

ARTICLE OPEN



Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation

Stefano Romano¹✉, George M. Savva¹, Janis R. Bedarf^{1,2}, Ian G. Charles^{1,3}, Falk Hildebrand^{1,4}✉ and Arjan Narbad¹

The gut microbiota is emerging as an important modulator of neurodegenerative diseases, and accumulating evidence has linked gut microbes to Parkinson's disease (PD) symptomatology and pathophysiology. PD is often preceded by gastrointestinal symptoms and alterations of the enteric nervous system accompany the disease. Several studies have analyzed the gut microbiome in PD, but a consensus on the features of the PD-specific microbiota is missing. Here, we conduct a meta-analysis re-analyzing the ten currently available 16S microbiome datasets to investigate whether common alterations in the gut microbiota of PD patients exist across cohorts. We found significant alterations in the PD-associated microbiome, which are robust to study-specific technical heterogeneities, although differences in microbiome structure between PD and controls are small. Enrichment of the genera *Lactobacillus*, *Akkermansia*, and *Bifidobacterium* and depletion of bacteria belonging to the Lachnospiraceae family and the *Faecalibacterium* genus, both important short-chain fatty acids producers, emerged as the most consistent PD gut microbiome alterations. This dysbiosis might result in a pro-inflammatory status which could be linked to the recurrent gastrointestinal symptoms affecting PD patients.

npj Parkinson's Disease (2021)7:27; <https://doi.org/10.1038/s41531-021-00156-z>

INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease¹. Globally, it has an incidence of 10–50/100,000 person/year and a prevalence of 100–300/100,000 people, and due to the progressive aging of the world population, the number of people with PD is expected to double by 2030¹. PD affects predominantly dopaminergic neurons in the brain, leading to decreased dopamine levels and motor impairments, such as tremor, rigidity, balance difficulties, and loss of spontaneous movement (akinesia)². Its pathological hallmark has long been considered to be the intracellular deposition of aggregated α -synuclein, leading to neuronal cell death and neuroinflammation³. PD is now considered a multi-systemic disease, affecting the central as well as the peripheral nervous system (CNS, PNS), resulting in several non-motor symptoms, often including gastroparesis or constipation. Due to the early involvement of the gastrointestinal tract, often preceding motor symptoms for years⁴, changes in gut-microbiota composition have been studied in relation to the pathophysiology of PD. The potential role of gut microbiota in PD³ and other neurodegenerative diseases⁵ is supported by animal studies⁶, showing that the microbiota can affect α -synucleinopathy as well as neuroinflammation. Thus, the microbiota is a putative therapeutic target and has the potential for developing diagnostic biomarkers.

PD patients can have increased gut permeability and inflammation^{7,8}, and these have been hypothesized to be linked to low gastrointestinal short-chain fatty acids (SCFA) concentrations⁹. SCFA are the end products of bacterial fermentation of dietary components and play a pivotal role in fueling and maintaining the integrity of the colonic epithelium. Low levels of SCFA have been considered to be a consequence of a decreased abundance of SCFA-producing taxa in PD patients^{10,11}. To date, more than 20 case-control studies have investigated the composition of the PD gut microbiota. Over 100 differently abundant taxa between PD

patients and controls have been reported^{10–19}, and in some cases, an association between taxa abundances and disease severity has been detected^{11,12,16}. Several studies suggested that individuals with PD have an altered gut microbiota compared to controls, even though findings are often inconsistent and a consensus on the taxa associated with the disease is still lacking. Across most studies, the genus *Akkermansia* and the Verrucomicrobiaceae family have been found to be enriched in PD patients, while bacteria belonging to the Lachnospiraceae family are depleted. On the other hand, various inconsistencies have been found among the different sampling cohorts. For example, the Lactobacillaceae family has been generally detected to be enriched in PD in the Western cohorts but never in Chinese studies^{18,20,21}. Similarly, conflicting results have been obtained for bacteria within the Prevotellaceae family. Several studies detected these taxa to be highly depleted in PD patients^{16,17,22,23}, compared to controls, whereas others found no differences in abundances¹¹ or found these taxa enriched in PD patients^{13,20}.

Inconsistencies amongst studies might arise from variations in study designs and methods used for producing and analyzing 16S rRNA-gene amplicon data, as well as from the natural variability of the gut microbiota across populations, lifestyles, and diets²⁴. To further elucidate the significance of changes in the intestinal microbiota composition in PD and to evaluate its potential as a biomarker for PD risk, diagnosis, and prognosis it is important to perform cross-study comparisons and identify disease-specific alterations. Here, we provide an individual patient data meta-analysis (pooled re-analysis) of all ten available studies that described the gut microbiome in PD through 16S rRNA-gene amplicon sequencing. We apply a standardized workflow to analyze each study individually and combined different statistical approaches to identify the major changes affecting the gut microbiome of PD patients across sampling cohorts.

¹Quadram Institute Bioscience, Norwich Research Park, Norwich, UK. ²Department of Neurology, University of Bonn, Bonn, Germany. ³University of East Anglia, Norwich Research Park, Norwich, UK. ⁴Earlham Institute, Norwich Research Park, Norwich, UK. ✉email: stefano.romano@quadram.ac.uk; falk.hildebrand@quadram.ac.uk

RESULTS

Study selection

We identified a total of 22 studies that investigated the PD-associated gut microbiome using 16S rRNA-gene amplicon sequencing (March 2020; Supplementary Table 1). Of these, only ten made raw sequencing data available and could be re-analyzed in our study. These ten studies covered nine different cohorts (one was reported at baseline then at follow-up 2 years later), across six countries (Table 1). Overall, this resulted in 1269 samples (1211 after filtration of samples with low read counts) (Fig. 1a) all obtained from case-control studies. Cases were usually selected from clinics local to the investigating teams, were at different stages of the disease, and almost all were using some form of PD medication (Table 1). Controls were typically sampled by convenience from the local population or from families of the PD patients. All studies except one¹⁶ applied the UK brain bank criteria to define PD. Cases had an average age of between 60 and 70 years in all studies, with controls typically well-matched in age. Some studies matched on sex, while for others there were significantly more males in the case compared to control groups^{12,13}, and significantly more cases than controls (Fig. 1a).

Various sampling protocols were used across studies, with considerable variation in the methods adopted to preserve the samples before processing (Table 1). In some cases, samples were kept at room temperature for up to 48 hours before analysis¹⁴, in others, samples were stored either in DNA preservative^{16,19} or on ice¹⁸. DNA extractions and sequencing strategies also varied across studies (Table 1). The Illumina MiSeq platform was the most used sequencing technology, but the chosen variable region of the 16S gene and sequencing strategy (paired-end vs single-end) varied considerably (Table 1). Considering the heterogeneity across studies, we first re-analyzed every single dataset individually, then we used a combination of statistical approaches to obtain a consensus overview of the gut microbiome in PD accounting for the heterogeneity between studies (Fig. 1b). Two studies were based on the same cohort measured at different time points, hence, we performed a sensitivity analysis by comparing the results obtained considering both datasets with those obtained after omitting the baseline samples¹⁵.

The gut microbiome differs significantly between PD patients and controls

Measures of microbial alpha-diversity and abundances of rare taxa, based on species profiles, were higher in PD samples compared to controls in three out of ten studies (Supplementary Fig. 1). Interestingly, these three datasets were the only ones using single-end sequencing approaches, suggesting that this might influence the estimation of bacterial diversity. These differences in diversity were statistically significant when we pooled estimates across studies using random-effect meta-analysis (Fig. 2 and Supplementary Fig. 1). Specifically, PD samples had a higher overall richness as indicated by a significantly higher number of observed species and higher Fisher's alpha, ACE, and Chao1 indices (Fig. 2 and Supplementary Fig. 1). Our analyses suggest that this higher diversity might derive from a decrease in the abundance of dominant species and an increase in rare/low abundant ones, as dominance indices were lower and a rarity index was significantly higher in PD patients (Fig. 2 and Supplementary Fig. 1). Previous studies reported a higher abundance of Firmicutes in control samples compared to PD¹¹, and the Firmicutes to Bacteroidetes ratio (F/B ratio) has been frequently used to assess gut health. Therefore, we calculated F/B ratios across all studies. Only in the study of Keshavarzian et al.¹¹, we observed a significant difference in the F/B ratio between PD and control, but this difference was not significant overall (Supplementary Fig. 2). Similarly, Aho et al.¹⁶ reported that controls had an increased *Prevotella* to *Bacteroides* ratio (P/B ratio)

in the baseline samples¹⁵ of their longitudinal study. We confirmed this result and detected an increased P/B ratio in one additional cohort but not in the rest of the datasets, and there was only weak evidence for a higher P/B ratio in controls when results were pooled (Supplementary Fig. 2). Omitting the baseline samples from the longitudinal Finnish cohort did not alter the conclusions of the alpha-diversity analyses, and attenuated the P/B ratio difference between PD and controls (data not shown).

The genera *Bacteroides* and *Prevotella* and the Firmicutes phylum are key gut-microbiota taxa and their abundances have been used to define three enterotypes, ET_B, ET_P, and ET_F, representing the *Bacteroides*, *Prevotella*, and Firmicutes enterotypes, respectively²⁵. To verify the prevalence of these gut microbiome types among PD patients, we assigned each microbiome to one of the three known enterotypes, classifying 592 samples. The distribution of assigned enterotypes varied enormously across studies, but there was no significant difference between PD and control samples in any individual study and no trend toward specific enterotypes when studies were considered together (Supplementary Fig. 3). For example, the majority of PD samples from the Finnish cohort^{15,16} were assigned to the ET_F, whereas in the study of Hill-Burns et al.¹² and Pietrucci et al.¹⁹ most of the PD samples were classified as ET_B. Interestingly, only in the Finnish cohorts was the ET_P more common among control samples, in agreement with the authors' finding of *Prevotella* being enriched in the control groups.

Considering the variability among studies and the potential data-dependent effect of different microbiome analysis workflows²⁶, we used a thorough and comprehensive approach to investigate the structure of the bacterial communities associated with PD (Fig. 1b). We used three independent normalization strategies (Variance Stabilizing Transformation, VST; Total Sum Scaling, TSS, Centered Log-Ratio, CLR) combined as appropriate with three beta-diversity distances (Bray-Curtis, BC; Jensen-Shannon divergence, JSD; Euclidean) and statistical testing via permutational multivariate analysis of variance using distance matrices (PERMANOVA). We applied these strategies to all three taxonomy ranks we considered (species, genus, and family). In most studies, disease status significantly explained the differences within the data, even though it accounted only for a limited portion of data variability (from <1% to <13%; R^2 expressed in percentage; Fig. 3). When the datasets were pooled, both study and disease status statistically significantly explained the separation of the samples, but the proportion of variance explained by the disease status was in all cases <1% (Fig. 3), whereas the study explained between 28 and 53% of the variance.

