**Pathway-centric remodeling of the gut microbiome in constipated Parkinson’s disease: PD-only metagenomic analysis**

Siddiqui, Ayan M.

College of Medicine, Burnett Honors College, University of Central Florida, Orlando, Florida, USA

Correspondence: [ay153866@ucf.edu](mailto:ay153866@ucf.edu)

**Abstract**

Constipation is a prevalent non-motor feature of Parkinson’s disease (PD), yet microbiome signatures within PD remain unclear. We analyzed a PD-only shotgun metagenomic cohort (*n* = 490; non-constipated = 282, constipated = 208) using a covariate-adjusted, compositional framework to test associations between constipation and microbial diversity, composition, function, and enterotypes. Shannon and Gini-Simpson (1−D) α-diversity did not differ by group, whereas observed richness was modestly higher in constipated PD (median 94 vs. 89; *P =* 0.002). β-diversity showed small but significant separation (PERMANOVA *R*² = 0.008 [Bray–Curtis], 0.005 [Aitchison]; *q =* 0.001), with dispersion differences confined to Bray–Curtis. At the species level (CLR + limma), one taxon met Benjamini-Hochberg (BH) FDR < 0.05, *Ruthenibacterium lactatiformans*, higher in cases, while several taxa displayed non-significant trends. Functional analysis (voom–limma) revealed broader changes: of 510 pathways tested, 64 were BH FDR-significant overall (52 enriched and 12 depleted in cases), led by lipid/cofactor and central-carbon modules including phosphopantothenate/CoA biosynthesis, fatty-acid initiation, cis-vaccenate biosynthesis, and chorismate-related pathways. Data-driven enterotypes derived by mclust/BIC produced *k* = 2 (seed = 42) with a modest distribution shift (Fisher’s exact *P* = 0.0105). In sum, constipation within PD is marked by slight compositional differences and prominent remodeling of microbial functions. A pathway-centric lens may better capture symptom-linked biology than species lists and suggests testable metabolic targets for biomarker development and dietary or probiotic interventions.

**Introduction**

Parkinson’s disease (PD) is the second most common neurodegenerative disorder worldwide, affecting over 10 million people1. PD, as a complex neurodegenerative disorder, is defined by the gradual loss of dopaminergic neurons in the substantia nigra, and its symptoms often manifest with typical motor features: bradykinesia, resting tremor, rigidity, and postural instability2. Beyond the motor symptoms, PD also includes a broad spectrum of non-motor symptoms, which often include sleep disturbance, depression, anosmia, and gastrointestinal (GI) dysfunction, many of which arise years before clinical diagnosis3.

Constipation is among the earliest and most prevalent non-motor symptoms of PD, affecting up to an estimated 80% of patients and often preceding motor onset by decades4,5. These observations generally align with the Braak hypothesis, which proposes the idea that PD pathology may originate in the GI tract and ascend to the central nervous system (CNS) via the vagus nerve through misfolded α-synuclein (α-syn) that aggregates in the enteric nervous system (ENS) and then propagates to the brain6. Studies of α-syn inclusions in intestinal tissue before motor symptoms appear further support a gut-brain axis contribution to PD pathogenesis7.

The human gut microbiome, which includes trillions of microorganisms inhabiting the GI tract, can modulate neuroinflammation, epithelial barrier integrity, and α-syn biology8. Across numerous studies, PD is associated with some level of microbial dysbiosis, often featuring shifts in short-chain fatty acid (SCFA) producers and increases in taxa linked to pro-inflammatory signaling9. SCFAs (e.g., butyrate, propionate, acetate), products from microbial fermentation of dietary fiber, support gut balance and immune signaling across the gut-brain axis10. Reduced SCFA activity has been connected to increased intestinal permeability, sometimes known as a “leaky gut,” and systemic inflammation, processes relevant to PD11.

Despite the rapid progress made concerning microbiome studies, including the importance of Wallen et al.12, most studies of this type contrast PD cases with neurologically healthy controls. Such designs are informative for disease-control differences but are less suited to interrogate intra-disease heterogeneity, especially with early and clinically impactful non-motor symptoms like constipation. Because the microbiome reflects and potentially contributes to host physiology, a PD-only, symptom-anchored comparison may reveal associations diluted or obscured in traditional case-control analyses.

Here, we focus on PD cases to test whether constipation status is associated with shifts in gut microbial diversity, community structure, taxonomic composition, and microbial functional pathways. Using a publicly available, high-quality shotgun metagenomic dataset from BioProject PRJNA834801 processed by prior work12, we quantify: (i) α-diversity (Shannon, Gini–Simpson, observed richness) and (ii) β-diversity (Bray–Curtis and Aitchison/CLR); (iii) differentially abundant taxa; (iv) differentially abundant metabolic pathways with attention to SCFA-related functions; and (v) the distribution of data-driven enterotypes within PD by constipation status. By focusing on constipated vs. non-constipated PD, we aim to delineate symptom-associated microbial signatures that may refine mechanistic hypotheses along the gut-brain axis and inform future biomarker or intervention studies.

**Methods**

**Data sources and cohort**

Shotgun metagenomic data were obtained from the NCBI Sequence Read Archive (SRA) BioProject PRJNA834801. Wallen et al. curated the raw reads and released processed derivative tables12. We exclusively used those processed outputs: (i) species-resolved relative-abundance tables (MetaPhlAn-style) for taxonomic analyses and (ii) HUMAnN pathway count tables for functional analyses (not RPK/relative). Our analysis focused only on people with PD; neurologically healthy controls were not analyzed. After sample-ID alignment across metadata and abundance tables, the PD-only analysis set comprised *n* = 490 individuals (control = non-constipated PD, *n* = 282; case = constipated PD, *n* = 208).

**Case definition, covariates, and ID alignment**

Constipation status was extracted from the released metadata. We adjusted the variable to a binary Group factor with levels control = non-constipated PD and case = constipated PD; standard affirmative strings (e.g., “Y/Yes/TRUE/1”) mapped to the case level. Prespecified covariates were Age at collection, Sex, BMI, Laxatives, and collection\_method (stool kit/protocol).

