Long-term life history predicts current elderly gut microbiome

Jeroen Raes (jeroen.raes@kuleuven.vib.be)
KU Leuven https://orcid.org/0000-0002-1337-041X

Jiyeon Si
VIB KU Leuven Center for Microbiology https://orcid.org/0000-0002-4356-5592

Ann Gregory
VIB-KU Leuven Center for Microbiology

Jorge Vázquez-Castellanos
VIB-KU Leuven Center for Microbiology

Lindsey Decommer
VIB-KU Leuven Center for Microbiology

Leen Rymenans
University of Leuven

Sebastian Proost
VIB-KU Leuven Center for Microbiology

Martin Weger
Medizinische Klinik II. Klinikum Ingolstadt

Marlene Notdurfter
Department of Internal Medicine, Hospital of Bruneck

Christoph Leitner
Department of Internal Medicine, Hospital of Bruneck

Peter Santer
Department of Laboratory Medicine, Hospital of Bruneck

Gregorio Rungger
Department of Neurology, Hospital of Bruneck

Johann Willeit
Innsbruck Medical University https://orcid.org/0000-0003-4083-0466

Peter Willeit
Department of Neurology, Medical University Innsbruck

Raimund Pechlaner
Medical University of Innsbruck

Felix Grabherr
Medical University Innsbruck

Stefan Kiechl

Innsbruck Medical University

Herbert Tilg

Medical University Innsbruck https://orcid.org/0000-0002-4235-2579

Biological Sciences - Article

Keywords: Microbiome Composition, Bruneck Study, Flemish Gut Flora Project, Community Richness, Low-abundance Bacteria, Multivariate Analysis

Posted Date: January 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-130812/v1

Abstract

Extensive scientific and clinical microbiome studies have explored contemporary variation and dynamics of the gut microbiome in human health and disease[1–4], yet the role of long-term life-history effects is underinvestigated. Here, we analyzed the current microbiome composition in the elderly Bruneck Study cohort (n = 304; age 65–98) with extensive clinical, demographic, lifestyle, and nutritional data collected over the past 26 years. Combined analysis with the Flemish Gut Flora Project cohort (FGFP; n = 2,215; age 18–85) showed community richness increasing during aging linked to increased observation of low-abundance bacteria. Multivariate analysis of historical variables indicated that medication history, historical physical activity, past dietary habits, and specific past laboratory parameters explain a significant fraction of current elderly quantitative microbiome variation, enlarging the explanatory power of contemporary covariates by 33.4%. Prediction of current enterotype by past host variables revealed good levels of predictability (AUC > 0.7) for the Prevotella and dysbiotic Bacteroides 2 enterotypes with information from up to 15 years past. These findings demonstrate long-term life history effects on the microbiota and provide first insights into lifestyle variables and their role in maintaining a healthy gut microbiota in later life.

Main Text

Structure, function and dynamics of the human gut microbiome are generally studied in cross-sectional or short-term longitudinal settings. Contemporary microbiome variation is partially explained by host variables such as age, gender, stool consistency/transit time, health status, diet, and medication[1]. However, the gut environment is a dynamic ecosystem, continuously perturbed by daily dietary intake and egestion or occasional exposures to medication and disease[5]. Isolated events as well as long-term lifestyle choices can permanently alter the microbiome[2], yet long-term temporal effect are understudied. A recent study revealed that diet only allowed prediction of future microbiome up to two days after food consumption[3]. On the other hand, incomplete recovery of the original gut microbiota even after six months of antibiotic exposure implies that, when strong enough, perturbation effects can last for longer terms[4]. As host health and lifestyle continuously impact the microbiome environment over time, a consistent, enduring collection of host data is necessary to study long-term cumulative effects of life history, especially for long-lived hosts like humans.

Here, we capitalized on the community-based North-Italian Bruneck Study cohort (n=304), which prospectively collected long-term individualized host metadata (i.e. food intake, lifestyle, medication, blood chemistry and clinical assessments) over 26 years (1990 to 2016) in 5-year intervals[6]. Fecal samples collected in 2016 (age 65-98) were subjected to quantitative microbiota profiling (QMP) enabling association of current absolute microbiome abundances with historical metadata[7]. Using this unique dataset, we explored (*i*) the aging microbiome, combining Bruneck and Belgian Flemish Gut Flora Project (FGFP; n = 2,215; aged 18 to 85) population cohorts, (*ii*) associations of historical variables and current microbiome, and (*iii*) predictive capacity of lifestyle history on current microbiome.

Microbiota diversity change with age is a debated subject. Prior assessments of alpha-diversity during aging gave conflicting results (with both positive[8,9] and negative[10,11] associations) attributed to confounding by residence type[9] and/or frailty[12]. In our combined cohort (age range 18-98, combined n=2,519; relative microbial profiling (RMP)), we observed a significant positive correlation between age and microbial richness (Spearman rho range 0.12 to 0.18, $p < 0.05$; Figure 1a). As only three subjects were in nursing homes in the combined cohort, the positive correlation observed here is likely independent of residence type. Further analyses of Bray-Curtis dissimilarities found elderly (age≥65) community composition significantly different from young adults (18≤age<30) and adults (30≤age<65; pairwise Adonis R^2 [0.003:0.01], FDR = 0.001; Extended Data Fig. 1a and Supplementary Table 1). Taxa increasing with age were mostly from the rare tail of the abundance curve (Extended Data Fig. 1b and Supplementary Tables 2, 3). On the contrary, top-ranked core taxa such as *Faecalibacterium*, *Roseburia,* and *Anaerostipes* were negatively correlated with age in elderly (Spearman rho [-0.28:-0.27], adjusted for gender, BMI, stool moisture, and antibiotics intake, FDR<0.1; Figure 1b and Extended Data Fig. 1c, d). Reduction of these major butyrate-producers might suggest unfavorable metabolic changes in the elderly gut.

We next studied whether the distinct elderly gut microbiome could be due to cumulative exposures to various perturbations. Prior to the determination of long-term life-history effects on the elderly microbiome, we quantitatively investigated contemporary microbial community covariates in the Bruneck Study cohort based on a db-RDA analysis after removing collinear variables (as in [1]; Supplementary Table 4). Highly ranked contemporary covariates linked with microbial community composition were qualitative and quantitative proxies of transit time like current stool moisture, defecation frequency, hard stools, and obstipation (db-RDA, adjusted R^2 [1.5:2.4%], FDR < 0.1; Figure 2b, and Supplementary Table 5a). These variables were consistently found associated with microbial community composition in different age groups (Extended Data Fig. 2a), and previously identified as top covariates in the FGFP cohort[1]. In total, we observed a 7.0% non-redundant cumulative effect size from 11 significant contemporary covariates (db-RDA, FDR < 0.1; Supplementary Table 5a) using a forward stepwise redundancy analysis (Supplementary Table 5b), in line with estimates in adult populations[1]. Analysis of combined effect size of significant covariates per category revealed that current health parameters (e.g. liver stiffness) and bowel habits (stool moisture, defecations, and hard stools) had similar (redundant) cumulative explanatory power for elderly community variation (12.4 to 12.8% of variation; Supplementary Table 6a). Total additive effect size including further categories (medication, diet, anthropometric feature, and lifestyle) was 16.6%.

