

 Open access • Posted Content • DOI:10.1101/2020.08.10.20171397

Meta-analysis of the gut microbiome of Parkinson's disease patients suggests alterations linked to intestinal inflammation — [Source link](#)

Stefano Romano, George M. Savva, Janis R. Bedarf, Janis R. Bedarf ...+4 more authors

Institutions: Norwich Research Park, University of Bonn, University of East Anglia

Published on: 12 Aug 2020 - medRxiv (Cold Spring Harbor Laboratory Press)

Topics: Microbiome, Dysbiosis, Gut flora, Lachnospiraceae and Akkermansia

Related papers:

- [Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation.](#)
- [Gut-Brain Axis: Potential Factors Involved in the Pathogenesis of Parkinson's Disease.](#)
- [Gut Vibes in Parkinson's Disease: The Microbiota-Gut-Brain Axis.](#)
- [Gut microbiota in Parkinson disease in a northern German cohort.](#)
- [Gut microbiota differences between healthy older adults and individuals with Parkinson's disease: A systematic review.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/meta-analysis-of-the-gut-microbiome-of-parkinson-s-disease-30zcyo6zzj>

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

Stefano Romano^{a*}, George M. Savva^a, Janis R. Bedarf^{b,a}, Ian G. Charles^{a,c}, Falk Hildebrand^{a,d*},
4 Arjan Narbad^a

Running title: Meta-analysis of the Parkinson's disease gut microbiome

6

^aQuadram Institute Bioscience, Norwich Research Park, Norwich, UK

8 ^bDepartment of Neurology, University of Bonn, Bonn, Germany

^cUniversity of East Anglia, Norwich Research Park, Norwich, UK.

10 ^dEarlham Institute, Norwich, Norfolk, Norwich Research Park, UK

12

To whom correspondence should be addressed: stefano.romano@quadram.ac.uk,

14 falk.hildebrand@quadram.ac.uk

16

18

20 Key words: Parkinson's disease, microbiome, meta-analysis, metagenomics, gut-brain axis

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

2 1

22 **Abstract**

24 The gut microbiota is emerging as an important modulator of neurodegenerative diseases, and
26 accumulating evidence has linked gut microbes to Parkinson’s disease (PD) symptomatology and
28 pathophysiology. PD is often preceded by gastrointestinal symptoms and alterations of the enteric
30 nervous system accompany the disease. Several studies have analyzed the gut microbiome in PD,
32 but a consensus on the features of the PD-specific microbiota is missing. Here, we conduct a meta-
34 analysis re-analyzing 10 currently available 16S microbiome datasets to investigate whether
36 underlying alterations in the gut microbiota of PD patients exist. We found consistent alterations in
PD-associated microbiome, which are significant and robust to confounders across studies, although
differences in microbiome structure between PD and controls are limited. Enrichment of the genera
Lactobacillus, *Akkermansia*, and *Bifidobacterium* and depletion of bacteria belonging to the
families Lachnospiraceae and Ruminococcaceae, which are 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 explain the recurrent gastrointestinal symptoms
affecting PD patients.

38

40

42

44

46

Introduction

48 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
50 of 100-300/100,000 people, and due to the increase in aging population the number of PD patients
is expected to double by 2030¹. PD affects predominantly dopaminergic neurons in the brain,
52 leading to decreased dopamine levels and motor impairments². Its pathological hallmark has long
been considered to be the intracellular deposition of aggregated α -synuclein, leading to neuronal
54 cell death and neuro-inflammation³. PD is now considered a multi-systemic disease, affecting the
central as well as the enteric nervous system (CNS, ENS), resulting in several non-motor
56 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
58 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
60 that the microbiota can affect α -synucleinopathy as well as neuro-inflammation. Thus, the
microbiota is a putative therapeutic target and has the potential for developing diagnostic
62 biomarkers.

PD patients can have increased gut permeability and inflammation^{7,8}, and these have been
64 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
66 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
68 patients^{10,11}. To date, more than 20 cohort-studies have investigated the composition of the PD gut
microbiota. Over 100 differently abundant taxa between PD patients and controls have been
70 reported¹⁰⁻¹⁹, and some studies detected association between taxa abundances and disease
severity^{11,12,16}. Several studies suggested that PD patients have an altered gut microbiota compared
72 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
74 *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
76 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

78 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
80 controls, whereas others found no differences in abundances¹¹ or found these taxa enriched in PD
patients^{13,20}.

82 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
84 variability of the gut microbiota across populations, lifestyles, and diets. To further elucidate the
significance of changes in the intestinal microbiota composition in PD and to evaluate its potential
86 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 a thorough meta-analysis
88 (pooled re-analysis) of all ten available studies that described the gut microbiome in PD through
16S rRNA amplicon sequencing. We apply a standardized workflow to analyze each study
90 individually and combined different statistical approaches to identify the major changes affecting
the gut microbiome of PD patients across sampling cohorts.

92 **Results**

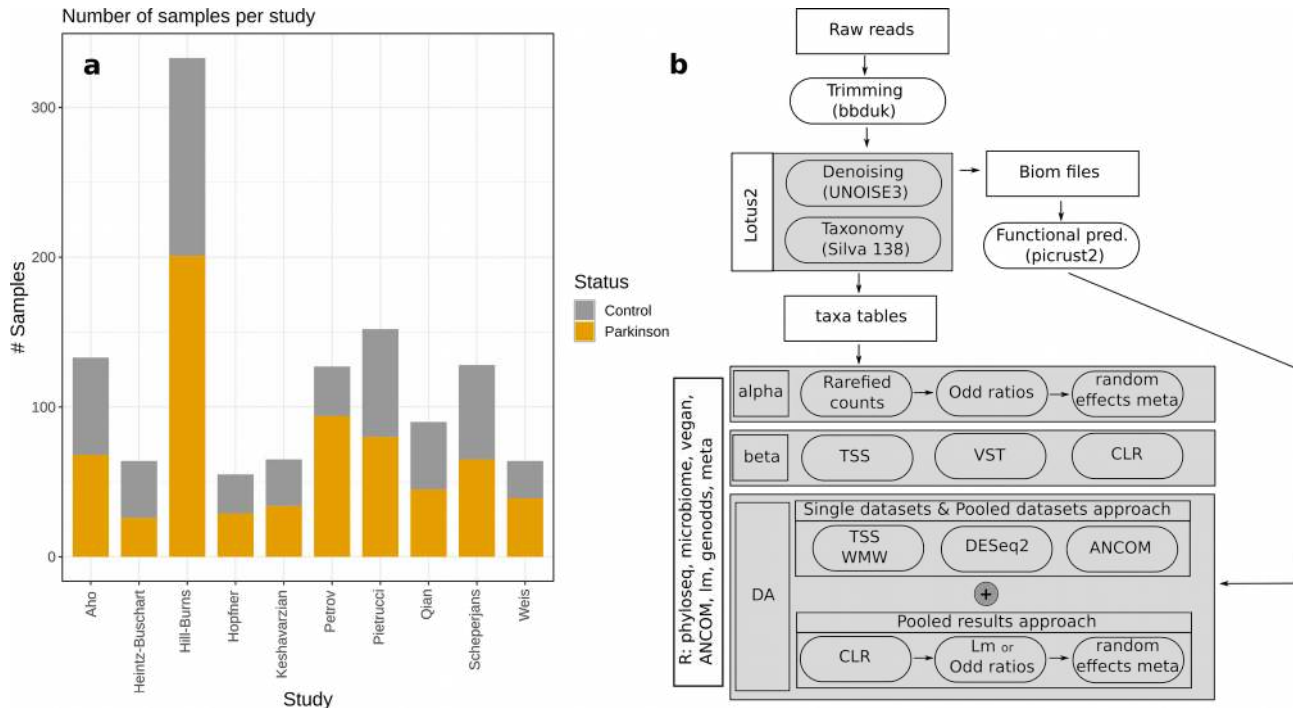
Study selection

94 We identified a total of 22 studies that investigated the PD-associated gut microbiome using
16S rRNA gene amplicon sequencing (Supplementary Table 1). Of these, ten made raw sequencing
96 data available and were re-analyzed in our study. These ten studies covered nine different cohorts
(one was reported at baseline then at follow-up two years later), across six countries (Table 1).
98 Overall, this resulted in 1,211 samples (Fig 1A) obtained from all case-control studies. Cases were
usually selected from clinics local to the investigating teams, were at different stages of the disease,
100 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
102 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,
104 while for others there were significantly more males in case compared to control groups^{12,13}.