To avoid mismatches across files, we normalized sample-IDs (trimming read suffixes, punctuation, and case) and restricted analyses to the intersection of IDs across metadata and abundance tables. A diagnostic report of matched/unmatched IDs is exported.

**α-diversity analysis**

α-diversity was computed on the relative-abundance (RA) table. For sample *i* let the species composition be with and .

Observed richness (count of taxa present):

Shannon index (natural log; with convention ):

Gini–Simpson index (1 - dominance; probability two draws differ in taxon):

Two-sided Wilcoxon rank-sum tests assessed between-group differences (case vs. control). Exact group medians and *P* values are reported.

**β-diversity, ordination, and PERMANOVA with dispersion checks**

Between-sample dissimilarity was assessed using Bray–Curtis on RA (bounded in [0,1]; sensitive to abundance differences):

where is the relative abundance of taxon *g* in sample *i*.

Aitchison distance on centered log-ratio (CLR) features (appropriate for compositions). To avoid ln(0), we add a small pseudocount to RA () and compute the CLR (all ordinations and permutations used seed = 42):

Aitchison distance is Euclidean in CLR space (both forms are shown):

Principal coordinates analysis (PCoA) was used for visualization. Group differences were tested by PERMANOVA (vegan::adonis2, 999 permutations, by = "terms") with the model:

To assess the PERMANOVA assumption of homogeneous multivariate dispersions, we used betadisper (vegan package) with permutation tests; dispersion *P* values are reported alongside PERMANOVA *R*2 and FDR-adjusted *q*.

**Mathematical conventions and transformations**

1. CLR on RA vs. counts: CLR is defined on compositions. We therefore work on RA and add a tiny pseudocount to handle zeros before CLR. Using counts with a pseudocount and then CLR is algebraically equivalent to an additive constant that cancels under CLR centering; operating on RA keeps the interpretation explicitly compositional.
2. Aitchison distance: the compact norm form and the expanded square-root-of-sum-of-squares form above are mathematically identical; both are provided for clarity.
3. Log bases: we use the natural log for Shannon and for CLR (standard in information theory and compositional data analysis), and base-2 logs only for voom outputs (log2-CPM; hence pathway effects are reported as log2 fold-change).

**Differential abundance of taxa (compositional modeling)**

Because RA data are compositional, we modeled CLR-transformed taxa with limma:

1. Transform: starting from the RA table , add and compute:
2. Prevalence filter: retain taxa with of samples and non-zero CLR variance.
3. Design alignment: build the design matrix *X* with Group + covariates (Age, Sex, BMI, Laxatives, collection\_method); use complete-case rows so .
4. Model and test: fit gene-wise linear models with empirical-Bayes moderation (limma), testing the Group coefficient (case vs. control):

where is the CLR abundance of taxon *g* in sample *i*, are covariates, and estimates the mean difference (case - control) on the natural-log CLR scale.

Effect sizes are reported as limma logFC (here, a difference on the ln-CLR scale; positive = higher in cases). Multiple testing used Benjamini-Hochberg (BH) FDR across taxa; exceedingly small *P*/*q* are printed as “<1e−16”. The present analysis used CLR + limma on relative abundance; no DESeq2 path is implemented in this pipeline.

**Differential abundance of microbial pathways**

HUMAnN pathway count tables were analyzed with voom–limma:

1. Align samples and design as above.
2. Library-size normalization (edgeR TMM), then voom to obtain precision-weighted log2-CPM.
3. Fit limma models and test the Group coefficient, reporting log2 fold-change (log2FC) and BH-adjusted *q*.

For interpretability, we also report group-wise mean normalized abundances (CPM). Positive log2FC indicates higher in cases (constipated PD).

**Enterotype inference and group comparison**

We summarized community structure into enterotypes using model-based clustering (mclust) on the CLR matrix reduced by PCA to the smallest number of PCs explaining ≥ 8 of variance (capped at 10 PCs). The Bayesian Information Criterion (BIC) selected the number of clusters *k*, all clustering used seed = 42. If model-based clustering failed to converge, we used *k*-means (centers ) with model selection by mean silhouette; the chosen approach and selection metrics were exported.

Differences in enterotype distributions between groups were tested using Fisher’s exact test on the enterotype X Group contingency table.

**Multiple testing, reporting, and diagnostics**

All families of feature-level tests (taxa, pathways) controlled BH FDR; we report both *P* and , with treated as statistically significant. The number of features tested after filtering is stated in each results table. The pipeline exports diagnostic reports for sample-ID alignment and design-matrix complete-case drops.

**Use of a large language model (LLM)**

An LLM, ChatGPT-4.5 (OpenAI), was used solely during the analytical phase to assist with R troubleshooting (identifying and correcting coding errors, debugging scripts, clarifying error messages, and annotating sections of the pipeline script for ease of reading). The LLM was not used for study design, hypothesis generation, interpretation of results, statistical decision-making, or substantive manuscript drafting. No AI-generated images were used in this study; all visualizations were derived from code and source data. No participant-level or sensitive data were shared with the LLM; all the interactions were limited to de-identified code snippets and error text. All LLM suggestions were independently reviewed and verified by the authors and incorporated only when reproduced in version-controlled scripts.

**Software and reproducibility**

Analyses were performed in R version 4.4.2 using vegan (ordination, PERMANOVA, betadisper), phyloseq (data handling), limma/edgeR (voom–limma), mclust, and cluster (enterotypes), and tidyverse/patchwork (data wrangling/visualization). The exact script (versioned) produces all figures (PNG + PDF), main tables (CSV + HTML), and supplementary files; repository links and data locations are provided in the Data availability and Code availability sections.