Next, we sought to assess the influence of the extensive array of historical parameters collected during prior Bruneck Study evaluations (1990-2016) on the current microbiome composition. We first performed a db-RDA as above using the historical parameters of each year as explanatory variables (Supplementary Table 7a), identifying historical variables contributing significantly to a cumulative model that also included present variables (Supplementary Table 7b). Overall, significant historical variables were mostly linked to beta blocker, blood parameters, and diet (adjusted R^2 [0.60:0.80%], FDR<0.1). Interestingly,

inclusion of these significant historical parameters of each evaluation (db-RDA, FDR<0.1; Supplementary Table 7a) significantly increased cumulative non-redundant effect size to 8.5% (likelihood ratio test $p < 0.05$) and total additive effect size to 20.7% indicating potential explanatory power of long-term historical covariates on current microbiome (Supplementary Table 6b).

To better capture long-term lifestyle and health effects, we further investigated overall historical trends of variables by using the average across all years and difference (i.e. delta) for continuous variables and counts of event occurrence for categorical variables between each year and the year 2016. Analysis of averaged covariates across the years gave very few results; only average dumpling intake.1995-2016 was significant (adjusted R^2 = 0.75%, FDR < 0.1; Supplementary Table 8a), likely as a proxy for more traditional lifestyle. Covariate analysis of change (delta) in historical host parameters identified multiple non-colinear parameters independent of the time period covered (i.e., beta blocker usage, hemoglobin, alanine transaminase (ALT), and non-sport physical activity; db-RDA, adjusted R^2 [0.63:1.11%], FDR<0.1; Supplementary Table 8b and Extended Data Fig. 2b). These were again analyzed with 11 significant contemporary covariates to calculate non-redundant cumulative effect sizes. Beta blocker.1990-2016, non-sport physical activity.2005-2016, hemoglobin.1990-2016, and Alanine transaminase.2005-2016 were shown to have significant explanatory power in addition to contemporary covariates, significantly raising the cumulative non-redundant effect size to 8.5% (likelihood ratio test $p < 0.05$; Supplementary Table 8c).

Finally, we combined all significant contemporary and historical features (Supplementary Tables 5a, 7a, 8a, and 8b), in one comprehensive db-RDA analysis and found a final set of 14 variables significantly explaining the current microbiome variation, of which three were historical ones from 2010 (gamma-glutamyl transferase (GGT), caloric intake, cereal fiber score, one was average dumpling intake (1995-2016), and four were duration of beta blocker treatment (1990-2016), long-term changes in non-sport physical activity (2005-2016), hemoglobin (1990-2016), and alanine transaminase (ALT; 2005-2016). All together, they significantly increased the final cumulative non-redundant effect size to 10.4% (likelihood ratio test $p < 0.05$; Figure 2a and Supplementary Table 8d) and total additive effect size to 25.5% (Figure 2c and Supplementary Table 6c). Overall, this shows that the inclusion of historical data resulted in a 33.4% increase in non-redundant explanatory power for global microbiota variation.

We further deepened the relationship of these historical variables with the current microbiome by focusing on current absolute taxonomic group abundances as well as community enterotype based on Dirichlet Multinomial Modeling-based clustering previously validated across multiple cohorts[13–15]. Previous studies detected four enterotypes[7], dominated by either *Bacteroides* (B1 and B2 enterotypes, with B2 having a.o. lower microbial load and abundance of *Faecalibacterium* compared to B1)[16], *Prevotella* (P), and *Ruminococcaceae* (R). All four enterotypes were present in the Bruneck cohort. Using the significant historical variables identified in the above RDA analyses (db-RDA, FDR<0.1; Figure 2a and Supplementary Table 8d), we first analyzed beta blocker treatment in association with community diversity. By dividing subjects into three groups (chronic (treatment of beta blocker both in 1990 and

2016), current, and none (not medicated in 1990 and 2016), we found that beta blocker treatment was linked to a significant compositional shift (beta-diversity; Adonis r^2 = 0.013, p < 0.001; Extended Data Fig. 2c and supplementary Table 9a), but not to alpha-diversity (Kruskal-Wallis test, p > 0.05; Extended Data Fig. 2d). Prevalence of enterotype was found to be significantly different between the three groups (Fisher's exact test, FDR < 0.1 for B2 versus B1 and P enterotypes; Supplementary Table 9b). Especially, B2 enterotype was more prevalent in beta blocker treated individuals than other enterotypes (Kruskal-Wallis test, FDR < 0.1; Extended Data Fig. 2e). Further analysis of specific taxonomic associations identified a list of bacteria more abundant in subjects who did not use beta blockers, which can be potential targets for remediation strategies (Generalized linear model (GLM), FDR<0.1, adjusted for age and stool moisture; Supplementary Table 10) if future studies confirm a causal link for this association.

Analysis of average dumpling intake (1995-2016), a historical covariate with the second-largest effect size, showed that P enterotype has had higher dumpling intake than B1 enterotype (Kruskal-Wallis test, FDR < 0.1; Extended Data Fig. 2f and Supplementary Table 11). This could be explained by the association of the P enterotype with long-term dietary patterns[15] and perhaps impact of traditional dietary intake or -lifestyle on the current microbiome given that they are a key component of traditional food in the Bruneck region.

We next looked at change in non-sport physical activity between the year 2005 and 2016. We first identified taxa associated with both physical activity shifts (i.e., the change from the past to the present) as well as current levels of physical activity. Although no genera were associated with both variables, butyrate-producing bacteria (i.e., *Roseburia*, *Faecalibacterium*, and *Butyricicoccus*) significantly increased with long-term physical activity (Spearman rho [0.18:0.21], FDR<0.1, adjusted for age and stool moisture; Supplementary Tables 12-13). The positive influence of exercise on gut health has gained recent attention with elevated abundances of *Roseburia* and *Faecalibacterium* reported in fit individuals and those who perform regular exercise[17–20]. In order to study effects of changing physical activity, we clustered subjects into four categories: high activity in the past and at present (cluster 1), high in the past and low at present (cluster 2), low in the past and high at present (cluster 3), low in the past and at present (cluster 4). Interestingly, subjects who have recently increased physical activity as well as those who have consistently maintained high activity exhibited reduced ratio of (dysbiotic) B2 to non-B2 enterotypes. This suggests a beneficial role of physical activity for the healthy elderly gut ecosystem (pairwise Chi-Square test, FDR < 0.1; Figure 2e and Supplementary Table 14a).

