nature portfolio

Peer Review File

Habitual Coffee Intake Shapes the Gut Microbiome and Modifies Host Physiology and Cognition

Corresponding Author: Professor John Cryan

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This is a comprehensive, timely and very interesting investigation into how coffee intake shapes the gut microbiome and modifies host physiology and cognition. However, I believe the study lacks robust measurements of some basic outcomes that are required to interpret the results.

One major point of critique on this paper is the lack of measurement of stool transit time. The investigators rightly mention that both caffeinated and decaffeinated coffee stimulates the contractility of ileal and colonic smooth muscle, and thus speeds up transit time, which in turn is believed to have a significant effect on the composition, diversity and function of the microbiome. Considering the fact that both caffeinated and decaffeinated coffee stimulate contractility, this effect is may not be due to caffeine per se. In any case, effects of coffee consumption (with or without caffeine) on the microbiome should be considered against a background of changing transit time, as this may be an essential element of the causal mechanistic pathway. In addition, effects of coffee consumption (with or without caffeine) on the microbiome should be considered against the background of the full habitual diet (not just its macronutrients), as the diet may have a larger effect on the microbiome than a short-term shift in the number of (de)caffeinated drinks. Both elements (stool transit time and background diet prior to stool sampling) were not properly considered in this research, and therefore it will be difficult to interpret the results, especially on the microbiome, properly. A final major point of critique is the lack of an appropriate power analysis, which probably requires a more robust research question and pre-identified validated outcomes. It is not clear whether the study would have had sufficient power to find significant effects on individual or group level, and many results are presented without indicating any level of significance.

The text contains a large number of inaccuracies and vague descriptions, and the figures are not always informative and easy to summarise/interpret, other than showing results on a high level within the paper, and with many more detailed figures in the supplement, which does not necessarily help to interpret the results. My comments are outlined in more detail below.

Abstract

The abstract indicates differences were found between groups, what they were linked to, and metabolites involved, but no indication as to the direction of effects is given which would make the abstract more informative.
The concluding sentence suggests the microbiome could be a biomarker for coffee consumption but this was not stated as the aim of the study. Using the microbiome for habitual coffee consumption seems very cost- and effort- ineffective considering that there are easier ways to monitor coffee consumption already. Instead, the authors may wish to elude to whether the gut microbiome could ‘predict’ coffee consumption, indicating a close relationship between individual coffee consumption patterns and the microbiome.

Introduction

The first two paragraphs of the introduction are not clear on the level and strength of evidence in relation to acute and habitual consumption of coffee, e.g. coming from intervention or observational studies, with data being generated in mouse models or obtained in human patients, which gives an incomplete picture of the evidence available. In addition, some of the evidence from studies that are being referred to is not of very high quality, and therefore any conclusions about efficacy

should be put in the right perspective.

I am not sure what the purpose is of the third paragraph; if all evidence is from in vitro studies this basically means we are not clear on what coffee consumption does to gut microbiota composition and generation of secondary metabolites. That should be made clear in relation to the research question introduced in the fourth paragraph. A clear topic sentence starting this paragraph would help.

Lines 50-52: I'm not sure if references 15-17 substantiate the claim that coffee may delay cognitive decline in aging. Improvements in short term outcomes are not necessarily associated with reduction in cognitive decline.

Lines 75-78 state 'Thus, for the first time, we investigate the effect of coffee consumption on cognition, mood, behaviour, and gut microbiota in healthy, adult, moderate coffee drinkers compared with non-coffee drinkers, using shotgun metagenomics, targeted and untargeted metabolomics.' This is not the first study to investigate the effect of coffee on cognition and mood, and metagenomics and metabolomics are not methods for assessing cognition and mood (although they are for gut microbiota). This sentence should be reworked, maybe indicating also how cognition and mood were assessed.

In general, the introduction introduces coffees influence on some cognitive outcomes and some gut microbiome outcomes, but I don't think a clear link is made between these two areas that would help complete the narrative that underpins the reason for the study investigating these two areas. The hypothesis stated on lines 81-84 makes this link a bit clearer, but citations may be needed to back up why this is the hypothesis.

Methods:

Lines 552-554: "Dietary intake was monitored and quantified by using a 7-day food diary. The food diary was completed three times, before the baseline visit, before the pre-intervention visits and before the post-intervention visit". How exactly was the food diary executed, and was food intake weighted?

Lines 595-597: "During the screening visit, both NCD and CD participants completed a 7-day caffeine consumption diary. In the diary, each participant reported all the caffeinated beverages consumed in the previous 7 days". Whilst I realise that this questionnaire was only used for caffeine consumption quantification, reporting frequency of single food consumption is really difficult to remember beyond 4-5 days – how accurate do researchers think the outcomes for this diary are, and why did they not use a (weighted) caffeine consumption diary over 7 days?

Generally, the participants had to complete a large number of self-reported questionnaires and neuropsychological assessments, which presents a significant participant burden. How robust and validated are each of these measurements? It is likely that this amount of tests could have led to reporting fatigue – and I don't think the authors tested for this?

Line 610 and lines 690-696, I believe the STAI questionnaire has two arms, a trait arm and a state arm. I think it should be mentioned whether one or both of these were used. The trait arm seems most appropriate for some aspects but the state arm seems more appropriate for others, especially the use of the STAI questionnaire before and after the SECPT task.

Lines 625 to 637, I don't find the description of the ModRey task to be very clear. Two forms are mentioned suggesting these are different but what makes forms 1 and 2 unique is not stated. The test is also described as being split into two parts, however, there appears to be three parts described: a short-delayed recall, a long-delayed recall, and a recognition trial. This paragraph needs to be much clearer.

Lines 670-672, what makes the forms of PASAT different?

Lines 949-950: g*Power was used for power calculations considering an α-error probability (false positive) of 0.05 and a power of 80%. Allowing for attrition, a minimum of 60 volunteers were recruited. The description of the power analysis is not clear enough. What is the treatment effect the investigators were looking for, for which outcome(s), and what was the expected variability in this outcome(s). How much attrition did the investigators allow for?

Lines 951-952: "Outliers have been identified using Grubb's test ($\alpha = 0.05$) and removed in every set of data". Again, the description of outlier analysis is not clear enough. Based on what criteria were outliers removed? How many data (and in which dataset) were date removed?