We wanted to verify whether underlying differences, unrelated to the origin of the sampling cohorts, existed between the gut microbiome of PD patients and controls. Moreover, we aimed at identifying which study-specific factors most defined the differences across datasets. To reach these goals, we used the normalization-distance pairs which best captured the variability of the data (Fig. 3; JSD-TSS) to perform a distance-based redundancy analysis (dbRDA) on the pooled data (Supplementary Figs 4–6). First, we ordinated the combined data without constraints and without accounting for the variability introduced by the study. In accordance with the previous PERMANOVA analyses, the separation of the samples was driven by the study of origin (Supplementary Figs 4–6). We then inferred the degree of similarities between studies using the sample coordinates in the dbRDA (Supplementary Fig. 7, see "Methods" for details). The only four strongly divergent datasets were from Weis et al.²⁷, who used the sequencing platform IonTorrent; Hopfner et al.¹⁴, who maintained the samples at room temperature up to 48 h and analyzed them using the V1–V2 variable region of the 16S rRNA gene; Keshavarzian et al.¹¹, who collected samples in anaerobic pouches; and from Heintz-Buschart et al.¹³, who immediately flash-frozen the samples after collection and used a lab-specific

Table 1. Technical details of the studies included in the meta-analysis.

Study	Design	Disease duration and medications ^a	Country	Sampling	DNA extr.	16S region	Seq. Tech. ^b	PE vs SE ^c	Accession number
Hill-Burns et al. ¹²	Case-control	13.7 ± 6.5 years, 93.9% medicated	USA	Swabs, delivered at RT	MO BIO's PowerMag Soil Kit	V4	MiSeq	SE	ERP016332 ^d
Hopfner et al. ¹⁴	Case-control	11.2 ± 4.8 years, all medicated	Germany	Home or Hospital collection, delivered at RT	Power Soil Kit	V1-V2	MiSeq	PE	PRJEB14928 ^e
Keshavarzian et al. ¹¹	Case-control	6.4 ± 4.7 years, 68.4% medicated	USA	Home collection, delivered in GasPak	FastDNA Soil Kit	V4	MiSeq	SE	PRJNA268515 ^e
Petrov et al. ¹⁷	Case-control	NA	Russia	NA	Mechanical disruption	V3-V4	Miseq	SE	Obtained from the authors
Pietrucci et al. ¹⁹	Case-control	8.1 ± 4.5 years, 82.5% on l-dopa	Italy	Home collection, DNA stabilizer	PSP-Spin Stool Kit	V3-V4	MiSeq	PE	PRJNA510730 ^e
Qian et al. ¹⁸	Matched case-control	5.7 ± 4.1 years, all medicated	China	Home collection, transferred in Ice	QIAm DNA stool Mini Kit	V3-V4	MiSeq	PE	PRJNA391524 ^e
Aho et al. ¹⁶	Matched case-control	Median 8.5 years, 73% on l-dopa	Finland	Home collection, DNA stabilizer, stored in the fridge	PSP-Spin Stool Kit	V3-V4	MiSeq	PE	PRJEB27564 ^d
Scheperjans et al. ¹⁵	Matched case-control	Median 6.5 years, all but 2 using dopamine agonist	Finland	Home collection, DNA stabilizer, stored in the fridge	PSP-Spin Stool Kit	V3-V4	MiSeq	PE	PRJEB27564 ^d
Heintz-Buschart et al. ¹³	Case-control	72 ± 31 months, 85% on l-dopa agonist	Germany	Home collection, flash-frozen	Modified Qiagen Allprep	V4	MiSeq	PE	PRJNA381395 ^e
Weis et al. ²⁵	Matched case-control	82 ± 56 months, 85.3% medicated	Germany	MED AUXIL fecal collector set	FastDNA Spin Kit	V4-V5	IonTorrent	PE	PRJEB30615 ^d

The table reports the following information: study design (a: disease duration at sampling reported as mean and respective standard deviation, and proportion of patients receiving PD-medications), sample collection, DNA extraction, and sequencing techniques (b: paired-end vs single-end). NA indicates that the information was not mentioned in the original article. RT indicates room temperature. In many studies, the proportion of medicated patients was reported for the individual types of drugs. Hence, we could not estimate the total percentage of patients undergoing pharmaceutical treatments, and we report here only the type of drugs with the highest proportions of medicated patients. Accession numbers refer to d: European Nucleotide Archive (ENA), or e: Sequence Read Archive (SRA).

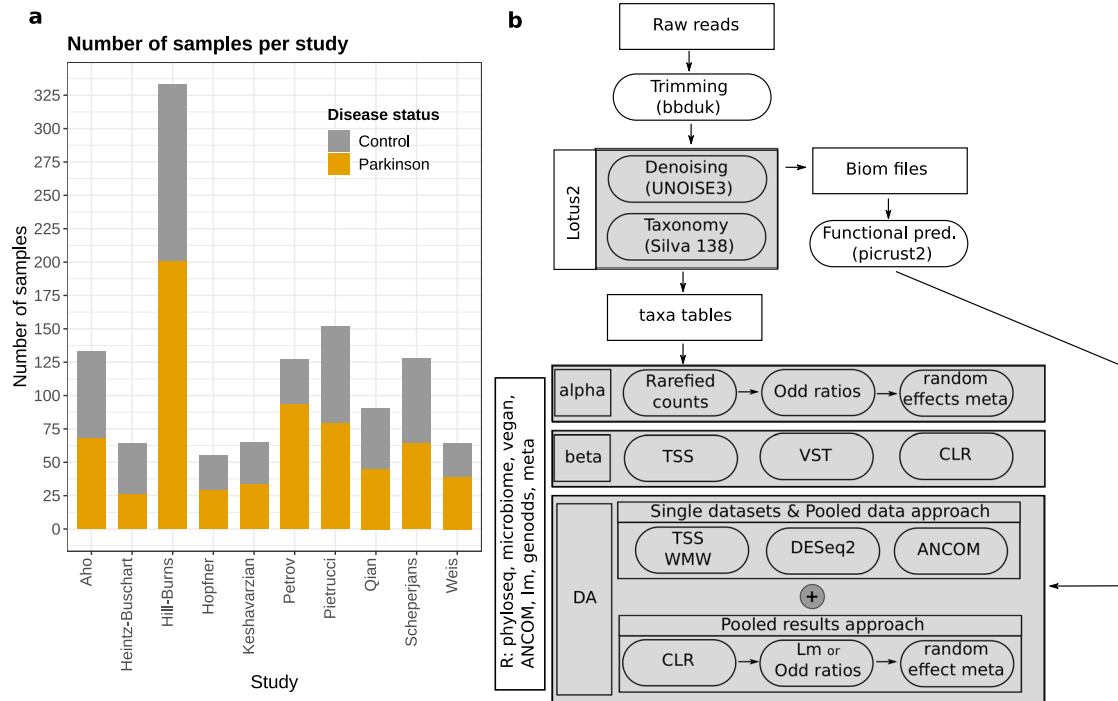


Fig. 1 Sample distribution across studies and bioinformatic workflow adopted in our study. **a** The number of control and PD samples refers to the data that could be recovered from the Sequence Read Archive (SRA) or the European Nucleotide Archive (ENA) and passed the quality filtering we applied. **b** Overview of the bioinformatic workflow adopted in our study. TSS total sum scaling, VST variance stabilizing transformation, CLR centered log ratios, WMW Wilcoxon–Mann–Whitney test, ANCOM analysis of the composition of microbiomes, lm linear models, genodds Agresti’s generalized odd ratios, meta random-effect meta-analysis.

DNA extraction protocol. We then verified which study-specific aspects (e.g., sequencing strategy, DNA extraction) most influenced the structure of the bacterial communities. We created additional dBRDAs constraining the data by country of origin and other known study-specific technical factors. Each factor significantly shaped the clustering of the data when considered individually, and this was observed for all taxonomic ranks and normalization approaches we used (Supplementary Data 1). In general DNA extraction protocols, country of origin, and 16S variable region were the factors that explained most of the variance within the data (Supplementary Data 1). Finally, we compared the dBRDA models constrained by all confounding factors and disease status with the one constrained only by the disease status and study and verified that both models explained the same proportion of data variability (Supplementary Data 1). Hence, removing the influence of the study will simultaneously eliminate the effects of other known study-specific technical confounders.

Accounting for the variability introduced by the study within the dBRDA drastically decreased the batch effect, irrespectively of the normalization-distance pair used (Supplementary Figs. 4–6). However, samples did not cluster according to the disease status, suggesting that the environmental variability is higher than the variability explainable by the disease. Therefore, we constrained the dBRDAs conditioned for study by disease status, to maximize the divergence between PD and control samples. We used this approach to determine the major taxa driving the separation between conditions. In accordance with the above results, the newly created constrained axis along which PD and controls diverged, significantly explained the clustering of the data (ANOVA-like permutation test; Fig. 4, Supplementary Figs. 4–6 and 8), but accounted for only <1% of the data variability. Since the constrained ordination obtained for the TSS–JSD pair explained a slightly higher proportion of variance, we selected

this approach to identify taxa that strongly influenced the separation of the samples. The divergence between PD and controls was mainly driven by the bacterial families Bifidobacteriaceae and Akkermansiaceae, which were more enriched in PD, and Lachnospiraceae, which was more abundant in control samples (Supplementary Fig. 8). Similarly relevant, but with a minor difference between conditions were the families Rikenellaceae, Porphyromonadaceae, Christensenellaceae, Oscillospiraceae, and the *Clostridium methylpentosum* group in the Oscillospirales order, all of which were more enriched in PD (Supplementary Fig. 8). These results were mirrored in the dBRDA performed using genus and species abundances, which revealed that species in the *Akkermansia* and *Bifidobacterium* genera were strongly enriched in PD, whereas several species belonging to the Lachnospiraceae family caused the divergence of control samples (Supplementary Fig. 8 and Fig. 4). When we omitted the baseline samples of the longitudinal Finnish cohort¹⁵, the overall results did not change, and only minor differences were observed (some genera and species in the Lachnospiraceae family were not detected as main drivers of sample separation, whereas the *Faecalibacterium* genus and a species within this genus were additionally detected as influencing the divergence of controls).

