



**HAL**  
open science

## Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography

Mélanie Deschasaux, Kristien Bouter, Andrei Prodan, Evgeni Levin, Albert K. Groen, Hilde Herrema, Valentina Tremaroli, Guido Bakker, Ilias Attaye, Sara-Joan Pinto-Sietsma, et al.

► **To cite this version:**

Mélanie Deschasaux, Kristien Bouter, Andrei Prodan, Evgeni Levin, Albert K. Groen, et al.. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nature Medicine*, 2018, 24 (10), pp.1526-1531. 10.1038/s41591-018-0160-1 . hal-02276326

**HAL Id: hal-02276326**

**<https://hal.science/hal-02276326>**

Submitted on 25 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography**

Authors: Mélanie Deschasaux (1), Kristien E. Bouter (2), Andrei Prodan (2), Evgeni Levin (2), Albert K. Groen (2), Hilde Herrema (2), Valentina Tremaroli (3), Guido J. Bakker (4), Ilias Attaye (4), Sara-Joan Pinto-Sietsma (1,4), Daniel H. van Raalte (2), Marieke B. Snijder (1,5), Mary Nicolaou (5), Ron Peters (6), Aeilko H. Zwinderman (1), Fredrik Bäckhed (3,7), Max Nieuwdorp (2,3,4)

### Affiliations:

(1) Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam, The Netherlands;

(2) Amsterdam Diabetes Center, Department of Internal Medicine, Academic Medical Center-VU Medical Center, Amsterdam, The Netherlands;

(3) Wallenberg Laboratory, Sahlgrenska University of Gothenburg, Gothenburg, Sweden;

(4) Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands

(5) Department of Public Health, Academic Medical Center, Amsterdam, The Netherlands

(6) Department of Cardiology, Academic Medical Center, Amsterdam, The Netherlands

(7) Novo Nordisk Foundation Center for Basic Metabolic Research, Section for Metabolic Receptology and Enteroendocrinology, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark.

**Correspondence to:** M. Deschasaux ([m.c.deschasaux@amc.uva.nl](mailto:m.c.deschasaux@amc.uva.nl)) or M. Nieuwdorp ([m.nieuwdorp@amc.uva.nl](mailto:m.nieuwdorp@amc.uva.nl)).

## INTRODUCTORY PARAGRAPH

Trillions of microorganisms inhabit the human gut and are regarded as potential key factors for health<sup>1,2</sup>. Characteristics such as diet, lifestyle or genetics can shape the composition of the gut microbiota<sup>2-6</sup> and are usually shared by individuals from comparable ethnic origin. So far, most studies assessing how ethnicity relates to the intestinal microbiota compared small groups living at separate geographical locations<sup>7-10</sup>.

Here we show, using fecal 16S rRNA gene sequencing in 2084 participants of the HELIUS study<sup>11,12</sup>, that individuals living in the same city tend to share similar gut microbiota characteristics with others of their ethnic background. Ethnicity contributed to explain the inter-individual dissimilarities in gut microbiota composition, with 3 main poles primarily characterized by OTUs classified as *Prevotella* (Moroccans/Turks/Ghanaians), *Bacteroides* (African Surinamese/ South-Asian Surinamese) and Clostridiales (Dutch). The Dutch exhibited the greatest gut microbiota alpha-diversity and the South-Asian Surinamese the smallest, with corresponding enrichment/depletion in numerous OTUs. Ethnic differences in alpha-diversity and inter-individual dissimilarities were independent of metabolic health and only partly explained by ethnic-related characteristics including sociodemographics, lifestyle or diet. Hence, the ethnic origin of individuals may be an important factor to consider in microbiome research and its potential future applications in ethnic-diverse societies.

## MAIN

The gut microbiota composition is unique to every individual but shaped by common factors including diet, lifestyle, medication, early-life determinants, environment and genetics<sup>2-5</sup>, most of these likely influenced by ethnicity. To date, variations in gut microbiota composition across ethnicities were observed in small groups, mostly comparing individuals living in different geographical areas with marked lifestyle differences<sup>7,13,14</sup>. We therefore aimed to depict the fecal microbiota composition in individuals with varied ethnic backgrounds but living in the same city. This has not previously been studied in such a large cohort of adults<sup>7,8,10</sup>.

The HELIUS study enrolled ~25,000 adults from the six largest ethnic groups living in Amsterdam, NL<sup>12</sup>. We included 2084 participants who donated a feces sample at inclusion: 439 Dutch, 367 Ghanaians, 280 Moroccans, 197 Turks, 443 African Surinamese and 358 South-Asian Surinamese (**Methods, Supplementary Table 1**). The fecal microbiota composition was profiled by sequencing the 16S rRNA gene, V4 region (**Methods**).

We found that ethnicity significantly contributed to the inter-individual dissimilarities in OTU-level gut microbiota composition (beta-diversity; Bray-Curtis index, PERMANOVA,  $P=0.001$ ). Main variations were visualized using principal coordinate analysis (PCoA) (**Figure 1**). Although no clear separation was observed, individuals sharing the same ethnicity tended to group together, suggesting that they shared a more similar gut microbiota. In this respect, Dutch, South-Asian Surinamese and Moroccans were the most discriminant groups (three main poles), while Ghanaians appeared more spread over the graph, probably resulting from similarities with Moroccans/Turks or African Surinamese, likely echoing similarities in diet and ancestries (**Supplementary Figure 1**).

Similarly, gut microbiota alpha-diversity differed between ethnic groups (ANOVA, richness:  $R^2=0.19$ , Shannon index:  $R^2=0.13$ , Faith's PD:  $R^2=0.17$ , all  $P<2.2\times 10^{-16}$ , **Supplementary**

**Table 3A**) and was greatest in the Dutch and smallest in the South-Asian Surinamese, while Ghanaians, Turks and African Surinamese ranked in between (**Figure 2**). Moroccans exhibited high richness and phylogenetic diversity but an uneven distribution of OTUs (intermediate Shannon index).