**Results**

**Cohort used in analyses (PD-only set)**

After sample-ID alignment across metadata and abundance tables, we analyzed *n* = 490 individuals with PD: control (non-constipated PD), *n* = 282; case (constipated PD), *n* = 208 (Table 1). Alignment/complete-case diagnostics are provided in the Supplementary information section (Supplementary Table S6; ID and design-drop reports).

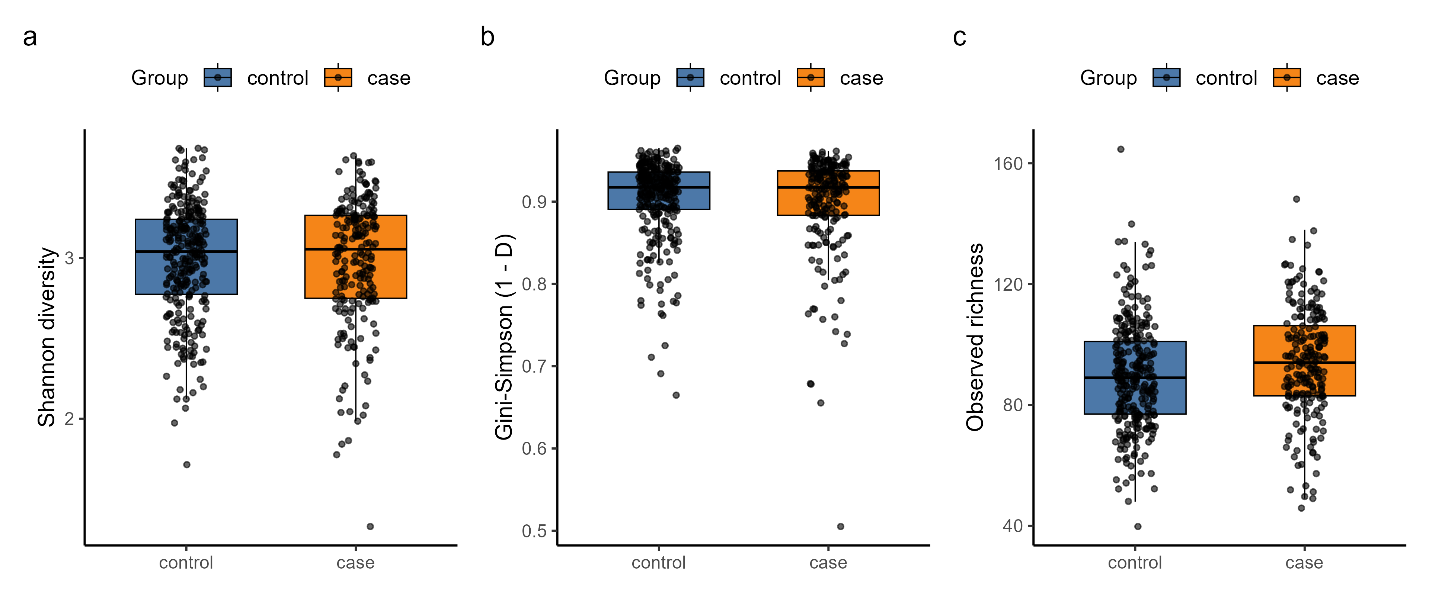
A screenshot of a computer

AI-generated content may be incorrect.

**Table 1 | PD-only cohort counts used in analyses.** Control = non-constipated PD; case = constipated PD. Counts reflect the intersection of sample IDs across metadata and abundance tables.

**α-diversity is broadly similar; observed richness is modestly higher in constipated PD**

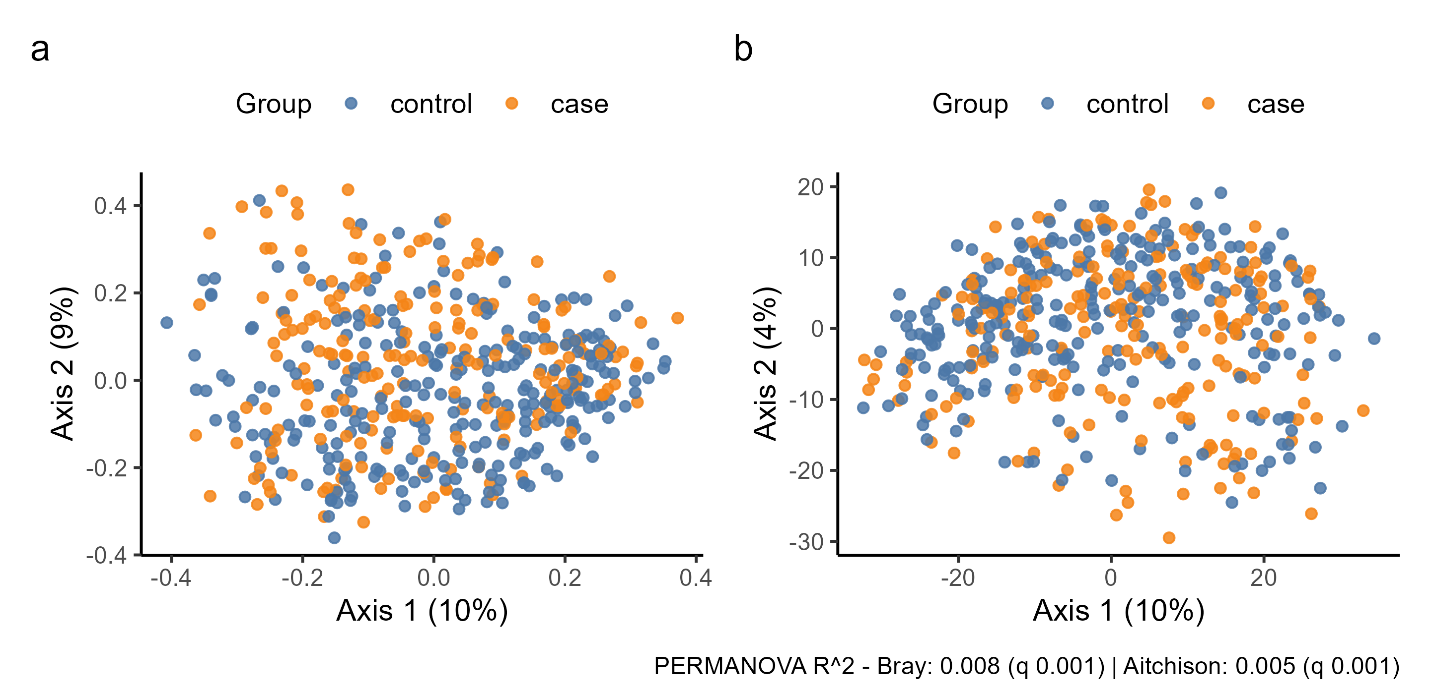
Across Shannon and Gini–Simpson (1−D) indices, groups were similar (Wilcoxon, two-sided: Shannon *P =* 0.802; Gini–Simpson *P =* 0.984). Observed richness (presence/absence on the relative-abundance table) was modestly higher in constipated PD (median 94 vs. 89; *P =* 0.002) (Supplementary Table S1). Representative distributions are shown in Fig. 1a–c.