The gut microbiome of PD patients and controls are enriched in different bacterial groups

In the first instance, we analyzed all ten datasets individually using three separate approaches. The number of taxa that showed a statistically significant difference in abundance between PD and controls varied greatly across studies and methods (Supplementary Fig. 9 and Supplementary Data 2). Among all studies, the highest number of significant taxa were detected in the datasets of Hill-Burns et al.¹² and Petrov et al.¹⁷, whereas the lowest number was observed in Qian et al.¹⁸ and Hopfner et al.¹⁴

(Supplementary Fig. 9 and Supplementary Data 2). To obtain a generalizable overview of the PD-associated microbiome, we combined two independent approaches that we refer to as the Pooled data and the Pooled results approach. In the first, we

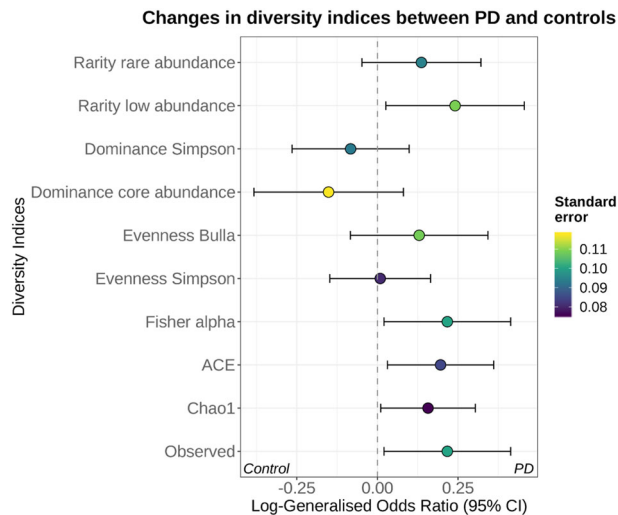


Fig. 2 Alpha-diversity indices are significantly different between PD patients and controls. Indices were calculated at the species level for each dataset. Results were then combined using a random-effect meta-analysis approach. The log-generalized Odds Ratios indicate the degree of variation of each index between controls and PD. The richness of the samples was estimated using the observed number of species and the indices Chao1, ACE, and Fisher's alpha. To estimate evenness, which indicates how different the species abundances in a community are from each other, we used the Bulla and Simpson indices. Finally, we estimated dominance, which describes how much one or few species dominate the community, and rarity, which assesses the number of species with low abundance in the samples. The data suggest that the gut microbiota of PD patients is more diverse (higher richness) than controls and this is likely a consequence of an increase in rare taxa (rarity).

pooled the count tables obtained for each study and used the same three methods we applied to the individual datasets and statistically accounted for the variability introduced by the study. Taxa were considered differentially abundant between PD and controls when detected as statistically significantly different by two out of three methods. This first list of taxa was then merged with the outcome of the Pooled results approach, in which we first estimated the differences in abundance for each taxon in the individual datasets and then used random-effect meta-analysis to pool the results (Fig. 5, Supplementary Figs. 10, 11, and Supplementary Data 3).

After obtaining the first list of differentially abundant taxa, we used the metadata made available by five of the ten studies we re-analyzed to verify whether age and/or gender influenced the abundances of the taxa we detected to be enriched or depleted in PD. By comparing generalized linear mixed models (GLMMs) with and without accounting for the disease status, we identified taxa that showed a significant alteration in abundance in PD irrespective of age and gender (Supplementary Data 4). In the majority of the cases, the disease status was required to best explain taxa abundances, underlining the robustness of our approaches. For 14 species, 12 genera, and 6 families, models not containing the factor disease status were selected among the best GLMMs ($\Delta AIC < 2$). Hence, we concluded that the available data do not allow to clearly establish whether the disease status is essential to explain the abundances of these taxa (Supplementary Data 4), and we did not consider them further. For example, the Prevotellaceae family and species within this family (Supplementary Data 4) were detected differentially abundant between PD and controls. However, the abundance of these taxa was also well explained by gender and/or age alone (Supplementary Data 4 and Supplementary Fig. 12).

Among the remaining genera, after controlling for metadata, *Roseburia*, *Blautia*, *Fusicatenibacter*, *Faecalibacterium*, *Moryella*, *Anaerostipes*, and three other unknown genera belonging to the family Lachnospiraceae were strongly enriched in the control samples (Fig. 5b) in several datasets (Fig. 5c). Consistently, the Lachnospiraceae family and species within this family and affiliated to the *Fusicatenibacter*, *Blautia*, and *Roseburia* genera were strongly enriched in control samples (Supplementary Figs. 10, 11).

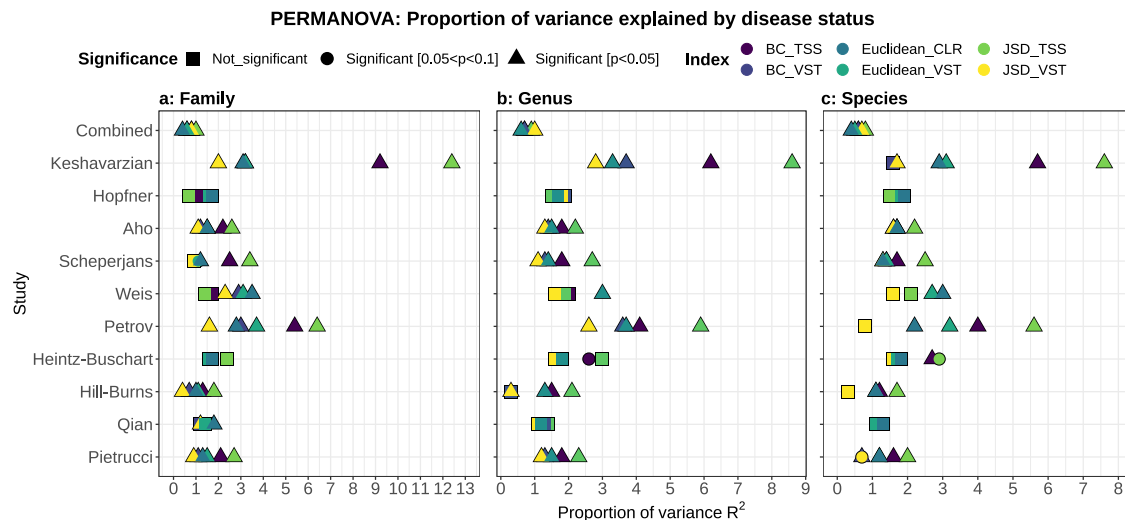


Fig. 3 The gut microbiome structure differs significantly between PD patients and controls. Data were normalized using three independent approaches (VST variance stabilizing transformation, TSS total sum scaling, CLR centered log-ratio) and beta-diversity was estimated using three indices (Bray–Curtis, BC; Jensen–Shannon divergence, JSD; Euclidean). The effect of the disease status on the clustering of the data was assessed using a permutational analysis of variance (PERMANOVA). In the majority of the studies and approaches considered, and across all taxonomic ranks (a, b, c), the gut microbiome of PD patients resulted significantly different from the one of controls. The disease status explains only a small fraction of the data variability ($< 13\% R^2$), indicating that other environmental factors might have a stronger role in shaping the bacterial communities. The dataset obtained by pooling all ten studies is referred to as “Combined” in the figure.

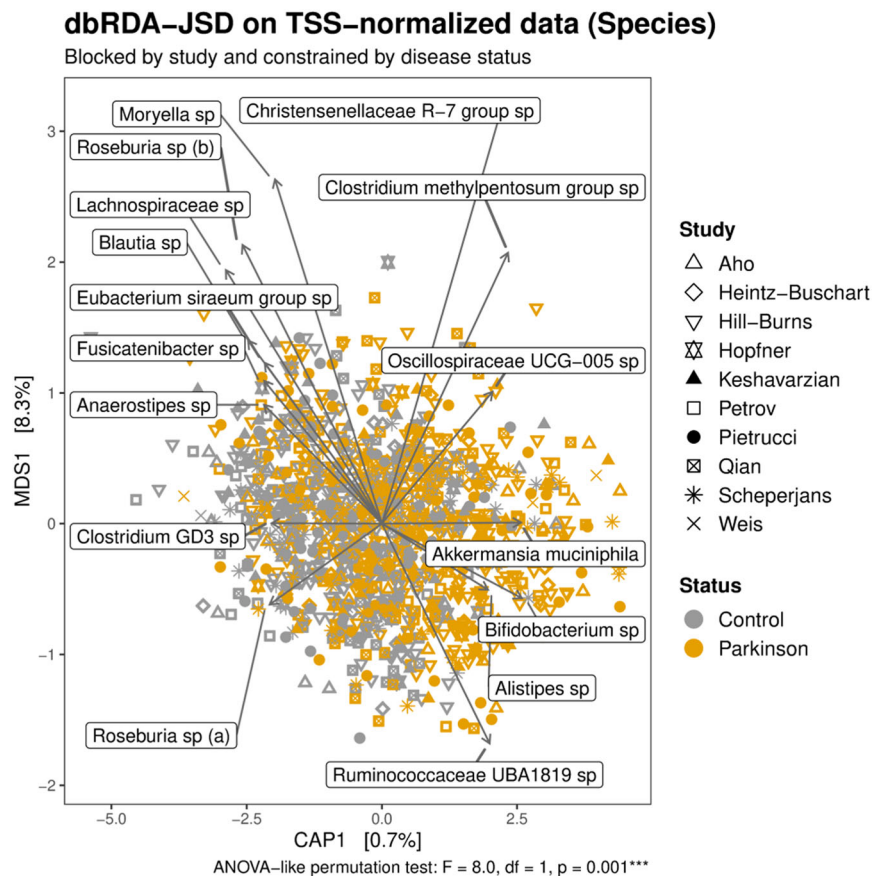


Fig. 4 Most important species driving the divergence of the gut microbiome between PD patients and controls. Distance-based redundancy analysis (dbRDA) was performed on Jensen–Shannon divergence (JSD) calculated on data normalized through total sum scaling (TSS). dbRDA was conditioned (blocked) by study and constrained by disease status. Data refer to species abundances. The limited proportion of data variability explained by the axis constrained for disease status (CAP1) indicates that environmental factors have a major influence in shaping the bacterial communities. However, the influence of the disease status on the community structure is statistically significant (ANOVA-like permutation test). Only taxa showing a significant association with the clustering of the samples and the strongest abundance variation between conditions are reported.

