eAppendix 1

Measurements of fecal and plasma levels of short-chain fatty acids

We collected 10 mL of venous blood and fresh feces from each participant. Blood was sampled in the morning after participants had fasted for at least 10 h. All samples were collected in tubes containing ethylenediaminetetraacetic acid (1 mg/mL) and glutathione (1 mg/mL) and immediately centrifuged (3000 rpm for 10 min). Plasma samples were stored at -80°C before analysis. Within two weeks of blood collection, participants provided two fresh fecal samples, which were mostly self-collected by participants at home. One would be transferred by the participant into a plastic tube and the other one would be collected using a stool specimen collection tube containing DNA stabilizer (Sarstedt, Fisher Scientific, USA). Total fecal DNA was extracted using a QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany). Fecal SCFA analysis was performed using the fresh fecal sample collected in a plastic tube within 24 hours after collection.

Fecal short-chain fatty acid (SCFA) analysis was performed with gas chromatography-mass spectrometry (GC-MS) as previously described.[1] One gram of crude feces was subjected to lyophilization and reweighed after lyophilization. The sample weight after lyophilization was used to calculate the concentrations of fecal SCFAs, expressed in μmol/g. The lyophilized fecal sample was suspended in 5 mL of

a 0.5% phosphoric acid solution. After sonication for 5 min and centrifugation at 3000 rcf for 10 min, 60 μL of the supernatant was diluted with 240 μL of a 0.5% phosphoric acid aqueous solution. Extraction of SCFAs with 300 μL of butanol was subsequently performed using the Geno/Grinder 2010 (SPEX, Metuchen, NJ, USA). Isotopically labeled internal standards were then added before GC-MS analysis, which was carried out on an Agilent 7890A gas chromatograph equipped with a MultiPurpose Sampler MPS (GERSTEL, Mülheim an der Ruhr, Germany) coupled to a Pegasus GC-TOFMS system (Leco Corporation, St. Joseph, MI, USA). A polar VF-WAXms capillary column (30 m × 0.25 mm i.d. × 0.25 μm film thickness) (Agilent Technologies, Santa Clara, CA, USA) was used for the separation. The helium carrier gas flow rate was set at 1 mL/min. One microliter of the sample was injected into the split mode at a ratio of 1:10. The oven temperature was initially held at 70°C for 1 min and then increased to 170°C at a rate of 10°C/min, to 240°C at a rate of 25°C/min, and finally maintained at 240°C for 2 min (total run time, 15.8 min). The electron impact ionization was 70 eV, and the data were acquired in full scan mode with a mass range of m/z 40–550.

Plasma SCFA levels were measured using liquid chromatography tandem mass spectrometry (LC-MS/MS) as previously described.[2] The volume of plasma sample was used to calculate the concentrations of plasma SCFAs, expressed in μM. Sixty

microliters of plasma were extracted with 420 μL of methanol using a Geno/Grinder 2010 (SPEX SamplePrep, Metuchen, NJ, USA). Derivatization of the supernatant was performed with 30 μL of 0.2 M 3-nitrophenylhydrazine and 30 μL of 0.12 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (40°C, 20 min). Isotopically labeled internal standards were then added before LC-MS/MS analysis. Plasma SCFAs and the corresponding internal standards were analyzed using an Agilent 6470 triple quadrupole mass spectrometer (Agilent Technologies, Santa Clara, CA, USA). The separation was performed on a ZORBAX Eclipse Plus C18 column (2.1 mm × 100 mm, 1.8 μm) with a mobile phase consisting of 0.1% formic acid in water (solvent A) and 0.1% formic acid in 2-propanol:ACN (3:1, v:v) (solvent B) at a flow rate of 0.35 mL/min. The LC gradient profile was as follows: 0–0.5 min, 0% B; 0.5–1.7 min, 0%–15% B; 1.7–3.5 min, 15% B; 3.5–9.0 min, 15%–30% B; 9.0–9.1 min, 30%–40% B; 9.1–16.0 min, 40%–72% B; 16.0–16.1 min, 72%–80% B; 16.1–16.2 min, 80%–100% B; 16.2–18.2 min, 100% B; and 18.2–20.0 min, 0% B. The injection volume was 5 μL. A negative electrospray ionization mode was used with the following parameters: 325°C for drying gas temperature, 7 L/min for drying gas flow, 45 psi for nebulizer pressure, 325°C for sheath gas temperature, 11 L/min for sheath gas flow rate, and 2500 V for capillary voltage. Nozzle voltage was set at 1500 V. Concentrations of each analyte in samples were determined from calibration curves using the peak area ratio of the analyte to its

corresponding isotope internal standard. All measurements were performed twice in each sample and the reported results were the mean of the two measurements.