Analyses of features characterizing the gut microbiota across ethnicities showed that Firmicutes was the most abundant phylum in all groups, followed by Bacteroidetes, Actinobacteria and Proteobacteria. Firmicutes and Bacteroidetes were enriched and depleted, respectively, in the Dutch, while Actinobacteria was enriched in the South-Asian Surinamese (**Supplementary Table 2C**). *Prevotella*, *Faecalibacterium* and *Bacteroides* were the dominant genera overall (**Supplementary Figure 3, Supplementary Table 2B.1-2**). *Bacteroides* and *Prevotella* were respectively enriched and depleted in Surinamese, while we observed the reverse in Moroccans and Turks, and comparable proportions in the Dutch. In line, the top two contributive OTUs to inter-individual dissimilarities (**Figure 1, Supplementary Table 2A.1**), Otu1\_*Prevotella copri* ( $R^2=0.23$ ) and Otu5\_*Bacteroides* ( $R^2=0.11$ ), exhibited clear gradients in opposite directions (**Supplementary Figure 4**). Overall, 21 OTUs were shared by more than 95% of individuals (“core”), irrespective of their ethnicities. These OTUs pertained to 10 genera previously identified as “core”<sup>5</sup> and support the idea of a shared and conserved gut microbiota for all individuals. Nevertheless, when defined separately for each ethnic group, the 95% core OTUs exhibited ethnic-specific profiles (**Figure 3**).

Ethnicity was a significant factor in the variation of the relative abundance of 559/744 OTUs (Kruskal-Wallis tests, Bonferroni correction,  $P<6.7\times 10^{-5}$ ). Consistent with observations made for dissimilarities and alpha-diversity, Dutch, Moroccans and South-Asian Surinamese exhibited more differentially abundant OTUs (Wilcoxon-Mann-Whitney tests, Bonferroni correction,  $P<3.2\times 10^{-6}$ ; **Supplementary Table 2A.1-2**), with a majority of enriched OTUs in

Dutch (266/287) and Moroccans (142/185), and of depleted OTUs in South-Asian Surinamese (189/215).

We then explored several characteristics potentially underlying the ethnicity-gut microbiota associations. With a primary focus on diet, we used detailed information collected through food frequency questionnaires on ~50% participants (n=1032), except Ghanaians (**Supplementary Table 1B, Supplementary Figure 5**), and derived four dietary patterns (DPs) from unconstrained principal component analysis. Participants adherence to three of these patterns related to ethnicity: DP1-Moroccans/Turks (e.g. red meat, olive oil, fast foods, sweets), DP2-Surinamese (e.g. chicken, rice/noodles, sugary drinks), DP4-Dutch (e.g. processed meat, coffee, alcohol), while adherence to DP3, a “prudent/healthy” pattern (e.g. fish, fruits, vegetables, legumes), was similarly distributed across ethnicities (**Supplementary Table 5A-B-C**). Gut microbiota alpha-diversity was higher with higher scores for DP1 and 4 (univariate ANOVA), and DP3 (ethnicity-adjusted ANOVA), and lower with higher scores for DP2 (univariate and ethnicity-adjusted ANOVA). All patterns were determinants of gut microbiota beta-diversity, with significant but weakened contributions when ethnicity-adjusted (PERMANOVA, **Supplementary Table 3B**). Specifically, the most representative DP (DP with the highest score across the 4 DPs) resulted in similar groupings as ethnicity on the PCoA plot (**Supplementary Figures 5-6**), but with a lower contribution to inter-individual dissimilarities (PERMANOVA, DP:  $R^2=0.026$ , ethnicity:  $R^2=0.059$ ). In addition to diet, we considered sex, age, education, BMI, alcohol, smoking, physical activity, area of residence, collection season, collection site and sequencing run. All these factors were associated with ethnicity in our study (**Supplementary Tables 1A-B**) and showed associations with gut microbiota (**Supplementary Tables 3A-B**), even though most were weakened (or disappeared) when ethnicity-adjusted (e.g. alcohol, education, area of residence). No factor individually reached the effect size observed for ethnicity and ethnicity

remained the strongest determinant of gut microbiota alpha- and beta-diversity in all models including non-dietary and/or dietary factors. The combination of all factors but ethnicity (including the most representative DP) explained less variability in gut microbiota alpha-diversity (richness/Shannon/PD:  $R^2=0.13/0.14/0.11$ ) and inter-individual dissimilarities ( $R^2=0.083$ ) than the same combination including ethnicity ( $R^2=0.23/0.19/0.21$  and  $0.091$ ).

Markers for metabolic health differed between ethnic groups: e.g. compared to the Dutch, higher prevalence of type-2 diabetes (DM2) and metabolic syndrome in South-Asian Surinamese and of obesity in Moroccans. Considering the proposed associations between the gut microbiota and these metabolic diseases<sup>15</sup>, we next investigated whether the observed ethnicity-gut microbiota associations would merely reflect metabolic health-gut microbiota associations. We thus repeated our analyses on a restricted sample of 646 participants without prevalent DM2, metabolic syndrome or excessive body fat (**Supplementary Table 1C**) and obtained similar ethnic patterns in alpha- and beta-diversity (**Supplementary Figure 7**).

Our study is one of the few to compare the gut microbiota composition between Western and non-Western individuals living in the same city. Previous results usually displayed a lower gut microbiota alpha-diversity in Western vs. non-Western populations, however likely confounded by lifestyle (urban Western vs. traditional/rural non-Western)<sup>7,16,17</sup>. Here, the Dutch exhibited the highest gut microbiota alpha-diversity, a proposed marker for a “healthier” gut microbiome<sup>15,16,18,19</sup>, and enrichments in many OTUs (**Supplementary Table 2A.1-B.1**). Among these, we highlight a particular set classified as *Oscillospira*, Christensenellaceae, Mogibacteriaceae and Rikenellaceae, conversely particularly depleted in the South-Asian Surinamese. These features have been previously associated with leanness and metabolic health<sup>6,18,20–22</sup>. Consistently, some OTUs highlighted above (Otu15-53-103-117-173-309\_*Oscillospira*; Otu89\_Christensenellaceae; Otu257\_Mogibacteriaceae; Otu57-158-383\_Rikenellaceae) were decreased in participants with metabolic syndrome, even