Study	Design	Disease duration & medications ^a	Country	Sampling	DNA extr.	16S Region	Seq. Tech. ^b	PE vs SE ^c
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
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
Keshavarzian et al. ¹¹	Case-control	29.7±11 years, 66.4% medicated	USA	Home collection, delivered in GasPak	FastDNA Soil Kit	V4	MiSeq	SE
Petrov et al. ¹⁷	Case-control	NA	Russia	NA	Mechanical disruption	V3-V4	Miseq	SE
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
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
Aho et al. ¹⁶	Matched case-control	Median 8.5 years, 73% on l-dopa	Finland	Home collection, DNA stabilizer, stored in fridge	PSP-Spin Stool Kit	V3-V4	MiSeq	PE
Scheperjans et al. ¹⁵	Matched case-control	Median 6.5 years, all but 2 using medications (77.8% on dopamine agonist)	Finland	Home collection, DNA stabilizer, stored in fridge	PSP-Spin Stool Kit	V3-V4	MiSeq	PE
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
Weis et al. ²⁴	Matched case-control	82 ± 56 months, 85.3% medicated	Germany	MED AUXIL fecal collector set	FastDNA Spin Kit	V4-V5	IonTorrent	PE

Table 1 | Technical details of the studies included in the meta-analysis. The table reports the following information: study design (a: disease duration at sampling reported as mean and respective standard deviation, and proportion of medicated patients), sample collection, DNA extraction, and sequencing techniques (b: sequencing platform; c: paired-ends vs single-ends). NA indicates that the information was not reported in the original article. RT indicate room temperature. In many studies the proportion of medicated patients was calculated for individual type of drugs. Hence, we could not estimate the total portion of patients undergoing pharmaceutical treatments, and we report here only the type of drugs with the highest proportions of medicated patients.

Various sampling protocols were used across studies, with considerable variation in the method used 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 16S variable region and sequencing strategy (paired-ends vs single-ends) varied considerably (Table 1). Considering the heterogeneity across studies, we first re-analyzed each single dataset individually, then we used a combination of statistical approaches to obtain a consensus overview of gut microbiome structure in PD accounting for the heterogeneity between studies. 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.

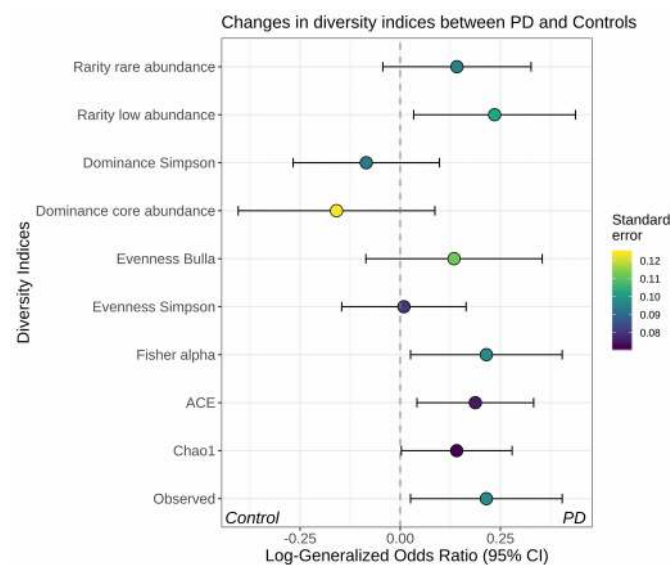


124 **Figure 1 | Sample distribution across studies (a) and bioinformatic work-flow adopted in our study (b).** The
 125 number of control and PD samples is reported and refers to the data that could be recovered from the Sequence
 126 Read Archive (SRA) or the European Nucleotide Archive (ENA). TSS = Total Sum Scaling; VST = Variance
 127 Stabilizing Transformation; CLR = Centered Log-Ratios; WMW = Wilcoxon-Mann-Whitney test; ANCOM =
 128 ANalysis of COmposition of Microbiomes; Lm = linear models.

The gut microbiome differs significantly between PD patients and controls

130 Measures of microbial alpha-diversity, based on species profiles, were higher in PD samples
 131 compared to controls in three out of ten studies (Supplementary Figure 1). Interestingly, these three
 132 datasets were the only ones using single-end sequencing approaches, suggesting that this might
 133 influence the estimation of bacterial diversity. These differences were still overall significant when
 134 we pooled estimates across studies using random-effects meta-analysis (Fig 2, Supplementary
 135 Figure 1). Specifically, PD samples had a higher overall richness as indicated by a significantly
 136 higher number of observed species and higher Fisher's alpha, ACE, and Chao1 indices (Fig 2;
 137 Supplementary Figure 1). Our analyses suggest that this higher diversity might derive from a
 138 decrease in the abundance of dominant species and an increase in rare/low abundant ones, as
 139 dominance indices were lower and rarity indices were significantly higher in PD patients (Fig 2;
 140 Supplementary Figure 1). Previous studies reported a higher abundance of Firmicutes in control
 141 samples compared to PD¹¹, and the Firmicutes to Bacteroidetes ratio (F/B ratio) has been frequently
 142 used to assess gut-health. Therefore, we calculated F/B ratios across all studies. Only in the study of

Keshavarzian et al¹¹ did we observe a significant difference in the F/B ratio between PD and
144 control, but this difference was not overall significant (Supplementary Figure 2). Similarly, Aho et
al¹⁶ reported that controls had an increased *Prevotella* to *Bacteroides* ratio (P/B ratio) in the baseline
146 samples¹⁵ of their longitudinal study. We confirmed this result and detected an increased P/B ratio in
two other studies, but did not detect the same association in the rest of the datasets, and there was
148 only weak evidence for a higher P/B ratio in controls when results were pooled (Supplementary
Figure 2). Omitting the baseline samples from the longitudinal Finnish cohort did not alter the
150 conclusions of the alpha-diversity analyses, and led to a decrease in the P/B ratio difference
between PD and controls (data not shown).



152 **Figure 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-effects meta-analysis
154 approach. The log-generalised 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 Chao1, ACE, and
156 Fisher's alpha indices. To estimate evenness, which indicates how different the abundances of the species in a
community are from each other, we used the Bulla and Simpson indices. Finally, we estimated dominance, which
158 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
160 (higher richness) than controls and this is likely a consequence of an increase in rare taxa (rarity).

The genera *Bacteroides* and *Prevotella* and the Firmicutes phylum are key gut microbiota
162 taxa that have different abundances in the three proposed enterotypes (ET_B, ET_P, and ET_F,
representing the *Bacteroides*, *Prevotella* and Firmicutes enterotypes, respectively)²⁵. To verify the
164 prevalence of these gut microbiome types amongst PD patients, we assigned each microbiome to

one of the three known enterotypes, and managed to classify 589 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 Figure 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.