Finally, we studied changes in hemoglobin between 1990 and 2016. Analysis of taxonomic association with both current hemoglobin as well as changes showed that another butyrate-producing bacterial genus, *Coprococcus,* was significantly associated with high levels of current hemoglobin as well as hemoglobin increase over time (Spearman rho [0.15:0.21], FDR<0.1, adjusted for age and stool moisture; Figure 2f and Supplementary Table 15-16). The oxygenation-dependent metabolic state of colonocytes was found associated to gut butyrate producers, also linked to outgrowth of opportunistic aerobic species[21]. Therefore, our results suggest that not only current but also historical adequate levels of hemoglobin and oxygen could be of value in retaining a healthy microbiome in later life. Using the

previous clustering approach, the B2/non-B2 ratio was significantly higher in hemoglobin cluster 2 (high-to-low) compared to cluster 4 (low-to-low) (pairwise Chi-Square test, FDR < 0.1; Figure 2g and Supplementary Table 14b) implying that rather a drop of hemoglobin is associated with dysbiosis than having consistently low levels throughout the years. Higher abundance of *Coprococcus* (butyrate producer) in cluster 4 versus cluster 2 is in line with this finding (GLM, FDR < 0.1, adjusted for age and stool moisture; Supplementary Table 17). As an accurate reflection of anemia, low hemoglobin levels can be a feature of recent major bleedings or surgery or of inflammatory, infectious, or neoplastic disorder[22]. Therefore, a drop in hemoglobin levels could be an indication of many conditions as well as poor health. Analysis of changes in ALT between 2005 and 2016 using the same approaches as above showed that only the current ALT levels were significantly associated with *Methanobrevibacter* but not with the enterotype ratio (Spearman rho = -0.18, FDR<0.1, adjusted for age and stool moisture; Supplementary Table 18, Extended Data Fig. 2g).

Next, we estimated the predictive power of past lifestyle and physiological parameters on current microbiome composition. We investigated both effect sizes of single markers and predictive potential of more complex machine learning-based models. To explore single markers, we first investigated long-term predictability by focusing on the three significant individual historical variables from the year 2010 (db-RDA, FDR < 0.1; Figure 2a and Supplementary Table 8d) and enterotypes, but no findings emerged (Kruskal-Wallis, p>0.05; Supplementary Table 19). The lack of statistical significance for these variables implied a limited predictive power by single parameters. Therefore, we sought to determine the predictive power of life history using a combination of variables to predict current microbiome enterotype, as well as to investigate how far back we can use this combined information. To this aim, we applied a Random Forest classifier with feature selection to predict the current enterotype for each past sampling year based on diet, blood parameters, health, anthropometrics, and lifestyle using only variables that were available across all years for parallel comparison. Models derived from a random training dataset (70% of the total) were applied to test data (the remaining 30%) in order to estimate predictive power and avoid overfitting. Models gave fair levels of prediction for the Prevotella (P; for 2000 and 2016) and Bacteroides2 enterotype (B2; for 2010; Figure 3a and Extended Data Fig. 3b). The year 2005 has overall lower predictive power possibly because it is a switching year in age demographics (see Extended Data Fig. 3a). Parallel 10-fold cross-validation identified additional fair levels of prediction for the R enterotype in the year 2010 (Extended Data Fig. 3c). Interestingly, the prediction variables selected for each sampling year showed distinct patterns for the various enterotypes (Figure 3b and Extended Data Fig. 3d). For example, variables belonging to multiple categories were selected for B2, while for the P enterotype a much larger contribution of dietary variables was visible across the years. Unlike the two other enterotypes, in B2 more variables from the medication class were selected in the three recent time points implying an association with health deterioration (Extended Data Fig. 3e and Supplementary Table 20). Overall, our study provides first evidence that the current gut microbiome is indeed predictable by past variables. Nonetheless, further validation in independent cohorts with a similar long-term sampling protocol would be warranted to confirm these results.

In conclusion, we show that an individual's life history has long-term effects on the assembly of the gut microbiome. We report, for the first time, the predictability of the current gut microbiome by historical host parameters along with distinct characteristics of the elderly microbiome using a quantitative approach. In a large-scale, lifespan covering combined cohort (n = 2,519), we confirm a positive correlation between community richness and age. There is compelling evidence that the gut microbiome rapidly expands from birth until early childhood and stabilizes in adulthood[23–25]. Our results indicate that the gut remains dynamic at high ages and is influenced by the host's life history. Specifically, we found that changes in an individual's medication history, non-sport physical activity, and hemoglobin levels over time were linked to the individual's current microbiome. Further, we could predict an individual's current enterotype based on host parameters from as early as 15 years prior. Overall, these results suggest that long-term history of host laboratory parameters, medication, diet, and lifestyle can exert significant impacts on the current microbiome, highlighting first key variables that are important for maintaining a healthy gut at a later life stage.

Main References

1. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
2. Schmidt, T. S. B., Raes, J. & Bork, P. The human gut microbiome: from association to modulation. *Cell* **172**, 1198–1215 (2018).
3. Johnson, A. J. *et al.* Daily sampling reveals personalized diet-microbiome associations in humans. *Cell Host Microbe* **25**, 789-802.e5 (2019).
4. Palleja, A. *et al.* Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nat. Microbiol.* **3**, 1255–1265 (2018).
5. Falony, G., Vieira-Silva, S. & Raes, J. Richness and ecosystem development across faecal snapshots of the gut microbiota. *Nat. Microbiol.* **3**, 526–528 (2018).
6. Kiechl, S. & Willeit, J. In a nutshell: findings from the Bruneck study. *Gerontology* **65**, 9–19 (2019).
7. Vandeputte, D. *et al.* Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551**, 507–511 (2017).
8. Odamaki, T. *et al.* Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiol.* **16**, 90 (2016).
9. Kong, F. *et al.* Gut microbiota signatures of longevity. *Curr. Biol.* **26**, R832–R833 (2016).
10. Biagi, E., Candela, M., Fairweather-Tait, S., Franceschi, C. & Brigidi, P. Aging of the human metaorganism: the microbial counterpart. *Age (Dordr).* **34**, 247–267 (2012).
11. Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184 (2012).
12. Jackson, M. A. *et al.* Signatures of early frailty in the gut microbiota. *Genome Med.* **8**, 8 (2016).
13. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).

14. Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the human body. *Nature* **509**, 357–360 (2014).
15. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science (80-.).* **334**, 105 LP – 108 (2011).
16. Vieira-Silva, S. *et al.* Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. *Nat. Microbiol.* **4**, 1826–1831 (2019).
17. Monda, V. *et al.* Exercise modifies the gut microbiota with positive health effects. *Oxid. Med. Cell. Longev.* **2017**, 3831972 (2017).
18. Codella, R., Luzi, L. & Terruzzi, I. Exercise has the guts: How physical activity may positively modulate gut microbiota in chronic and immune-based diseases. *Dig. Liver Dis.* **50**, 331–341 (2018).
19. Estaki, M. *et al.* Cardiorespiratory fitness as a predictor of intestinal microbial diversity and distinct metagenomic functions. *Microbiome* **4**, 42 (2016).
20. Bressa, C. *et al.* Differences in gut microbiota profile between women with active lifestyle and sedentary women. *PLoS One* **12**, e0171352 (2017).
21. Byndloss, M. X. *et al.* Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science (80-.).* **357**, 570 LP – 575 (2017).
22. Means Jr, R. T. Advances in the anemia of chronic disease. *Int. J. Hematol.* **70**, 7–12 (1999).
23. Cheng, J. *et al.* Discordant temporal development of bacterial phyla and the emergence of core in the fecal microbiota of young children. *ISME J.* **10**, 1002–1014 (2016).
24. Rinninella, E. *et al.* What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* **7**, (2019).
25. Jeffery, I. B., Lynch, D. B. & O'Toole, P. W. Composition and temporal stability of the gut microbiota in older persons. *ISME J.* **10**, 170–182 (2016).

Methods

Study cohort

Faecal samples were collected in the Bruneck Study, a prospective population-based study on the epidemiology and pathogenesis of atherosclerosis launched in 1990 in Bruneck, northwest Italy[6]. The participation and follow-up rates were above 90%. The study has extensive metadata on all individuals since 1990 with comprehensive evaluations every five years up to 2016. The 2016 evaluation was performed with a 6-month delay in spring 2016 rather than in autumn 2015 (as usual) due to delays in ethics approval. The study protocol was approved by the ethics committees of Bolzano and Verona and conforms to the Declaration of Helsinki. All study subjects provided written informed consent. Stool samples (n = 325) were collected at the most recent time point during the 2016 evaluation when study participants were 65 to 98 years old. Metadata collected include anthropometric information, each individual's physician-confirmed medical history and diseases, food intake, lifestyle, vascular risk factors,

medication, and laboratory parameters[6,26–30]. In the survey area, virtually all inhabitants are referred to one local hospital that closely works together with the general practitioners, which allows retrieval of full medical information. Accordingly, in this study information on clinical diseases (current and past) and morbidities as well as medication does not rely on the subject's self-report but was validated by medical records and based on standard diagnostic criteria.

Dietary intake was evaluated by quinquennial (1995, 2000, 2005, 2010, and 2015) dietician-administered 118-item food-frequency questionnaires (FFQ) based on the gold-standard FFQ by Willett and Stampfer[31] and adapted to the dietary peculiarities in the survey area[26,30]. Dieticians made use of illustrative photos of foods when exploring aphasic patients and of information provided by spouses, caregivers, and nursing homes. For each item in the FFQ, a common unit or portion size was specified, and we instructed participants to customize how often on average they had consumed that amount in the past years. The nine response categories ranged from 'never' to 'six or more times a day'. We calculated nutritional intake by assigning a weight proportional to the frequency of use for each food (once per day equals a weight of one), multiplying this weight by the nutrient value for the specified size, and summing the contribution of all foods. Nutrient composition data for foods were based on the US Department of Agriculture Nutrient Database (Release 23) (U.S. Department of Agriculture, Agricultural Research Service. 2010. USDA National Nutrient Database for Standard Reference, Release 23. Nutrient Data Laboratory Home Page, http://www.ars.usda.gov/ba/bhnrc/ndl). We dissected complex foods into component foods utilizing common recipes. Estimates of nutrient intakes were calorie-adjusted. For that purpose, we used the residuals obtained by regressing polyamine or other nutrient intake on total energy intake[30,32]. Reproducibility and validity of the original FFQ are well documented[31] and extend to its application in the Bruneck Study, in which it was compared against nine-day diet records[26,30]. The Alternate Healthy Eating Index (AHEI), a measure of diet quality, significantly associated with the risk of major chronic diseases in a large number of studies, was calculated as described previously[33]. We did not consider the 'duration of multivitamin use component' because multivitamin supplementation was almost absent in our cohort. Accordingly, this index has eight components in our study (vegetable score, fruit score, cereal fiber score, alcohol score, meat ratio score, nuts and soy score, trans-fat score, polyunsaturated-to-saturated fatty acids ratio)[33]. Physical activity was quantified using the Baecke questionnaire[34], rated activity intensities according to the compendium of physical activities, and used these data to calculate average metabolic-equivalent hours per week (overall and separated in sports and non-sports physical activity). Individuals were coded as current smokers or non-smokers (including former smokers) with assessment of pack-years of smoking[29]. Alcohol intake was quantified in grams per day. Body mass index was calculated as weight in kilograms divided by height squared in meters. Systolic and diastolic blood pressures were taken after the participant had been sitting for at least ten minutes, and the mean of three independent measurements was calculated. Hypertension was defined as systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg or the use of antihypertensive drugs. Socioeconomic status was defined on a three-category scale (low, medium, high) based on information about occupational status and educational level of the person with the highest income in the household. Blood samples were taken

in the morning hours after an overnight fast and 12 hours of abstinence from smoking and immediately processed or stored at -70 °C. Diabetes mellitus was diagnosed when fasting plasma glucose exceeded 126 mg/dl or when participants were on anti-diabetic medication. Laboratory parameters were assessed by standard methods in certified labs as detailed previously[6,26–30]. All study participants underwent ultrasound and transient elastography (Fibroscan®, Echosens, France) examination to evaluate hepatic steatosis and liver stiffness. Out of 325, 20 subjects were excluded due to missing data of laboratory parameters, liver stiffness, stool features, and visceral fat thickness (Supplementary Information). Missing data less than 5 was replaced by the cohort mean, otherwise removed throughout the analysis (variables removed: muscle mass (%), metabolic rate, Bristol stool score, and fat mass (kg)). The FGFP cohort used in the present study (n = 2,215) is an expanded version of the first round of sampling completed in 2014 (n = 1,106)[1,35]. FGFP procedures were approved by the medical ethics committee of the University of Brussels–Brussels University Hospital (approval 143201215505, 5/12/2012). A declaration concerning the FGFP privacy policy was submitted to the Belgian Commission for the Protection of Privacy. Written informed consent was obtained from all participants.