though differences were small and above the multiple-testing correction threshold in ethnicity-adjusted models. In contrast, ethnic groups having migrated from non-Western countries exhibited a lower gut microbiota alpha-diversity and enrichments in features reportedly associated to unhealthy states. Notably, the South-Asian Surinamese gut microbiota was enriched in Otu291\_*Eggerthella lenta*, previously associated with DM2<sup>23</sup>, and in *Blautia* OTUs, Otu6 being also increased in metabolic syndrome (although above the multiple-testing correction threshold), consistent with published associations of *Blautia* with increased central adiposity<sup>18</sup> and impaired lipid and glucose metabolism<sup>24</sup>. The Ghanaians showed enrichment for Otu81\_*Haemophilus parainfluenzae* and Otu105\_*Veillonella*, previously associated to lipopolysaccharides (LPS), an inflammation-driving marker<sup>25</sup>. Other OTUs found to be differentially abundant both across ethnicities and metabolic syndrome status (all decreased in metabolic syndrome, although not always significant after multiple-testing correction in ethnicity-adjusted models) mostly followed the same trend: enrichment in the Dutch (some also enriched in Moroccans/Turks), depletion in the other groups (mainly South-Asian Surinamese). These OTUs included Otu13\_*Faecalibacterium prausnitzii*, Otu34\_*Akkermansia muciniphila*, Otu42-256-276-277\_Firmicutes, Otu212-461\_Clostridia, Otu350\_Clostridiales, Otu41\_Clostridiaceae, Otu80-116-120-130\_Lachnospiraceae, Otu77-215-250\_Ruminococcaceae, Otu193-510\_*Ruminococcus*, Otu107\_*Coprococcus*. All OTUs highlighted above remained differentially abundant between ethnic groups after adjustment for confounding factors (although not always below the multiple-testing correction threshold). It is therefore tempting to speculate that individuals from ethnicities with no or limited exposure to Western lifestyle are more sensitive to environmental and dietary exposures in Western societies.

One main distinctive feature of gut microbiota composition in our population was the typical “trade-off” between abundances of *Prevotella* OTUs (mainly *P. copri*; higher in Moroccans,



Turks and Ghanaians) and *Bacteroides* OTUs (higher in Surinamese)<sup>7</sup>, completed by a third ‘Clostridiales’ direction (*Oscillospira*, *Ruminococcus bromii*, *Coprococcus*; higher in Dutch), previously coined “enterotypes”<sup>26,27</sup>. *Prevotella* species are usually associated with plant-based diets rich in dietary fibers<sup>7,28,29</sup>, while *Bacteroides* abundance is usually seen as an indicator of Western/urbanized diets, rich in fat, animal proteins and sugars<sup>7,28</sup>. While fiber intake was indeed higher in Moroccans/Turks and lower in Surinamese, fat intake was inconsistently lower in Surinamese and higher in Moroccans/Turks (**Supplementary Table 5C**). Still, we observed a functional enrichment for bile acid biosynthesis (PICRUSt inferred, **Methods**) in South-Asian Surinamese and corresponding depletion in Moroccans/Turks (**Supplementary Table 6**), consistent with *Bacteroides* high capacity to deconjugate bile acids<sup>30</sup> and bile acid toxicity to *Prevotella*<sup>31</sup>. Importantly, the average abundance of *Prevotella* in Moroccans/Turks (23.6%) and *Bacteroides* in South-Asian Surinamese (12.5%) ranked in between countries with characteristically high abundance of *Prevotella* (Malawi 32%) or *Bacteroides* (USA, 38%), exhibiting respectively more marked plant-based or high-fat diets<sup>7</sup>.

The sole contribution of ethnicity to inter-individual dissimilarities in gut microbiota composition was ~6% (2.5-3% when multi-adjusted, including for diet), which is relatively large compared to the reported 16.4% and 18.7% (7.6% in non-redundant models) overall contribution of classical determinants of inter-individual differences in gut microbiota composition, found in recent large studies in Caucasians<sup>4,5</sup>. Ethnicity comprises many different aspects – genetics, cultural habits, migration (e.g. socioeconomic status, health care/antibiotics use, early-life environment) – which may all contribute to shape the gut microbiota in specific fashion (nutrients/xenobiotics exposure, immune system regulation, initial/further microbial colonization)<sup>7,11,32,33</sup>. None of the characteristics we studied as potential major confounders entirely accounted for ethnicity, nor were they as strong

predictors as ethnicity alone. The additional explaining effect of ethnicity in the models suggest the involvement of other factors than those studied, notably genetics in our population from distinct ancestries (**Methods**). Although the influence of genetics is likely low (2-8% heritability<sup>10</sup>) compared to environmental factors, especially diet<sup>7,10</sup>, it may participate in building diverse profiles of gut microbiota that would be further refined by the environment<sup>6,16,22</sup>. Supporting an involvement of genetics, ethnic-specific abundances were observed for Christensenellaceae OTUs (Dutch: enrichment, South-Asian Surinamese: depletion) and Otu4\_*Bifidobacterium* (South-Asian Surinamese: enrichment, Dutch: depletion). Christensenellaceae is a highly heritable family associated with higher alpha-diversity and likely causally associated with leanness<sup>6,18,22</sup>. In line, we found inverse associations with BMI and positive associations with alpha-diversity, even when ethnicity-adjusted (ANOVA, BMI: F=6.1, P=0.01; richness/Shannon/PD: F=375.8/308.5/356.7, all  $P < 2.2 \times 10^{-16}$ ). *Bifidobacterium* abundance may result from an interaction between the host ability to digest lactose (LCT gene: lactase persistence selected in traditional milk-consuming communities<sup>34</sup>) and dairy consumption, with lower abundances of the lactose-consuming *Bifidobacterium* following lower amounts of lactose reaching the large intestine in lactase persisters (Dutch) compared to non-persisters after dairy consumption (**Supplementary Table 5C**).

Finally, ethnicity and metabolic health are likely in complex (and bidirectional) interactions with gut microbiota. However, here, we found that the influence of ethnicity on alpha- and beta-diversity went beyond current metabolic health. Different gut microbiota profiles may reflect specific adaptations to environmental conditions<sup>14</sup>. Most non-Dutch participants were first generation migrants (94%) who arrived in the Netherlands in adulthood<sup>35</sup> (arrival dates linked to their country of origin, **Supplementary Table 1**). Despite sharing the same environment for a long period, our participants showed ethnic-specific gut microbiota

profiles, probably reflecting the composition they acquired before migration (no robust association between gut microbiota and age at migration or residence duration, but these strongly correlated to ethnicity: **Supplementary Table 3A, Supplementary Figure 8**). Changes accompanying migration (e.g. food supply/consumption, lifestyle, bacterial/chemical environment) may not have been sufficient to trigger major modifications in the resilient gut microbiome<sup>15</sup>, potentially leading to a lack of adaptation to its environment.