Figure 1A in its current form it is too small – generally the font size of all figures is too small and unreadable. The numbers in the images of the people are illegible and why are the people all different colours? The figure does also not match with the written description of the study. The washout period for CDs is described as being a caffeine and coffee washout in the text but only a caffeine washout in the figure. I assume the washout was for all coffee and other caffeine as compounds other than coffee were also of interest in this study, so the figure should be corrected. I don't think the decaffeinated, caffeinated, and non-coffee drinker symbols accurately represent the words and the use of the words doesn't represent the study as described. The decaf symbol looks like no coffee, and the non-coffee drinker symbol still looks like coffee. These could be replaced simply with the CD and NCD abbreviations and maybe just DC for decaf periods. The caffeinate symbol is not correct for the baseline participants who drank coffee, they are described as coffee drinkers, not just caffeine consumers. I think the day numbers could be written as study day numbers, not section day numbers, to improve clarity.

Why does the intervention bar overlap with the washout bar? On lines 524 to 530, the washout period was described as two

weeks, then there was a pre-intervention visit, and then the intervention period began. No overlap was described.

Results

Table 1: were these data taken at baseline? That should be pointed out in the title. The outcome of 'daily caffeine given up for the study' is provided for both NCD and CD, but this outcome is not relevant for NCD as they did not participate in any study interventions? Also, Adenosine A2A Receptor Genotype distribution was different for NCD and CD. The TT genotype has been linked to increased coffee intake, possibly due to these individuals experiencing less negative side effects of caffeine consumption. However, the opposite was found in this study. This should be discussed.

Lines 105-107: "Participants' food intake was monitored for 7 days before each study visit. No major dietary changes were reported, except for slightly higher trans-fat consumption by CD participants compared to NCD participants (Supplementary Table 2a)." Assuming that NDC participants were only involved in the baseline measurements, this sentence implies that a dietary data 'comparison' is only provided for baseline data?

Measuring food intake is of real importance for interpreting the effects on the microbiome, as changes in the microbiome are more likely to be the result of an entire diet rather than a few changes in coffee intake, and this should be tested in the study. What happened to the results of the 7-day dietary intake monitoring of other study visits? Dietary intake data should be reported in the main study, and linked directly to changes in the gut microbiome, and evaluated against whether participants were drinking caffeinated versus decaffeinated coffee. I am unclear why the authors only report on macronutrients – they should present a full picture of the diet.

Lines 110-111: "With no significant dietary differences between CD and NCD throughout the study" - again, NCD participants were only used for baseline measurements? I am now starting to think that the acronym NCD was used for other participants also and this is starting to be terribly confusing.

Line 150-152 "When comparing ModRey scores between the start of the study, at baseline, and the end of the study, post intervention significant improvement was seen only in NCDs (Supplementary Table 8)." This sentence gives the impression that NCD participants were tested beyond the baseline visit, which I don't think can be the case?

Lines 152-154 and 161-164, I feel like these are more points for the discussion, not the results, however, why are differences between the caffeinated group and decaffeinated group explained by a learning effect for the ModRey task, but non-caffeine compounds in coffee for the PAL task? In the methods, it was stated that the ModRey task has different forms to reduce impact of learning, but the PAL task does not, and even still, both tasks were only used on visits two and four, so surely the PAL task is at greater risk of learning bias given it does not have different forms?

Lines 178-183, this result and justification is unclear. In figure s9, it appears that at various time points during the intervention, headache, "decreased sociability, motivation to work", mood disturbance, and drowsiness/fatigue were increased in decaffeinated group compared to the caffeinated group, which doesn't tally with 'These findings suggest coffee consumption improves mood and reduces withdrawal symptoms, regardless of caffeine content', it suggests that caffeinated coffee maybe has an advantage with regards to these mood components.

Lines 185-186: "Overall, we saw that coffee influences cognitive processes, both impulsiveness and emotionality when compared with non-coffee drinkers" – this is vague and needs to contain a more focussed description of main results.

Lines 187-189, figure 4 does not show 'reintroduction of caffeinated coffee, and decaffeinated coffee had distinct and specific effects on several cognitive and physical measures'. It shows associations between microbial species, metabolites, and tasks.

Discussion

The authors present a significant amount of data and outcomes (cognitive function, stress levels, inflammation, faecal metabolites, changes in the gut microbiome etc), but it remains unclear how this work has moved beyond the state-of-the-art, and how this study has delivered new insights. Many outcomes in relation to coffee consumption have been reported before, and are not new.

Reviewer #2

(Remarks to the Author)
This is an original, and novel investigation by Boscaini et al, titled "Habitual Coffee Intake Shapes the Gut Microbiome and Modifies Host Physiology and Cognition", with the purpose of probing coffee and caffeine's effects on the microbiota-gut-brain axis in healthy females and males using a sophisticated multiomics approach. The study directly compared thirty-one non-coffee drinkers (NCD) with thirty-one coffee drinkers (CD) using a cross-sectional comparison. This comparison was followed by a two-week coffee abstinence wash-out period in the CD's only, at which time they were block randomized into either a caffeinated or decaffeinated coffee consumption group in a double-blinded, parallel design for an additional three weeks. All participants underwent comprehensive collection for biological blood samples, urinary and faecal metabolomics, faecal gut microbiome, cognitive, impulsivity, emotion reactivity, urgency, memory, as well as physiological and psychological stress testing, and dietary intake. The authors conclude that coffee drinkers exhibit higher impulsivity and