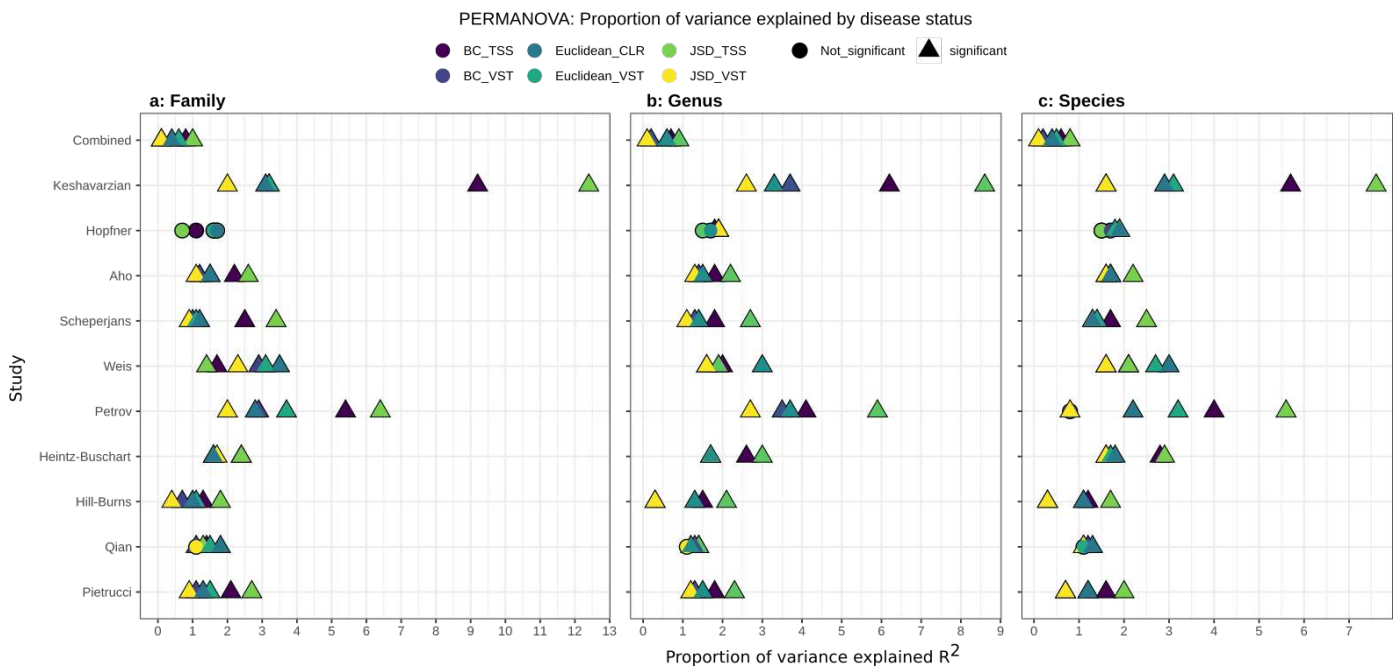


Figure 3 | The gut microbiome structure differs significantly between PD patients and controls. Data were normalized using three independent approaches (Variance Stabilizing Transformation, VST; Total Sum Scaling, TSS; Centered Log-Ratio, CLR) and beta-diversity was estimated using three indices (Bray-Curtis, 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 microbiota 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.

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
186 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
188 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
190 most studies, irrespective of the normalization-distance combinations, disease status significantly
explained the differences within the data ($p < 0.05$), even though it accounted only for a limited
192 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 significantly explained the separation of the
194 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 54% of the variance.

196 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
198 identifying which study-specific factors most defined the differences across datasets. We used the
normalization-distance pairs which best captured the variability of the data (Fig 3) to perform a
200 distance-based Redundancy Analysis (dbRDA) on the pooled data (Supplementary Figures 4-6).
First, we ordinated the combined data without constraints and without accounting for the variability
202 introduced by the study. In accordance with the previous PERMANOVA analyses, the separation of
the samples was driven by the study of origin (Supplementary Figures 4-6). We then inferred the
204 degree of similarities between studies using the sample coordinates in the dbRDA (Supplementary
Figures 7, see methods for details). The only four strongly divergent datasets were from: Weis et
206 al.²⁴, 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
208 gene; Keshavarzian et al.¹¹ who collected samples in anaerobic pouches; and from Heintz-Buschart
et al.¹³ who immediately flash-froze the samples after collection and used a lab-specific DNA
210 extraction protocol. We then verified which study-specific aspect (e.g. sequencing strategy, DNA
extraction) most influenced the structure of the bacterial communities. We created additional
212 dbRDA constraining the data by each potential confounding factor. Each factor significantly shaped
the clustering of the data when considered individually, and this was observed for all taxonomic
214 ranks and normalization approaches we used (Supplementary Table 2). In general DNA extraction
protocols, country of origin, and 16S variable region were the factors that explained most of the
216 variance within the data (Supplementary Table 2). Finally, we compared the dbRDA models

constrained by all confounding factors and disease status with the one constrained only by the
 218 disease status and study and verified that both models explained the same proportion of data
 variability (Supplementary Table 2). Hence, removing the influence of the study will
 220 simultaneously eliminate the effects of other known study-specific confounding factors.

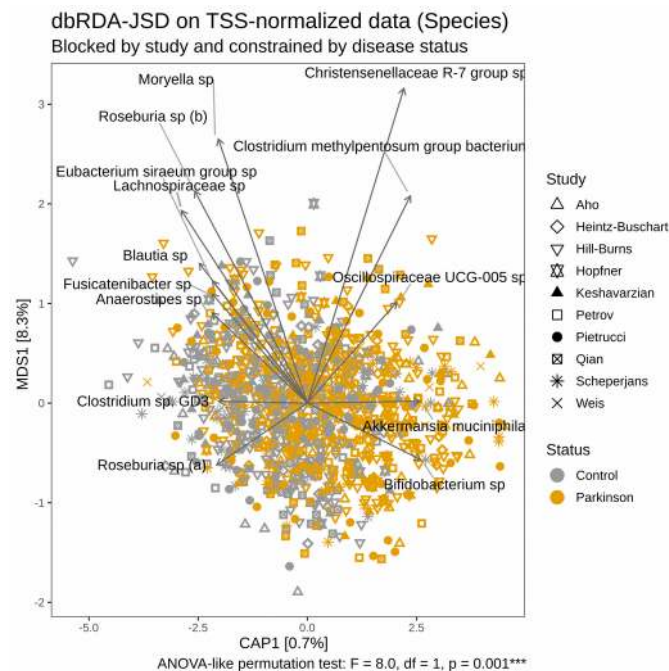


Figure 4 | Most important species driving the divergence of the gut microbiota between PD patients and controls. Distance-based Redundancy analysis (dbRDA) was performed on Jensen-Shannon divergence (JSD) calculated on data normalized through total sum scaling (TSS). dbRDA was conditioned by study and 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 the conditions are reported.

Accounting for the variability introduced by the study within the dbRDA drastically
 230 decreased the batch effect, irrespectively of the normalization-distance pair used (Supplementary
 Figures 4-6). However, samples did not cluster according to the disease status, suggesting that the
 232 environmental variability is higher than the variability explainable by the disease. Therefore, we
 constrained the dbRDA conditioned for study by disease status, to maximize the divergence
 234 between PD and control samples. We used this approach to determine the main taxa driving the
 separation between conditions. In accordance with the above results, the newly created constrained
 236 axis along which PD and controls diverged, significantly explained the clustering of the data

(ANOVA-like permutation test; Fig 4, Supplementary Figure 5-6), but accounted for only <1% of
238 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
240 influenced the separation of the samples. The divergence between PD and controls was mainly
driven by the family Bifidobacteriaceae and Akkermansiaceae, which were more enriched in PD,
242 and the family Lachnospiraceae, which was more abundant in control samples (Supplementary
Figure 8). Similarly relevant, but with a minor difference between conditions were the families
244 Rikenellaeae, Porphyromonadaceae, Christensenellaceae, and the *Clostridium methylpentosum*
group in the Oscillispirales order, all of which were more enriched in PD (Supplementary Figure 8).
246 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,
248 whereas several species belonging to the Lachnospiraceae family caused the divergence of control
samples (Supplementary Figure 7, Fig 4). When we omitted the dataset of Scheperjans et al¹⁵, the
250 overall results did not change, and only minor differences were observed (some genera and species
in the Lachnospiraceae and Christensenellaceae family were not detected as main drivers of sample
252 separation).

