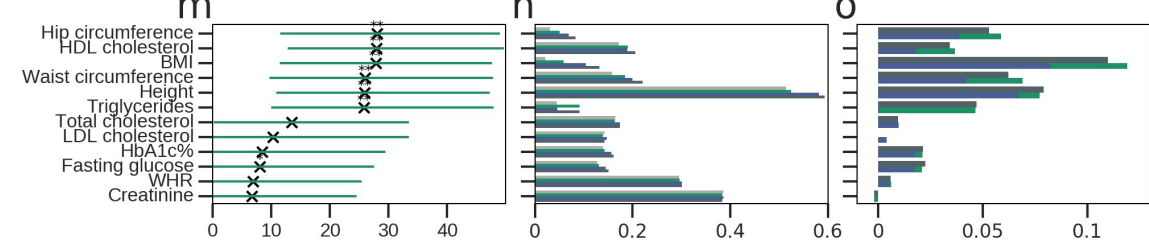
Phylum level



Species level



LLD genes level



Microbiome-association index (% explained variance)

Prediction (R²)

Contribution to prediction (R²)

Basic: Age+Gender+Calories

Basic

Basic+Microbiome

Basic+Genetics

Basic+Genetics+Microbiome

Microbiome

Genetics

Genetics+Microbiome

Extended Data Figure 5 | b² estimates and phenotype prediction results when using various data sources. Each row is similar to Fig. 4c–e, but is based on a different data source. **a–c**, Relative abundance of genera, obtained from 16S rRNA gene sequencing (using *n* = 464 individuals). **d–f**, Relative abundance of genera, obtained from metagenomic sequencing (using *n* = 715 individuals). **g–i**, Relative abundance of phyla (using *n* = 715 individuals). **j–l**, Relative abundance of species (using

n = 715 individuals). **m–o**, Relative abundance of bacterial genes in the LLD cohort (using *n* = 836 individuals). Note that two phenotypes that were analysed in the Israeli cohort (lactose consumption and glycaemic status) were not available for the LLD cohort, and two phenotypes available for the LLD cohort and shown here (LDL cholesterol and triglycerides) were not available for the Israeli cohort. The *P* value ranges for all panels are: *FDR < 0.05, **FDR < 0.01 and ***FDR < 0.001.

Extended Data Table 1 | Baseline characteristics of the cohort

Property	Value (mean±SD)
Number of participants	1,046
Age	42.6±12.7
Gender (%female)	61%
Median calories consumed daily (kcal)	1867±530
Median carbohydrates consumed daily (gr)	210.7±66.1
Median fat consumed daily (gr)	74.2±27.6
Median protein consumed daily (gr)	69.1±24.8
BMI	26.6±5.2
Waist circumference (cm)	87.7±13.6
Hip Circumference (cm)	104.6±13.6
Waist-hip ratio	0.84±0.09
Height (cm)	167±9.1
Total cholesterol (mg/dl)	186±37
HDL cholesterol (mg/dl)	58.1±17.1
HbA1c%	5.44±0.47
Fasting glucose (mg/dl)	92.05±11.3
Creatinine (mg/dl)	0.84±0.19
Lactose consumption (gr)	562±604

The mean and standard deviation of all properties used as covariates or as investigated phenotypes are shown. Dietary properties are based on information recorded in real time by study participants on their smartphones (see Methods).

Extended Data Table 2 | No significant association between ancestral or genetic similarity and the gut microbiome

		Ancestry (categorical)	Ancestry (proportions)	Genetic kinship
Host genetics (control)	Kinship	<0.0001	<10 ⁻⁵	-
Microbiome	β-diversity	>0.06	0.73	0.59
	α-diversity	0.57	0.22	0.61
	Specific taxa	>0.05	>0.05	>0.05

Each cell contains the *P* value of a single or of multiple statistical tests, testing whether individuals who are more similar according to ancestry or genetic kinship (in columns) are also more similar according to (i) microbiome β-diversity (using Bray–Curtis dissimilarity); (ii) microbiome α-diversity (using Shannon diversity); (iii) abundance of specific taxa; or (iv) genetic kinship (in rows). The first column includes *n* = 582 non-admixed individuals, the second includes 946 individuals and the third includes 715 unrelated genotyped individuals. *P* values in the first column are based on Kruskal–Wallis tests (using the top 5 microbiome PCOs for Bray–Curtis dissimilarity, and the top 5 genetic principal components for genetic kinship); *P* values in the other columns are based on Mantel tests (Methods).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

▶ Experimental design

1. Sample size

Describe how sample size was determined.

As this is an observational study based on voluntary participation, we used all the samples that were collected throughout the study collection phase. We verified that our sample was well powered to answer our research questions using extensive power analyses, as described in detail in the Supplementary Information.

2. Data exclusions

Describe any data exclusions.

Exclusion criteria consisted of: (i) pregnancy; (ii) usage of antibiotics within three months prior to participation; (iii) chronically active inflammatory or neoplastic disease in the three years prior to enrollment; (iv) chronic gastrointestinal disorder, including Inflammatory Bowel Disease and Celiac disease; (v) skin disease, including contact dermatitis, precluding proper attachment of the continuous glucose monitor; (vi) active neuropsychiatric disorder; (vii) myocardial infarction or cerebrovascular accident in the 6 months prior to participation; (viii) chronic immunosuppressive medication usage; (ix) pre-diagnosed type I or type II diabetes mellitus.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Experimental replication was performed by repeating all analyses on an independent cohort, namely the LifeLines-DEEP cohort.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

As this is an observational study, randomization is not relevant and was not employed. We controlled for potential sources of confounding by (a) excluding participants who are at least third degree relatives with another participants (unless stated otherwise for specific analyses); (b) excluding participants self-reported to share a household with another participant (unless stated otherwise for specific analyses); (c) including covariates encoding gender, age, stool collection method, self-reported daily median caloric, fat, protein and carbohydrates consumption, and the top five principal component of the host genotypes (unless stated otherwise for specific analyses); (d) when testing for associations between specific SNPs and specific taxa, we used a linear mixed model, which controls for confounding due to population structure or subtle genetic relatedness.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

As this is an observational study, blinding is not relevant and was not employed.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

Trimmomatic (v.0.32)
Bowtie2 (v.2.1.0)
MetaPhlan 2.2
RL-SKAT 1.0
FIESTA 1.0
QIIME 1.9.1
Pathoscope 2.0
R GENESIS package 2.4.0
Python package FaST-LMM 0.1
GCTA 1.26.0
plink 1.90b3.44
KING-robust 2.0
R vegan package 2.4-2
Python package scikit-learn 0.18.2
Python package scipy 0.19.1
R DBF.test function (https://wwwf.imperial.ac.uk/~gmontana/software/dbf/dbf_test.R)
Impute 2 2.3.2
Eagle 2 2.3.2
Python package xgboost 0.6

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No eukaryotic cell lines were used

▶ **Animals and human research participants**Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Study participants were healthy individuals aged 18–70 able to provide informed consent and operate a glucometer. Anthropometric and blood pressure measurements were taken by a CRA or a certified nurse, as well as a blood test and weights.