emotional reactivity, improved immune function, reduced neuroactive compounds, and increased alpha diversity of gut microbiome compared to non-coffee drinkers. Further, attention and vigilance increased whereas impulsivity and emotional reactivity decreased upon withdrawal of caffeine. Coffee reintroduction following abstinence, whether caffeinated or decaffeinated, decreased perceived stress and impulsivity and self-reported depression. Only reintroduction of caffeinated coffee reduced anxiety. The authors also conclude periods of coffee withdrawal impact mood, immune markers, and microbial diversity due to caffeine and/or other coffee containing compounds. The study uses cutting-edge methodologies to explore unique impacts of coffee on physiological, psychological, immunological, and microbial composition, which is novel and of great value. However, the basis of the comparison between NCD and CD participants has significant limitations and this influences the interpretation of all the different analyses.
Major Comment:
The major limitation with the study is the lack of appropriate control of caffeine prior to the baseline testing day. This compromises the integrity of the remaining analyses performed in all outcome measures. Specifically, the NCD's were consuming almost 70mg of caffeine/day prior to baseline testing and abstained from nearly all of this prior to the baseline visit. This suggests they experienced a drastic change in their caffeine intake during baseline were essentially in caffeine withdrawal. Whereas, CD's were consuming 4x this amount on a daily basis and only abstained 1/10th (28mg) of their usual intake. Thus, the baseline data is comparing caffeine withdrawal in NCD's and caffeine habituation in CD's. The authors need to explain this within the context of their overall findings prior to any further consideration.
Minor Comments:
line
24 this assumes the gut aspects are different from the gut microbiome aspects?
44 the introduction is focused primarily on coffee, independent of any mention of caffeine. Given the age-related pharmokinetics, metabolic, and cardiovascular effects of caffeine, how might this affect the outcomes? This should be discussed more thoroughly in the introduction for perspective. Further, it appears the authors transition, unintentionally between coffee and caffeine throughout this section. Is the emphasis coffee or is it caffeine?
66-72 This appears to both address coffee's potential as a prebiotic by different mechanisms. The authors should state these differences, as such.
87 Was heart rate, blood pressure, fasting blood glucose, body weight/composition, etc. measured to determine physical wellbeing?
94 This seems like a major limitation with the study. NCD's were consuming almost 70mg of caffeine/day and abstained from nearly all of this prior to baseline visit. This suggests they experienced a drastic change in their caffeine intake during baseline. Whereas, CD's were consuming 4x this amount on a daily basis and only abstained 1/10th (28mg) of their usual intake. Thus, the baseline data is comparing caffeine withdrawal in NCD's and caffeine habituation in CD's. The authors need to explain this within the context of their overall findings prior to any further consideration.
121 interesting that cognitive performance improved following coffee washout period. If it was due to task repetition, performance would be improved following the coffee intervention in both groups. Was it?
144 NCD had improved sleep and physical activity. This seems like a significant difference and benefit.
148 Here again, NCD improved memory compared to CD, which is notable. If it was a practice effect why did only NCD improve and not CD?
185 There seems to be little doubt that NCD performs better on most important health outcomes than CD. This should be explained further.
312 what would cause the increase in velonella parvula in CD similar to increased levels in NCD?
341 what would account for the increased velonella in NCD then?
342 delete either "associated" or "correlated".
361 How do you explain CD having higher impulsivity and emotional reactivity, yet greater stress tolerance and anti-inflammatory responses?
406 What explains the lower GABA in CD, while reduced anxiety? It would suggest these should respond in opposition to each other.
410 This sentence is unclear. Please reword.
424 How is it that coffee abstinence does not influence diversity, yet species level changes occur after abstinence? Please explain.
427 Is this Day 2 of the Intervention phase? Please specify
436 It's assumed the authors are referring to Cryptobacterium curtum here, correct?
438 The authors should consider organizing the Discussion similar to how it's presented in the figure.
441 This is a very long sentence and rather confusing. Please rephrase.
446 This statement appears redundant as the authors state caffeine metabolites change which are associated with microbial species and cognitive/behavioral outcomes in lines 441-445..
461 The authors state their data agrees with the hypothesis that the gut microbiome may influence the variability in phenolic metabolism. Yet, they later state this connection is far more connected with the urinary metabolome than the gut microbiome. This discrepancy needs addressing.
1109 While groups differed in total caffeine intake, the NCD group still consumed the equivalent of one cup of coffee per day. Thus, they were not caffeine naive. How does this impact the results and conclusions?
1110 Given pharmokinetic differences in caffeine matabolism between ethnicities and sexes and the fact there were 10 times more "other" ethnicities between groups and more women, how might this influence the outcomes?

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
In the revised manuscript, the authors provide thorough responses to both reviewers' concerns. Specifically, stool transit time was not accounted for in the original manuscript, as noted by Reviewer 1. However, the authors added BSS and GI-VAS to the revised manuscript and present comprehensive data suggesting similar transit times between and among groups (Results; Page 6, paragraph 1). The authors also addressed the lack of sufficient power analysis in the original manuscript and provide detailed explanations for choosing the use of effect sizes over p -values for most of the outcome variables, including coffee-related metabolites, cognitive performance, response to an acute stressor, peripheral blood inflammatory profile and the cortisol awakening response. Indeed, given the abundance of microbial taxa within and between individuals, the authors highlight the limitation of using power analysis for gut microbiome outcomes.
In addition, the major limitation of the lack of appropriate control of caffeine prior to baseline testing between groups was sufficiently addressed by the authors. While this still remains a major concern of the study, the authors have adequately addressed this issue in the Discussion as a limitation with their study design and the impact it has on the findings.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This is a comprehensive, timely and very interesting investigation into how coffee intake shapes the gut microbiome and modifies host physiology and cognition.
We thank the Reviewer for their positive and very constructive comments on our paper

However, I believe the study lacks robust measurements of some basic outcomes that are required to interpret the results.
1. One major point of critique on this paper is the lack of measurement of stool transit time. The investigators rightly mention that both caffeinated and decaffeinated coffee stimulates the contractility of ileal and colonic smooth muscle, and thus speeds up transit time, which in turn is believed to have a significant effect on the composition, diversity and function of the microbiome. Considering the fact that both caffeinated and decaffeinated coffee stimulate contractility, this effect is may not be due to caffeine per se. In any case, effects of coffee consumption (with or without caffeine) on the microbiome should be considered against a background of changing transit time, as this may be an essential element of the causal mechanistic pathway.
We thank the Reviewer for raising this important point and understand their concerns regarding stool transit time, or contractility and the fact that we did not measure it in this study. While we now identify this as a limitation of our study in the discussion [Page 16, last paragraph], we feel that to measure gut contractility in a clinically meaningful way would require using a wireless capsule or alternatively, radio-opaque markers, which would require clinical supervision at each timepoint and for each participant. Furthermore, we had participants giving samples at multiple timepoints throughout the study, and thus to schedule contractility measurement at each one, depending on bowel habit would reduce participant compliance significantly. Additionally, both methods would influence microbiome composition and would require additional control groups in the study to mitigate against these potential differences given that microbiome composition was one of our main readouts from the study.

However, we do have data from the Bristol Stool Scale (BSS) and elaborate on these findings in the revised paper as the BSS has long been proposed as a surrogate marker of stool transit (Heaton et al. 1992; Saad et al. 2010; Degen and Phillips 1996)**. One study identified BSS as an accurate predictor of delayed transit when compared to wireless motility capsule and radio-opaque marker. Overall, a BSS value less than 3 predicted delayed whole-gut transit with a sensitivity of 85% and specificity of 82% and delayed transit with a sensitivity of 82% and specificity of 83%** (Saad et al. 2010)**. BSS is widely used in clinical practice and is extensively used in the diagnosis of constipation or diarrhoea subtypes in patients with Irritable Bowel Syndrome (IBS),** (Longstreth et al. 2006) **with high scores (6 or 7) corresponding to loose stools and fast transit, while lower scores (1 and 2) indicating hard stools and longer colon transit times** (O'Donnell et al. 1990; Saad et al. 2010) **with BSS scores of 3 and 4 indicating ideal stool form** (Lewis and Heaton 1997)**.**

In the context of our own data, whilst we show that there is a modest difference between the BSS score of NCD (Median, 4.00, 25% 3.00, 75% 5.00) and CD

(Median, 3.00, 25% 2.00, 75% 4.00) at the baseline visit, both medians are within what is considered the normal range for BSS scores.