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

254 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
256 between PD and controls varied greatly across the studies and methods we used (Supplementary
Figure 9, Supplementary Table 3). Amongst all studies, the highest number of significant taxa were
258 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 Figure 9, Supplementary Table 3). To
260 obtain a generalizable overview of the PD-associated microbiome, we combined two independent
approaches that we refer to as Pooled data and Pooled results approach. In the first, we pooled the
262 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
264 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
266 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,

268 Supplementary Figures 10, 11, Supplementary Table 4).

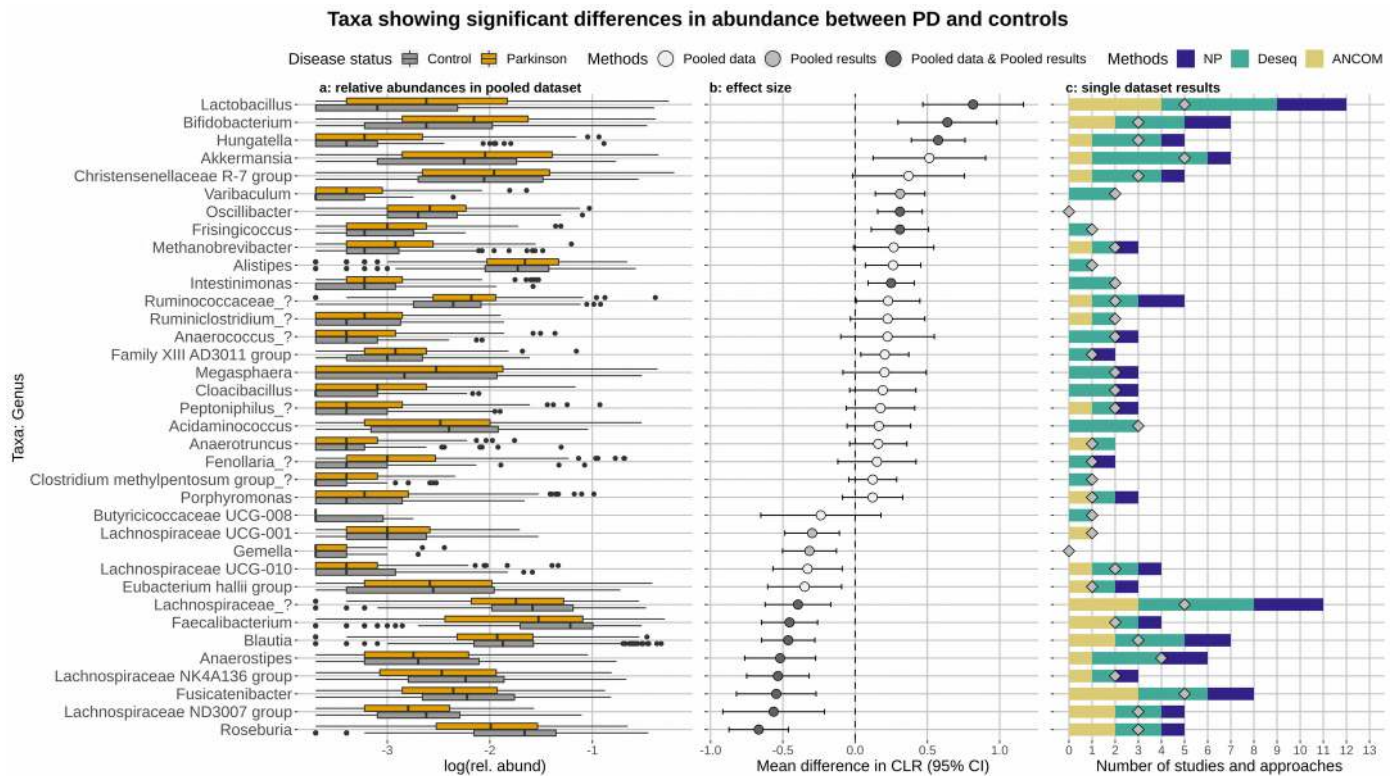


Figure 5 | Genera showing a significant difference in abundance between PD patients and controls. The relative abundances of the genera retrieved from the pooled data are reported in panel a. Effect size was estimated via the mean difference in CLR (b) using a random-effect meta-analysis approach on all taxa resulting differentially abundant in the Pooled results or Pooled data approaches. The color of the dots indicate which of the above two 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. C shows the number of times each genus was detected differentially abundant between PD patients and control samples across studies (diamonds) and approaches (bars). We used ten studies and three approaches, hence the maximum number of times a taxon can be detected differentially abundant is 30.

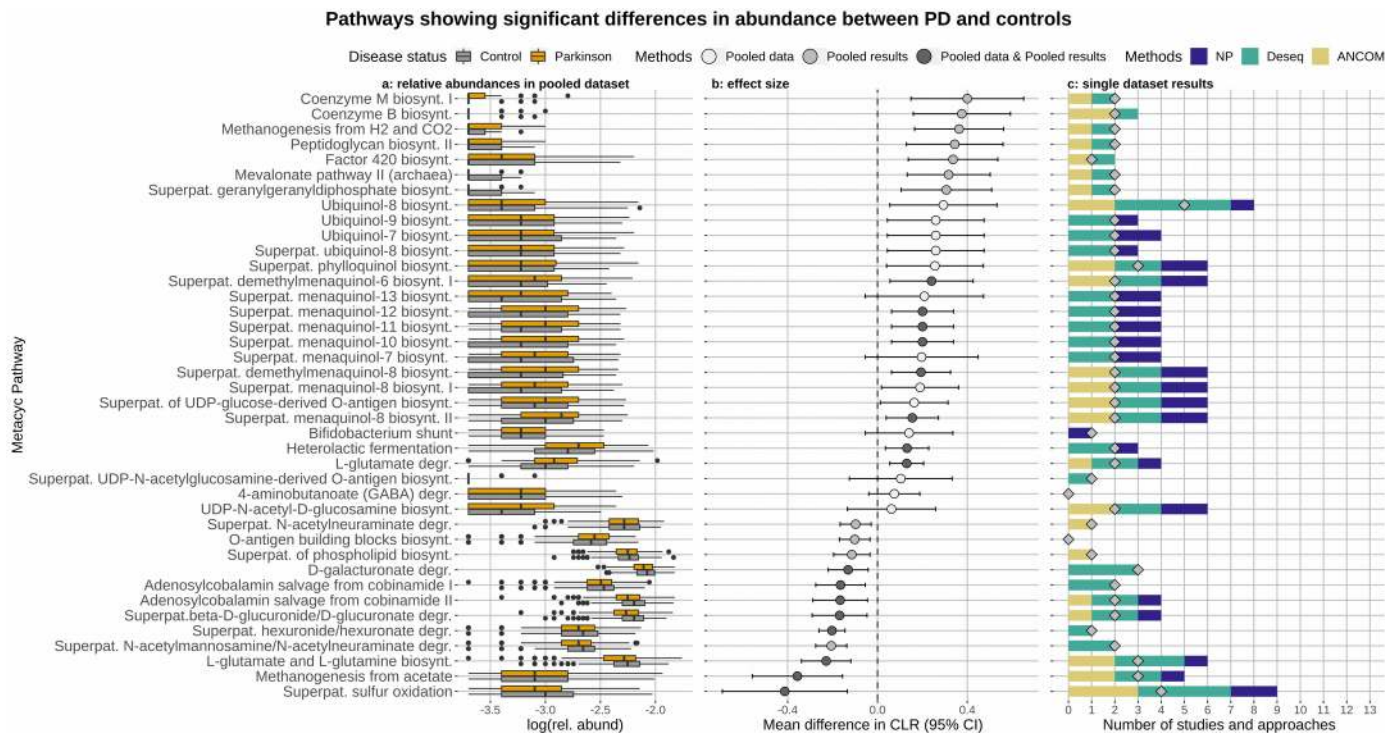
After obtaining a first list of differentially abundant taxa, we used the metadata made available by five of the ten studies we re-analyzed to verify whether age and/or gender influenced the abundances of the taxa we detected to be enriched or depleted in PD. By comparing generalized linear mixed models (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 Table 5). For most of the taxa, the disease status was required to best explain taxa abundances, underlining the robustness of our approaches. We did not detect any taxa for which gender and/or age alone could best describe the data. However, for 11 Species, 13 Genera, and 7

286 Families the available data did not allow to clearly establish whether the disease status was an
essential factor to explain taxa abundances (Supplementary Table 5). Hence, we did not consider
288 these taxa further. For example, the Prevotellaceae family and a species within this family
(Supplementary Table 5) were both detected differentially abundant between PD and controls.
290 However, the abundance of both taxa was also influenced by gender and age (Supplementary Table
5, Supplementary Figure 12).