Gut microbiota composition by shotgun metagenomic sequencing

For shotgun metagenomic sequencing, the purity of DNA was examined using an Epoch Microplate Spectrophotometer (BioTek, USA), the quantity of DNA was measured with Quant-iT PicoGreen dsDNA Assay (ThermoFisher Scientific, USA), and Illumina DNA Prep kits (Illumina, USA) were used for constructing libraries. The prepared libraries with target insert sizes of approximately 350 bp were sequenced on an Illumina NovaSeq 6000 sequencer using S4 flow cells. The metagenomic Illumina paired-end reads for both case and control samples are trimmed using Trimmomatic v0.39 with parameters "PE -phred33 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:12 TRAILING:12 SLIDINGWINDOW:4:15 MINLEN:36".[3] The trimmed paired-end reads of the samples are then searched for their prokaryotic species profiles using Kraken2 v2.0.6-beta 49.[4] To maximize the search efficiency, the maxikraken2 database (total size 140 GB; constructed by Daniel Fischer at Natural Resources Institute Finland and available at https://lomanlab.github.io/mockcommunity/mc_databases.html was used instead of the default kraken2 database. Measurements of α diversity in terms of Chao1 index,

Shannon index, and Simpson index, were calculated at species level using the "*vegan*" package[5] (https://cran.r-project.org/web/packages/vegan/) in the R environment. Principal coordinates analysis was conducted using Bray–Curtis distances, and analysis of similarities (ANOSIM) using the method implemented in R package "*vegan*" was used to test the difference of microbial communities between groups. The correlations between the relative abundance of gut microbiota and fecal or plasma levels of SCFAs were performed using Spearman's rank sum tests with Benjamini and Hochberg adjustment for multiple tests. $P < 0.05$ was considered significant. We performed all analyses with R software 3.4.1.[6] We then performed the functional profiling of the microbial community for the same non-human reads using HUMAnN2 (v0.11.1) integrated with the DIAMOND alignment tool (version 0.8.22), uniref90 protein database (version 0.1.1), and the ChocoPhlAn pangenome database (version 0.1.1).[7]

References for supplementary methods

1. Hsu YL, Chen CC, Lin YT, et al. Evaluation and Optimization of Sample Handling Methods for Quantification of Short-Chain Fatty Acids in Human Fecal Samples by GC-MS. Journal of proteome research 2019;18(5):1948-1957.

2. Liao HY, Wang CY, Lee CH, Kao HL, Wu WK, Kuo CH. Development of an

Efficient and Sensitive Chemical Derivatization-Based LC-MS/MS Method for Quantifying Gut Microbiota-Derived Metabolites in Human Plasma and Its Application in Studying Cardiovascular Disease. Journal of proteome research 2021.

3. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30(15):2114-2120.

4. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol 2019;20(1):257.

5. Jari Oksanen FGB, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner. Community Ecology Package. 2020.

6. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2020.

7. Franzosa EA, McIver LJ, Rahnavard G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods.* 2018;15(11):962-968; doi: 10.1038/s41592-018-0176-y.

Supplemental Table 1. Dietary intakes of patients with PD and normal controls based on food frequency Questionnaires in the current study.