Furthermore, we see that abstinence from coffee in the CD group results in a median value of 4, while reintroduction of either caffeinated coffee or decaffeinated coffee results in a median value of 3, irrespective of group.

In addition, each participant, at each visit completed a Gastrointestinal Symptoms Visual Analogue Scales (GI-VAS) which included questions about the presence of abdominal pain, bloating, bowel urgency and overall bowel habit satisfaction. The results of this questionnaire mirror those of the BSS. Suggesting that overall, there was little change in two different surrogate measures of bowel contractility. We have amended the discussion to reflect this.

In addition, effects of coffee consumption (with or without caffeine) on the microbiome should be considered against the background of the full habitual diet (not just its macronutrients), as the diet may have a larger effect on the microbiome than a short-term shift in the number of (de)caffeinated drinks. Both elements (stool transit time and background diet prior to stool sampling) were not properly considered in this research, and therefore it will be difficult to interpret the results, especially on the microbiome, properly.
We thank the Reviewer for suggesting a more in-depth dietary intake analysis to strengthen the results of our study. To this end we repeated and expanded the dietary analysis to take this into account. Although magnesium (for caffeinated coffee drinkers) and selenium (for decaffeinated group) intake where statistically different from their baseline compared to their final timepoint, the magnitude of observed differences in nutrient intakes was small relative to the high variability within groups, as shown in the descriptive tables (Supplementary Tables 2a-e). This substantial overlap suggests that these differences may reflect minor fluctuations or inaccuracies in self-reported dietary intake, well-known limitations of all self-reported dietary assessment instruments (Park et al. 2018). No other differences in macro- and micronutrient intakes were detected at baseline and within groups. Additionally, food-group–based analyses provide a more robust characterization of dietary intake (Cotillard et al. 2022) and are more reliable in linking diet with gut microbiota features. In the present study, no food-group comparisons reached statistical significance at baseline (CD vs NCD) or during the intervention period for both groups (Caffeinated vs Decaffeinated). Moreover, the large number of comparisons, further increases the risk of type I errors, and the chance of detecting noise in the data as opposed to actual variations of intake. Taken together, we believe that the absence of differences at the food-group level and the inherent variability in self-reported nutrient data indicate that the few small observed differences are unlikely to have meaningful implications for our study outcomes. We have amended the results section [Page 6, paragraph 1], and Supplementary Table 2 a-e] to reflect the more comprehensive analysis suggested by the reviewer.

A final major point of critique is the lack of an appropriate power analysis, which probably requires a more robust research question and pre-identified validated outcomes. It is not clear whether the study would have had sufficient power to find significant effects on individual or group level, and many results are presented without indicating any level of significance.

We agree that the initial manuscript did not give enough information regarding the power analysis and pre-identified validated outcomes. For clarity, we have included this information in the manuscript [Methods: Statistical Analysis]. Our primary outcome measure was microbiota composition and function (see NCT05927038); thus, this was pre-identified and validated [Figure 3]. Secondary outcome measures included gut microbial metabolites (including SCFAs) and coffee-related metabolites along with cognitive performance, response to an acute stressor, peripheral blood inflammatory profile and the cortisol awakening response.

There are several issues in calculating power statistics in microbiome studies. Microbiome data often includes thousands of microbial taxa, many of which are rare or absent in most samples. Microbiome data are compositional with only relative abundances observed. The microbiome also varies greatly between individuals even within the same treatment group which can reduce power and increases the samples size needed to detect meaningful differences. Furthermore, studies often test thousands of taxa individually for associations, thus correcting for multiple comparisons (using FDR) can reduce statistical power dramatically. Moreover, microbial taxa are not independent; they often co-occur or are functionally related. In addition, we look at the microbiome longitudinally.

We have amended the Statistical Analysis section to provide more information. ***Our primary outcome measure was microbiota composition and function (see NCT05927038) while secondary outcome measures included gut microbial metabolites (including SCFAs) and Coffee-related metabolites. Other secondary outcomes included cognitive performance, response to an acute stressor, peripheral blood inflammatory profile and the cortisol awakening response. g*Power was used for power calculations considering an α-error probability (false positive) of 0.05 and a power of 80%. Effect sizes of 0.25 will be able to be differentiated using n=18 per group (ANOVA repeated measures) for microbiota composition and function. Allowing for attrition, a minimum of sixty volunteers were recruited. The minimum required samples size per group was eighteen, thus, thirty-six participants were required for analysis. We recruited sixty-two in the end; thus, we allowed for an estimated attrition rate of 42%.***

In the other comparisons, using a sensitivity analysis when comparing coffee drinkers with non-coffee drinkers, using an α of 0.05, a power of 0.8 (80%) and a sample size of 32 per group, we would have the power to detect a Cohen's d of 0.63, a medium to large effect size. When assessing power requirements for the longitudinal study of coffee drinkers consisting of 32 participants over 5 different timepoints, using an α of 0.05, a power of 0.8 (80%) we have the power to detect an effect size of 0.22, a small to medium effect size, we will list this fact as a limitation of the study in the discussion.

Throughout this paper, we have indicated levels of significance by reporting either p values or effect sizes—both valid measures of statistical analysis. However, because the microbiome was the primary focus of this study and due to its longitudinal design, we prioritized reporting effect sizes over p values. While p values indicate whether an effect is statistically significant, effect sizes provide a more meaningful measure of the practical or real-world impact of our findings.

Given the generalizability of our data—especially considering the widespread consumption of coffee—we believed that effect sizes offered a more informative and impactful way to present our results. In the microbiome and metabolomics data, all that is shown are indeed the significant features; thus, all would be significant. Moreover, all microbiome data from this study will be made publicly available to support future research and meta-analyses. Effect sizes are more conducive to this goal than p values, which can limit comparability across studies. Finally, since effect sizes are less sensitive to sample size, and our study involved a more modest sample compared to recent work (Manghi et al. 2024) this approach offered a more appropriate and robust method of interpretation.