292 Among the remaining genera, after controlling for available metadata, *Roseburia*, *Blautia*,
Fusicatenibacter, *Faecalibacterium*, *Anaeorostipes* and other two unknown genera belonging to the
294 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
296 the *Fusicatenibacter*, *Blautia*, and *Roseburia* genera were strongly enriched in control samples
(Supplementary Figures 10, 11). The family Butyricocccaceae was also more abundant in controls,
298 even though it was detected differentially abundant in fewer studies (Supplementary Figure 10). PD
samples were instead most often enriched in the genera *Lactobacillus*, *Bifidobacterium*, *Hungatella*,
300 and *Akkermansia* (Fig 5). Additionally, the R-7 group of the Christensenellaceae family, the genera
Methanobrevibacter, *Oscillobacter*, *Frisingicoccus* and *Varibaculum* were also detected more
302 abundant in PD, but with a smaller effect size (Fig 5). PD samples were enriched in species
belonging to different taxonomic groups e.g. Ruminococcaceae, Christensenellaceae,
304 *Bifidobacterium*, *Lactobacillus*, *Hungatella*, and *Alistipes* (Supplementary Figure 11). Other
species, such as *Intestinimonas* sp, *Oscillibacter* sp also resulted more abundant in PD, but in fewer
306 studies (Supplementary Figure 12). In contrast, the majority of species enriched in controls
belonged to the families Lachnospiraceae and Ruminococcaceae (Supplementary Figure 12). The
308 shifts in taxa abundances outlined above were robust to the sensitivity analysis we performed
omitting the baseline data of the longitudinal Finnish cohort, which resulted in minor differences
310 affecting only taxa having a small effect size (Supplementary Figure 13).

Finally, for each dataset, we obtained hypothetical functional prediction based on the 16S
312 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
314 related to ubiquinone (Coenzyme Q; CoQ) and menaquinone (vitamin K2) biosynthesis. 4-
aminobutanoate (GABA) degradation and glutamine-glutamate metabolism, methanogenesis, and
316 lactic-type fermentation (Supplementary Table 3, 4; Fig 6, Supplementary figure 14). Instead, the

control samples were enriched in pathways involved in the biosynthesis of cobalamin (vitamin
 318 B12), degradation of glucuronate and galactoglucuronate, and methane production from acetate
 degradation (Supplementary Table 3, 4; Fig 6, Supplementary Figure 14).



320 **Figure 6 | Pathways showing a significant difference in abundance between PD patients and controls.** Only
 321 selected relevant pathways are shown (a full overview is reported in Supplementary Figure 13). The relative
 322 abundances of the pathways retrieved from the pooled data are reported in panel a. Effect size was estimated via
 323 the mean difference in CLR (b) using a random-effect meta-analysis approach on all taxa resulting differentially
 324 abundant in the Pooled results or Pooled data approaches. The color of the dots indicate which of the above two
 325 approaches detected the pathway differentially abundant. Pathways more abundant in controls have an effect size
 326 shifted to the left, whereas pathways more abundant in PD have an effect size shifted to the right. C shows the
 327 number of times each genus was detected differentially abundant between PD patients and control across studies
 328 (diamonds) and approaches (bars). We used ten studies and three approaches, hence the maximum number of
 times a taxon can be detected differentially abundant is 30.

330 Discussion

In recent years, several studies have analyzed the gut microbiome of PD patients and
 332 reported 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
 334 contrasting results have been reported. Hence, we performed a meta-analysis using all publicly

available 16S rRNA gene amplicon datasets comparing PD patients with healthy controls, to
336 address reproducibility and identify targets to foster further experimental work. By integrating
different statistical methods we re-analyze all data and offer a comprehensive and robust consensus
338 of the most consistent features of the gut microbiome associated with PD. Studies were
heterogeneous in sampling populations and methodological approaches (Table 1, Supplementary
340 Table 2; Supplementary Figure 7), and the inter-study variability was the main factor driving
bacterial community structures (Fig 3, Supplementary Figures 4-7; Table 1, Supplementary Table
342 2), as observed in previous microbiome-based meta-analyses performed in the context of diet and
colorectal cancer^{27,28}. Our analyses suggest that particular methodological approaches, such as
344 sample collection and transport, sequencing platform, and the chosen 16S variable region might be
the main reasons for the heterogeneity across the datasets we considered (Supplementary Figure 7,
346 Table 1, Supplementary Table 2). Variability across the studies was also a reflection of the different
sampling populations used in the different studies (Table 1). For example, the proportion of PD
348 patients using medication as well as the duration of the disease varied across datasets. Also, controls
were sometimes closely matched but sometimes differed considerably from cases in terms of age,
350 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.

352 By stratifying the analysis by study, we could simultaneously exclude the effect of all other
known study-specific confounding aspects, such as country of origin, DNA extraction, and
354 sequencing (Supplementary Table 2, Supplementary Figures 4-6). In agreement with previous
studies^{11-13,15,18,19}, we show that the gut microbiome of PD patients significantly differs from the one
356 of controls (Fig 3, Supplementary Figure 8). Although PD can explain only a limited portion of data
variability (Fig 3, Supplementary Figures 4-6; Supplementary Table 2) the observed differences are
358 robust to technical confounders and 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
360 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 (Fig 1, Supplementary Figure 1), which is a
362 typical alteration observed in dysbiosis associated with inflammatory bowel disease and irritable
bowel syndromes (IBD, IBS)²⁹. An increase in bacterial diversity in the gut microbiome of PD
364 patients has been previously reported both in studies we re-analyzed and studies for which data
were not available^{11,18,21,30}. Similarly, a recent work reported no differences in OTU-based alpha-
366 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
368 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*,
370 *Anaerostipes* (Lachnospiraceae family), and *Faecalibacterium* (Ruminococcaceae family) (Fig 5,
Supplementary Figures 10-12; Supplementary Tables 3, 4), a finding seen in other neuro-
372 inflammatory and rheumatologic disorders³²⁻³⁴. Most of these taxa are abundant and widespread
bacteria in the gut microbiota of healthy individuals, they are major butyrate-producers and have
374 often been found depleted in IBD²⁹. Similarly depleted in IBD are bacteria of the Butyricicoccaceae
family³⁵, which are important butyrate producers and were highly depleted in PD in our analyses
376 (Supplementary Figure 10). The depletion of these taxa suggests a low level of butyrate in the gut of
PD patients. Butyrate is a fundamental energy source for intestinal epithelial cells and has been
378 reported to reinforce the intestinal epithelium as well as preventing inflammation and
carcinogenesis³⁶. Our findings are consistent with previous studies showing low levels of butyrate
380 and increased gut permeability and inflammation in PD patients^{7,9}. SCFA are not only relevant for
gut health, but they can also influence the ENS, have systemic anti-inflammatory properties,
382 promote normal microglia development, and potentially affect epigenesis in the CNS³. Importantly,
PD patients have been shown to have increased levels of various cytokines in both the colon and
384 serum, suggesting that they suffer from systemic inflammation which could result in microglial
activation driving disease progression³⁷. Altogether these data suggest that the alteration in SCFA
386 production in the gut combined with an increase gut permeability and inflammation might have
systemic implications in PD. To the best of our knowledge, only Aho et al¹⁶ identified bacteria of
388 the Butyricicocaceae enriched in controls. Similarly, only Aho et al¹⁶ and Weis et al²⁴ detected
Fusicatenibacter to be significantly depleted in PD patients. Specifically, Weis et al. report that the
390 decrease of this genus together with *Faecalibacterium* was correlated to the degree of gut
inflammation²⁴. Interestingly, both these genera were low in abundance in IBS and ulcerative
392 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
394 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
396 health as underlined by retrospective studies indicating that the overall risk of developing PD in
IBD was significantly higher, reaching 28% and 30% increase in patients with Crohn's and
398 ulcerative colitis, respectively³⁷.