Food and nutrients	Healthy controls (n=85) Median (IQR)	PD patients (n=96) Median (IQR)	*p* value
Meat, servings	0.8 (0.3, 1.7)	0.9 (0.4, 1.9)	0.82
Fish, servings	0.5 (0.2, 1.2)	0.4 (0.2, 1.1)	0.73
Eggs, servings	0.4 (0.2, 0.7)	0.4 (0.1, 0.8)	0.62
Soy, servings	1.1 (0.6, 1.8)	1.2 (0.5, 2.0)	0.53
Vegetables, servings	5.4 (2.7, 8.3)	5.3 (3.2, 7.9)	0.49
Fruits, servings	1.3 (0.7, 2.2)	1.2 (0.6, 2.1)	0.73
Nuts, servings	0.3 (0.1, 1)	0.2 (0, 0.6)	0.73
Energy, kcal	2000	2200	0.08
Protein (% energy)	11 (10, 13)	12 (11, 13)	0.10
Animal protein, g	20 (12, 35)	21 (13, 839)	0.28
Plant protein, g	51 (42, 60)	54 (48, 61)	0.32
Fat (% energy)	25 (20, 33)	25 (19, 30)	0.88
Saturate fat, g	10 (7, 15)	11 (6, 16)	0.79
Monounsaturated fat, g	14 (11, 24)	15 (10, 21)	0.13
Polyunsaturatd fat, g	14 (10, 20)	13 (8.5, 19)	0.73
Carbohydrate (% energy)	63 (57, 69)	64 (58, 70)	0.42
Dietary fiber, g	28 (22, 34)	27 (21, 33)	0.51

Abbreviations: IQR = interquartile range;

A serving of meat, fish, eggs, soy is defined as 7g protein; a serving of dairy is defined as 8g protein; a serving of vegetable is defined as 100g; a serving of fruit is defined as 15g carbohydrates; a serving of nuts is defined as 45 kcal.

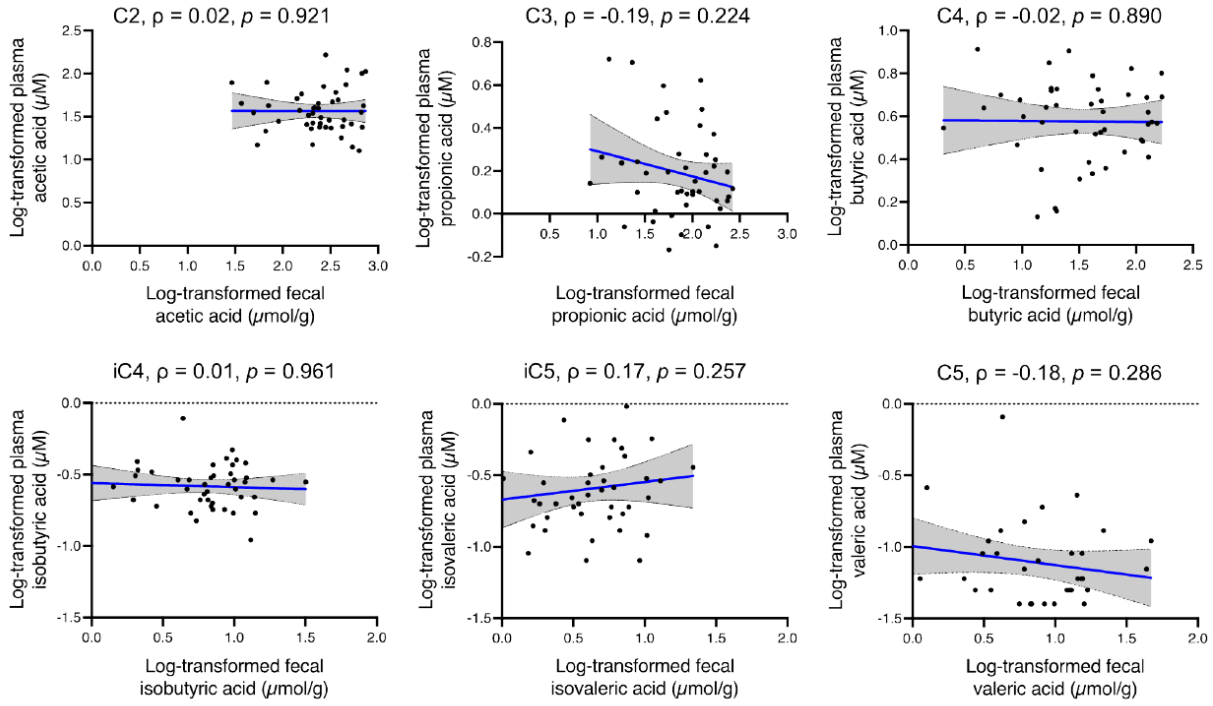
Supplementary Table 2. The associations between fecal and plasma levels of individual types of SCFAs and the risk of PD.