The text contains a large number of inaccuracies and vague descriptions, and the figures are not always informative and easy to summarise/interpret, other than showing results on a high level within the paper, and with many more detailed figures in the supplement, which does not necessarily help to interpret the results. My comments are outlined in more detail below.
We appreciate your detailed review of our manuscript and have attempted to address your comments below.

Abstract
The abstract indicates differences were found between groups, what they were linked to, and metabolites involved, but no indication as to the direction of effects is given which would make the abstract more informative.
We appreciate the Reviewer's comments on our abstract and have amended it to make it more informative.

Coffee influences multiple physiological processes, including gut function, stress, cognition, and the microbiome. However, the mechanisms underlying these effects remain poorly understood. In this study, we examined coffee's impact on the microbiota–gut–brain axis—a bidirectional communication pathway between the gut microbiome and the brain—and assessed whether these effects occur independently of caffeine in healthy participants. Significant group differences emerged in faecal microbiome composition, with coffee drinkers showing increased relative abundance of Cryptobacterium and Eggerthella species, alongside reduced levels of the metabolite's indole-3-propionic acid, indole-3-carboxyaldehyde, and the neurotransmitter γ-aminobutyric acid. Behaviourally, coffee drinkers exhibited greater impulsivity and emotional reactivity, whereas non-coffee drinkers demonstrated better memory performance. Some alterations in the faecal metabolome were reversible following coffee abstinence, and reintroduction triggered acute microbiome changes independent of caffeine. An integrated model identified nine key metabolites—including theophylline, caffeine, and selected phenolic acids—strongly linked to microbial species and cognitive measures. These findings reveal previously unrecognised effects of coffee on the microbiota–gut–brain axis, suggesting that microbiome profiles could potentially predict coffee consumption patterns and highlighting a close association between coffee intake and gut microbial composition.

The concluding sentence suggests the microbiome could be a biomarker for coffee consumption, but this was not stated as the aim of the study. Using the microbiome for

habitual coffee consumption seems very cost- and effort- ineffective considering that there are easier ways to monitor coffee consumption already. Instead, the authors may wish to allude to whether the gut microbiome could 'predict' coffee consumption, indicating a close relationship between individual coffee consumption patterns and the microbiome.

We appreciate the authors comments on the concluding sentence in the abstract and have modified the sentence to reflect the Reviewer's comments [Page 2].

Introduction

The first two paragraphs of the introduction are not clear on the level and strength of evidence in relation to acute and habitual consumption of coffee, e.g. coming from intervention or observational studies, with data being generated in mouse models or obtained in human patients, which gives an incomplete picture of the evidence available. In addition, some of the evidence from studies that are being referred to is not of very high quality, and therefore any conclusions about efficacy should be put in the right perspective.

I am not sure what the purpose is of the third paragraph; if all evidence is from in vitro studies this basically means we are not clear on what coffee consumption does to gut microbiota composition and generation of secondary metabolites. That should be made clear in relation to the research question introduced in the fourth paragraph. A clear topic sentence starting this paragraph would help.

In line with the Reviewer's comments, we have restructured the introduction to make it both clearer and more suitable for the data presented. We have added additional references and removed those considered less suitable [Introduction Page 3 and 4].

Lines 50-52: I'm not sure if references 15-17 substantiate the claim that coffee may delay cognitive decline in aging. Improvements in short term outcomes are not necessarily associated with reduction in cognitive decline.

We appreciate the reviewers' comments and have amended this text to improve the accuracy. [Page 3 paragraph 2].

Lines 75-78 state 'Thus, for the first time, we investigate the effect of coffee consumption on cognition, mood, behaviour, and gut microbiota in healthy, adult, moderate coffee drinkers compared with non-coffee drinkers, using shotgun metagenomics, targeted and untargeted metabolomics.' This is not the first study to investigate the effect of coffee on cognition and mood, and metagenomics and metabolomics are not methods for assessing cognition and mood (although they are for gut microbiota). This sentence should be reworked, maybe indicating also how cognition and mood were assessed.

We appreciate the Reviewer's comments and have restructured this section of the paper to better reflect the importance of this study [Page 4 line 83-90].

In general, the introduction introduces coffees influence on some cognitive outcomes and some gut microbiome outcomes, but I don't think a clear link is made between these two areas that would help complete the narrative that underpins the reason for the study investigating these two areas. The hypothesis stated on lines 81-84 makes this link a bit clearer, but citations may be needed to back up why this is the hypothesis.

We agree with the Reviewer's opinion and have made a concerted effort to make this connection clearer in the introduction [Page 4 last paragraph].

Methods:
Lines 552-554: "Dietary intake was monitored and quantified by using a 7-day food diary. The food diary was completed three times, before the baseline visit, before the pre-intervention visits and before the post-intervention visit". How exactly was the food diary executed and was food intake weighted?
The 7-day food diary was a paper-based document that was completed at the timepoints mentioned, and the information was then manually entered into Nutritics™, a food data management software. Food intake was weighted. To clarify this point we have updated this section of the methods. [Methods section, "Dietary intake quantification", Page 19]. We will include a copy of the food diary, including how participants were to estimate food portion size in the supplementary material.

Lines 595-597: "During the screening visit, both NCD and CD participants completed a 7-day caffeine consumption diary. In the diary, each participant reported all the caffeinated beverages consumed in the previous 7 days". Whilst I realise that this questionnaire was only used for caffeine consumption quantification, reporting frequency of single food consumption is really difficult to remember beyond 4-5 days – how accurate do researchers think the outcomes for this diary are, and why did they not use a (weighted) caffeine consumption diary over 7 days?
We acknowledge the Reviewer's comment regarding the questionnaire used to measure coffee consumption. Firstly, we would like to point out that the caffeine diary recorded the type of food or beverage, the amount prepared (if required) and the method of preparation. Coffee is consumed in a highly habitual manner, where it is either consumed daily, or not at all (Manghi et al. 2024; Salvini et al. 1989)**. To probe the reliability of our coffee consumption data obtained from participants we performed an Intraclass Correlation Coefficient (ICC) analysis using the 7-day coffee consumption data. Intraclass correlation coefficient (ICC) is a widely used reliability index in test-retest, intrarater, and interrater reliability analyses. Using the Two-way random effects, absolute agreement, single rater/measurement model we saw that coffee drinkers had an ICC value of 0.77 and non-coffee drinkers had an ICC value of 0.87 which indicate statistically reliable and consistent data** (Koo and Li 2016)**.**