The family Butyricocccaceae was also more abundant in controls, even though it was detected differentially abundant in fewer studies (Supplementary Fig. 10). PD samples were instead most often enriched in the genera *Lactobacillus*, *Bifidobacterium*, *Hungatella*, and *Akkermansia* (Fig. 5). Additionally, the R-7 group of the Christensenellaceae family, the genera *Methanobrevibacter*, *Oscillobacter*, *Frisingicoccus*, and *Varibaculum* were also detected more abundant in PD, but with smaller effect size and higher variability across studies (Fig. 5). PD samples were enriched in species belonging to different taxonomic groups e.g., Ruminococcaceae, Christensenellaceae, *Bifidobacterium*, *Lactobacillus*, *Hungatella*, and *Alistipes* (Supplementary Fig. 11). Other species, such as *Intestinimonas* sp, species within the Oscillospiraceae family were also more abundant in PD, but in fewer studies (Supplementary Fig. 11). In contrast, the majority of species enriched in controls belonged to the families Lachnospiraceae, Eubacteriaceae, and Ruminococcaceae (Supplementary Fig. 11). The shifts in taxa abundances outlined above were robust to the sensitivity analysis we performed omitting the baseline data of the longitudinal Finnish cohort, which overall resulted in minor differences affecting only taxa having a small effect size (Supplementary Fig. 13). When this dataset has removed the genus *Alistipes* and the species *Massiliomicrobiota timonensis* did not result in enriched in PD, which instead had higher abundances of bacteria within the *Coproccillus* genus (Supplementary Fig. 13).

Finally, for each dataset, we obtained hypothetical functional predictions based on the 16S profiles. Differential abundance testing of predicted-pathways between PD and controls was performed as for the taxonomic data. The majority of predicted-pathways enriched in PD were related to ubiquinone (Coenzyme Q; CoQ) and menaquinone (vitamin K2) biosynthesis, glutamate degradation, methanogenesis, and lactic-type fermentation (Supplementary Data 2, 3; Fig. 6 and Supplementary Fig. 14). Instead, the control samples were enriched in pathways involved in the biosynthesis of cobalamin (vitamin B12) and glutamine/glutamate, degradation of glucuronate and galactoglucuronate, and methane production via acetate degradation (Supplementary Data 2, 3; Fig. 6 and Supplementary Fig. 14). Omitting the baseline of the Finnish cohort just affected some pathways involved in the menaquinone and ubiquinone biosynthesis which had higher variability and small effect size (Supplementary Fig. 13).

DISCUSSION

In recent years, several studies have analyzed the gut microbiome of PD patients reporting various degrees of alteration compared to healthy controls. We wanted to verify whether consistent changes in the gut microbiome of PD patients can be identified across studies, as often contrasting results have been reported. Hence, we performed a meta-analysis re-analyzing all publicly available

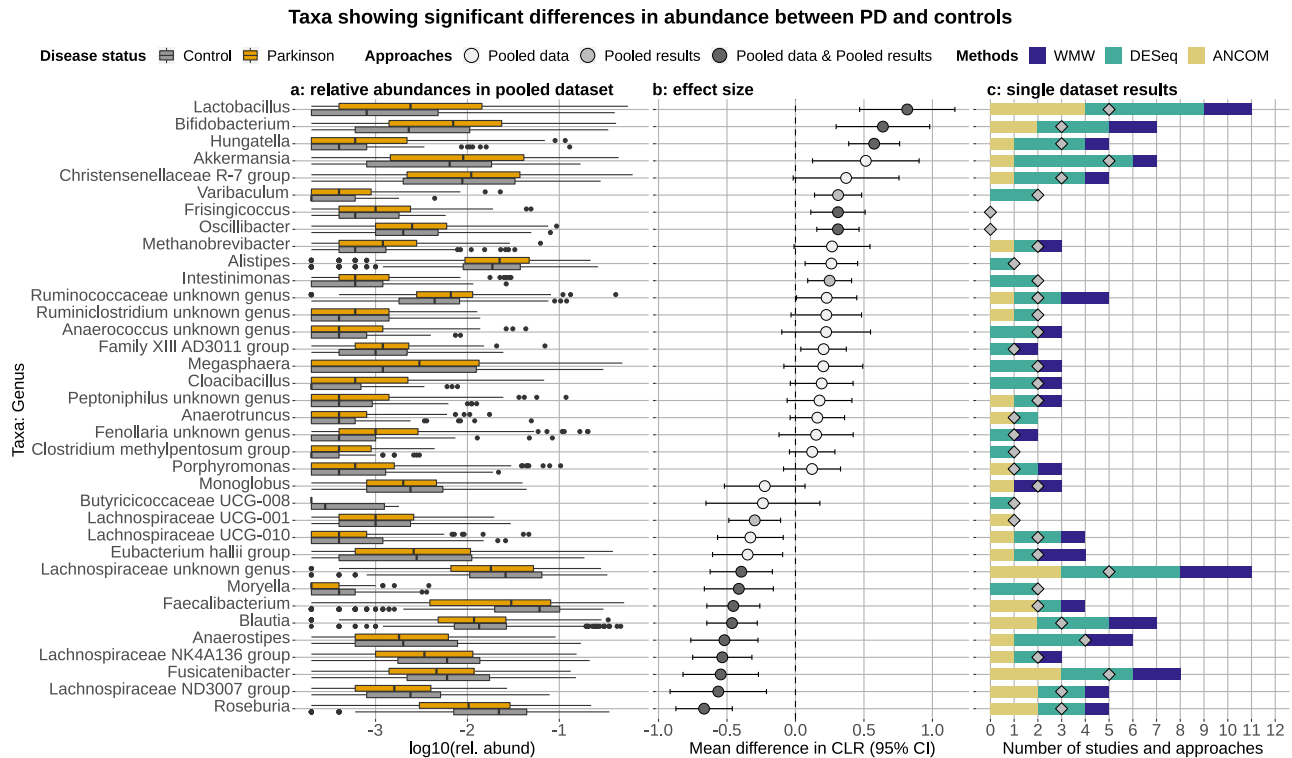


Fig. 5 Genera showing a significant difference in abundance between PD patients and controls. The relative abundances of the genera retrieved from the rarefied pooled data are reported in panel **a**. Effect sizes were estimated via the mean difference in CLR (panel **b**) using a random-effect meta-analysis approach (Pooled results approach). This was calculated for all taxa resulting differentially abundant in the Pooled results or Pooled data approaches. The color of the dots indicates which of the two above approaches detected the taxa differentially abundant. Taxa more abundant in controls have an effect size shifted to the left, whereas taxa more abundant in PD have an effect size shifted to the right. Panel **c** shows the number of times each genus was detected differentially abundant between PD patients and control samples across studies (diamonds) and approaches (bars), hence the maximum number of times a taxon can be detected differentially abundant is 30.

16S rRNA-gene amplicon datasets obtained by studies that compared PD patients with healthy controls, to offer a comprehensive and robust description of the PD-associated gut microbiome. Studies were heterogeneous in methodological approaches, and the interstudy variability was the main factor driving bacterial community structures, as observed in previous microbiome-based meta-analyses performed in the context of diet and colorectal cancer^{28,29}. Our analyses suggest that particular methodological approaches, such as sample collection and transport, sequencing platform, and the chosen variable region of the 16S rRNA gene might be the main reasons for the heterogeneity across the datasets we considered. Moreover, we show here that some differences between bioinformatic approaches exist, and we believe that the most robust and comparable results were obtained using the V3–V4 region of the 16S rRNA gene and data normalized via TSS or CLR. Variability across the studies we re-analyzed was also a reflection of the difference in the sampling populations, and the criteria used for recruiting participants. For example, the use of antibiotics was not always adopted as an exclusion criterion, and when included, different time intervals were chosen between the last treatment and sample collection (i.e. 1 month, 3 months). Confounding factors and reverse causation might also explain the observed associations. In particular, the vast majority of PD patients were using PD-related medication (e.g., Levodopa), most cases were progressed in the disease, and in some studies, controls differed considerably from cases in terms of age, sex and ethnicity. Hence, it is important to stress that using the available data it is impossible to determine whether the associations reported are

causally linked to PD. Future studies should concentrate on standardizing the recruitment criteria and preferentially focus on early-stage drug-naïve patients to evaluate the causality links between microbiome alterations and PD.

By stratifying the analysis by study, we could simultaneously account for the variability introduced by all study-specific factors and methodologies, such as country of origin, DNA extraction, and sequencing. In agreement with previous studies^{11–13,15,18,19}, we show that the gut microbiome of PD patients significantly differs from those of controls. Although PD can explain only a limited portion of data variability the observed differences are robust to the technical heterogeneities across sampling cohorts, indicating that the alteration in the gut microbiome of PD patients is a general phenomenon. The analysis of the bacterial alpha-diversity suggests that such alterations might be explained by a decrease in the abundances of the most abundant species and an increase in the rare ones. An increase in bacterial diversity in the gut microbiome of PD patients has been previously reported both in the studies we re-analyzed and studies for which data were not available^{11,18,21,30}. Similarly, a recent study reported no differences in OTU-based alpha-diversity but found that in controls 98% of OTUs could be assigned to the four dominant Phyla, whereas only 88% of OTUs belonged to these Phyla in PD³¹. This suggests a decrease in dominant taxa and an increase in less abundant ones as underlined by our results.

PD samples had a lower abundance of the genera *Roseburia*, *Fusicatenibacter*, *Blautia*, *Anaerostipes* (Lachnospiraceae family), and *Faecalibacterium* (Ruminococcaceae family), which have also been detected depleted in other neuro-inflammatory and