**Figure 1 | α-diversity in PD (constipated vs. non-constipated).** (a) Shannon; (b) Gini–Simpson (1−D); (c) Observed richness. Points are individual samples; boxes show IQR and median. Wilcoxon two-sided tests; PD-only *n*: control = 282, case = 208. Complete statistics in Supplementary Table S1.

**β-diversity shows small but detectable group separation; dispersion differs for Bray–Curtis**

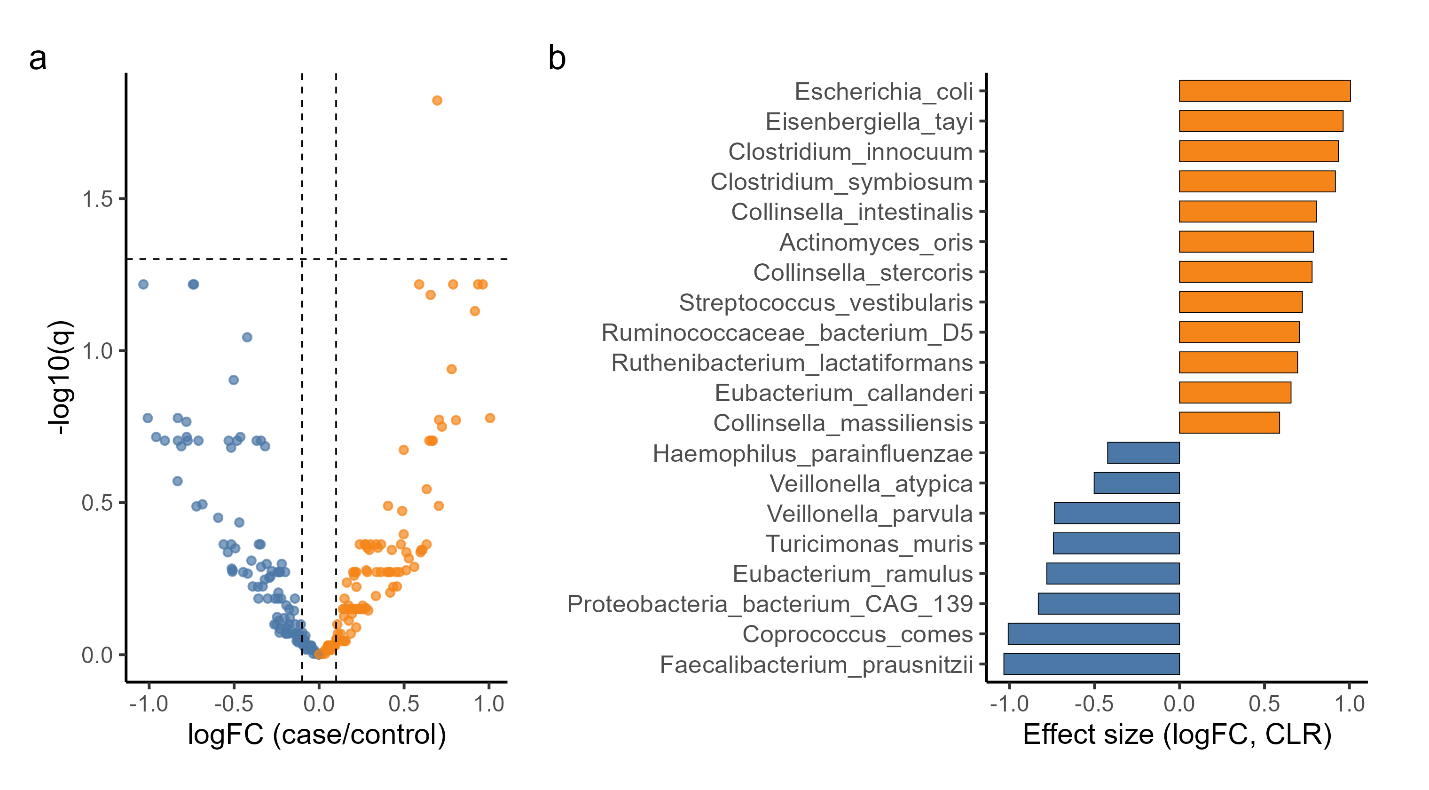
PCoA ordinations show broad overlap between groups for both Bray–Curtis and Aitchison (CLR) distances (Fig. 2a–b). In PERMANOVA models adjusting for Age, Sex, BMI, Laxatives, and collection\_method, Group explained a small but statistically significant proportion of variance (Bray–Curtis *R*² = 0.008, *q =* 0.001; Aitchison *R*² = 0.005, *q =* 0.001; 999 permutations; Supplementary Table S4). Homogeneity of dispersions (betadisper) differed for Bray–Curtis but not for Aitchison (permutation *P* values in Supplementary Table S4). Thus, group differences are minor in magnitude and partly confounded by dispersion for Bray–Curtis; the Aitchison result is less sensitive to this assumption.