	Coefficient	OR	95% CI	*p* value
Fecal level of SCFAs				
Acetic acid (C2)	-1.80	0.16	0.04 – 0.57	0.006**
Propionic acid (C3)	-1.59	0.20	0.06 – 0.60	0.006**
Butyric acid (C4)	-1.30	0.27	0.10 – 0.67	0.007**
Isobutyric acid (iC4)	0.37	1.45	0.49 – 4.28	0.499
Valeric acid (C5)	-0.24	0.79	0.39 – 1.55	0.493
Isovaleric acid (iC5)	0.55	1.73	0.55 – 5.60	0.351
Plasma level of SCFAs				
Acetic acid (C2)	-0.68	0.51	0.11 – 2.31	0.379
Propionic acid (C3)	2.67	14.4	1.63 – 159	0.021*
Butyric acid (C4)	2.88	17.8	2.50 – 152	0.006**
Isobutyric acid (iC4)	2.80	16.4	0.97 – 368	0.062
Valeric acid (C5)	2.16	8.70	1.49 – 67.3	0.025*
Isovaleric acid (iC5)	0.49	1.63	0.37 – 7.63	0.522

Multivariate logistic regression models were applied to examine the associations between SCFAs and the risk of PD development. In the models, presence of PD was set as dichotomous endpoint. The covariates included age, sex, and log transformed levels of SCFAs.CI, confidence interval; OR, Odds ratio; PD, Parkinson's disease; SCFA, short chain fatty acid. * $p < 0.05$; ** $p < 0.01$.

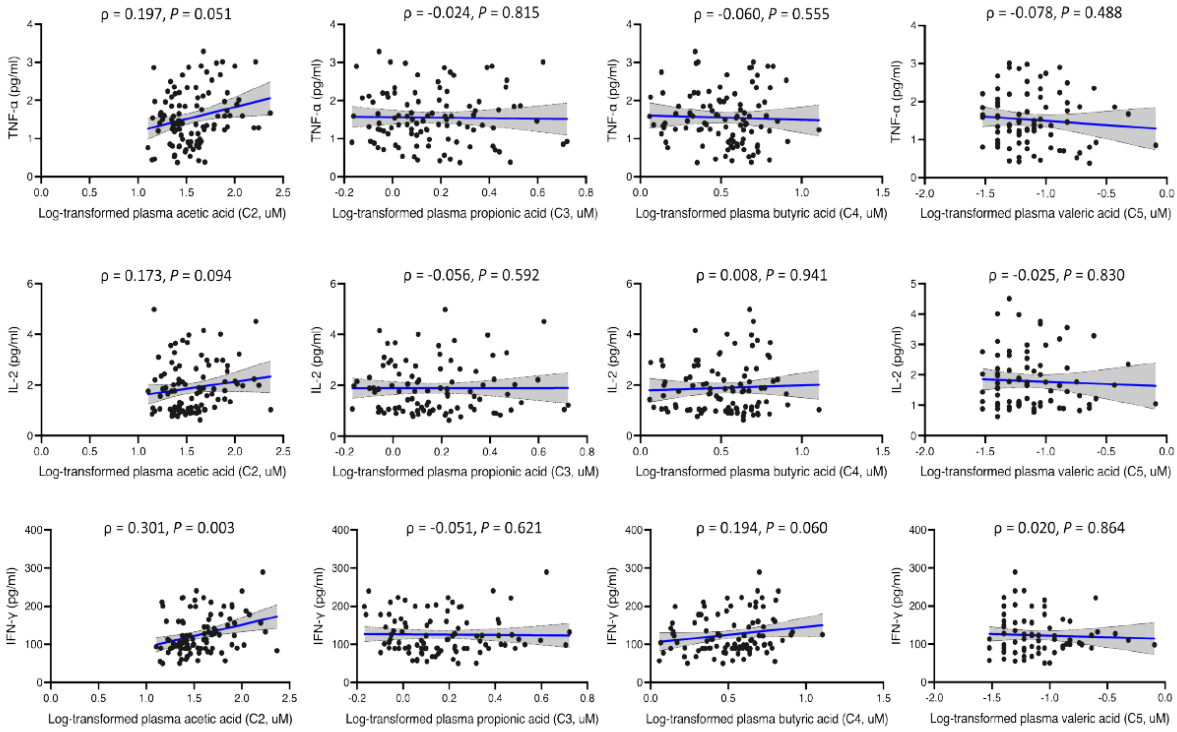
eFigure 1. Correlations between fecal and plasma levels of different types of SCFAs in all participants.