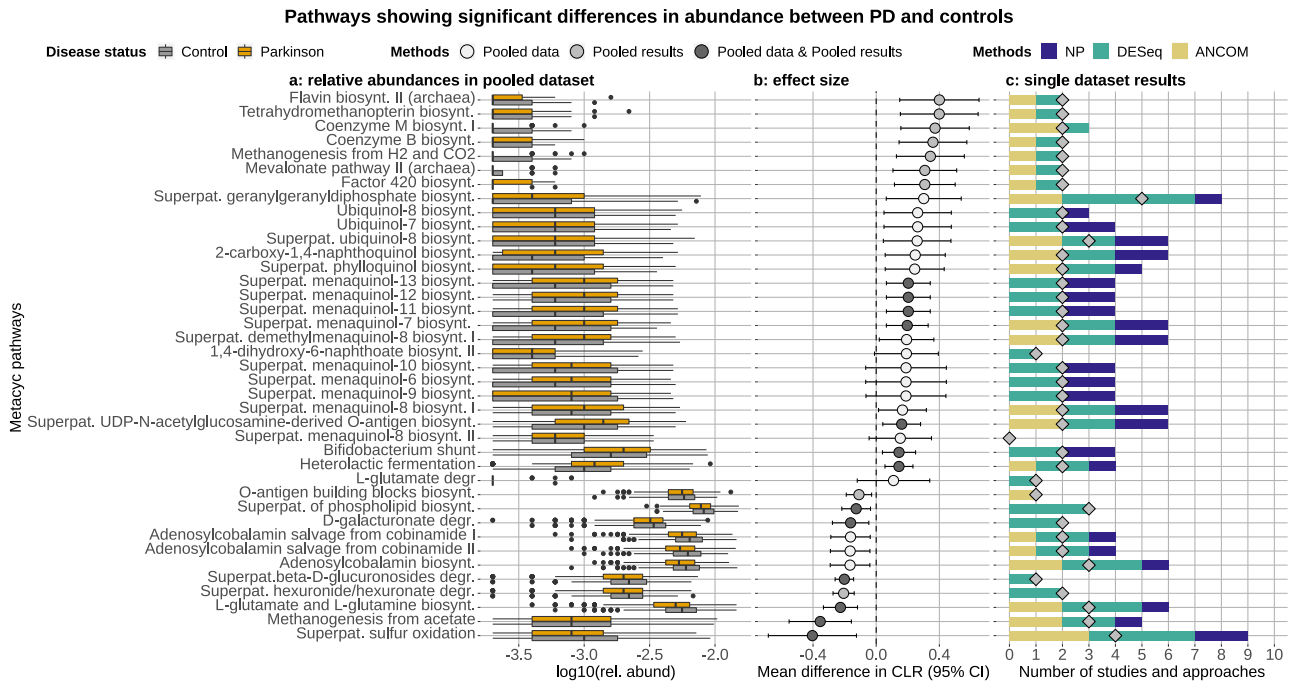


Fig. 6 Metabolic pathways showing a significant difference in abundance between PD patients and controls. Only selected relevant pathways are shown (a full overview is reported in Supplementary Fig. 14). The relative abundances of the pathways retrieved from the rarefied pooled data are reported in panel **a**. Effect sizes were estimated via the mean difference in CLR (panel **b**) using a random-effect meta-analysis (Pooled Results approach). This was calculated for all pathways resulting differentially abundant in the Pooled results or Pooled data approaches. The color of the dots indicates which of the two above approaches detected the pathway differentially abundant. Pathways more abundant in controls have an effect size shifted to the left, whereas pathways more abundant in PD have an effect size shifted to the right. Panel **c** shows the number of times each pathway was detected differentially abundant between PD patients and controls across studies (diamonds) and approaches (bars). We used ten studies and three approaches, hence the maximum number of times a pathway can be detected differentially abundant is 30.

neurodegenerative disorders (e.g., multiple sclerosis)^{32,33}. Most of these taxa are abundant and widespread bacteria in the gut microbiota of healthy individuals, they are major butyrate producers and have often been found depleted in IBD³⁴. Similarly depleted in IBD are bacteria of the Butyricicoccaceae family³⁵, which are important butyrate producers and were highly depleted in PD in our analyses. The depletion of these taxa could indicate a lowered level of butyrate in the gut of PD patients. Butyrate is a fundamental energy source for intestinal epithelial cells and can reinforce the intestinal epithelium lowering the risk of inflammation and carcinogenesis³⁶. Our findings are consistent with previous studies showing low levels of butyrate and increased gut permeability and inflammation in PD patients^{7,9}. Butyrate and other SCFA are not only relevant for gut health, but they can also influence the enteric nervous system (ENS), have systemic anti-inflammatory properties, promote normal microglia development, and potentially affect epigenesis in the CNS³. Importantly, PD patients have been shown to have increased levels of various pro-inflammatory cytokines in both the colon and serum, suggesting that they suffer from systemic inflammation which could result in microglial activation driving disease progression³⁷. To the best of our knowledge, only Aho et al.¹⁶ identified bacteria of the Butyricicoccaceae enriched in controls. Similarly, only Aho et al.¹⁶ and Weis et al.²⁷ detected *Fusicatenibacter* to be significantly depleted in PD patients. Specifically, Weis et al. report that the decrease of this genus together with that of *Faecalibacterium* was correlated to the degree of gut inflammation²⁷. Interestingly, both of these genera were low in abundance in IBS and ulcerative colitis^{38,39}, and *Faecalibacterium* showed strong anti-inflammatory and protective effects in an acute colitis mouse models⁴⁰. Our analysis suggests that the depletion of taxa playing a key role in maintaining gut health is widespread in PD across populations.

Such depletion resembles dysbiosis observed in other gastrointestinal dysfunctions (e.g., IBD) and supports the link between PD and gut health as underlined by retrospective studies indicating that the overall risk of developing PD in IBD was significantly higher. Specifically, the risk increased by 28% and 30% in patients with Crohn's disease (CD) and ulcerative colitis, respectively³⁷.

Our results indicate a higher abundance of the genera *Lactobacillus*, *Akkermansia*, *Hungatella*, and *Bifidobacterium* in PD gut microbiome. Interestingly, *Lactobacillus* and *Bifidobacterium* have been reported to be enriched in the gut of CD patients, and this increase coincided with a decreased abundance of *Faecalibacterium* sp⁴¹, similarly to what we have observed in PD. These taxa are commonly considered to be beneficial bacteria, and whether they are influencing PD or are simply well adapted to thrive in a pro-inflammatory gut environment remains to be elucidated. The genus *Lactobacillus*, and the Lactobacillaceae family, were the most strongly enriched taxa in PD across the studies we re-analyzed, in line with the previous findings^{11,12,17,19,42}. *Lactobacillus* strains are low abundant members of the gut microbiota and their abundance varies greatly across human disease and chronic conditions⁴¹. Some strains of *Lactobacillus* are able to produce enzymes that can degrade levodopa into dopamine, suggesting that their abundances might be a consequence of the use of this medication in PD^{43,44}. Levodopa is absorbed in the small intestine, but it has been reported that 10–20% can reach the large intestine⁴⁵ and could thus help these bacteria to proliferate.

Akkermansia has been repeatedly shown to be more abundant in PD compared to controls^{11–13,20,22}. *Akkermansia* spp is considered beneficial for human health and is potential probiotics, as they fortify the integrity of the epithelial cell layer and can

modulate the immune system^{46,47}. However, contrasting results regarding the influence of *Akkermansia* spp on gut health exist⁴⁸. Recently, safety concerns have been raised about the use of *A. muciniphila* as a probiotic, as its enrichment in neurodegenerative diseases (e.g., Alzheimer's disease, multiple sclerosis) could contribute to the progression of neural pathologies by degrading mucin, increasing gut inflammation and permeability and finally leading to higher endotoxaemia, and systemic inflammation^{33,49}. Intriguingly, constipated individuals have been shown to have a gut microbiome enriched in *Akkermansia*^{50–52}, and constipation is one of the major non-motor symptoms in PD, often starting decades before motor symptoms arise. The increase in *Akkermansia* could be a consequence of constipation, even though animal studies suggest that this genus might contribute to an increased transit time. *Akkermansia* spp is mucin-degrading bacteria and studies on mice suggest that they might lead to a depletion of the intestinal mucus-layer, drier stools, decreased number of goblet cells, and impaired intestinal barrier function when the gut microbiota is unbalanced^{53,54}. The unbalanced microbiota observed in PD patients might lead to a proliferation of *Akkermansia* spp, which in turn might lead to decreased mucus thicknesses and constipation. It is important to point out that multiple strains belonging to the same *Akkermansia* species can co-exist in the gut and the modulation of host-response can be strain-specific^{55,56}. For example, different *A. muciniphila* strains have different effects on the differentiation of Regulatory T cells (Tregs) and SCFA production⁵⁶, both factors altered in blood and gut, respectively, of PD patients^{6,9,57}. Altogether these data indicate that the increased abundance of *Akkermansia* spp in PD might be linked to alterations in the immune response and constipation. These effects might be strain-specific and more in-depth strain-resolved metagenomics are needed to elucidate these aspects in PD.

Among the most abundant taxa in PD, there were bacteria belonging to the Christensenellaceae family, in line with the previous reports^{11,12,21,42}. These bacteria are widespread in the gut of healthy individuals and their abundances positively correlate to the intestinal transit time⁵⁸. Via the production of H₂ *Christensenella* spp. can support the proliferation of the archaeon *Methanobrevibacter*⁵⁹, the major hydrogenotrophic methane producer in the human gut. This explains the positive correlation between the abundance of these two taxa⁵⁹, which we also observed in our dataset (Spearman rank test: PD, $Z = 10.3$, P value = 0.0005; controls, $Z = 8.8$, P value = 0.0005; Supplementary Fig. 15). Consistently, *Methanobrevibacter* and the 16S-predicted metabolic pathways for the production of methane from H₂ and CO₂, and for the synthesis of key co-factors involved in methanogenesis (coenzyme B and M, Factor 420) were more abundant in PD patients compared to controls. These data suggest an increased production of methane in the gut of PD patients that could influence the intestinal transit, as *Methanobrevibacter* is enriched in constipated patients^{51,52}, just as *Akkermansia*, and methane can decrease peristaltic movements⁶⁰. Surprisingly, to the best of our knowledge, only one other study reported an enrichment in *Methanobrevibacter* in PD²⁰. It is worth noting that another 16S-predicted pathway for the production of methane through acetate degradation was enriched in controls. This pathway is mainly found in Archaea of the genus *Methanosarcina*. However, we detected these taxa enriched only in one dataset¹⁸ and only by a single method (Supplementary Data 2). Hence, these data need to be interpreted with caution as they might be an artifact of the 16S-based predictions. It is important to specify that in the human gut-microbiota Archaea are considerably less abundant than Bacteria, and current methodologies (DNA extraction, primers used for 16S amplification) strongly discriminate against Archaea⁶¹. Hence, it is possible that the abundances and the diversity of these microorganisms are currently poorly represented in the available datasets.

At the functional level, the depletion of the glutamate/glutamine biosynthesis pathway and the higher abundance of the glutamate degradation pathway in PD suggests an alteration in the enteric production of these neurotransmitters. The gut microbiota has been previously suggested to alter the glutamate–glutamine–GABA cycles in schizophrenia and autism^{62,63} and alterations in the level of this transmitter have been found in brains of PD patients^{64,65}. Hence, it is intriguing to speculate that the gut microbiota might influence the concentrations of these chemicals in PD patients. Further experimental work will be required to verify whether these metabolic changes in the PD microbiota can induce alterations in the CNS. Surprisingly, the majority of the predicted-pathways enriched in the PD microbiota were related to ubiquinone (CoQ) and menaquinone (vitamin K2) biosynthesis. Data from animal and pre-clinical studies showed that both CoQ and vitamin K have a crucial role in avoiding the mitochondrial dysfunctions observed in PD^{66,67}. Therefore, the increased biosynthetic capacity we observed in the PD-associated microbiota is surprising. Although these findings would need to be confirmed via e.g., shot-gun metagenomics/metabolomics, it is tempting to speculate that the potential increase of vitamin K2 production in the gut might increase systemic concentrations of these chemicals in PD potentially influencing disease development. In fact, vitamin K plays an important role in the biosynthesis of sphingolipids⁶⁸ which are emerging as an important determinant in PD development⁶⁹. These data suggest novel mechanisms through which the gut microbiota might potentially influence PD development.