**Figure 2 | PCoA of community structure in PD.** (a) Bray–Curtis; (b) Aitchison/CLR. PERMANOVA (by terms) with covariates as listed; 999 permutations. Caption annotations show R² and BH-adjusted q for Group. Dispersion permutation P values are provided in Supplementary Table S4.

**One taxon differs after FDR correction; several others show trends**

After prevalence filtering (RA > 1×10⁻⁴ in ≥5% of samples), 230 taxa were tested using CLR + limma with covariate adjustment. One taxon met FDR (*q*) < 0.05: *Ruthenibacterium lactatiformans* (logFC = 0.696, *q =* 0.015; higher in cases). Several taxa exhibited non-significant trends (e.g., *Faecalibacterium prausnitzii* and *Coprococcus comes* lower in cases; *Escherichia coli* higher), summarized in Fig. 3a–b and Table 2. The complete results (all 230 taxa) are in Supplementary Table S2.



**Figure 3 | Differential taxa in PD (case vs. control; CLR + limma).** (a) Volcano plot (dashed lines: |logFC| = 0.1, q = 0.05); orange = higher in cases, blue = higher in controls. (b) Bar plot of Top-20 taxa ranked by FDR (q) and |logFC| (direction colored by group). Taxa logFC is on the CLR (natural-log) scale; positive = higher in cases. Models adjust for Age, Sex, BMI, Laxatives, and collection\_method. n = 230 taxa tested; BH FDR applied. Full table in Supplementary Table S2.

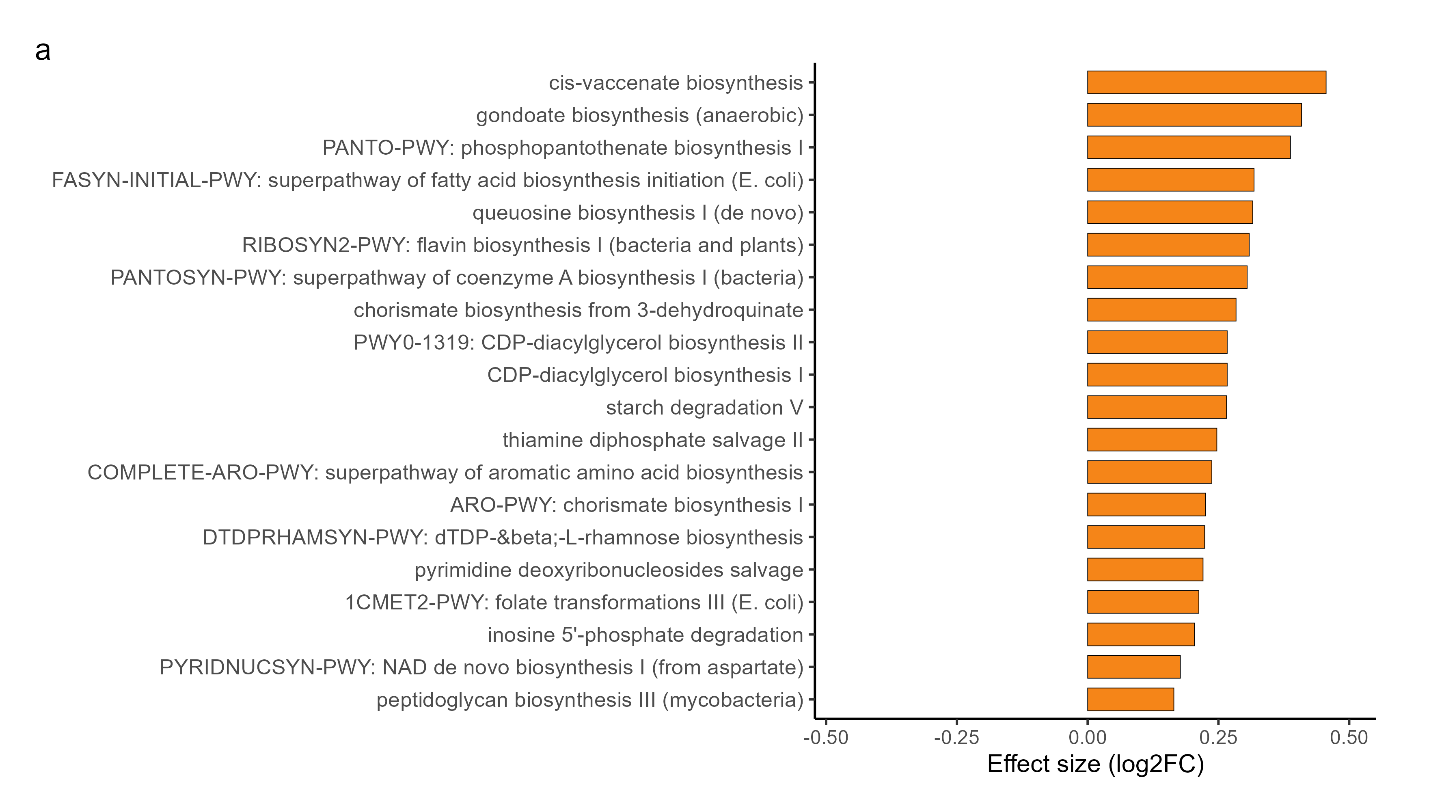
A screenshot of a computer

AI-generated content may be incorrect.

**Table 2 | Top 20 differentially abundant taxa (PD-only; case vs. control).** Columns include Taxon, logFC (case/control), FDR-adjusted *q*, Direction, Prevalence, and mean relative abundance by group. Footnotes document filtering and covariates. The full feature-level table is in Supplementary Table S2. Taxa effects are on the ln-CLR scale (limma); positive = higher in cases.