Our results thus show that within a single urban population, the ethnic origin of individuals could be a marker for differences in fecal microbiota composition. This and other reports of ethnic-specific microbiota composition at other body sites<sup>7,36</sup>, suggest that ethnicity may be an important factor to consider in future investigations relating microbiota profiles to health (particularly when designing interventions), especially considering the health disparities observed in ethnic-diverse societies<sup>11</sup>.

## **ACKNOWLEDGEMENTS**

The authors thank R. Jakubowicz and M. Krämer for DNA extractions, PCR amplification and sequencing. The HELIUS study is conducted by the Academic Medical Center Amsterdam and the Public Health Service of Amsterdam. Both organizations provided core support for HELIUS. The HELIUS study is also funded by the Dutch Heart Foundation, the Netherlands Organization for Health Research and Development (ZonMw), the European Union (FP-7), and the European Fund for the Integration of non-EU immigrants (EIF). We gratefully acknowledge the AMC Biobank for their support in biobank management and high-quality storage of collected samples. We are most grateful to the participants of the HELIUS study and the management team, research nurses, interviewers, research assistants and other staff who have taken part in gathering the data of this study.

The study reported here was additionally supported by Le Ducq consortium grant 17CVD01 (to M. Nieuwdorp and F.B.), JPI-HDHL MICRODIET consortium grant (to M. Nieuwdorp and F.B., on which I.A. is appointed) and Novo Nordisk Foundation GUT-MMM grant (to M. Nieuwdorp and F.B.). M. Nieuwdorp is supported by a personal ZONMW-VIDI grant 2013 [016.146.327] (on which G.J.B. is appointed) and a Dutch Heart Foundation CVON IN CONTROL Young Talent Grant 2013 (on which A.P. is appointed). F.B. is Torsten Söderberg professor in medicine and recipient of a European Research Council Consolidator Grant (615362 - METABASE). D.H.V.R. is supported by a junior fellowship of the Dutch Diabetes Foundation (2015.81.1840) and by a global fellowship Marie Curie Action (708193). The funders had no role in the study design, the collection, analysis, and interpretation of data, the writing of the report, and the decision to submit the article for publication.

## **AUTHOR CONTRIBUTIONS**

M.D. analyzed and interpreted the data and drafted the manuscript (M. Nieuwdorp supervised); M. Nieuwdorp, F.B., A.H.Z., R.P., M. Nicolaou, M.B.S., V.T., G.J.B., A.P., E.L.: contributed to the acquisition of the data; K.E.C., A.P., E.L., A.K.G., H.H., V.T., G.J.B., I.A., S.-J. P.-S., D.H.V.R., M.B.S., M. Nicolaou, R.P., A.H.Z., F.B. contributed to the data interpretation and critically reviewed the manuscript. All authors had access to the study data, reviewed, and approved the final manuscript.

## **COMPETING FINANCIAL INTERESTS STATEMENT**

M. Nieuwdorp is in the Scientific Advisory Board of Caelus Pharmaceuticals, the Netherlands; F.B. is in the Scientific Advisory Board of MetaboGen AB, Sweden. None of these are directly relevant to the current paper. There are no patents, products in development or marketed products to declare. The other authors declare no competing financial interests.

## REFERENCES

1. Schroeder, B. O. & Bäckhed, F. Signals from the gut microbiota to distant organs in physiology and disease. *Nat. Med.* **22**, 1079–1089 (2016).
2. Lynch, S. V. & Pedersen, O. The Human Intestinal Microbiome in Health and Disease. *N. Engl. J. Med.* **375**, 2369–2379 (2016).
3. Findley, K., Williams, D. R., Grice, E. A. & Bonham, V. L. Health Disparities and the Microbiome. *Trends Microbiol.* **24**, 847–850 (2016).
4. Zhernakova, A. *et al.* Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
5. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
6. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).
7. Gupta, V. K., Paul, S. & Dutta, C. Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity. *Front. Microbiol.* **8**, 1162 (2017).
8. Miller, G. E. *et al.* Lower Neighborhood Socioeconomic Status Associated with Reduced Diversity of the Colonic Microbiota in Healthy Adults. *PLOS ONE* **11**, e0148952 (2016).
9. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
10. Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
11. Stronks, K. *et al.* Unravelling the impact of ethnicity on health in Europe: the HELIUS study. *BMC Public Health* **13**, 402 (2013).
12. Snijder, M. B. *et al.* Cohort profile: the Healthy Life in an Urban Setting (HELIUS) study in Amsterdam, The Netherlands. *BMJ Open* **7**, e017873 (2017).
13. Yatsunenkov, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
14. Schnorr, S. L. *et al.* Gut microbiome of the Hadza hunter-gatherers. *Nat. Commun.* **5**, (2014).
15. Bouter, K. E., van Raalte, D. H., Groen, A. K. & Nieuwdorp, M. Role of the Gut Microbiome in the Pathogenesis of Obesity and Obesity-Related Metabolic Dysfunction. *Gastroenterology* **152**, 1671–1678 (2017).
16. Sonnenburg, J. L. & Bäckhed, F. Diet–microbiota interactions as moderators of human metabolism. *Nature* **535**, 56–64 (2016).
17. De Filippo, C. *et al.* Diet, Environments, and Gut Microbiota. A Preliminary Investigation in Children Living in Rural and Urban Burkina Faso and Italy. *Front. Microbiol.* **8**, (2017).
18. Beaumont, M. *et al.* Heritable components of the human fecal microbiome are associated with visceral fat. *Genome Biol.* **17**, 189 (2016).
19. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).