Supplementary Figure 1



eFigure 2. Correlations between plasma levels of different types of SCFAs and systemic inflammatory markers in patients with PD.

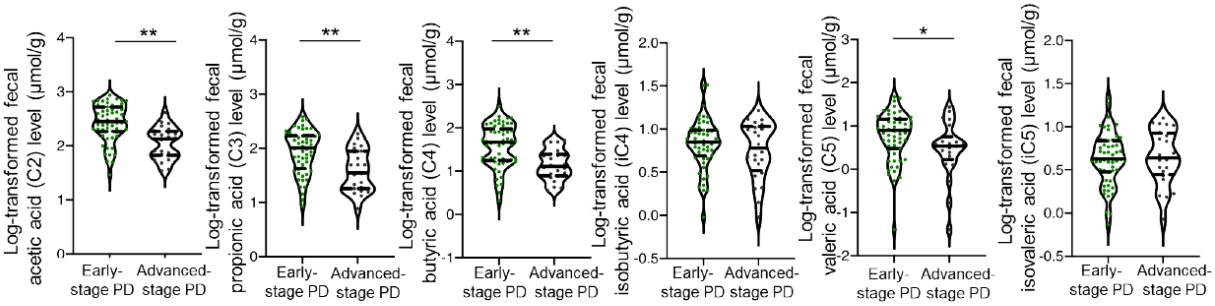
Supplementary Figure 2



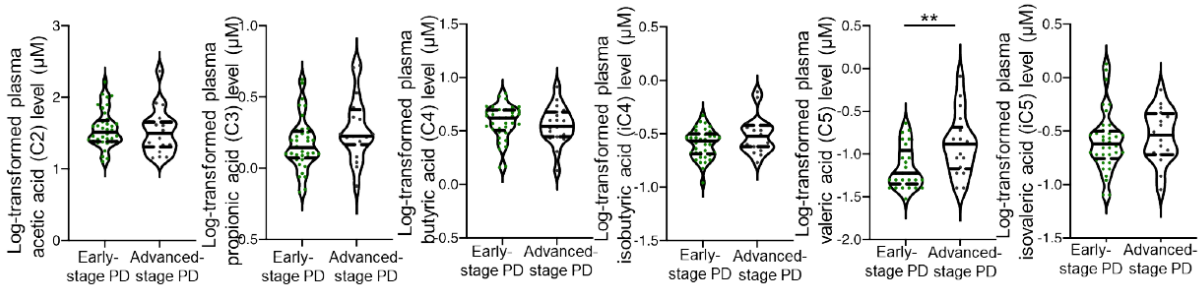
eFigure 3. Comparison of fecal and plasma concentrations of SCFAs in PD patients with different motor severity. (A) Fecal levels of individual types of SCFAs in patients with early-stage or advanced-stage PD. (B) Plasma levels of individual types of SCFAs in patients with early-stage or advanced-stage PD. All data points are shown in the violin plot. The median with interquartile range (solid and dashed horizontal lines) are shown. * $p < 0.05$, ** $p < 0.01$. (C) Scatter plots show a trend to a positive correlation between MDS-UPDRS part III motor scores and plasma levels of propionic acid. PD, Parkinson's disease; MDS-UPDRS, Movement Disorder Society-Unified Parkinson's Disease Rating Scale.