DNA extraction and sequencing

Faecal DNA extraction and sequencing were performed as described previously[1]. Briefly, DNA was extracted from 150-200mg of the frozen samples using MagAttract PowerMicrobiome DNA/RNA KF kit (QIAGEN) following the manufacturer's instructions. The V4 region of 16S rRNA genes was amplified using the 515F /806R primer pair and purified using the QIAquick PCR Purification Kit. Sequencing was performed using the Illumina MiSeq platform (MiSeq Reagent Kit v2) and HiSeq 2500 System (151 base pair paired-end) for the Bruneck Study and the FGFP cohorts, respectively.

Microbial load measurement by flow cytometry

Microbial load of the study cohort was measured as described previously[7]. Briefly, 200-250 mg frozen (-80°C) faecal aliquots were diluted in saline solution (0.85% NaCl; VWR International, Germany) and filtered using a sterile syringe filter (pore size of 5 µm; Sartorius Stedim Biotech GmbH, Germany). Next, 1 mL of the microbial cell suspension obtained was stained with 1 µL SYBR Green I (1:100 dilution in DMSO; Thermo Fisher Scientific, Massachusetts, USA) and incubated for 15 min in the dark at 37°C. The flow cytometry analysis was performed using a C6 Accuri flow cytometer (BD Biosciences, New Jersey, USA) based on Prest et al. [14]. Fluorescence events were monitored using the FL1 533/30 nm and FL3 >670 nm optical detectors. The BD Accuri CFlow software was used to gate and separate the microbial fluorescence events on the FL1/FL3 density plot from the faecal sample background. A threshold value of 2000 was applied on the FL1 channel. Based on the exact weight of the aliquots analyzed, cell counts were converted to microbial loads per gram of faecal material.

Relative and quantitative microbiome profiling

After demultiplexing with LotuS v1.565[36], fastq sequences were further processed following the DADA2 microbiome pipeline[37]. Briefly, sequence reads were first filtered and trimmed with the parameters: truncQ=11, truncLen=c(130,200), and trimLeft=c(30, 30). Filtered reads were denoised using the DADA2 algorithm, which infers the sequencing errors. After removing chimeras, amplicon sequence variants (ASVs) table was constructed, and taxonomy was assigned using the Ribosomal Database Project (RDP) classifier implemented in DADA2 (RDP trainset 16/release 11.5). To prepare quantitative microbiome profiling (QMP) table, the relative microbiome profiling (RMP) taxonomic table were then corrected for copy number and rarefied to even sampling depth - a division of sequencing depth by the cell counts – and subsequently multiplied by bacterial cell load to quantify the number of bacteria per gram of faecal sample as previously described in[7]. One subject was further excluded due to low read counts during the data conversion. This way, the sequencing data becomes proportional to the microbial loads in the samples. All analysis was performed based QMP unless otherwise noted.

Faecal moisture content

Moisture content was determined as the percentage of mass loss after lyophilization from 200-300 mg frozen aliquots of non-homogenized faecal material (-80°C). Lyophilization was performed for two days.

Faecal calprotectin measurement

Faecal calprotectin concentrations were determined using the fCAL ELISA kit (Bühlmann, Schönenbuch, Switzerland) on frozen faecal material (-80°C). The level of calprotectin was corrected for the amount of faecal samples used.

Microbiome and statistical analysis.

Statistical and microbiome analysis were performed in R (version 3.6.0)[38] using phyloseq[39], vegan[40], pairwiseAdonis[41], rcompanion[42], CoDaSeq[43], DirichletMultinomial[44], lm.beta[45], ppcor[46], randomForest[47], and Boruta[48] packages. For the associations of the microbiota with any host parameters, taxa found in less than 20% of the population were excluded for noise reduction and alleviation of multiple testing correction. Comparison of 2 groups was performed using Wilcoxon rank sum test, and Kruskal-Wallis test when analyzing more than 2 groups followed by post-hoc Dunn's test. Count data was analyzed by Fisher's exact test. Taxonomic association with host parameters were determined by partial correlation to

adjust for confounders using the R package ppcor[46]. All statistical tests were followed by multiple testing correction using the Benjamini–Hochberg method when testing more than two features.

Elderly microbiome analysis

Community diversity and compositions were evaluated using α- (observed richness, Shannon, and Pielou's index) and β-diversity (principal coordinates analysis on Bray-Curtis dissimilarity) at the genus level using phyloseq[39] and vegan[40] based on RMP. Partial correlation adjusting for gender, BMI, stool moisture, and antibiotic intake was used to analyze the associations of host age with the current microbiota. The correlation analysis was carried out based on centered log-ratio (CLR) transformed taxonomic table to remove compositional artifacts in the RMP data. Prior to the log transformation, zeros were imputed using the minimum relative abundance for each taxon[49]. Data transformation was performed independently for each cohort, and genera commonly found between the cohorts were further analyzed. Core taxa for the Bruneck cohort were defined as genera prevalent in greater than 95% of the population. As the core taxa are largely influenced by different sample size, 50 repetitions of subsampling were performed without replacement randomly selecting 15 subjects in each age category to match the small sample size of subject in their nineties (n =11).

Analysis of community variations using the current and past variables

Explanatory power of cohort covariates and their combined effect size for the microbial community variation was evaluated as described previously[1]. Briefly, distance-based RDA (db-RDA) was performed on genus level using Bray-Curtis dissimilarity as implemented in vegan[40]. Covariates (FDR < 0.1) found in this step were entered for forward stepwise model selection to measure their cumulative effect sizes. Prior to the analysis, collinearity of variables was checked by Spearman's rank correlation and Wilcoxon rank sum test for continuous and binary variables, respectively. One of the collinear variables was removed based on their representativeness and explanatory power if their effect size was > |0.8| (Supplementary Table 4). To identify variables driving the elderly gut community variation as the host become older, subjects were divided into equal size windows based on their age (n =156). To assess the effect of past events/host parameter shifts on the current microbiome variation, different approaches were performed for continuous and binary variables (infection, medication, and smoking). For continuous variables, variable shifts between each time point and the year 2016 were calculated by subtracting the values. History of the categorical binary variables was determined by summing the event occurred between the two-time points. Smoking was taken as smoking history if the subjects were current smokers at the time point. Comparison of past and present non-redundant effect size was performed by likelihood-ratio test.

Associations of the past with the current microbiome

Enterotyping based on the Dirichlet Multinomial Mixtures (DMM) approach was performed as described by Holmes et al.[50] on a genus-abundance RMP matrix using the R package DirichletMultinomial[44]. Taxonomic association analysis after adjusting for age and stool moisture was performed by fitting generalized linear model (GLM, link = logit). Beta blocker treatment and hemoglobin clusters were used as binary dependent variables and genera as independent variables. Standardized β coefficients were calculating using the R package lm.beta[45]. The significant associations of deconfounded genera with the host parameters were tested by performing likelihood-ratio test. Clustering of subjects was carried out by categorizing them as high or low based on the median values measured in 1990 and 2005 for hemoglobin and non-sport physical activity, respectively. A multiple linear regression was performed on the non-sport physical activity hemoglobin, and alanine transaminase regressing out the effect of age, gender/BMI. Prior to the regression, physical activity was transformed by inverse normal transformation to fit normal distribution.