**Functional pathways show widespread differences, dominated by lipid/cofactor biosynthesis; some pathways are lower in cases**

Using voom–limma on HUMAnN pathway counts (covariate-adjusted), 510 pathways were tested; 64 met FDR (*q*) < 0.05 (52 enriched in cases; 12 enriched in controls). The most substantial positive effects included phosphopantothenate/CoA biosynthesis (PANTO-PWY, log2FC = 0.388, *q =* 3.21×10⁻³⁴), fatty-acid biosynthesis initiation (FASYN-INITIAL-PWY, 0.318, 5.16×10⁻³²), cis-vaccenate biosynthesis (PWY-5973, 0.456, 1.15×10⁻³⁰), and chorismate/aromatic-amino-acid precursors (Fig. 4; Table 3). Significant depletions in cases (negative log2FC) were also present in the full table, including branched-chain amino-acid biosynthesis and glycolysis variants. The Top-20 figure displays enriched pathways; all significant depletions are tabulated in Supplementary Table S3.



**Figure 4 | Top pathways differentiating constipated vs. non-constipated PD.** Voom–limma log2FC; orange = higher in cases. Models adjust for Age, Sex, BMI, Laxatives, and collection\_method. *n =* 510 pathways tested; BH FDR applied. Full results are in Supplementary Table S3.

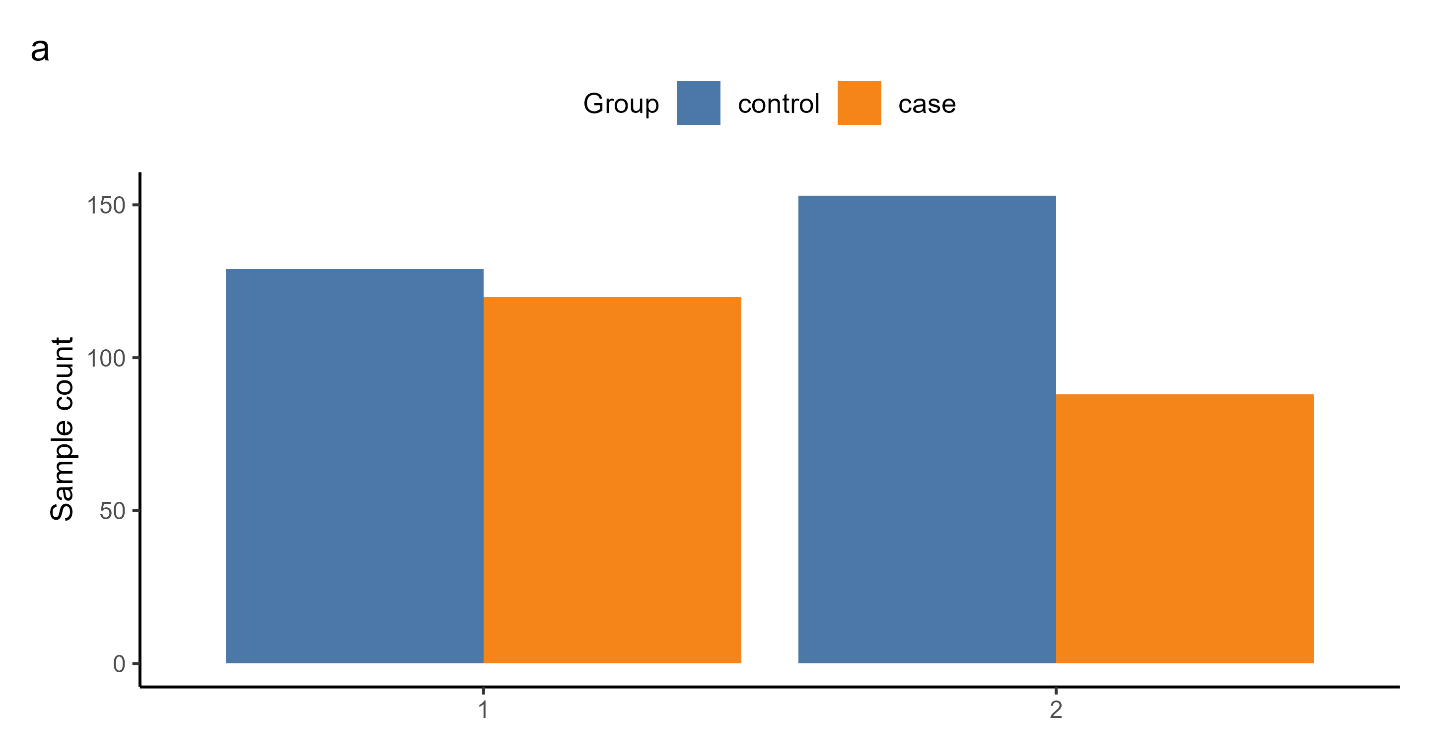
A screenshot of a computer

AI-generated content may be incorrect.

**Table 3 | Top 20 differentially abundant pathways (PD-only; case vs. control).** Columns include Pathway ID, Description, log2FC (case/control), FDR-adjusted *q*, Direction, and group-wise mean CPM. Pathways are derived from published HUMAnN outputs (not re-run). Full results in Supplementary Table S3. Pathway effects reported as log2FC (voom–limma); positive = higher in cases.

**Enterotypes (*k =* 2 by BIC) differ modestly by constipation status**

Data-driven enterotypes from mclust/BIC selected *k =* 2 after CLR→PCA reduction. Enterotype counts were control 129|153 vs. case 120|88 (clusters 1|2), indicating a modest but significant distribution shift (Fisher’s exact *P =* 0.0105; Fig. 5). Assignments and model-selection metrics are provided in Supplementary Table S5.



**Figure 5 | Enterotype counts by constipation status.** Model selection chose *k =* 2 by BIC on PCA-reduced CLR features; counts are shown for control (non-constipated PD) and case (constipated PD). Fisher’s exact test compares the enterotype × Group table (*P =* 0.0105). Complete assignments and selection metrics in Supplementary Table S5.