20. Stenman, L. K., Burcelin, R. & Lahtinen, S. Establishing a causal link between gut microbes, body weight gain and glucose metabolism in humans – towards treatment with probiotics. *Benef. Microbes* **7**, 11–22 (2016).
21. Oki, K. *et al.* Comprehensive analysis of the fecal microbiota of healthy Japanese adults reveals a new bacterial lineage associated with a phenotype characterized by a high frequency of bowel movements and a lean body type. *BMC Microbiol.* **16**, 284 (2016).
22. Goodrich, J. K. *et al.* Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe* **19**, 731–743 (2016).
23. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
24. Lippert, K. *et al.* Gut microbiota dysbiosis associated with glucose metabolism disorders and the metabolic syndrome in older adults. *Benef. Microbes* **8**, 545–556 (2017).
25. Luthold, R. V., Fernandes, G. R., Franco-de-Moraes, A. C., Folchetti, L. G. D. & Ferreira, S. R. G. Gut microbiota interactions with the immunomodulatory role of vitamin D in normal individuals. *Metabolism.* **69**, 76–86 (2017).
26. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
27. Knights, D. *et al.* Rethinking ‘Enterotypes’. *Cell Host Microbe* **16**, 433–437 (2014).
28. Wu, G. D. *et al.* Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science* **334**, 105–108 (2011).
29. Tremaroli, V. & Bäckhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **489**, 242–249 (2012).
30. Jones, B. V., Begley, M., Hill, C., Gahan, C. G. M. & Marchesi, J. R. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc. Natl. Acad. Sci.* **105**, 13580–13585 (2008).
31. Nakayama, J. *et al.* Diversity in gut bacterial community of school-age children in Asia. *Sci. Rep.* **5**, (2015).
32. Fortenberry, J. D. The uses of race and ethnicity in human microbiome research. *Trends Microbiol.* **21**, 165–166 (2013).
33. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M. & Relman, D. A. The application of ecological theory toward an understanding of the human microbiome. *Science* **336**, 1255–1262 (2012).
34. Walter, J. & Ley, R. The Human Gut Microbiome: Ecology and Recent Evolutionary Changes. *Annu. Rev. Microbiol.* **65**, 411–429 (2011).
35. Prideaux, L. *et al.* Impact of Ethnicity, Geography, and Disease on the Microbiota in Health and Inflammatory Bowel Disease: *Inflamm. Bowel Dis.* **19**, 2906–2918 (2013).
36. Borgdorff, H. *et al.* The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. *PLOS ONE* **12**, e0181135 (2017).

## FIGURE LEGENDS

**Figure 1: Dissimilarities in gut microbiota composition in the multi-ethnic HELIUS study (n=2084 participants), represented by unconstrained Principal Coordinate Analysis (Multidimensional scaling – PcoA/MDS) with Bray-Curtis index calculated on unscaled OTU relative abundances.** Ethnicity explained 5.7% of the dissimilarities in gut microbiota composition (PERMANOVA, P=0.001). The centroids of each ethnic group are featured as the group name on the graph (vegan::envfit). The top 20 contributive OTUs to the dissimilarities in gut microbiota composition (determined with R<sup>2</sup> from PERMANOVA, Supplementary Table 2A.1) are displayed as arrows, with a length proportional to the correlation between the variable and the PCoA ordination (vegan::envfit): clockwise from bottom left to bottom right – Otu1\_*Prevotella copri*, Otu27\_*Prevotella*, Otu308\_*Prevotella copri*, Otu550\_*Prevotella copri*, Otu10\_*Dialister*, Otu15\_*Oscillospira*, Otu7\_*Ruminococcus bromii*, Otu37\_*Coprococcus*, Otu21\_*Coprococcus*, Otu16\_*Bacteroides uniformis*, Otu92\_*Clostridiaceae Clostridium*, Otu11\_*Lachnospiraceae*, Otu2\_*Faecalibacterium prausnitzii*, Otu4\_*Bifidobacterium*, Otu9\_*Blautia*, Otu5\_*Bacteroides*, Otu12\_*Bacteroides*, Otu14\_*Coprococcus*, Otu6\_*Blautia*, Otu3\_*Roseburia faecis*. The first PCoA axis (MDS1) represented 18.4% of variability. The second PCoA axis (MDS2) represented 6.3%. Similar ethnic grouping patterns were observed in PCoA performed with Bray-Curtis dissimilarities calculated on scaled/unscaled genus or scaled OTU relative abundances (Supplementary Figure 2A-C), in PCoA constrained for ethnicity (Bray-Curtis index, unscaled OTU relative abundances, Supplementary Figure 2D) or when excluding participants who declared probiotics (n=110 participants), antibiotics (n=198 participants), oral diabetes medication (n=200 participants), PPI medication (n=256 participants) or diarrhea (n=202 participants) [data not shown].

**Figure 2: Richness, Shannon index and Faith's phylogenetic diversity index at the OTU level across ethnic groups, HELIUS study (n=2084 participants).** D: Dutch, G: Ghanaian, M: Moroccan, AS: African Surinamese, SAS: South-Asian Surinamese, T: Turkish. Box-plots feature the median (center line), the upper and lower quartiles (box limits), 1.5x the interquartile range (whiskers) and outliers (points). P-values from two-sided pairwise t-tests between the 6 ethnic groups are shown in Supplementary Table 4. Similar results were observed when excluding participants who declared probiotics (n=110 participants), antibiotics (n=198 participants), oral diabetes medication (n=200 participants), PPI medication (n=256 participants) or diarrhea (n=202 participants) [data not shown].

**Figure 3: Overall and ethnic-specific 95% core OTUs, HELIUS study (n=2084 participants).** Colors indicate the median relative abundance of the OTUs found in more than 95% of individuals overall and within each ethnic group (group-specific medians) from dark red: maximum value in the table to dark green: minimum value in the table. Exact median values and corresponding 25<sup>th</sup> and 75<sup>th</sup> percentiles can be found in Supplementary Table 2A.1. Blank cells indicate that this OTU is not part of the 95% core of the group considered.

## ONLINE METHODS

### Study population: the HELIUS study

This work was carried out on a subsample from the prospective HELIUS study (HEalthy LIfe in an Urban Setting, [www.heliusstudy.nl](http://www.heliusstudy.nl)). Details of the rationale, recruitment and data collection have been previously published<sup>11,12</sup>. Briefly, the HELIUS study was set up to investigate the causes of the unequal burden of disease across ethnic groups and includes participants belonging to the six largest ethnic groups living in Amsterdam, The Netherlands: (i) Dutch: Northwestern European ancestry, (ii) Ghanaian: Western African ancestry, (iii) Moroccan: Northern African, Mediterranean and Middle Eastern ancestries, (iv) African Surinamese: ‘Creoles’ descending from former slaves, mixed Western African and Dutch ancestries, (v) South-Asian Surinamese: ‘Hindustanis’ descending from workers arriving to the Suriname post slavery, South-Asian/Indian ancestry, and (vi) Turks: Mediterranean, Caucasus and Middle Eastern ancestries<sup>11,12,37</sup>. Participants aged 18-70 years were randomly sampled, stratified by ethnic origin, through the municipal registry of Amsterdam (MRA) between 2011 and 2015, and a total of nearly 25,000 participants were included at baseline. Country of birth was used as a basis to identify the ethnic origin of participants: a person was considered as of non-Dutch origin if she/he was born outside the Netherlands and has at least one parent who was born outside the Netherlands (first generation); or if she/he was born in the Netherlands but both parents were born outside the Netherlands (second generation)<sup>38</sup>. For the Dutch sample, we invited people who were born in the Netherlands and whose parents were born in the Netherlands. After inclusion, participants of Surinamese ethnic origin were further classified according to self-reported ethnic origin (obtained by questionnaire) into ‘African’, ‘South-Asian’, or ‘other’. The Surinamese migrated from Suriname (South America), a former Dutch colony, in 1975-80 following decolonization and political instability. Turks and Moroccans arrived starting the 1960s-1970s through encouraged labor