Our results indicate a higher abundance of the genera *Lactobacillus*, *Akkermansia*,
400 *Hungatella*, and *Bifidobacterium* in PD gut microbiome (Fig 5, Supplementary Figures 10, 11;
Supplementary Table 3, 4). The fact that the more abundant genera in PD belong to different
402 families and even orders, support the idea of increased diversity in the gut-microbiota of PD
patients (Fig 1, Supplementary Figure 1). The genus *Lactobacillus*, and the Lactobacillaceae
404 family, were the most strongly enriched taxa in PD across the studies we re-analyzed, in line with
previous findings^{11,12,17,19,41}. *Lactobacillus* strains are low abundant members of the gut microbiota
406 and their abundance varies greatly across human disease and chronic conditions⁴². Some strains of
Lactobacillus and *Enterococcus*, also enriched in PD (Supplementary Figure 11), are able to
408 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
410 intestine, but it has been reported that 10-20% can reach the large intestine⁴⁵, indicating that a
substantial amount of this molecule can be readily available for gut bacteria and could help these
412 bacteria to proliferate. Interestingly, none of the Chinese studies conducted so far detected
Lactobacillus or the family Lactobacillaceae enriched in PD^{18,20,21,23,46}. Consistently, these taxa were
414 not enriched in PD samples in the only Chinese study included in our meta-analysis. The majority
of *Lactobacillus* species have been found in food (e.g. dairy, milk, fermented food, probiotics)⁴².
416 Hence, it is possible that differences in diet between regions could explain the difference between
the Chinese cohorts and the others.

418 *Akkermansia* has been repeatedly shown to be more abundant in PD compared to controls<sup>11-
13,20,22</sup>. In general *Akkermansia* spp. are considered beneficial for human health, as they fortify the
420 integrity of the epithelial cell layer and can modulate the immune system^{47,48}. For example, a recent
study reported that these bacteria ameliorate age-related decline in colonic mucus thickness and
422 attenuate immune activation in accelerated aging mice⁴⁹. However, contrasting results regarding the
influence of *Akkermansia* spp. on gut health exist⁵⁰. Intriguingly, constipated individuals have been
424 shown to have a gut microbiome enriched in *Akkermansia*⁵¹⁻⁵³, and constipation is one of the major
non-motor symptoms in PD, often starting decades before motor symptoms arise. The increase in
426 *Akkermansia* could be a consequence of constipation, even though animal studies suggest that this
genus might contribute to an increased transit time. *Akkermansia* spp. are mucin-degrading bacteria
428 and they can lead to a depletion of the intestinal mucus-layer when the gut microbiota is
unbalanced^{54,55}. Mice receiving stool from chronically constipated patients showed drier stools,
430 decreased number of goblet cells, and impaired intestinal barrier function in association with an

increase in *Akkermansia* spp⁵⁴. The unbalanced microbiota observed in PD patients, might lead to a
432 proliferation of *Akkermansia* spp, which in turn might lead to decreased mucus thicknesses, drier
stools, and constipation. It is important to point out that multiple strains belonging to the same
434 *Akkermansia* species can co-exist in the gut and the modulation of host-response can be strain-
specific^{56,57}. For example, different *Akkermansia munichipalia* strains have different effects on the
436 differentiation of Regulatory T cells (Tregs) and SCFA production⁵⁷, both factors altered in blood
and gut, respectively, of PD patients^{6,9,58}. Altogether these data indicate that the increased abundance
438 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
440 to elucidate these aspects in PD.

Among the most abundant taxa in PD there were bacteria belonging to the
442 Christensenellaceae family (Fig 5, Supplementary Figure 10, 11), in line with previous
reports^{11,12,21,41}. This family is widespread in the gut of the human population and it is generally
444 associated with healthy phenotypes, even though their abundances positively correlate to the
intestinal transit time⁵⁹. *Christensenella* spp. can efficiently support the proliferation of
446 *Methanobrevibacter smithii* via H₂ production, explaining the recurrent co-occurrence between
these two taxa⁶⁰. *Methanobrevibacter* belongs to the Archaea domain and it is the major
448 hydrogenotrophic methane producer in the human gut, and was also more abundant in the PD
samples we re-analyzed (Fig 5). In both controls and PD patients the abundances of these taxa were
450 positively correlated (Spearman rank test: PD, Z = 10.3, p-value 0.0005; controls, Z = 8.8, p-value
0.0005; Supplementary Figure 15). Moreover, the 16S-based functional prediction we performed
452 showed that the pathways for the formation of methane from H₂ and CO₂ and for the synthesis of
key co-factors involved in methanogenesis (Coenzyme B and M, Factor 420) are strongly enriched
454 in PD (Fig 6). The increased abundance of these two co-occurring genera in PD patients might
contribute to the production of methane, which in turn could influence the intestinal transit. In fact,
456 *Methanobrevibacter* is enriched in constipated patients^{52,53}, just as *Akkermansia*, and growing
evidence indicates that methane decreases peristaltic movements slowing down intestinal transit
458 time⁶¹. Surprisingly, to the best of our knowledge, only one other study reported an enrichment in
Methanobrevibacter in PD²⁰. It is worth noting that in the 16S-based functional prediction another
460 pathway that produces methane through the degradation of acetate was enriched in controls (Fig 6).
This pathway is mainly found in Archaea of the genus *Methanosarcina*. However, we did not detect
462 these taxa enriched in controls. They were identified as significantly more abundant in controls only

in the dataset of Qian et al¹⁸ and only by a single method (Supplementary Table 3). Hence, these
464 data need to be interpreted with caution as they might be an artifact of the 16S-based predictions. It
is important to specify that the abundance of Archaea in the human gut microbiota is considerably
466 lower than that of Bacteria, and current methodologies (DNA extraction, primers used for 16S
amplification) strongly discriminate against Archaea⁶². Hence, it is possible that the abundances and
468 the diversity of these microorganisms is currently poorly represented in the available datasets.

In agreement with the only shot-gun metagenomic study so far available¹⁰, we identified
470 several predicted-pathways involved in galacturonate and glucuronate degradation depleted in the
PD-microbiome (Fig 6, Supplementary Figure 14). The enrichment of GABA degradation pathways
472 and glutamate/glutamine biosynthesis pathways in PD suggests an alteration in the enteric
production of these neurotransmitters. The gut microbiota has been previously suggested to alter the
474 glutamate-glutamine-GABA cycles in schizophrenia and autism^{63,64}, and alterations in the level of
this transmitter have been found in brains of PD patients^{65,66}. Hence, it is intriguing to speculate that
476 the gut microbiota might play a role in modulating these chemicals in PD patients. Further
experimental work will be required to verify whether these metabolic changes in the PD microbiota
478 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.
480 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^{67,68}. Hence, the increased biosynthetic
482 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
484 the potential increase of vitamin K2 production in the gut might increase systemic concentrations of
these chemicals in PD. Interestingly, vitamin K plays an important role in the biosynthesis of
486 sphingolipids⁶⁹ which are emerging as an important determinant in PD development⁷⁰. These data
suggest novel mechanisms through which the gut microbiota might potentially influence PD
488 development.

In summary, our analyses reveal underlying consistent differences in the average gut
490 microbiota composition between PD patients and controls. Variations among studies are the
strongest factor in shaping the data structure, but by accounting for the variability derived by the
492 sampling cohorts we were able to show that the alteration of the gut microbiome in PD is consistent
across studies and countries. The differences in taxa abundances between PD and controls indicate

494 that the gut microbiota of PD patients shares similarities with those of other neurological and
inflammatory gastrointestinal diseases. Taxa important in maintaining gut integrity and health via
496 the production of SCFA are depleted in PD and this together with the growing evidence of gut and
systemic inflammation in PD, point towards an important role of the gut microbiota in modulating
498 the immune function in this disease. Moreover, we were able to identify previously overlooked taxa
enriched in PD such as *Methanobrevibacter*, Butyricicocaceae, and identified some potentially
500 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
502 dysbiotic gut microbiota could influence host immune function and be linked to the gastrointestinal
symptoms often recurring in PD patients.

504 **Materials and Methods**

Study selection

506 *Search strategy*

On the 29th of March 2020 Google Scholar was searched for publications that contained all
508 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
510 the article. This resulted in 1,010 entries. Titles were then manually screened and if they contained
the words “microbiome” or “microbiota” and “Parkinson’s disease” the abstracts were further
512 consulted. Moreover, the Sequence Read Archive (SRA) in NBCI was queried with the following
term “Parkinson” [AND] “microbiome”, resulting in two additional studies (Bioprojects):
514 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.