Prediction of the current microbiome based on the life history

In order to construct a prediction model, multinomial random forest was performed using default parameters as implemented in the R package randomForest[47]. The model was trained with randomly selected 70% of the entire dataset and validated using the remaining 30% of the dataset. Variables entered were the ones available across all years from 1995 to 2016. Additional steps of significant feature selection for each enterotype were performed to using the Boruta package[48], which implements the random forest classifier and identifies significant features by random shuffling and comparing z-scores. Boruta presented the features by 3 categories (i.e., confirmed, tentative, and rejected), and only the confirmed ones were presented in the study. Tenfold cross-validation was used to validate the results obtained from the training and test datasets (R package randomForest[47], ntree= 500 and default mtry).

Data availability

Raw 16S data is available through managed access at the European Genome/Phenome Archive (https://ega-archive.org) under accession number EGAS00001004453. Derived species abundance counts and transformed microbial trait data can be found in Supplementary Table 21. Bruneck host metadata from this study are available in accordance and in consent with ethical permission through managed access, and organized via Principal Investigator Herbert Tilg, as follows: Upon data request by email to Herbert.tilg@i-med.ac.at the Bruneck data access committee will evaluate access permission, which will be granted upon signature of a data use agreement/material transfer agreement between the governing legal entities.

Additional information

Extended Data, supplementary information, and source data are available for this paper.

Methods references

26. Eisenberg, T. *et al.* Cardioprotection and lifespan extension by the natural polyamine spermidine. *Nat. Med.* **22**, 1428—1438 (2016).
27. Kiechl, S. *et al.* Blockade of receptor activator of nuclear factor-κB (RANKL) signaling improves hepatic insulin resistance and prevents development of diabetes mellitus. *Nat. Med.* **19**, 358–363 (2013).
28. Kiechl, S. *et al.* Toll-like receptor 4 polymorphisms and atherogenesis. *N. Engl. J. Med.* **347**, 185–192 (2002).
29. Kiechl, S. *et al.* Active and passive smoking, chronic infections, and the risk of carotid atherosclerosis: prospective results from the Bruneck Study. *Stroke* **33**, 2170–2176 (2002).
30. Kiechl, S. *et al.* Higher spermidine intake is linked to lower mortality: a prospective population-based study. *Am. J. Clin. Nutr.* **108**, 371–380 (2018).
31. Willet, W. C. *et al.* Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am. J. Epidemiol.* **122**, 51–65 (1985).
32. Willett, W. & Stampfer, M. J. Total energy intake: implications for epidemiologic analyses. *Am. J. Epidemiol.* **124**, 17–27 (1986).
33. McCullough, M. L. *et al.* Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *Am. J. Clin. Nutr.* **76**, 1261–1271 (2002).
34. Baecke, J. A., Burema, J. & Frijters, J. E. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am. J. Clin. Nutr.* **36**, 936–942 (1982).
35. Hughes, D. A. *et al.* Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. *Nat. Microbiol.* **5**, 1079–1087 (2020).
36. Hildebrand, F., Tadeo, R., Voigt, A. Y., Bork, P. & Raes, J. LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome* **2**, 30 (2014).
37. Callahan, B. J. *et al.* DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581 (2016).
38. Team, R. C. R: a language and environment for statistical computing. R Foundation for Statistical Computing. (2019).
39. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).

40. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
41. Barnier, J. pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4. (2020).
42. Mangiafico, S. Functions to support extension education program evaluation. (2020).
43. Gloor, G. B. & Reid, G. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can. J. Microbiol.* **62**, 692–703 (2016).
44. Morgan, M. DirichletMultinomial: dirichlet-multinomial mixture model machine learning for microbiome data. R package version 1.26.0. (2019).
45. Behrendt, S. Add standardized regression coefficients to lm-objects. R package version 1.5-1. (2015).
46. Kim, S. ppcor: an R Package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. methods* **22**, 665–674 (2015).
47. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R news* **2**, 18–22 (2002).
48. Kursa, M. B., Jankowski, A. & Rudnicki, W. R. Boruta–a system for feature selection. *Fundam. Informaticae* **101**, 271–285 (2010).
49. Valles-Colomer, M. *et al.* The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat. Microbiol.* **4**, 623–632 (2019).
50. Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* **7**, e30126 (2012).

Declarations

Acknowledgments

We would like to thank all study participants for their valuable contribution. The Bruneck Study is supported by the “Pustertaler Verein zur Vorbeugung von Herz- und Hirngefässerkrankungen,” the “Gesundheitsbezirk Bruneck,” and the “Sanitätsbetrieb Südtirol,” province of Bolzano, Italy, and received support from the excellence initiative (Competence Centers for Excellent Technologies – COMET) of the Austrian Research Promotion Agency FFG: “Research Center of Excellence in Vascular Ageing” (K-Project No. 843536 and K-Centre 868624) funded by the Austrian Ministry for Transport, Innovation and Technology, the Austrian Ministry for Digital and Economic Affairs and the federal states Tyrol (via Standortagentur), Salzburg and Vienna (via Vienna Business Agency). This work was co-funded by VIB, the Rega institute for Medical Research, KU Leuven and the FWO/F.R.S.-FNRS under the Excellence of Science (EOS) programme (MiQuant/30770923).

Author contributions

H.T., S.K., J.S., A.G., and J.R. conceived the study objectives and study design. H.T., F.G., M.W., M.N., C.L., P.S., G.R., J.W., P.W., R.P., and S.K. coordinated recruitment and sample collection. L.D. assisted in project

coordination. L.R. carried out faecal microbial DNA extraction and sequencing. Faecal moisture, calprotectin, and cell counts were measured by J.S. Data preprocessing was done by S.P. and J.S. Statistical analyses were designed and executed by J.S. and J.V. The draft manuscript was prepared by J.S. and J.R, and revised by all authors.

Competing interests

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to J.R. and H.T.