**Discussion**

We examined gut-microbiome features associated with constipation within PD using a PD-only cohort (*n =* 490) and a covariate-adjusted, compositional framework aligned to prespecified objectives. In brief, we observed (i) broadly similar Shannon and Gini–Simpson diversity with a modest increase in observed richness among constipated PD (Fig. 1), (ii) small but statistically significant between-group separation in community structure (PERMANOVA) with dispersion effects limited to Bray–Curtis (Fig. 2), (iii) one species meeting FDR < 0.05 (*Ruthenibacterium lactatiformans*, higher in cases) with several additional trends that did not survive FDR (Fig. 3; Table 2), (iv) widespread functional differences dominated by lipid/cofactor and central-carbon pathways (e.g., phosphopantothenate/CoA biosynthesis, fatty-acid initiation, cis-vaccenate and chorismate routes), with both enrichments and depletions in cases (Fig. 4; Table 3), and (v) a modest but significant shift in enterotype distribution with *k =* 2 by BIC (Fisher’s *P* = 0.0105; Fig. 5). Collectively, constipation within PD is associated with subtle compositional changes, marked pathway remodeling, and a small redistribution of community archetypes.

Across PD microbiome studies, α-diversity differences are typically small or inconsistent, and effect sizes for between-group separation are modest (patterns our symptom-anchored, PD-only analysis reproduces)13,14,15. Case-control cohorts frequently emphasize functional reprogramming (energy, lipid, xenobiotic metabolism) over stable species lists, again concordant with our pathway-centric signals (CoA/fatty-acid/central-carbon modules)14,16,17. Some studies link bowel symptoms (especially constipation) to lower *Prevotella* and altered fiber fermentation, whereas others report variable taxa shaped by diet, medication, and geography14,16,18. Our within-PD design, explicit compositional handling (CLR/Aitchison; voom–limma), covariate adjustment, and dispersion checks align with recommended practice for small-effect microbiome data15 and help reconcile why we see sparse species-level hits yet coherent functional shifts.

We observed no material differences in Shannon or Gini–Simpson diversity; observed richness was modestly higher in constipated PD (Fig. 1). Within PD, a plausible explanation is slower transit, enabling more low-abundance taxa to persist above detection thresholds without significant changes in evenness. Overall, bulk richness/evenness appears less informative for constipation status than composition and function.

Community structure differed slightly between groups (PERMANOVA *R*² ≈ 0.5–0.8%; Fig. 2), magnitudes typical for human microbiome cohorts and best interpreted cautiously15. Dispersion differed for Bray–Curtis but not for Aitchison/CLR, so the CLR-space signal is the more robust indicator of subtle compositional differences. This pattern, slight separation with a dispersion caveat, is consistent with multi-center PD cohorts where lifestyle and technical heterogeneity dilute global shifts while preserving symptom-linked structure14,18.

At the species level (CLR + limma; covariate-adjusted), only *Ruthenibacterium lactatiformans* reached FDR < 0.05 (higher in cases); several taxa often discussed in PD cohorts (e.g., *Faecalibacterium prausnitzii*, *Coprococcus comes*, lower in cases; *Escherichia coli* higher) remained trends (Fig. 3; Table 2). Species-level findings vary widely across PD studies due to geography, diet, medication exposure, and processing pipelines13,14,18. Our uniform covariate adjustment (Age, Sex, BMI, Laxatives, collection\_method) and PD-only contrast likely reduce confounding that can inflate unadjusted taxa signals, yielding a conservative species panel.

In contrast, pathway analysis revealed broad, internally consistent differences: 64 pathways met FDR < 0.05 overall, with most enriched in cases and a minority depleted (Fig. 4; Table 3; Supplementary Table S3). The leading signals: phosphopantothenate/CoA biosynthesis, fatty-acid initiation, cis-vaccenate biosynthesis, and chorismate/aromatic-amino-acid precursors, implicate remodeling of lipid and cofactor metabolism alongside central-carbon flow. These functional modules may reflect altered substrate availability and redox demands with constipation (e.g., slower transit, differential fiber/lipid utilization), and they echo prior PD reports in which function tracks clinical features more reliably than any fixed species list14,17. We did not directly quantify metabolites (e.g., SCFAs, bile acids, quinones), so these inferences should be tested with paired metabolomics.

Enterotypes derived by model-based clustering (mclust/BIC) resolved two archetypes whose distribution differed modestly by constipation status (Fisher’s exact *P* = 0.0105; Fig. 5). The effect size is small, but statistically credible (accords with the prespecified BIC-based approach in Methods). Even so, the functional signals (Fig. 4) carry more apparent biological coherence than these coarse archetypes, suggesting pathway-level readouts will be more useful for symptom stratification and intervention design.

This study has several strengths: (i) a PD-only design targeting symptom heterogeneity; (ii) consistent covariate adjustment across all models; (iii) compositional handling via CLR/Aitchison and voom–limma; (iv) explicit dispersion checks qualifying PERMANOVA; and (v) fully reproducible outputs with alignment and complete-case diagnostics. Limitations reflect the retrospective, cross-sectional dataset: constipation was not uniformly phenotyped with stool-form or transit measures; diet, disease duration, and dopaminergic medications were incompletely available; and causality cannot be inferred. Shotgun metagenomics infers function from gene content rather than measured metabolites; paired stool/serum metabolomics (SCFAs, bile acids, redox cofactors) would materially strengthen mechanistic claims. Finally, effect sizes are small (typical in human microbiome studies) and warrant external replication, sensitivity analyses with alternative compositional models15, and cross-cohort comparisons under harmonized processing14,18.

In conclusion, a PD-only framework reveals that constipation is associated with subtle but statistically credible shifts in community composition, pronounced remodeling of microbial lipid/cofactor and central-carbon pathways, and a modest redistribution of enterotypes. These data refine the microbiome-constipation link in PD from a species-list narrative toward a pathway-centric view and motivate next steps: longitudinal PD cohorts to test temporal ordering of pathway signatures, integrated metagenome-metabolome studies to validate biochemical outputs, and dietary or probiotic interventions that target reproducible functional modules rather than wholesale community types.