Generally, the participants had to complete a large number of self-reported questionnaires and neuropsychological assessments, which presents a significant participant burden. How robust and validated are each of these measurements? It is likely that this amount of tests could have led to reporting fatigue – and I don't think the authors tested for this?
We understand the reviewers' concerns relating to reporting fatigue. We chose the Food Frequency Questionnaire (FFQ) specifically to reduce participant burden, while acknowledging that FFQ has limitations. The FFQ allows us to capture long term dietary patterns ("Reproducibility and Validity of Food Frequency Questionnaires" 2012)**, are low cost and easy to administer to participants** (Cade et al. 2002)**, and allow nutritional pattern analysis** (Linseisen et al. 2009)**. We felt that the brevity of the FFQ would allow us to capture important dietary information without burdening the participant too much.**

We agreed that there is a large amount of self-reported questionnaires in this study. Of the questionnaires administered to participants, Cohen's Perceived Stress Scale (Cohen et al. 1983)**, Emotional and Reactivity Scale,** (Nock et al. 2008)**, Urgency Premeditation Perseverance Sensation Seeking Positive Urgency Impulsive Behaviour Scale (UPPS-P)** (Cyders et al. 2014)**, Hopkins Symptom Checklist 90 (SCL 90-R),** (Lipman et al. 1979)**, Beck Depression Inventory-II (BDI),** (Arnarson et al. 2008)**, State Trait Anxiety Inventory, (STAI),** (Ferreira and Murray 1983)**, Pittsburgh Sleep Quality Index, (PSQI),** (Buysse et al. 1989)**, International Physical Activity Questionnaire (IPAQ),** (Booth 2000) **are all validated, references are included at the end of this document and in parentheses for clarity. In addition, while the use of visual analogue scales for gastrointestinal symptoms has been used in patient populations such as Irritable Bowel Syndrome** (Bengtsson et al. 2011)**, it is not validated in healthy populations, thus, we will include this as a limitation of the use of this, in the discussion.**

We do acknowledge that there may be some fatigue amongst participants, but all questionnaires and cognitive tests were conducted in the same sequence across all study visits, so fatigue, if any should be consistent. In addition, all cognitive tests were conducted at the beginning of the study visit, to ensure that fatigue did not affect the results. Moreover, the questionnaires assess predominantly trait measures which in theory are not affected by state factors such as fatigue or mood. Furthermore, we have quantified the number of unanswered questionnaires in the study at the specific timepoint of the study and have added them in a table format below. From this data we can see that dropout or failure to submit questionnaires was quiet low throughout the study. Indeed, all questionnaires were returned during Visit 4 at the end of the study.

Table 1. Number of Unanswered Questionnaires at Specific Timepoints During the Study.

Visit Two	Not answered	Visit 3	Not answered	Visit 4	Not answered	Intervention	Not answered
PSSQv2	0	PSSQv3	0	PSSQv4	0	QCC Intervention t2	1
BDIv2	0	BDIv3	0	BDIv4	0	QCC Intervention t4	2
GI-VASv2	0	GI-VASv3	0	GI-VASv4	0	QCC Intervention t14	4
UPPSv2	0	UPPSv3	0	UPPSv4	0	CWSQ Intervention t2	3
ERSv2	0	ERSv3	0	ERSv4	0	CWSQ Intervention t4	1
STAIv2	0	STAIv3	0	STAIv4	0	CWSQ Intervention t14	3
HSCLv2	0	HSCLv3	0	SCLv4	0		
		QCC v3	0	CWSQ V4	0		
		CWSQ v3	1				

		QCC washout 2	0				
		QCC washout4	2				
		CWSQ washout 2	0				
		CWSQwashout4	3				

PSS = perceived stress scale, Q = Questionnaire, V = visit, BDI = Beck's depression inventory, GI-VAS = gastrointestinal visual analogue scale, QCC = questionnaire of caffeine craving, CWSQ = caffeine withdrawal symptoms questionnaire, UPPS = impulsive behaviour scale, ERS = emotional reactivity scale, STAI = state-trait and anxiety scale, HSCL = Hopkins symptoms checklist.

Line 610 and lines 690-696, I believe the STAI questionnaire has two arms, a trait arm and a state arm. I think it should be mentioned whether one or both of these were used. The trait arm seems most appropriate for some aspects but the state arm seems more appropriate for others, especially the use of the STAI questionnaire before and after the SECPT task.

We agree with the Reviewer's point that the STAI information is not as clear as it should be. During the baseline, post-withdrawal, and post-intervention phases of the study the STAI trait arm was used while before and after the SECPT challenge, the STAI state arm was used. In the methods section we have clarified this point [Page 21, paragraph 1] as well as throughout the study where STAI is mentioned, the precise arm is defined.

Lines 625 to 637, I don't find the description of the ModRey task to be very clear. Two forms are mentioned suggesting these are different but what makes forms 1 and 2 unique is not stated. The test is also described as being split into two parts, however, there appears to be three parts described: a short-delayed recall, a long-delayed recall, and a recognition trial. This paragraph needs to be much clearer.

We agree with the Reviewer and have rewritten this section to improve clarity and readability.

Lines 670-672, what makes the forms of PASAT different?

The PASAT forms differ based on the numbers used in each. Thus, the numbers in Form A (Visits 2 and 4) are different from Form B, used in Visit 3. We have included a clarifying sentence in the manuscript [Page 22 paragraph 3].

Lines 949-950: g*Power was used for power calculations considering an α-error probability (false positive) of 0.05 and a power of 80%. Allowing for attrition, a minimum of 60 volunteers were recruited. The description of the power analysis is not clear enough. What is the treatment effect the investigators were looking for, for which outcome(s), and what was the expected variability in this outcome(s). How much attrition did the investigators allow for?

We agree that this section was lacking important information, and thus have added the following detail [Methods section, Statistical Analysis, Page 30]

Our primary outcome measure was microbiota composition and function (see NCT05927038) while secondary outcome measures included gut microbial metabolites (including SCFAs) and Coffee-related metabolites. Other secondary

outcomes included cognitive performance, response to an acute stressor, peripheral blood inflammatory profile and the cortisol awakening response. g*Power was used for power calculations considering an α-error probability (false positive) of 0.05 and a power of 80%. Effect sizes of 0.25 will be able to be differentiated using n=18 per group (ANOVA repeated measures) for microbiota composition and function. Allowing for attrition, a minimum of sixty volunteers were recruited. The minimum required samples size per group was eighteen, this, thirty-six participants were required for analysis. We recruited sixty-two in the end; thus, we allowed for an estimated attrition rate of 42%.