Figures

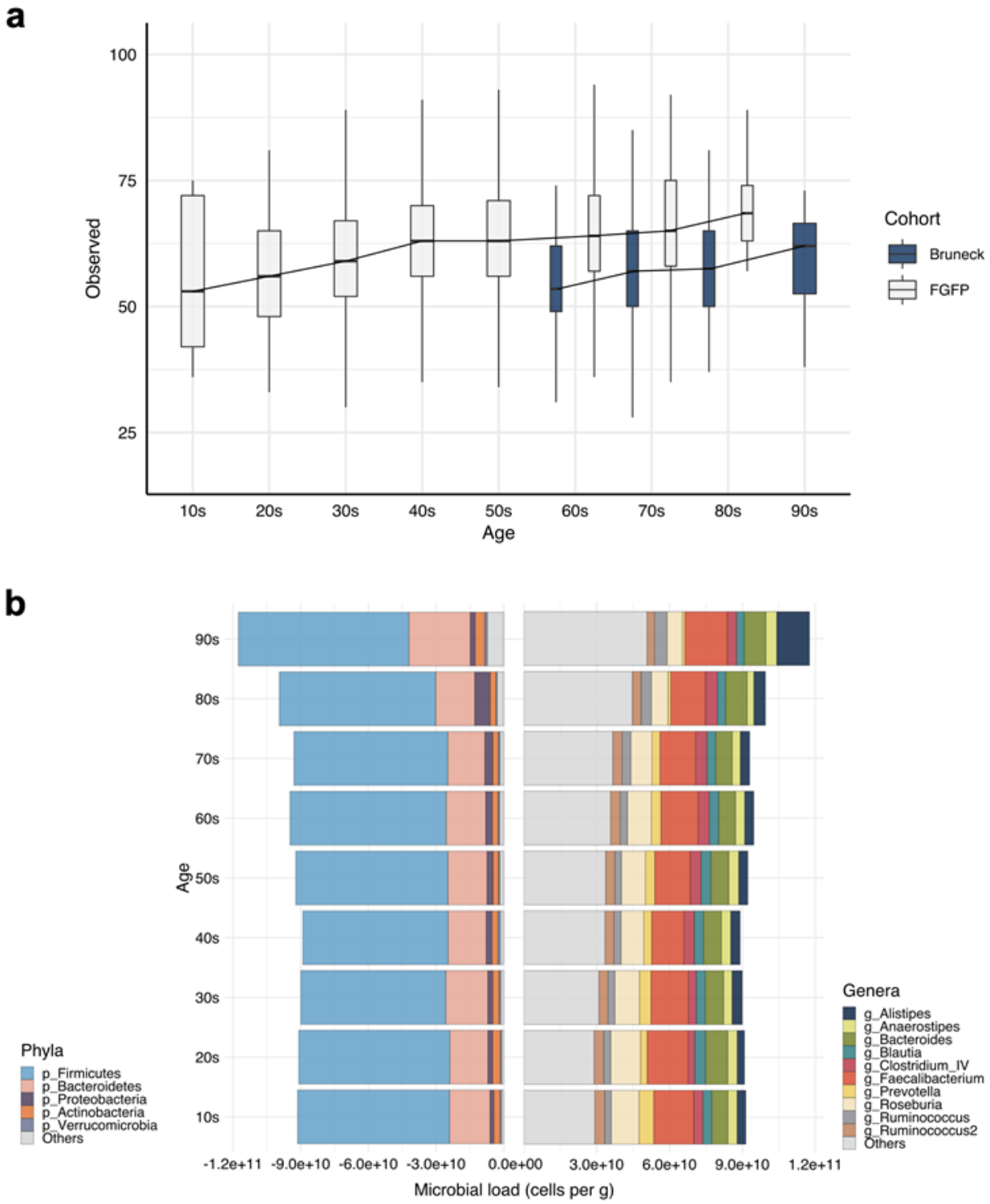

Figure 1

Microbial composition of population cohorts. (a) Correlation of community richness with age using the Bruneck Study cohort n = 304 (rho = 0.12, p = 0.03) and the Flemish Gut Flora Project cohort (FGFP) cohort n = 2,215 (rho = 0.18, p < 0.001) Boxes represent the 25th percentile, median, and 75th percentile. Whiskers represent the lowest and highest values of the richness. (b) Bacterial abundances at the level of

phylum and genus across decadal age groups in the Bruneck Study and FGFP cohorts. Analysis were based on relative microbiome profiling (RMP).