516 *Inclusion criteria*

We included all studies comparing the composition of the gut microbiota between patients
518 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-
520 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
522 that cataloged the gut microbiome of PD patients using metagenomics (Supplementary Table 1). 9

of these studies did not make the raw data publicly available. We were unsuccessful in obtaining the
524 raw reads from the authors, as data were either protected by ethical restriction or the authors did not
answer our requests. In other cases the raw reads were available, but it was impossible to associate
526 the data with the disease status as this information was not reported in the metadata. The samples
from Scheparjans et al¹⁵, originally sequenced using a 454 technology, were recently re-sequenced
528 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
530 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

532 Raw reads were downloaded from SRA or the European Nucleotide Archive (ENA).
Adapters were removed using the bbttools suit⁷¹. Data were analyzed using Lotus⁷² and the
534 UNOISE3⁷³ 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
536 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 Table
538 6). For the datasets of Petrov et al¹⁷ and Weis et al²⁴ we had to decrease the accepted minimum error
due to the low quality of the sequencing data (Supplementary Table 6). 16S-based functional
540 predictions were obtained using the default settings in *picrust2* v2.3.0-b⁷⁴ and the Metacyc database.
In this analysis, the dataset of Qian et al.¹⁸ was not included, as with the default cutoffs the
542 sequences aligned poorly with the reference database used. Count tables for species, genera,
families, and functional predictions were then analyzed using R v3.6.2⁷⁵. Datasets were processed
544 using the *phyloseq* R-package⁷⁶, samples with < 4,500 reads were removed, and taxa with < 5
counts and predicted-functionalities with < 20 counts in < 2.5% of samples were removed. This
546 filtration steps left a total of 1,211 (530 control, and 681 PD samples) and 1,121 samples (485
control and 636 PD samples) for the taxonomic and predicted-function data, respectively. Finally,
548 enterotypes were predicted using rarefied relative abundances of genera via the
<https://enterotypes.org/> web-platform.

550

552 Statistical analyses of single studies

Analysis of alpha-diversity

554 Alpha-diversity indices at the species level were calculated using the *microbiome* R-
package⁷⁷, after rarefying without re-sampling at the even depth of 5,000. Due to rarefaction 8
556 samples were further removed, leaving a total of 1,203 samples (523 control and 680 PD samples).
We measured richness using the number of Observed species, the Chao1, Fisher's alpha, and ACE
558 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
560 50%, 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
562 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. Additionally, we calculated the
564 ratios of Firmicutes to Bacteroidetes phyla and *Prevotella* to *Bacteroides* genera, as \log_2 ratios of
their relative abundances. In each dataset, the differences in alpha-diversity between control and PD
566 samples were assessed using Agresti's generalized odd ratios using the *genodds* function in the
genodds R package⁷⁸. This statistic, based on ranks and analogous to the U statistic underlying the
568 Mann-Whitney test, does not make strong assumptions about the distributions of measures and is
comparable between measures of diversity with different scales.

570 *Beta-diversity and differential abundance analyses*

For each dataset, beta-diversity and differential abundance analyses were performed using
572 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
574 Wilcoxon-Mann-Whitney (WMW) tests; ii) variance stabilizing transformation (VST) and DA
inference using DESeq2⁷⁹; iii) compositional approach based on centered log-ratios (CLR) and DA
576 inference using analysis of composition of microbiomes (ANCOM)⁸⁰. We then reported the number
of times each taxon showed a significant difference in abundance between PD and controls across
578 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 Fig 5,
580 Supplementary Figures 10, 11, 14). Differential abundances of *picrust2* predicted functionalities
between PD and controls were inferred using the same approach outlined above. The rarefaction

582 used in the TSS approach did not result in a loss of samples for the 16S-based predicted
functionalities.

584 *Total sum scaling (TSS) and non-parametric tests*

After rarefying without re-sampling at the even depth of 5,000, data were normalized by
586 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
588 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
590 the *adonis2* (analysis of variance using distance matrices, ADONIS) function in the *vegan* R-
package⁸¹. DA analysis was performed using a two-sided WMW test, using Benjamini Hochberg
592 (BH) p-value correction.

Variance stabilizing transformation (VST) and DESeq2 analyses

594 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 ratios⁸².
596 Euclidean, BC, and JSD distances were used as beta-diversity estimators after normalizing the data
via VST. Statistical differences between control and PD groups were tested using the *adonis2*
598 function as specified above. DAs were calculated using default DESeq2 parameters that include a
negative binomial GLM fitting and a Wald test⁷⁹. Multiple testings were accounted for using BH p-
600 value correction.

Compositional analysis: Centered log-ratios (CLR) and ANCOM

602 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
604 *zCompostions* R package⁸³. Euclidean distances, that for such data correspond to Aitchison
distances, were then calculated⁸⁰. Statistical differences between control and PD groups were tested
606 using the *adonis2* function as specified above. DA analysis was performed via the ANCOM
approach as implemented in the R script *ancom_v2.0*⁸⁴ using a 0.95 zero-cutoff and significance at
608 the 0.6 percentile.

610 **Statistical analyses of the combined studies.**

Analysis of alpha-diversity

612 The Agresti's generalized odd ratios estimated for each alpha-diversity index and each
individual study were pooled using a random effect meta-analysis using the function *metagen* in the
614 R package *meta*⁸⁵.

Analysis of beta-diversity

616 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
618 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
620 measure that captured a highest fraction of the variability in the pooled dataset to compute distance-
based Redundancy Analyses (dbRDA). dbRDAs were performed using the "CAP" option in
622 *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
624 conditioning for study and constraining for disease status (PD vs Control):

$$distance \sim 1$$

626 $distance \sim 1 + Condition(study)$

$$distance \sim status + Condition(study)$$

628 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
630 study-dependent confounding factors such as country, sequencing platform (e.g. MiSeq vs
IonTorrent), sequencing approach (single-end vs paired-end), amplified region (e.g. V4 vs V1-V2),
632 and extraction method and type by creating additional dbRDAs and constraining the data for each
individual factors. The effect of each constraining variable was tested using an ANOVA-like
634 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
636 confounding factors. We compared the adjusted R^2 (R^2_{adj}) of a dbRDA obtained using the full model
 $distance \sim country + 16Sregion + ends + seq + extraction + extraction\ type + status$ with the one

638 of a reduced model including only disease status and study ($distance \sim study + status$). Similar R^2_{adj} ,
differences $\leq 0.1\%$, indicates that the two models are equivalent. The influence of confounding
640 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
642 constrained and conditioned dbRDA via the *envfit* function in the *vegan* package. We selected only
taxa significantly correlated with the clustering ($p\text{-value} < 0.01$), and showing the highest degree of
644 variation ($> |0.1|$ for genus and species, and $> |0.075|$ for family) along the constrained axis (CAP1).

Similarity amongst studies

646 We used the unconstrained and unconditioned dbRDA performed on the TSS normalized species
data to estimate dissimilarity amongst studies. We selected the coordinates of each sample across all
648 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
650 package *usedist*⁸⁶. Similarity among studies was then visualized using non-metric multidimensional
scaling (NMDS) via the *metaMDS* function in the *vegan* R package.

Differential abundance analysis

We combined two independent approaches to gather a consensus view on the taxa differentially
654 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
656 and processed with the same methods used for the single datasets: i) TSS normalization on rarefied
data and *independence_test* in the *coin* R package⁸⁷ blocking data for the study; ii) DESeq2
658 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 with study (via the formula
660 “*random.formula = ~1 + status|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
662 was set at ≤ 0.05 . If a taxon or pathway had a significant difference in abundance in 2 out of three
approaches, it was then retained (Consensus).

664 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
666 dataset using CLR after adding a pseudo-count of 1 to 0 values. We then selected all taxa and

668 pathways detected in at least 3 studies and estimated their shift in abundance between PD and
670 controls using linear models for family, genera, and 16S-based predicted functionalities and Agresti'
672 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
674 in the Pooled data (2 out of 3 methods referred to as Consensus) or Pooled results approach were
further considered. All taxa having abundances potentially influenced by age and/or gender were
676 then removed (see below). For each taxa/pathway the effect size and the respective 95% CI were
estimated using the Pooled results approach. Finally, correlation between the genera
678 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*
680 in the *coin* R package).