migration. The Ghanaians migrated between 1974 and 1983 (economic reasons) and in the early 1990s (drought, political instability)<sup>11,12</sup>. The HELIUS study has complied with all relevant ethical regulations in accordance with the Declaration of Helsinki and was approved by the Academic Medical Center (AMC) Medical Ethics Committee and all participants provided written informed consent.

### **Data collection**

Upon recruitment, participants filled out a questionnaire gathering information on migration-related variables (e.g. acculturation), socio-demographic characteristics, lifestyle (e.g. physical activity, smoking, alcohol use, dietary habits), general quality of life, and personal and family history of diseases (e.g. hypertension, cardiovascular diseases, DM2, mental health). Participants were invited to undergo a physical examination at a local research site to which they were asked to bring all currently used medications. The examination included measurements of anthropometric characteristics (e.g. height, weight, waist and hip circumferences, body fat percentage using bioelectrical impedance) and blood pressure, and collection of fasted blood samples. Samples of EDTA whole blood and heparin plasma were transported directly to the AMC Clinical Chemistry Laboratory for determination of hemoglobin, HbA1c, glucose, lipid profile, and creatinine levels. The remaining samples were transported to the AMC Biobank, where samples were checked, registered, and stored at -80°C.

### *Feces sample collection*

Feces samples were collected on a subset of voluntary participants who gave permission for storage of body material. They were given a feces collection tube and a safety bag (for transport) either (i) through mail prior to the physical examination visit or (ii) at the end of the visit if they agreed to. They were asked to bring a “fresh” stool sample within six hours after collection to the research location (on the morning of their physical examination visit or later

on if they received the collection kit during their physical examination visit). If not possible, they were instructed to keep the stool sample in their freezer overnight and to bring it in frozen state to the research location the next morning. At the research location, the samples were temporarily stored at -20°C until daily transportation to the AMC, where the samples were checked by a study nurse and stored at -80°C. No information was recorded on whether samples arrived either frozen or fresh at the research location. During the physical examination, participants who handed over or agreed upon handing over a stool sample were additionally asked if (i) they used probiotics (frequency, type), (ii) they used antibiotics in the past 3 months, (iii) they had diarrhea in the past week. About 6000 stool samples were collected.

Three research locations were used for the physical examination and feces sample collection, situated in three Amsterdam areas, at the convenience of participants: Nieuwlandhof – Amsterdam South-East, Noord – Amsterdam North, Slotervaart – Amsterdam West (Supplementary Table 1). Standardized procedures were used across the three collection sites with the same (written) protocols and the same training of the personnel (with many staff members working at all three locations). Quality checks on the staff/procedures were done regularly during the entire data collection period.

#### *HELIUS-Dietary Patterns study*

As described previously<sup>39,40</sup>, a subsample of ~5200 HELIUS participants (from Dutch, Moroccan, Turkish and Surinamese ethnic origin) were also enrolled in the HELIUS-Dietary Patterns study, with the objective to collect detailed information on their diet. Habitual dietary intakes were assessed using ethnic-specific semi-quantitative food frequency questionnaires (FFQs) with a reference period of 4 weeks. These FFQs were adapted from an existing Dutch FFQ and comprised about 200 items. Food items were collapsed into 37 food groups based on similarity in nutrient profile or culinary use.

## **Profiling of fecal microbiota composition**

### *Sequencing of the 16S rRNA gene*

The first 2170 feces samples among ~6000 collected were shipped to the Wallenberg Laboratory (Sahlgrenska University of Gothenburg, Sweden) for determination of the fecal microbiome. Total genomic DNA was extracted from a 150 mg aliquot using a repeated bead beating method as previously described<sup>41</sup>. Fecal microbiota composition was profiled by sequencing the V4 region of the 16S rRNA gene on an Illumina MiSeq instrument (Illumina RTA v1.17.28; MCS v2.5) with 515F and 806R primers designed for dual indexing<sup>42</sup> and the V2 Illumina kit (2x250 bp paired-end reads). 16S rRNA genes from each sample were amplified in duplicate reactions in volumes of 25 µl containing 1x Five Prime Hot Master Mix (5 PRIME GmbH), 200 nM of each primer, 0.4 mg/ml BSA, 5% DMSO and 20 ng of genomic DNA. PCR was carried out under the following conditions: initial denaturation for 3 min at 94°C, followed by 25 cycles of denaturation for 45 sec at 94°C, annealing for 60 sec at 52°C and elongation for 90 sec at 72°C, and a final elongation step for 10 min at 72°C. Duplicates were combined, purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) and quantified using the Quant-iT PicoGreen dsDNA kit (Invitrogen). Purified PCR products were diluted to 10 ng/µl and pooled in equal amounts. The pooled amplicons were purified again using Ampure magnetic purification beads (Agencourt) to remove short amplification products. Negative controls were included for each sample. The absence of detectable PCR products in these negative controls was confirmed with electrophoresis gel. Positive controls were not included in these runs but the protocol used to analyze the samples has been optimized using mock samples. All analytical procedures were blinded for ethnicity (but not randomized). The distribution of ethnic groups across runs of 16S rRNA sequencing are shown in Supplementary Table 1.