In summary, our analyses reveal consistent differences in the average gut-microbiota composition between PD patients and controls. The variation among studies is the strongest factor in shaping the data structure, but by accounting for the variability derived by the sampling cohorts we were able to show that the alteration of the gut microbiome in PD is consistent across studies and countries. Additional datasets would help to better quantify the differences observed between PD and controls, helping to identify country-specific effects that we could not resolve. Moreover, the heterogeneity in the available metadata did not allow us to account for more detailed confounding factors such as co-morbidities, drug therapies, and diet. These data would help to further refine the associations between PD and microbiome, and identify taxa that could be mechanistically linked to the disease. The differences in taxa abundances between PD and controls indicate that the gut microbiota of PD patients shares similarities with those of other neurological (e.g., multiple sclerosis) and inflammatory gastrointestinal diseases. Taxa important in maintaining gut integrity and health via the production of SCFA are depleted in PD and this together with the growing evidence of gut and systemic inflammation in PD, points towards an important role of the gut microbiota in modulating the immune function in this disease. Moreover, we were able to identify previously overlooked taxa enriched in PD such as *Methanobrevibacter* and Butyricicocaceae, and identified some potentially new metabolic routes through which the microbiota might influence PD. Our findings align with the accumulating evidence indicating gut and systemic inflammation in PD and suggest that the dysbiotic gut microbiota could influence host immune function and be linked to the gastrointestinal symptoms often observed in PD patients.

METHODS

Study selection

On March 29, 2020, Google Scholar was searched for publications that contained all the words "16S", "gut", "Parkinson", "metagenomic", the exact phrase "Parkinson's disease", at least one of the words "microbiota" [OR] "microbiome" [OR] "gut" [OR] "intestinal" anywhere in the article. This resulted in 1010 entries. Titles were then manually screened and if they contained the words "microbiome" or "microbiota" and "Parkinson's

disease” the abstracts were further consulted. Moreover, the Sequence Read Archive (SRA) in NCBI was queried with the following term “Parkinson” [AND] “microbiome”, resulting in two additional studies (Bioprojects): PRJNA530401 and PRJEB14928. We managed to match only the latter Bioproject ID to a published study¹⁴, hence we considered only this dataset in our analyses.

Inclusion criteria

We included all studies comparing the composition of the gut microbiota between patients with confirmed PD to a control population without PD, and that made the raw reads of the 16S rRNA-gene amplicon sequencing available. Studies with any design (e.g., cohort studies, case–control studies, or cross-sectional studies), and from any geographical area were included. Studies could use any method for the acquisition and analysis of samples. We identified a total of 23 studies that cataloged the gut microbiome of PD patients using metagenomics (Supplementary Table 1). Nine of these studies did not make the raw data publicly available. We were unsuccessful in obtaining the raw reads from the authors, as data were either protected by ethical restrictions or the authors did not answer our requests. In other cases, the raw reads were available, but it was impossible to associate the data with the disease status as this information was not reported in the metadata. The samples from Scheperjans et al.¹⁵, originally sequenced using a 454 technology, were recently re-sequenced in a follow-up study by the same group using Illumina MiSeq¹⁶. Hence, we only included in our analysis the most recent datasets. Finally, one study used shot-gun metagenomics¹⁰, and three studies were available only as pre-prints at the time of writing and the raw reads were not made public yet.

Data retrieval and zOTU picking

Raw reads were downloaded from SRA or the European Nucleotide Archive (ENA). Adapters were removed using the *bbtools* suite⁷⁰. Data were analyzed using *Lotus*⁷¹ and the *UNOISE3*⁷² algorithm for zOTUs calculation, bundled in a new *Lotus* version (*Lotus2*), currently under development. Due to the technical variability among datasets (e.g., 16S region, sequencing technology) the filtering parameters used by the *sdm* program called by *Lotus*, were adjusted for each dataset independently and are reported in the supplementary materials (Supplementary Data 5). For the datasets of Petrov et al.¹⁷ and Weis et al.²⁷, we had to decrease the accepted minimum error due to the low quality of the sequencing data (Supplementary Data 5). 16S-based functional predictions were obtained using the default settings in *picrust2*⁷³ and the *Metacyc* database. In this analysis, the dataset of Qian et al.¹⁸ was not included, as with the default cutoffs the sequences aligned poorly with the reference database used. Count tables for species, genera, families, and functional predictions were then analyzed using R v3.6.2⁷⁴ and processed using the *phyloseq* R package⁷⁵. We then retained all samples with >4500 reads, as well as taxa with >5 counts and predicted functionalities with >20 counts in at least 2.5% of the samples. These filtration steps left a total of 1211 (530 control, and 681 PD samples) and 1121 samples (485 control and 636 PD samples) for the taxonomic and predicted-function data, respectively. Enterotypes were predicted using rarefied relative abundances of genera via the <https://enterotypes.org/> web-platform.

Statistical analyses of single studies

Alpha-diversity indices at the species level were calculated using the *microbiome* R package⁷⁶ after rarefying without re-sampling at the even depth of 5000. Due to rarefaction eight samples were further removed, leaving a total of 1203 samples (523 control and 680 PD samples). We measured richness using the number of observed species, Chao1, Fisher's alpha, and ACE indices; evenness using the Bulla and Simpson indices, dominance using the core abundance, which measures the relative proportion of core species that exceed relative abundance of 0.2% in over 50% of the samples, and the Simpson's index of dominance. Finally, we estimated rarity using the low abundance index, which considers the relative proportion of the least abundant species below a detection level of 0.2%, and the rare abundance index, which estimates the relative proportion of the non-core species exceeding the detection level of 0.2% at 50% prevalence. In addition, we calculated the ratios of Firmicutes to Bacteroidetes phyla and *Prevotella* to *Bacteroides* genera, as log₂ ratios of their relative abundances. In each dataset, the differences in alpha-diversity between control and PD samples were assessed using Agresti's generalized odd ratios using the *genodds* function in the *genodds*

R package⁷⁷. This statistic, based on ranks and analogous to the U statistic underlying the Mann–Whitney test, does not make strong assumptions about the distributions of measures and is comparable between measures of diversity with different scales.

For each dataset, beta-diversity and differential abundance analyses were performed using three independent approaches (described in the sections below): (i) normalization via total sum scaling (TSS; i.e., relative abundances) and differential abundance (DA) inference through Wilcoxon–Mann–Whitney (WMW) tests; (ii) variance stabilizing transformation (VST) and DA inference using DESeq2⁷⁸; (iii) compositional approach based on centered log ratios (CLR) and DA inference using analysis of the composition of microbiomes (ANCOM)⁷⁹. We then reported the number of times each taxon showed a significant difference in abundance between PD and controls across studies and statistical approaches. For example, a taxon detected differentially abundant across all ten datasets and all three approaches would have a final score of 30 (panel c in Figs. 5 and 6 and Supplementary Figs. 10, 11, 14). Differential abundances of *picrust2* predicted functionalities between PD and controls were inferred using the same approach outlined above. The rarefaction used in the TSS approach did not result in a loss of samples for the 16S-based predicted functionalities.

Total sum scaling (TSS) and non-parametric tests

After rarefying without re-sampling at the even depth of 5000, data were normalized by dividing the counts of each taxon for the total counts of all taxa (total sum) in the sample. Beta-diversity matrices were calculated using the Bray–Curtis (BC) dissimilarity index and the Jensen–Shannon distances (JSD). Statistical differences between control and PD groups were then tested using the permutational multivariate analysis of variance (PERMANOVA) as implemented in the *adonis2* (analysis of variance using distance matrices, ADONIS) function in the *vegan* R package⁸⁰. DA analysis was performed using a two-sided WMW test, using the Benjamini–Hochberg (BH) *P* value correction.

Variance stabilizing transformation (VST) and DESeq2 analyses

Since the DESeq2 approach does not account for zero-inflated data, the correction factors were calculated using the GMPR method that is based on geometric means of pairwise ratios⁸¹. Euclidean, BC, and JSD distances were used as beta-diversity estimators after normalizing the data via VST through the DESeq2 package. Statistical differences between control and PD groups were tested using the *adonis2* function as specified above. DAs were calculated using default DESeq2 parameters that include a negative binomial GLM fitting and a Wald test⁷⁸. Multiple testings were accounted for using BH *P* value correction.

Compositional analysis: centered log ratios (CLR) and ANCOM

Data were transformed using CLR, after imputing zeros through Bayesian-multiplicative replacements via the count zero multiplicative approach (“CZM”) in the *cmultRepl* function of the *zCompositions* R package⁸². Euclidean distances, which for such data correspond to Aitchison distances, were then calculated⁷⁹. Statistical differences between control and PD groups were tested using the *adonis2* function as specified above. DA analysis was performed using the count tables and the ANCOM approach as implemented in the R script *ancom_v2.0*⁸³ using a 0.95 zero-cutoff and significance at the 0.6 percentile.

Statistical analyses of the combined studies

The Agresti's generalized odd ratios estimated for each alpha-diversity index and each individual study were pooled using a random-effect meta-analysis via the function *metagen* in the R package *meta*⁸⁴.