Influence of confounding factors on differential abundances

682 Of the ten studies we used only five had metadata available, and among the latter, only age
and gender were always reported. We assessed the influence of age and gender on the abundance of
684 the taxa we previously identified as significantly enriched or depleted in PD using generalized
linear mixed models (GLMMs) controlling for zero-inflation as implemented in the R package
686 *glmmTMB*⁸⁸:

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

688 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
690 Poisson distribution. First, we constructed zero-inflated and non zero-inflated models, and choose
the best model using the Akaike information criterion (AIC; $\Delta\text{AIC} \geq 2$). We then created reduced
692 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
694 = $(1 | \text{study})$). We then compared all models using the *model.sel* function and the AIC in the R
package MuMIn⁸⁹. If one of the best models (within a ΔAIC of 2) did not contain the variable

696 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 analyses. If all best models
698 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
700 were rarefied to a fixed depth of 10,000 to avoid overparameterization.

Code availability

702 The R code used in this study will be made publicly available on GitHub after the peer-review
process is completed.

704 Acknowledgments

We are grateful for the technical support provided by A. Telatin and the assistance from S.J. Green
706 and A. Keshavarzian in obtaining the raw reads of their study from SRA. This research was
supported in part by the NBI Computing infrastructure for Science (CiS) group through access to
708 High-Performance Computing infrastructures. The authors gratefully acknowledge the support of
the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded
710 by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and its
constituent project BBS/E/F/000PR10356. GS was supported by the BBSRC Core Capability Grant
712 BB/CCG1860/1. We acknowledge M. Mayer and R. Ansorge for constructive comments on the
manuscript.

714 Contributions

SR designed the study, performed bioinformatic and statistical analyses, interpreted the results and
716 wrote the paper. GMS assisted in statistical analyses and data interpretation. JRB helped in data
interpretation. FH provided support in bioinformatic and statistical analyses, and data interpretation.
718 AN and IGC helped in data interpretation.

Competing interest

720 The funding bodies had no role in the study design, execution of the analyses, and data
interpretation. The authors have no competing interest to declare.

722 References

1. Elbaz, A., Carcaillon, L., Kab, S. & Moisan, F. Epidemiology of Parkinson's disease. *Rev. Neurol. (Paris)* **172**, 14–26 (2016).
2. 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).
3. 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* vol. 252 357–450 (Elsevier, 2020).
4. Cersosimo, M. G. & Benarroch, E. E. Pathological correlates of gastrointestinal dysfunction in Parkinson's disease. *Neurobiol. Dis.* **46**, 559–564 (2012).
5. 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. (Berl.)* **136**, 345–361 (2018).
6. 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).
7. 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).
8. Clairembault, T. *et al.* Structural alterations of the intestinal epithelial barrier in Parkinson's disease. *Acta Neuropathol. Commun.* **3**, (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**, (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**, (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. 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).
25. MetaHIT Consortium *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. 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* S1931312819303026 (2019) doi:10.1016/j.chom.2019.06.013.
28. 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).
29. 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).
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. Swidsinski, A. *et al.* Reduced Mass and Diversity of the Colonic Microbiome in Patients with Multiple Sclerosis and Their Improvement with Ketogenic Diet. *Front. Microbiol.* **8**, 1141 (2017).
34. Bellocchi, C. & Volkmann, E. R. Update on the Gastrointestinal Microbiome in Systemic Sclerosis. *Curr. Rheumatol. Rep.* **20**, 49 (2018).
35. Eckhaut, V. *et al.* *Butyricoccus 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.* mds.28004 (2020) doi:10.1002/mds.28004.
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.* **105**,

- 16731–16736 (2008).
40. Martín, 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. Baldini, F. *et al.* Parkinson's disease-associated alterations of the gut microbiome can invoke disease-relevant metabolic changes. Preprint at <http://biorxiv.org/lookup/doi/10.1101/691030> (2019) doi:10.1101/691030.
42. 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).
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. Li, W. *et al.* Structural changes of gut microbiota in Parkinson's disease and its correlation with clinical features. *Sci. China Life Sci.* **60**, 1223–1233 (2017).
47. 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).
48. Derrien, M. *et al.* Modulation of Mucosal Immune Response, Tolerance, and Proliferation in Mice Colonized by the Mucin-Degrader *Akkermansia muciniphila*. *Front. Microbiol.* **2**, (2011).
49. van der Lugt, B. *et al.* *Akkermansia muciniphila* ameliorates the age-related decline in colonic mucus thickness and attenuates immune activation in accelerated aging *Ercc1- Δ 7* mice. *Immun. Ageing* **16**, 6 (2019).
50. 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).
51. 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).
52. 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).
53. 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).
54. Cao, H. *et al.* Dysbiosis contributes to chronic constipation development via regulation of serotonin transporter in the intestine. *Sci. Rep.* **7**, 10322 (2017).
55. 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).
56. Guo, X. *et al.* Different subtype strains of *Akkermansia muciniphila* abundantly colonize in southern China. *J. Appl. Microbiol.* **120**, 452–459 (2016).

57. 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).
58. 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).
59. Waters, J. L. & Ley, R. E. The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. *BMC Biol.* **17**, 83 (2019).
60. Ruaud, A. *et al.* Syntrophy via Interspecies H₂ Transfer between *Christensenella* and *Methanobrevibacter* Underlies Their Global Cooccurrence in the Human Gut. *mBio* **11**, e03235-19, /mbio/11/1/mBio.03235-19.atom (2020).
61. Triantafyllou, K., Chang, C. & Pimentel, M. Methanogens, Methane and Gastrointestinal Motility. *J. Neurogastroenterol. Motil.* **20**, 31–40 (2014).
62. 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, /mbio/8/6/mBio.00824-17.atom (2017).
63. 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).
64. 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).
65. 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**, (2014).
66. 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).
67. Vos, M. *et al.* Vitamin K2 Is a Mitochondrial Electron Carrier That Rescues Pink1 Deficiency. *Science* **336**, 1306–1310 (2012).
68. Shults, C. W. Therapeutic role of coenzyme Q10 in Parkinson's disease. *Pharmacol. Ther.* **107**, 120–130 (2005).
69. Denisova, N. A. & Booth, S. L. Vitamin K and Sphingolipid Metabolism: Evidence to Date. *Nutr. Rev.* **63**, 111–121 (2005).
70. 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).
71. BBMap. *SourceForge* <https://sourceforge.net/projects/bbmap/>.
72. Hildebrand, F., Tadeo, R., Voigt, A., Bork, P. & Raes, J. LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome* **2**, 30 (2014).
73. Edgar, R. C. *UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing*. Preprint at <http://biorxiv.org/lookup/doi/10.1101/081257> (2016) doi:10.1101/081257.
74. Douglas, G. M. *et al.* *PICRUSt2: An improved and customizable approach for metagenome inference*. Preprint at <http://biorxiv.org/lookup/doi/10.1101/672295> (2019) doi:10.1101/672295.
75. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for

Statistical Computing, 2019).

76. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **8**, e61217 (2013).
77. Lahti, L. & Shetty, S. *microbiome: Microbiome Analytics*. (Bioconductor version: Release (3.10), 2020). doi:10.18129/B9.bioc.microbiome.
78. Johns, H. *genodds: Generalized Odds Ratios*. (2019).
79. 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).
80. 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).
81. Oksanen, J. *et al. vegan: Community Ecology Package*. (2019).
82. Chen, L. *et al.* GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. *PeerJ* **6**, e4600 (2018).
83. 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).
84. Lin, F. H. *HuangLin/ANCOM: Third release of ANCOM*. (Zenodo, 2019). doi:10.5281/ZENODO.3577802.
85. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Health* (2019).
86. Bittinger, K. *usedist: Distance Matrix Utilities*. (2020).
87. Hothorn, T., Hornik, K., van de Wiel, M. A. & Zeileis, A. A Lego System for Conditional Inference. *Am. Stat.* **60**, 257–263 (2006).
88. Brooks, M., E. *et al.* glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R J.* **9**, 378 (2017).
89. Bartoń, K. *MuMIn: Multi-Model Inference*. (2020).