**Data availability**

The raw shotgun metagenomic reads used in this study are publicly available in the NCBI SRA under BioProject ID PRJNA834801 (<https://www.ncbi.nlm.nih.gov/bioproject/834801>). We did not process raw reads. Instead, we used the post-quality control (QC), post-profiling derivative tables released by Wallen et al.12, available at Zenodo (record 7246185; <https://zenodo.org/records/7246185>). From that release, we used the species-level relative-abundance tables (MetaPhlAn-style) and HUMAnN pathway count tables as the analysis inputs described in Methods.

The analysis-ready inputs (copies of the relevant Wallen et al. derivatives used here)12, the derived outputs (CSV tables for α/β-diversity, differential taxa and pathways, PERMANOVA, enterotype assignments; figure-ready PNG/PDF), and the versioned R script pipeline and configuration files required to reproduce the results are available without restriction at [GitHub repository link to be inserted]. The repository includes an organized list of files (results/README\_outputs.txt) and the exact software environment (results/tables/session\_info.txt). Data from third-party sources is redistributed under the terms set by the original providers. The additional information supporting the findings of this study is provided within the paper and in Supplementary information; further materials are available from the corresponding author upon request.

**Code Availability**

The upstream processing code used by Wallen et al. for sequence QC12, taxonomic profiling (MetaPhlAn), and functional profiling (HUMAnN) is publicly available at Zenodo, record 7246185 (<https://zenodo.org/records/7246185>). The code used in this study is available at [GitHub repository link to be inserted]. The repository contains the top-level script (constipation\_pd\_paper\_pipeline.R).

**Acknowledgements**

We gratefully acknowledge Wallen et al. for curating and openly releasing the post-QC taxonomic and functional derivative tables used in this study (Zenodo record 7246185)12, derived from raw reads deposited in the NCBI SRA BioProject PRJNA834801. We also thank the developers and maintainers of the open-source tools that made this work possible, including MetaPhlAn, HUMAnN, phyloseq, vegan, limma, edgeR, mclust, cluster, and the broader R/tidyverse ecosystem. For research funding transparency: this research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Author contributions**

A.S. conceived and designed the study; accessed and curated publicly available processed datasets; implemented the analysis pipeline; performed statistical analyses; generated figures and tables; interpreted the results; drafted the manuscript; revised and approved the final version.

**Competing Interests**

The author declares no competing interests.

**Supplementary information**

The supplementary tables S1–S6 and all other additional readable files are available at [GitHub repository link to be inserted]. A brief description for each supplementary item and a complete manifest (Supplement\_index.txt) are provided at the exact location. Source data for Figs. 1–5 and Tables 1–3 are included within these supplementary files.

**References**

1. Ou, Z. *et al.* Global Trends in the Incidence, Prevalence, and Years Lived With Disability of Parkinson’s Disease in 204 Countries/Territories From 1990 to 2019. *Front. Public Health* **9**, 776847 (2021).

2. Poewe, W. *et al.* Parkinson disease. *Nat. Rev. Dis. Primer* **3**, 17013 (2017).

3. Schapira, A. H. V., Chaudhuri, K. R. & Jenner, P. Non-motor features of Parkinson disease. *Nat. Rev. Neurosci.* **18**, 435–450 (2017).

4. Yu, Q.-J. *et al.* Parkinson disease with constipation: clinical features and relevant factors. *Sci. Rep.* **8**, 567 (2018).

5. Abbott, R. D. *et al.* Frequency of bowel movements and the future risk of Parkinson’s disease. *Neurology* **57**, 456–462 (2001).

6. Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson’s disease. *Neurobiol. Aging* **24**, 197–211 (2003).

7. Shannon, K. M. *et al.* Alpha‐synuclein in colonic submucosa in early untreated Parkinson’s disease. *Mov. Disord.* **27**, 709–715 (2012).

8. 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).

9. Romano, S. *et al.* Meta-analysis of the Parkinson’s disease gut microbiome suggests alterations linked to intestinal inflammation. *Npj Park. Dis.* **7**, 27 (2021).

10. Dalile, B., Van Oudenhove, L., Vervliet, B. & Verbeke, K. The role of short-chain fatty acids in microbiota–gut–brain communication. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 461–478 (2019).

11. Houser, M. C. & Tansey, M. G. The gut-brain axis: is intestinal inflammation a silent driver of Parkinson’s disease pathogenesis? *Npj Park. Dis.* **3**, 3 (2017).

12. Wallen, Z. D. *et al.* Metagenomics of Parkinson’s disease implicates the gut microbiome in multiple disease mechanisms. *Nat. Commun.* **13**, 6958 (2022).

13. Fu, S.-C. *et al.* Exploring the Causal Effect of Constipation on Parkinson’s Disease Through Mediation Analysis of Microbial Data. *Front. Cell. Infect. Microbiol.* **12**, 871710 (2022).

14. Kenna, J. E. *et al.* Changes in the Gut Microbiome and Predicted Functional Metabolic Effects in an Australian Parkinson’s Disease Cohort. *Front. Neurosci.* **15**, 756951 (2021).

15. Debelius, J. *et al.* Tiny microbes, enormous impacts: what matters in gut microbiome studies? *Genome Biol.* **17**, 217 (2016).

16. Mertsalmi, T. H. *et al.* More than constipation – bowel symptoms in Parkinson’s disease and their connection to gut microbiota. *Eur. J. Neurol.* **24**, 1375–1383 (2017).

17. Baldini, F. *et al.* Parkinson’s disease-associated alterations of the gut microbiome predict disease-relevant changes in metabolic functions. *BMC Biol.* **18**, 62 (2020).

18. Khedr, E. M. *et al.* Gut microbiota in Parkinson’s disease patients: hospital-based study. *Egypt. J. Neurol. Psychiatry Neurosurg.* **57**, 153 (2021).