Supplementary Figure 3

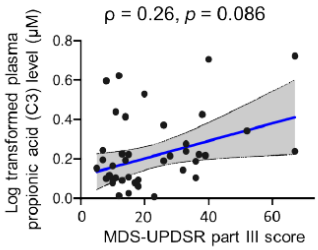
A



B

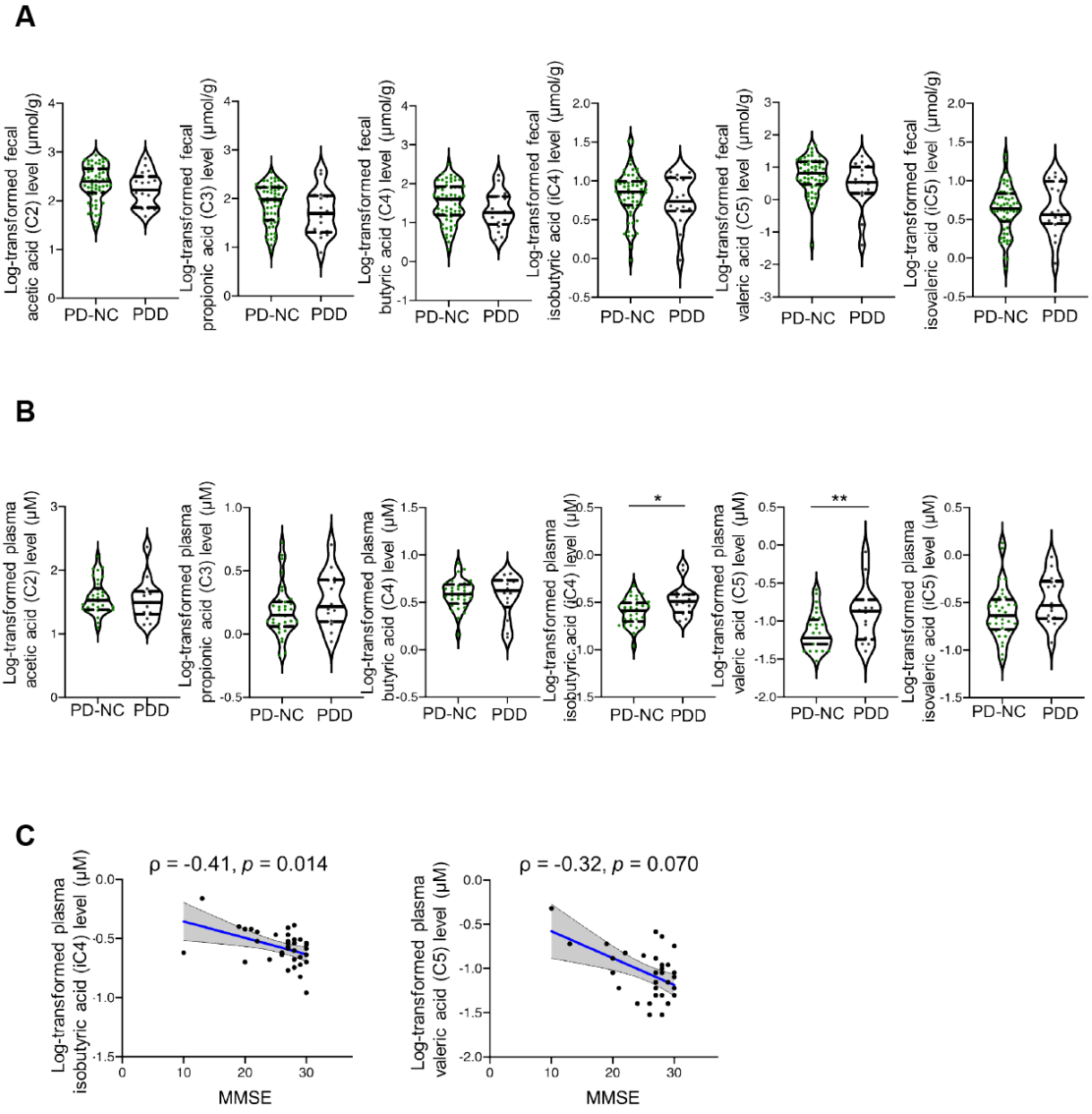


C



eFigure 4. Comparison of fecal and plasma concentrations of SCFAs in PD patients with different cognition severity, and correlations between SCFAs levels and MMSE scores. (A) Fecal levels of individual types of SCFAs in PD patients with normal cognition or PDD. (B) Plasma levels of individual types of SCFAs in PD patients with normal cognition or PDD. All data points are shown in the violin plot. The median with interquartile range (solid and dashed horizontal lines) are shown. * $p < 0.05$, ** $p < 0.01$. (C) Scatter plots show negative correlations between MMSE scores and plasma levels of isobutyric acid, and a trend to a negative correlation with plasma valeric acid levels. PD, Parkinson's disease; PD-NC, Parkinson's disease with normal cognition; PDD, Parkinson's disease dementia; MMSE, Mini-Mental State Examination.