### *Processing of 16S rRNA gene reads and OTU clustering*

The 2170 samples were processed all at once using the following pipeline. Illumina MiSeq reads (177,089,775 total raw reads in 2170 samples) were quality checked with FastQC (v.0.11.5)<sup>43</sup> and processed on a Mothur pipeline (version 1.39.5)<sup>44</sup>. After merging forward and reverse reads, contigs were screened to ensure absence of ambiguous bases and a length between 252 and 253 bases. Around 70% of reads survived the merging and quality control steps. Contigs were aligned to the Silva reference database (version 128), dereplicated and subsequently preclustered (allowing a maximum of 2 differences). Singletons (sequences with an abundance of 1 in the entire dataset) were removed (1.7 % of total reads). The remaining sequences were chimera-filtered with chimera.vsearch (removing a further 2.8% of total reads) and classified using a mothur-formatted Greengenes reference (version 13\_5\_99) with a 80% confidence cutoff. Sequences classified as mitochondria, chloroplasts, eukaryota, as well as un-classified sequences were removed (0.005% of total reads). Remaining sequences (65.6% of total initial reads) were clustered using the VSEARCH (v.2.6) implementation of a distance-based greedy algorithm (DGC), an open-source alternative to USEARCH<sup>45,46</sup>. A phylogenetic tree was constructed with the “double-precision” build of Fastree 2.1 using the abundance-based representative sequences of the OTUs<sup>47</sup> (representative sequences of the OTUs and corresponding taxonomy classification are shown in Supplementary Table 7). A PICRUSt-compatible biom file was created in mothur for PICRUSt analyses. Samples with less than 10,000 counts were removed from the dataset (8 samples) and the resulting OTU table was rarefied to 10,960 counts per sample (vegan package, R)<sup>48</sup>. OTUs with a mean relative abundance of less than 0.002% (i.e. a mean of less than 1 count per 5 samples) were filtered out (phyloseq package, R)<sup>49</sup>. The final dataset was comprised of 2162 samples and 744 OTUs. In the main text, when a particular OTU is referred to using taxonomy levels

above the species level, it means that this OTU could not be characterized to a lower taxonomical level.

### **Statistical analyses**

Gut microbiota composition was obtained for 2162 individuals, of which 78 were removed because they did not pertain to one of the 6 main ethnic groups under study (30 Javanese Surinamese, 42 Other/Unknown Surinamese and 6 Other/Unknown). Three samples were defined for the analyses: a full sample including 2084 participants, a “FFQ” sample composed of 1032 participants in whom also detailed dietary information was available (enrolled in the HELIUS-dietary pattern study) and a “healthy” sample comprising 646 participants resulting from the exclusion of participants with prevalent metabolic syndrome, DM2 or with a fat percentage indicating obesity ( $\geq 35\%$  in women and  $\geq 25\%$  in men). We used the body fat percentage to define obesity because BMI may be less accurate to define obesity in South-Asian populations<sup>50</sup>.

All analyses were performed using R (version 3.3.1) and particularly functions embedded in the *vegan*<sup>48</sup> and *phyloseq*<sup>49</sup> packages. Statistical tests were two-sided.

#### *Characteristics of gut microbiota composition*

Relative abundances of features (OTU, genus, family, phylum) were calculated by dividing the rarefied number of reads for this OTU by the total number of reads of the individual.

The dissimilarities in gut microbiota composition between individuals (beta-diversity) were assessed with the Bray-Curtis index calculated at the OTU and at the genus levels (R function *vegan::vegdist*). Unscaled and scaled (i.e. divided by standard deviation) relative abundances were used. Scaling of the relative abundances allows to obtain the same order of magnitude for the features (OTU/genus) relative abundances and their variations (standard deviation unit), thus giving less weight to the highly abundant features in the Bray-Curtis index. Bray-Curtis dissimilarities were plotted using unconstrained principal coordinate analysis (PCoA)

allowing to visualize the dimensions explaining most variability in the dissimilarity matrix (R function `vegan::capscale`). The contribution of ethnic origin and features (OTU, genus) to the PCoA ordination plots was assessed using the R function `vegan::envfit`. The obtained  $R^2$  gives the proportion of the main variability (i.e. main dimensions of the ordination) that can be attributed to the studied variables. The contribution of ethnic origin and features (OTU, genus) to the overall Bray-Curtis dissimilarities in gut microbiota composition was also assessed using permutational multivariate analysis of variance models (PERMANOVA, R function `vegan::adonis`, 999 permutations) which decompose the dissimilarity matrix into "variance" explained by each covariate. The obtained  $R^2$  gives the proportion of variability observed in the entire dissimilarity matrix that can be attributed to the studied variables.

The following potential confounders of the association between ethnicity and gut microbiota composition were tested using PERMANOVA models assessing the marginal effect of each covariate (R function `vegan::adonis2`, 999 permutations): sex, age, educational level, BMI, alcohol consumption (frequency), smoking status, physical activity, area of living, season of feces collection, collection site and 16S rRNA sequencing run.

The alpha-diversity of gut microbiota for each individual was assessed with several indices calculated at the OTU level: richness (number of unique OTUs: R function `vegan::estimateR`), richness and evenness of distribution (Shannon index: R function `vegan::diversity`) and phylogenetic diversity (Faith's PD: R function `picante::pd`). These indices were compared across the 6 ethnic groups using ANOVA (6 groups) and pairwise t-tests with Bonferroni correction, after checking assumptions, as appropriate. The following potential confounders of the association between ethnicity and alpha-diversity were tested using ANOVA models: sex, age, educational level, BMI, alcohol consumption (frequency), smoking status, physical activity, area of living, season of feces collection, collection site and 16S rRNA sequencing run.

The relative abundances of OTU were compared across the 6 ethnic groups using Kruskal-Wallis rank sum tests (1 test per 744 OTUs) and Wilcoxon-Mann-Whitney tests (6 comparisons between one group and all others combined and 15 pairwise comparisons between the 6 ethnic groups per 744 OTUs) with Bonferroni correction:  $P < 6.72 \times 10^{-5}$  (i.e.  $0.05/744$ ) for Kruskal-Wallis rank sum tests and  $P < 3.2 \times 10^{-6}$  (i.e.  $0.05/((6+15)*744)$ ) for Wilcoxon-Mann-Whitney tests. With our large sample size, even small differences were observed to be significant. Thus, the probability of concordance (AUC, number of concordant pairs/total number of possible pairs) and the effect size ( $z/\sqrt{N}$  with  $N$  the sample size) were computed and used in addition to the OTU prevalence to select the most relevant differences. As sensitivity analyses, we adjusted the comparison of OTU relative abundances across ethnic groups for sex, age, educational level, BMI, alcohol consumption (frequency), smoking status, physical activity, area of living, season of feces collection, collection site, 16S rRNA sequencing run with and without additional adjustment for the most representative dietary pattern (see below). General linear models with arcsin square root transformation of OTU relative abundances were run in a pairwise manner (6 ‘one vs all others’ and 15 ‘one vs one’ analyses). Finally, to assess the health relevance of the differentially abundant OTUs across ethnic groups, we compared OTU relative abundances across metabolic syndrome status (yes vs no) using Wilcoxon-Mann-Whitney tests and ethnicity-adjusted general linear models with arcsin square root transformation and Bonferroni correction for multiple testing ( $P < 6.72 \times 10^{-5}$ , i.e.  $0.05/744$ ).