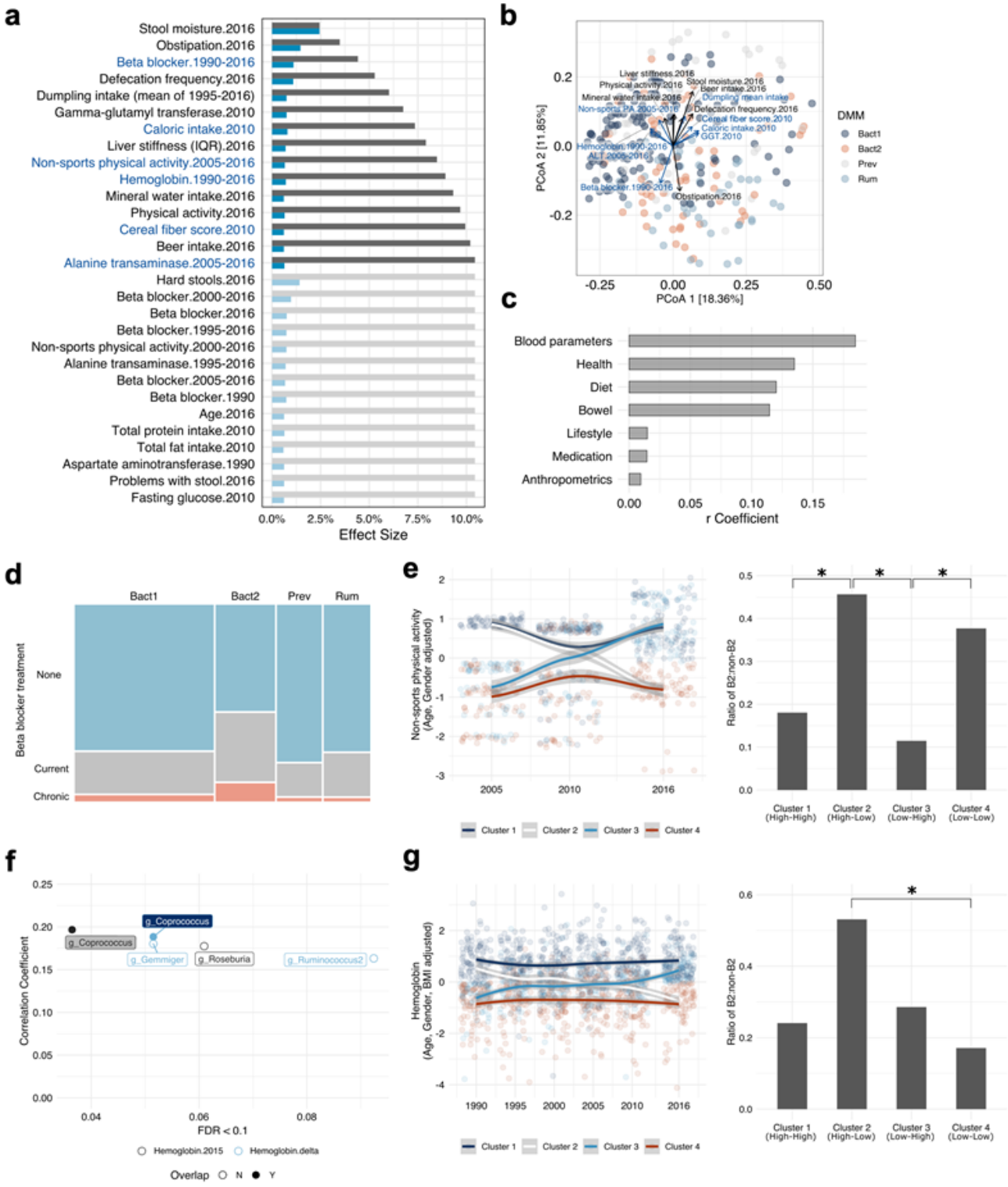

Figure 2

Explanatory variables for the microbial community variations in the Bruneck Study cohort. (a) Individual and cumulative effect size of contemporary and historical covariates. Dark colored upper bars indicate individual (upper bar) and cumulative (down bar) effect sizes of variables included in the forward

stepwise RDA model. Historical covariates are denoted in blue labels. (b) Principal coordinate analysis (PCoA) based on Bray-Curtis dissimilarity. Arrows indicate significant covariates that can significantly explain the current microbiome variation. (c) Combined effect size of significant contemporary and historical covariates. (d) Prevalence of enterotype by beta blocker treatment (Fisher's exact test p = 0.046). Chronic: treatment of beta blocker both in 1990 and 2016, current: currently medicated, and none: not medicated in 1990 and 2016. (e) Clusters of non-sport physical activity across the years. (f) Correlation of hemoglobin with current bacterial abundances after adjusting for age and stool moisture. Color filled labels indicate taxa overlapping between historical and current levels of hemoglobin (partial correlation, FDR < 0.1). (g) Clusters of hemoglobin across the years. Cluster 1: high in the past and at present; Cluster 2: high in the past and low at present; Cluster 3: low in the past and high at present; Cluster 4: low in the past and at present. Comparisons of ratio of B2 and non-B2 by clusters were plotted by bar graphs. *: pairwise Chi2 test, FDR < 0.1.

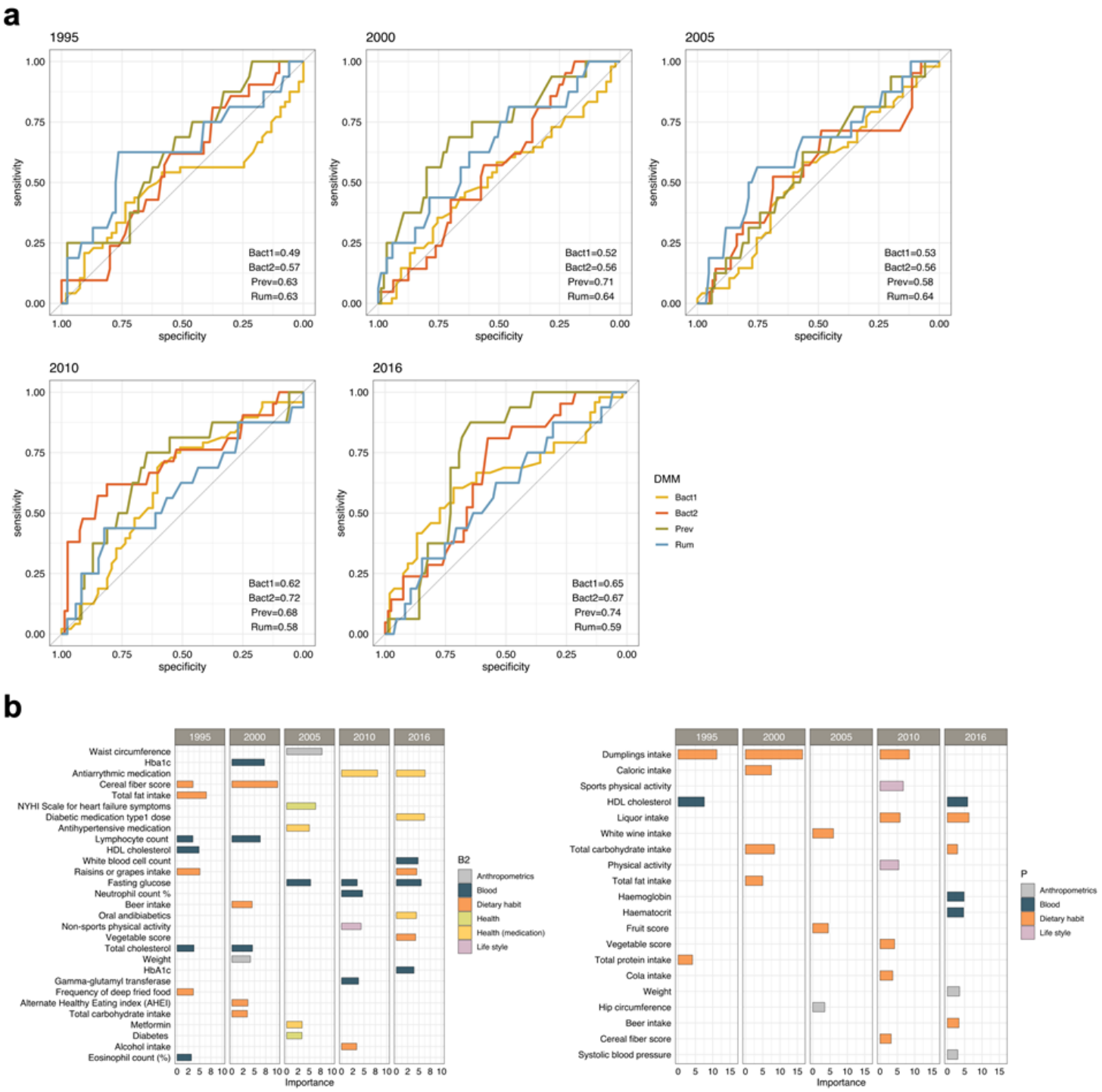

Figure 3

Prediction of current microbiome using past variables. (a) Receiver operating curve (ROC) for the evaluations 1995, 2000, 2005, 2010 and 2016 using the test dataset (n =101). (b) Variables selected for B2 and P enterotypes using the training dataset (n = 203). Colors indicate different categories of the variables.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Brunecksupplementary.infoJS.pdf
- Bruneckextended.dataJS.pdf
- Brunecksupplementary.tableJS.xlsx