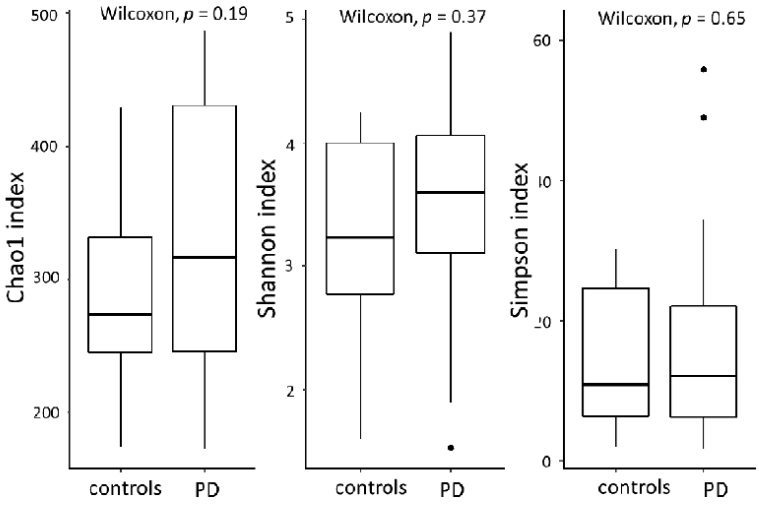
Supplementary Figure 4



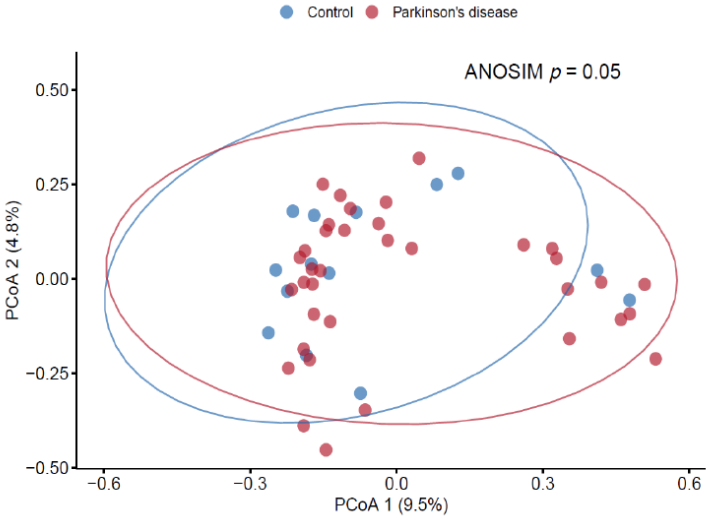
eFigure 5. The α-diversity and β-diversity indices of the fecal microbiome in the PD and control groups. (A) Boxplots depict differences in fecal microbiome diversity indices between the PD and control groups according to the Chao 1 index, Shannon index, and Simpson index based on OTU counts. Each boxplot shows the median, interquartile range, minimum, and maximum values. (B) Principal coordinates analysis plots of bacterial β-diversity based on the Bray–Curtis dissimilarity, analyzed according to health status. Patients with PD and age-matched controls are colored in red and blue, respectively. OTU, operational taxonomic units; PD, Parkinson's disease.