Lines 951-952: "Outliers have been identified using Grubb's test (α= 0.05) and removed in every set of data". Again, the description of outlier analysis is not clear enough. Based on what criteria were outliers removed? How many data (and in which dataset) were date removed?

We agree that the description about how outliers were handled during statistical analyses was unclear and have amended the text to make it clearer. Briefly, due to the modest sample size in the study, we only removed statistical outliers to satisfy the assumption of equal variances, or in situations when participant-reported data was clearly outside the range of acceptable responses. Furthermore, many of the outputs analysed do not have established reference ranges, thus we did not remove outliers unless they were deemed significant by the Grubb's test (Grubbs 1950)**. We would also like to clarify that outliers were not removed in every set of data, this was a typo and has been corrected in the text. Below we have included tables that contain the number of outliers removed from the analysis.**

Group	CRP (mg/L)	IFN-γ	IL-10	IL-6	IL-8	TNF-α
	Plasma					
NCD	1	2	1	1	1	1
CD Visit 2	1	1	1	2	0	0
CD Visit 3	1	2	0	1	0	0
CD Visit 4	1	2	0	1	0	0

Group	CRP (mg/L)	IFN-γ	IL-10	IL-6	IL-8	TNF-α
	Stimulate Blood					
NCD	0	0	0	0	0	1
CD Visit 2	0	0	0	3	0	0
CD Visit 3	0	0	0	0	0	0
CD Visit 4	0	0	2	0	2	0

Table 2: Number of Unanswered Questions per Study Visit

Behavioural and Cognitive Data	
Test and No. outlier removed.	**No. Outliers Removed**
BDI Total Visit 2	1
Visit 3 GIS VAS Total	1
Visit 3 UPPSTot	1
Visit 4 UPPS Premeditation	1
Visit 2 ERSTot	1
Visit 2 ERS Persistency	1
Visit 3 ERS Persistency	1
Visit 4 ERS Persistency	1
Visit 2 STAI TRAITTot	1
Visit 3 STAI TRAITTot	1
Visit 2 SCL obsessive	2
Visit 3 SCL depr	2
Visit 3 SCL PST	1
Visit 4 SCL GSI	2
Visit 2 PSQI GlobalScore	1
Visit 4 PSQI GlobalScore	1
Visit 4 MdR ListBDelayed F2	1
Visit 4 MdR ListAFalseAllarm F2	2
Visit 4 MdR Intrusion F2	1
Interv t2 VASFFatigue	1
Interv t14 VASFFatigue	1
Interv t14 VASFEnergy	1
Visit 4 VASFFatigue	1
Visit 4 VASFEnergy	1
Washout t2 CWSQ Factor 2	1
Washout t2 CWSQ Factor 3	1
Washout t2 CWSQ Factor 5	1

PSS = perceived stress scale, Q = Questionnaire, V = visit, BDI = Beck's depression inventory, GI-VAS = gastrointestinal visual analogue scale, QCC = questionnaire of caffeine craving, CWSQ = caffeine withdrawal symptoms questionnaire, UPPS = impulsive behaviour scale, ERS = emotional reactivity scale, STAI = state-trait and anxiety scale, HSCL = Hopkins symptoms checklist, MdR = ModRey.

Figure 1A in its current form it is too small – generally the font size of all figures is too small and unreadable. The numbers in the images of the people are illegible and why are the people all different colours? The figure does also not match with the written description of the study. The washout period for CDs is described as being a caffeine and coffee washout in the text but only a caffeine washout in the figure. I assume the washout was for all coffee and other caffeine as compounds other than coffee were also of interest in this study, so the figure should be corrected. I don't think the decaffeinated, caffeinated, and non-coffee drinker symbols accurately represent the words and the use of the words doesn't represent the study as

described. The decaf symbol looks like no coffee, and the non-coffee drinker symbol still looks like coffee. These could be replaced simply with the CD and NCD abbreviations and maybe just DC for decaf periods. The caffeinate symbol is not correct for the baseline participants who drank coffee, they are described as coffee drinkers, not just caffeine consumers.
I think the day numbers could be written as study day numbers, not section day numbers, to improve clarity.
We appreciate the Reviewer's comments regarding the study overview figure (Figure 1A). We have re-drafted this figure to make it clearer and easier to follow and have amended the title of the washout phase in Fig1A to reflect that it was indeed a coffee and caffeine washout. In addition, we have changed the day numbers to sequential numbers from start to finish in the image to better reflect the timing of the study parts.

Why does the intervention bar overlap with the washout bar? On lines 524 to 530, the washout period was described as two weeks, then there was a pre-intervention visit, and then the intervention period began. No overlap was described.
We apologise if this description was not clear enough, the last day of the washout phase of the study was also the first day of the intervention phase, with coffee drinkers on the last day of the washout being instructed to consume caffeinated or decaffeinated coffee for the remaining 2 weeks of the study. We have amended the description of this in the methods section [Pages 17 and 18 Line 574-575].

Results
Table 1: were these data taken at baseline? That should be pointed out in the title.
We have amended the title of this table to match the reviewer's suggestion. [Table 1].

The outcome of 'daily caffeine given up for the study' is provided for both NCD and CD, but this outcome is not relevant for NCD as they did not participate in any study interventions?
We included this data as a comparator for the reader, to compare with coffee drinkers.

Also, Adenosine A2A Receptor Genotype distribution was different for NCD and CD. The TT genotype has been linked to increased coffee intake, possibly due to these individuals experiencing less negative side effects of caffeine consumption. However, the opposite was found in this study. This should be discussed
We agree that this is a discrepancy. It is important to note though, that we are not sampling from the same population. In response to the reviewers comment, we have added a brief sentence discussing this fact (Rétey et al. 2007)**. [Page 5, results section paragraph 1]**
We inserted the following text.
In previous research, the T/T haplotype has been associated with higher coffee intake. This discrepancy may suggest that contextual, environmental, or sample-specific factors moderated the relationship between genotype and caffeine consumption in this population.

Lines 105-107: "Participants' food intake was monitored for 7 days before each study visit. No major dietary changes were reported, except for slightly higher trans-fat consumption by CD participants compared to NCD participants (Supplementary

Table 2a).” Assuming that NDC participants were only involved in the baseline measurements, this sentence implies that a dietary data ‘comparison’ is only provided for baseline data?
That is correct, dietary comparisons were only made between NCD and CD at the baseline visit or visit 2.