### *Dietary patterns*

Detailed dietary information from the HELIUS-dietary patterns study were available for ~50% of our study sample ( $n=1032$ ). A principal component analysis (PCA) with varimax rotation was performed on the consumption (g/day) of 37 food groups (see Supplementary Table 5 for the description of the groups, ethnic-specific food groups were not included). This

data-driven approach accounts for the correlations between the consumption of several food groups and thus leads to the identification of the main dietary patterns in the studied population. Intakes of food groups were obtained by collapsing food items assessed in the FFQ on the basis of similarity in nutrient profile, culinary use, or ethnic origin. Ethnic-specific food groups were not included in the analysis in order to derive the dietary patterns on a same basis for all ethnic groups. Four components (patterns) were retained, based on the observation of the scree plot and the interpretability of the patterns (see Supplementary Table 5A for the loading of the patterns). Food items were considered to contribute significantly to a component if they had an absolute factor loading  $\geq 0.3$ . A larger factor loading indicates a higher correlation of the food group to the respective component. A score was obtained for each participant and each component, reflecting the strength of adherence of each individual to each of the derived pattern. The maximum positive score across the 4 patterns was used to define the pattern that reflected most the individual's diet.

The influence of dietary patterns (continuous score, dichotomous variable with 1 if positive score and 0 if not, and predominant pattern) and of specific food groups/nutrients was tested using ANOVA (alpha-diversity measures) and PERMANOVA (Bray-Curtis dissimilarities) in univariate and ethnicity-adjusted models and in models adjusted for ethnicity, sex, age, educational level, BMI, alcohol consumption (frequency), smoking status, physical activity, area of living, season of feces collection, collection site and 16S rRNA sequencing run.

### *PICRUSt*

The functional potential of the gut microbiota community profiled with 16S rRNA gene sequencing was assessed using the PICRUSt methodology (phylogenetic investigation of communities by reconstruction of unobserved states)<sup>51</sup> on the dedicated Galaxy module online (<https://huttenhower.sph.harvard.edu/galaxy/>). Based on the OTU table, we obtained for each subject the derived KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways,



predicting the functional capacities of the gut microbiome. The obtained pathways were collapsed to a higher hierarchical level. The relative abundances of hierarchically classified pathways were compared using Kruskal-Wallis rank sum test (6 groups) and pairwise Wilcoxon-Mann-Whitney tests with Bonferroni correction.

## **Reporting Summary**

A Life Sciences Reporting Summary is available for this article.

## **Data Availability**

The 16S rRNA gene sequences and associated metadata (including ethnicity) have been deposited at the European Genotyping Agency ([www.ebi.ac.uk/ega/](http://www.ebi.ac.uk/ega/))—study no. EGAD00001004106. The data that support the findings of this study are available from the study coordinator upon reasonable request.

The HELIUS data are owned by the Academic Medical Center (AMC) in Amsterdam, The Netherlands. Any researcher can request the data by submitting a proposal to the HELIUS Executive Board as outlined at [www.heliusstudy.nl/en/researchers/collaboration](http://www.heliusstudy.nl/en/researchers/collaboration). Requests for further information and proposals can be submitted to Dr. Marieke Snijder, Scientific Coordinator and Data Manager of HELIUS, at [m.b.snijder@amc.uva.nl](mailto:m.b.snijder@amc.uva.nl), or to [info@heliusstudie.nl](mailto:info@heliusstudie.nl). The “heliusstudy” website and generic email address will continue to be actively managed in the event Dr. Snijder should leave her post. The HELIUS Executive Board will check proposals for compatibility with the general objectives, ethical approvals and informed consent forms of the HELIUS study, and potential overlap with ongoing work affiliated with HELIUS. There are no other restrictions to obtaining the data and all data requests will be processed in the same manner.

## METHODS-ONLY REFERENCES

11. Stronks, K. *et al.* Unravelling the impact of ethnicity on health in Europe: the HELIUS study. *BMC Public Health* **13**, 402 (2013).
12. Snijder, M. B. *et al.* Cohort profile: the Healthy Life in an Urban Setting (HELIUS) study in Amsterdam, The Netherlands. *BMJ Open* **7**, e017873 (2017).
37. Hellenthal, G. *et al.* A Genetic Atlas of Human Admixture History. *Science* **343**, 747–751 (2014).
38. Stronks, K., Kulu-Glasgow, I. & Agyemang, C. The utility of ‘country of birth’ for the classification of ethnic groups in health research: the Dutch experience. *Ethn. Health* **14**, 255–269 (2009).
39. Dekker, L. H. *et al.* A prospective cohort study of dietary patterns of non-western migrants in the Netherlands in relation to risk factors for cardiovascular diseases: HELIUS-Dietary Patterns. *BMC Public Health* **11**, (2011).
40. Beukers, M. H. *et al.* Development of the HELIUS food frequency questionnaires: ethnic-specific questionnaires to assess the diet of a multiethnic population in The Netherlands. *Eur. J. Clin. Nutr.* **69**, 579–584 (2015).
41. Mobini, R. *et al.* Metabolic effects of *Lactobacillus reuteri* DSM 17938 in people with type 2 diabetes: A randomized controlled trial. *Diabetes Obes. Metab.* **19**, 579–589 (2017).
42. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
43. Andrews S. FastQC: a quality control tool for high throughput sequence data. (2010). Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
44. Schloss, P. D. *et al.* Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
45. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinforma. Oxf. Engl.* **26**, 2460–2461 (2010).
46. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
47. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**, e9490 (2010).
48. Oksanen J *et al.* vegan: Community Ecology Package. R package version 2.4-3. (2017). Available at: <https://CRAN.R-project.org/package=vegan>.
49. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **8**, e61217 (2013).
50. World Health Organization. *Physical status: the use and interpretation of anthropometry. Report of a WHO expert committee.* 1–452 (World Health Organ Tech Rep Ser, 1995).
51. Langille, M. G. I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013).