Supplementary Figure 5

A

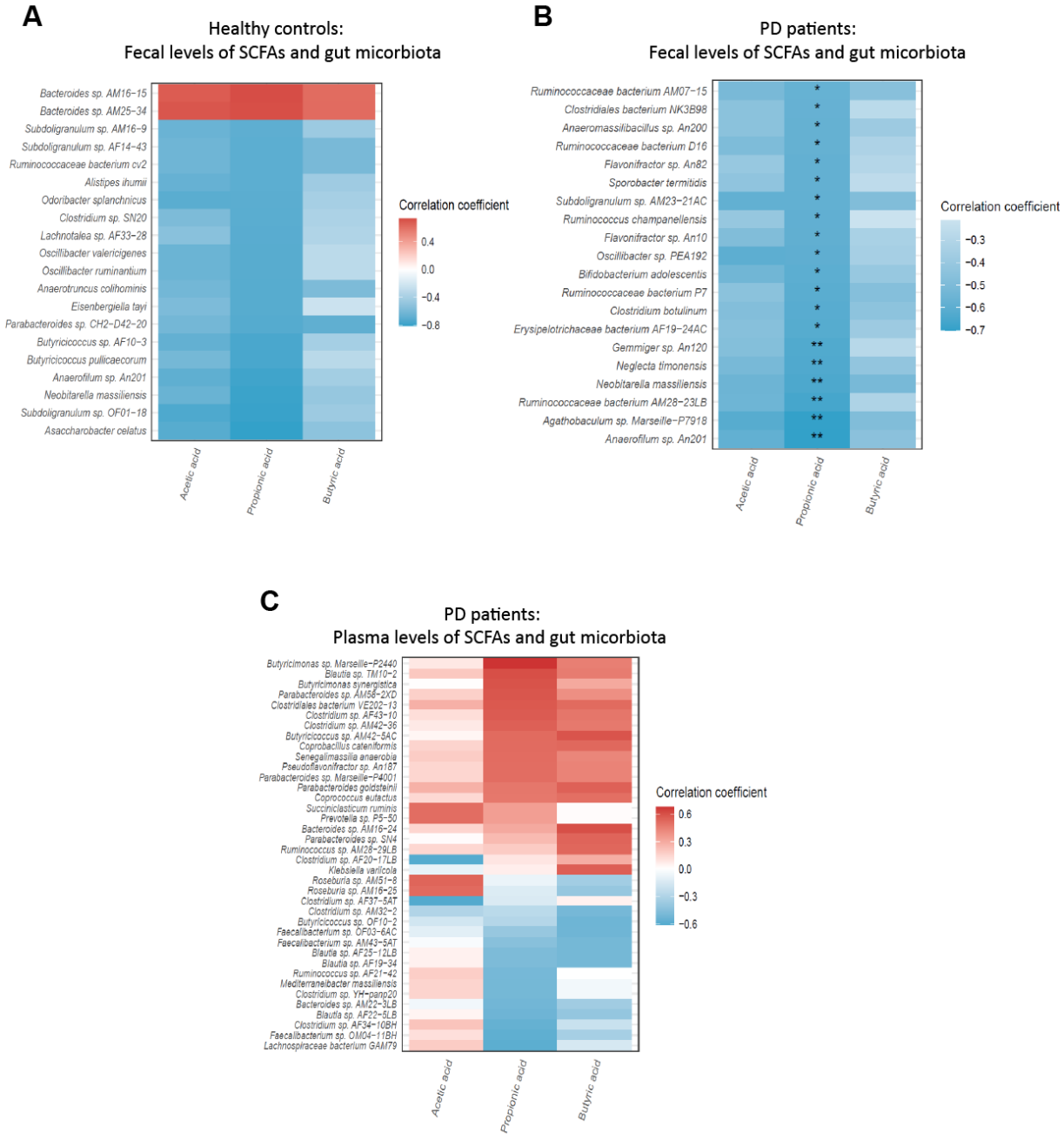


B



eFigure 6. Heat maps representing the Spearman correlation of the relative abundance of different bacteria and the concentrations of SCFAs in feces or plasma. Significant correlations between fecal levels of different types of SCFAs and the most abundant gut microbiota in (A) unaffected controls and (B) PD patients. (C) Correlations between plasma levels of different types of SCFAs in patients with PD. The *r* values are represented by gradient colors, where red and blue cells indicate positive and negative correlations, respectively. Significant differences were indicated by * $p < 0.05$ and ** $p < 0.01$, respectively. PD, Parkinson's disease; SCFAs, short-chain fatty acids.

Supplementary Figure 6



eFigure 7. Correlations between plasma levels of SCFAs and the functional profiling of the microbial community in patients with PD. Significant correlations between plasma levels of different types of SCFAs and the functional profiling pathways of the microbial community in PD patients. The *r* values are represented by gradient colors, where red and blue cells indicate positive and negative correlations, respectively. Significant differences were indicated by * $p < 0.05$ and ** $p < 0.01$.

Supplementary Figure 7