Count tables obtained for each dataset were pooled and beta-diversity analyses were performed using the three approaches described above (TSS-, VST-, CLR-based analysis). For each normalization approach, statistical differences between control and PD groups and the marginal effects of study and disease status were tested using the *adonis2* function. We then used the distance measure that captured a highest fraction of the variability to compute distance-based redundancy analyses (dbRDA). dbRDAs were performed using the “CAP” option in *phyloseq*, which calls the *capscale* function in the *vegan* package. Data were clustered without conditioning (blocking) for studies and without constraining, by conditioning for study, and by conditioning for study and constraining for disease

status (PD vs control):

Distance ~ 1

Distance $\sim 1 + \text{Condition}(\text{study})$

Distance $\sim \text{status} + \text{Condition}(\text{study})$

The significance of the constrain was tested using an ANOVA-like permutation test (*anova.cca* function in the *vegan* R package). For each normalization method, we investigated the effect of study-dependent factors such as country, sequencing platform (e.g., MiSeq vs IonTorrent), sequencing approach (single-end vs paired-end), the region of the 16S gene used (e.g., V4 vs V1–V2), and extraction methods by creating additional dbRDAs and constraining the data for each individual factor. The effect of each constraining variable was tested using an ANOVA-like permutation test. We then verified whether accounting for the variability introduced by the study alone will allow us to simultaneously account for the variation derived by the other technical factors. We compared the adjusted R² (R_{2adj}) of a dbRDA obtained using the full model *distance ~ country + 16S region + ends + seq + extraction + extraction type + status* with the one of a reduced model including only disease status and study (*distance ~ study + status*). Similar R_{2adj} differences $\leq 0.1\%$, indicates that the two models are equivalent. The influence of study-specific factors on microbial community structure was assessed at the species, genus, and family level. Finally, we used the TSS normalized data to correlate the relative abundance of the taxa to the constrained and conditioned dbRDA via the *envfit* function in the *vegan* package. We selected only taxa significantly correlated with the clustering (*P* value < 0.01), and showing the highest degree of variation ($\geq |0.095|$ for genus and species, and $\geq |0.07|$ for family) along the constrained axis (CAP1).

Similarity among studies

We used the unconstrained and unconditioned dbRDA performed on the TSS normalized species data to estimate dissimilarity among studies. We selected the coordinates of each sample across all axis that explained 90% of the data variance. These scores were then used to calculate Euclidean distances amongst samples. We then calculated distances between study centroids using the R package *usdist*⁸⁵. Similarity among studies was then visualized using non-metric multidimensional scaling (NMDS) via the *metaMDS* function in the *vegan* R package.

Differential abundance analyses of the combined datasets

We combined two independent approaches to gather a consensus view on the taxa/pathways differentially abundant between PD patients and controls. We refer to these two approaches as Pooled data and Pooled results. In the Pooled data approach, the count tables obtained for each dataset were pooled and processed with the same methods used for the single datasets: (i) TSS normalization on rarefied data and *independence_test* in the *coin* R package⁸⁶ blocking data for the study; (ii) DESeq2 approach adding the “study” variable as a covariate in the model; (iii) ANCOM performed using a mixed-effect model with the effect of PD allowed to vary across study (via the formula “*random.formula* = “ $\sim 1 + \text{status} | \text{study}$ ”), using a zero-cutoff 0.975 and significance at the 0.6 percentile. For all three methods, BH *P* value correction was used and the threshold for significance was set at ≤ 0.05 . If a taxon or pathway had a significant difference in abundance in two out of three approaches, it was then retained (Consensus).

To this first list of differentially abundant taxa/pathways, we added the data obtained from the Pooled results approach. In this approach, we normalized the count table of each individual dataset using CLR after adding a pseudo-count of 1 to 0 values. We then selected all taxa and pathways detected in at least three studies and estimated their shift in abundance between PD and controls using linear models for family, genera, and 16S-based predicted functionalities and Agresti’ generalized odd ratios for species. We then pooled these results using a random-effect meta-analysis via the *metagen* R function. The resulting *P* values were corrected using BH. All taxa/pathways showing an adjusted *P* value ≤ 0.05 and a 95% confidence interval (CI) not crossing 0 were retained.

Taxa and pathways showing significant differences in abundance between PD and controls in the Pooled data (two out of three methods referred to as Consensus) or Pooled results approach were further considered. All taxa having abundances potentially influenced by age and/or gender were then removed (see below). For each taxa/pathway, the effect size and the respective 95% CI were estimated using the Pooled results approach

(random-effect meta-analysis). Finally, the correlation between the genera Christensenellaceae R-7 group and *Methanobrevibacter* was calculated on the relative abundances of non-rarefied data using a Spearman correlation test by blocking the data by study (*spearman_test* in the *coin* R package).

Influence of confounding factors on differential abundances

The metadata made available by five studies were used to assess the influence of age and gender (the only two factors reported in all five studies) on taxa abundances. We used generalized linear mixed models (GLMMs) controlling for zero-inflation as implemented in the R package *glmmTMB*⁸⁷.

Abundance $\sim \text{status} * \text{gender} + \text{status} * \text{age} + (1 + \text{status} | \text{study})$

We created random slope and random intercept GLMMs for all taxonomic ranks we analyzed (species, genus, family). Models were fitted using either a negative binomial or a generalized Poisson distribution. First, we constructed zero-inflated and non-zero-inflated models, and choose the best model using the Akaike information criterion (AIC; $\Delta\text{AIC} > 2$). We then created reduced models omitting each of the predictors (status, age, gender), their interactions (status:gender, status:age), and considering a constant effect of the disease status across studies (i.e., *random effect = 1 | study*). We then compared all models using the *model.sel* function and the AIC in the R package *MuMIn*⁸⁸. If one of the best models (within a ΔAIC of 2) did not contain the variable disease status we concluded that the disease status might be not an essential factor needed to explain the taxon abundance. Hence, we removed these taxa from further discussion. If all best models contained the variable disease status, we consider PD as an essential factor shaping taxa abundances, thus we retained the taxa. For building the GLMMs, raw counts were used and data were rarefied to a fixed depth of 10,000 to avoid overparameterization.

Reporting summary

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

DATA AVAILABILITY

The data used in this paper have been previously published, and details are reported in Table 1.

CODE AVAILABILITY

The R code used in our analysis can be accessed at https://github.com/StfnRomano/PD_meta_analysis.

Received: 14 July 2020; Accepted: 8 January 2021;

Published online: 10 March 2021

REFERENCES

- Elbaz, A., Carcaillon, L., Kab, S. & Moisan, F. Epidemiology of Parkinson’s disease. *Rev. Neurol.* **172**, 14–26 (2016).
- Shulman, J. M., De Jager, P. L. & Feany, M. B. Parkinson’s disease: genetics and pathogenesis. *Annu. Rev. Pathol. Mech. Dis.* **6**, 193–222 (2011).
- Keshavarzian, A., Engen, P., Bonvegna, S. & Cilia, R. The gut microbiome in Parkinson’s disease: a culprit or a bystander? In *Progress in Brain Research* (eds Björklund, A. & Angela Cenci, M.) Vol. 252, 357–450 (Elsevier, 2020) <https://doi.org/10.1016/bs.pbr.2020.01.004>.
- Cersosimo, M. G. & Benarroch, E. E. Pathological correlates of gastrointestinal dysfunction in Parkinson’s disease. *Neurobiol. Dis.* **46**, 559–564 (2012).
- Pellegrini, C., Antonioli, L., Colucci, R., Blandizzi, C. & Fornai, M. Interplay among gut microbiota, intestinal mucosal barrier and enteric neuro-immune system: a common path to neurodegenerative diseases? *Acta Neuropathol.* **136**, 345–361 (2018).
- Sampson, T. R. et al. Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson’s disease. *Cell* **167**, 1469–1480.e12 (2016).
- Schwartz, A. et al. Fecal markers of intestinal inflammation and intestinal permeability are elevated in Parkinson’s disease. *Parkinsonism Relat. Disord.* **50**, 104–107 (2018).
- Clairembault, T. et al. Structural alterations of the intestinal epithelial barrier in Parkinson’s disease. *Acta Neuropathol. Commun.* **3**, 12 (2015).

9. Unger, M. M. et al. Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls. *Parkinsonism Relat. Disord.* **32**, 66–72 (2016).
10. Bedarf, J. R. et al. Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. *Genome Med.* **9**, 1–13 (2017).
11. Keshavarzian, A. et al. Colonic bacterial composition in Parkinson's disease: colonic microbiota in Parkinson's disease. *Mov. Disord.* **30**, 1351–1360 (2015).
12. Hill-Burns, E. M. et al. Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome: PD, medications, and gut microbiome. *Mov. Disord.* **32**, 739–749 (2017).
13. Heintz-Buschart, A. et al. The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder: nose and gut microbiome in PD and iRBD. *Mov. Disord.* **33**, 88–98 (2018).
14. Hopfner, F. et al. Gut microbiota in Parkinson disease in a northern German cohort. *Brain Res.* **1667**, 41–45 (2017).
15. Scheperjans, F. et al. Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov. Disord.* **30**, 350–358 (2015).
16. Aho, V. T. E. et al. Gut microbiota in Parkinson's disease: temporal stability and relations to disease progression. *EBioMedicine* **44**, 691–707 (2019).
17. Petrov, V. A. et al. Analysis of gut microbiota in patients with Parkinson's disease. *Bull. Exp. Biol. Med.* **162**, 734–737 (2017).
18. Qian, Y. et al. Alteration of the fecal microbiota in Chinese patients with Parkinson's disease. *Brain. Behav. Immun.* **70**, 194–202 (2018).
19. Pietrucci, D. et al. Dysbiosis of gut microbiota in a selected population of Parkinson's patients. *Parkinsonism Relat. Disord.* **65**, 124–130 (2019).
20. Li, C. et al. Gut microbiota differs between Parkinson's disease patients and healthy controls in Northeast China. *Front. Mol. Neurosci.* **12**, 171 (2019).
21. Jin, M. et al. Analysis of the gut microflora in patients With Parkinson's disease. *Front. Neurosci.* **13**, 1184 (2019).
22. Lin, C.-H. et al. Altered gut microbiota and inflammatory cytokine responses in patients with Parkinson's disease. *J. Neuroinflammation* **16**, 129 (2019).
23. Li, F. et al. Alteration of the fecal microbiota in North-Eastern Han Chinese population with sporadic Parkinson's disease. *Neurosci. Lett.* **707**, 134297 (2019).
24. Boertien, J. M., Pereira, P. A. B., Aho, V. T. E. & Scheperjans, F. Increasing comparability and utility of gut microbiome studies in Parkinson's disease: a systematic review. *J. Park. Dis.* **9**, S297–S312 (2019).
25. Arumugam, M. et al. Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
26. Weiss, S. et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
27. Weis, S. et al. Effect of Parkinson's disease and related medications on the composition of the fecal bacterial microbiota. *npj Park. Dis.* **5**, 28 (2019).
28. Bisanz, J. E., Upadhyay, V., Turnbaugh, J. A., Ly, K. & Turnbaugh, P. J. Meta-analysis reveals reproducible gut microbiome alterations in response to a high-fat diet. *Cell Host Microbe.* <https://doi.org/10.1016/j.chom.2019.06.013> (2019).
29. Wirbel, J. et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **25**, 679–689 (2019).
30. Barichella, M. et al. Unraveling gut microbiota in Parkinson's disease and atypical parkinsonism. *Mov. Disord.* **34**, 396–405 (2019).
31. Gorecki, A. M. et al. Altered gut microbiome in Parkinson's disease and the influence of lipopolysaccharide in a human α -Synuclein over-expressing mouse model. *Front. Neurosci.* **13**, 839 (2019).
32. Shi, Z. et al. Dysbiosis of gut microbiota in patients with neuromyelitis optica spectrum disorders: a cross sectional study. *J. Neuroimmunol.* **339**, 577126 (2020).
33. Freedman, S. N., Shahi, S. K. & Mangalam, A. K. The "Gut Feeling": breaking down the role of gut microbiome in multiple sclerosis. *Neurotherapeutics* **15**, 109–125 (2018).
34. Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M. & Owen, L. J. Dysbiosis of the gut microbiota in disease. *Microb. Ecol. Health Dis.* **26**, 26191 (2015).
35. Eeckhaut, V. et al. *Butyrivibrio pullicaecorum* in inflammatory bowel disease. *Gut* **62**, 1745–1752 (2013).
36. Hamer, H. M. et al. Review article: the role of butyrate on colonic function: review: role of butyrate on colonic function. *Aliment. Pharmacol. Ther.* **27**, 104–119 (2007).
37. Elfil, M., Kamel, S., Kandil, M., Koo, B. B. & Schaefer, S. M. Implications of the gut microbiome in Parkinson's disease. *Mov. Disord.* <https://doi.org/10.1002/mds.28004> (2020).
38. Takeshita, K. et al. A single species of clostridium subcluster XIVa decreased in ulcerative colitis patients. *Inflamm. Bowel Dis.* **22**, 2802–2810 (2016).
39. Sokol, H. et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105**, 16731–16736 (2008).
40. Martin, R. et al. The Commensal bacterium *Faecalibacterium prausnitzii* is protective in DNBS-induced chronic moderate and severe colitis models. *Inflamm. Bowel Dis.* **20**, 417–430 (2014).
41. Heeney, D. D., Gareau, M. G. & Marco, M. L. Intestinal *Lactobacillus* in health and disease, a driver or just along for the ride? *Curr. Opin. Biotechnol.* **49**, 140–147 (2018).
42. Baldini, F. et al. Parkinson's disease-associated alterations of the gut microbiome predict disease-relevant changes in metabolic functions. *BMC Biol.* **18**, 62, <https://doi.org/10.1186/s12915-020-00775-7> (2020).
43. Maini Rekdal, V., Bess, E. N., Bisanz, J. E., Turnbaugh, P. J. & Balskus, E. P. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* **364**, eaau6323 (2019).
44. van Kessel, S. P. et al. Gut bacterial tyrosine decarboxylases restrict levels of levodopa in the treatment of Parkinson's disease. *Nat. Commun.* **10**, 310 (2019).
45. Deleu, D., Northway, M. G. & Hanssens, Y. Clinical pharmacokinetic and pharmacodynamic properties of drugs used in the treatment of Parkinson's disease? *Clin. Pharmacokinet* **41**, 261–309 (2002).
46. Reunanen, J. et al. *Akkermansia muciniphila* adheres to enterocytes and strengthens the integrity of the epithelial cell layer. *Appl. Environ. Microbiol.* **81**, 3655–3662 (2015).
47. Derrien, M. et al. Modulation of mucosal immune response, tolerance, and proliferation in mice colonized by the mucin-degrader *Akkermansia muciniphila*. *Front. Microbiol.* **2**, 166 (2011).
48. Ring, C. et al. *Akkermansia muciniphila* strain ATCC BAA-835 does not promote short-term intestinal inflammation in gnotobiotic interleukin-10-deficient mice. *Gut Microbes* **10**, 188–203 (2019).
49. EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards) et al. Statement on the update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 12: suitability of taxonomic units notified to EFSA until March 2020. *EFSA J.* **18**, 6174, <https://doi.org/10.2903/j.efsa.2020.6174> (2020).
50. Gobert, A. P. et al. The human intestinal microbiota of constipated-predominant irritable bowel syndrome patients exhibits anti-inflammatory properties. *Sci. Rep.* **6**, 39399 (2016).
51. Blatchford, P. et al. Consumption of kiwifruit capsules increases *Faecalibacterium prausnitzii* abundance in functionally constipated individuals: a randomised controlled human trial. *J. Nutr. Sci.* **6**, e52 (2017).
52. Vandeputte, D. et al. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* **65**, 57–62 (2016).
53. Cao, H. et al. Dysbiosis contributes to chronic constipation development via regulation of serotonin transporter in the intestine. *Sci. Rep.* **7**, 10322 (2017).
54. Desai, M. S. et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339–1353.e21 (2016).
55. Guo, X. et al. Different subtype strains of *Akkermansia muciniphila* abundantly colonize in Southern China. *J. Appl. Microbiol.* **120**, 452–459 (2016).
56. Zhai, R. et al. Strain-specific anti-inflammatory properties of two *Akkermansia muciniphila* strains on chronic colitis in mice. *Front. Cell. Infect. Microbiol.* **9**, 239 (2019).
57. Álvarez-Luquín, D. D. et al. Regulatory impairment in untreated Parkinson's disease is not restricted to Tregs: other regulatory populations are also involved. *J. Neuroinflammation* **16**, 212 (2019).
58. Waters, J. L. & Ley, R. E. The human gut bacteria *Christensenellaceae* are widespread, heritable, and associated with health. *BMC Biol.* **17**, 83 (2019).
59. Raud, A. et al. Syntrophy via interspecies H₂ transfer between *Christensenella* and *Methanobrevibacter* underlies their global cooccurrence in the human gut. *mBio* **11**, e03235–19 (2020).
60. Triantafyllou, K., Chang, C. & Pimentel, M. Methanogens, methane and gastrointestinal motility. *J. Neurogastroenterol. Motil.* **20**, 31–40 (2014).
61. Koskinen, K. et al. First insights into the diverse human archaeome: specific detection of archaea in the gastrointestinal tract, lung, and nose and on skin. *mBio* **8**, e00824–17 (2017).
62. Zheng, P. et al. The gut microbiome from patients with schizophrenia modulates the glutamate-glutamine-GABA cycle and schizophrenia-relevant behaviors in mice. *Sci. Adv.* **5**, eaau8317 (2019).
63. Wang, M. et al. Alterations in gut glutamate metabolism associated with changes in gut microbiota composition in children with autism spectrum disorder. *mSystems* **4**, e00321–18 (2019).
64. Buchanan, R. J. et al. Changes in GABA and glutamate concentrations during memory tasks in patients with Parkinson's disease undergoing DBS surgery. *Front. Hum. Neurosci.* **8**, 81 (2014).
65. O'Gorman Tuura, R. L., Baumann, C. R. & Baumann-Vogel, H. Beyond dopamine: GABA, glutamate, and the axial symptoms of Parkinson disease. *Front. Neurol.* **9**, 806 (2018).
66. Vos, M. et al. Vitamin K2 is a mitochondrial electron carrier that rescues Pink1 deficiency. *Science* **336**, 1306–1310 (2012).

67. Shults, C. W. Therapeutic role of coenzyme Q10 in Parkinson's disease. *Pharmacol. Ther.* **107**, 120–130 (2005).
68. Denisova, N. A. & Booth, S. L. Vitamin K and sphingolipid metabolism: evidence to date. *Nutr. Rev.* **63**, 111–121 (2005).
69. Lin, G., Wang, L., Marcogliese, P. C. & Bellen, H. J. Sphingolipids in the pathogenesis of Parkinson's disease and parkinsonism. *Trends Endocrinol. Metab.* **30**, 106–117 (2019).
70. Bushnell B. B. BMap. SourceForge <https://sourceforge.net/projects/bbmap/>.
71. Hildebrand, F., Tadeo, R., Voigt, A., Bork, P. & Raes, J. LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome* **2**, 30 (2014).
72. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. <https://doi.org/10.1101/081257> (2016).
73. Douglas, G. M. et al. PICRUSt2: an improved and customizable approach for metagenome inference. <https://doi.org/10.1101/672295> (2019).
74. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
75. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
76. Lahti, L. & Shetty, S. microbiome: microbiome analytics. Bioconductor version: release (3.10). <https://doi.org/10.18129/B9.bioc.microbiome> (2020).
77. Johns, H. genodds: Generalized Odds Ratios. R package version 1.0.0. <https://CRAN.R-project.org/package=genodds> (2019).
78. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
79. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
80. Oksanen, J. et al. vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan> (2019).
81. Chen, L. et al. GMPR: a robust normalization method for zero-inflated count data with application to microbiome sequencing data. *PeerJ* **6**, e4600 (2018).
82. Palarea-Albaladejo, J. & Martín-Fernández, J. A. zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemom. Intell. Lab. Syst.* **143**, 85–96 (2015).
83. Lin, F. H. HuangLin/ANCOM: third release of ANCOM. *Zenodo* <https://doi.org/10.5281/ZENODO.3577802> (2019).
84. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evidence Based Ment. Health* **22**, 153–160 (2019).
85. Bittinger, K. usedist: Distance Matrix Utilities. R package version 0.4.0. <https://CRAN.R-project.org/package=usedist> (2020).
86. Hothorn, T., Hornik, K., van de Wiel, M. A. & Zeileis, A. A Lego system for conditional inference. *Am. Stat.* **60**, 257–263 (2006).
87. Brooks, M. E. et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R. J.* **9**, 378 (2017).
88. Bartoń, K. MuMIn: Multi-Model Inference. R package version 1.43.17. <https://CRAN.R-project.org/package=MuMIn> (2020).

ACKNOWLEDGEMENTS

We are grateful for the technical support provided by A. Telatin and the assistance from S.J. Green, A. Keshavarzian, and A.V. Petrov in obtaining the raw reads of their studies. This research was supported in part by the NBI Computing infrastructure for

Science (CIS) group through access to High-Performance Computing infrastructures. The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and its constituent project BBS/E/F/000PR10356. G.S. was supported by the BBSRC Core Capability Grant BB/CCG1860/1. We acknowledge M. Mayer, C. Hutchins, and R. Ansofor for constructive comments on the paper. The funding bodies had no role in the study design, execution of the analyses, and data interpretation.

AUTHOR CONTRIBUTIONS

S.R. designed the study, performed bioinformatics and statistical analyses, interpreted the results, and wrote the paper. G.M.S. assisted in statistical analyses and data interpretation. J.R.B. helped in data interpretation. F.H. provided support in bioinformatics and statistical analyses, and data interpretation. A.N. and I.G.C. helped in data interpretation.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41531-021-00156-z>.

Correspondence and requests for materials should be addressed to S.R. or F.H.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021