Measuring food intake is of real importance for interpreting the effects on the microbiome, as changes in the microbiome are more likely to be the result of an entire diet rather than a few changes in coffee intake, and this should be tested in the study. What happened to the results of the 7-day dietary intake monitoring of other study visits? Dietary intake data should be reported in the main study, and linked directly to changes in the gut microbiome, and evaluated against whether participants were drinking caffeinated versus decaffeinated coffee. I am unclear why the authors only report on macronutrients – they should present a full picture of the diet.
We thank the Reviewer for suggesting a more in-depth dietary intake analysis to strengthen the results of our study. We repeated the dietary analysis with the exclusion of an outlier reporting an unrealistic amount of food (>26’000 kcal/day on average). Although magnesium (for caffeinate group) and selenium (for decaffeinated group) intake where statistically different from their baseline compared to their final timepoint, the magnitude of observed differences in nutrient intakes was small relative to the high variability within groups, as shown in the descriptive tables. This substantial overlap suggests that these differences may reflect minor fluctuations or inaccuracies in self-reported dietary intake, well-known limitations of all self-reported dietary assessment instruments (Park et al. 2018). No other differences in macro- and micronutrient intakes were detected at baseline and within groups. Additionally, food-group–based analyses provide a more robust characterization of dietary intake (Cotillard et al. 2022) and are more reliable in linking diet with gut microbiota features. In the present study, no food-group comparisons reached statistical significance at baseline (CD vs NCD) or during the intervention period for both groups (Caffeinated vs Decaffeinated). Moreover, the large number of comparisons, further increases the risk of type I errors, and the chance of detecting noise in the data as opposed to actual variations of intake. Taken together, we believe that the absence of differences at the food-group level and the inherent variability in self-reported nutrient data indicate that the few small observed differences are unlikely to have meaningful implications for our study outcomes (Supplementary Table 2a-2e).

Lines 110-111: “With no significant dietary differences between CD and NCD throughout the study” - again, NCD participants were only used for baseline measurements? I am now starting to think that the acronym NCD was used for other participants also and this is starting to be terribly confusing.
We apologise for the lack of clarity and have amended the text to improve it.

Line 150-152 “When comparing ModRey scores between the start of the study, at baseline, and the end of the study, post intervention significant improvement was seen only in NCDs (Supplementary Table 8).” This sentence gives the impression that NCD participants were tested beyond the baseline visit, which I don’t think can be the case?
Indeed, the Reviewer is correct, we have amended the text to correct this sentence making it clearer.

When comparing ModRey scores between the start of the study, at baseline, and the end of the study, post-intervention significant improvement was seen only in participants receiving decaffeinated coffee during the reintroduction phase of the study (Supplementary Table 8).

Lines 152-154 and 161-164, I feel like these are more points for the discussion, not the results, however, why are differences between the caffeinated group and decaffeinated group explained by a learning effect for the ModRey task, but non-caffeine compounds in coffee for the PAL task? In the methods, it was stated that the ModRey task has different forms to reduce impact of learning, but the PAL task does not, and even still, both tasks were only used on visits two and four, so surely the PAL task is at greater risk of learning bias given it does not have different forms?
We have amended both sections of text and hope that this is now clearer. We appreciate the feedback [Page 7].

Lines 178-183, this result and justification is unclear. In figure s9, it appears that at various time points during the intervention, headache, "decreased sociability, motivation to work", mood disturbance, and drowsiness/fatigue were increased in decaffeinated group compared to the caffeinated group, which doesn't tally with 'These findings suggest coffee consumption improves mood and reduces withdrawal symptoms, regardless of caffeine content', it suggests that caffeinated coffee maybe has an advantage with regards to these mood components.
We appreciate the Reviewer's comments regarding this section of the paper and will attempt to clarify it. The data presented in Figure S9 is only related to coffee drinkers during the withdrawal phase of the study. In the text we have written more clearly what we feel is the adequate description of this data [Page 8].

Lines 185-186: "Overall, we saw that coffee influences cognitive processes, both impulsiveness and emotionality when compared with non-coffee drinkers" – this is vague and needs to contain a more focussed description of main results.
We agree with the Reviewer that this is vague and have added additional context to this section of the paper.

Lines 187-189, figure 4 does not show 'reintroduction of caffeinated coffee, and decaffeinated coffee had distinct and specific effects on several cognitive and physical measures'. It shows associations between microbial species, metabolites, and tasks.
We have re-written this section to provide a clearer explanation of the data presented. [Page 8 paragraph 3.]

Discussion
The authors present a significant amount of data and outcomes (cognitive function, stress levels, inflammation, faecal metabolites, changes in the gut microbiome etc), but it remains unclear how this work has moved beyond the state-of-the-art, and how this study has delivered new insights. Many outcomes in relation to coffee consumption have been reported before, and are not new.
We appreciate the reviewers' comments and accept that we may not have focussed our discussion enough on the novel findings of our study.
Although prior studies have examined associations between coffee consumption and behavioural or physiological outcomes, our study advances the field by employing a longitudinal, interventional, and multi-omics approach

that allows for causal inference, personalized insight, and mechanistic understanding. Unlike previous cross-sectional or observational research, we uniquely integrate:

1. **Controlled coffee withdrawal and blinded reintroduction (caffeinated vs. decaf) over 35 days,**
2. **Multi-omics integration —including gut microbiota, urinary and faecal metabolomics,**
3. **Cognitive, psychological, and HPA-axis metrics, all tracked within subjects.**

This comprehensive design enables the disaggregation of caffeine-specific effects from those of other coffee components (e.g., polyphenols), revealing previously unrecognized roles for decaffeinated coffee in enhancing memory, sleep, and emotional regulation. Moreover, by linking specific microbial strains and metabolites to behavioural and cognitive outcomes, this study uncovers a microbiota–metabolite–brain axis modulated by coffee, which has not been demonstrated in prior work. In addition, when compared with a recent paper that analysed over 22,000 metagenomes showing the consistent presence of *L. asaccharolyticus* in coffee drinkers, our study design has examined causality, an extremely important cornerstone of current microbiome research. These findings suggest novel mechanisms through which dietary bio-actives influence cognition and mood, laying groundwork for personalized nutritional strategies based on individual microbiome and genetic profiles.
We have rewritten the introduction and discussion to amplify the novelty of our experimental approach and findings

Reviewer #2 (Remarks to the Author):
This is an original, and novel investigation by Boscaini et al, titled “Habitual Coffee Intake Shapes the Gut Microbiome and Modifies Host Physiology and Cognition”, with the purpose of probing coffee and caffeine’s effects on the microbiota-gut-brain axis in healthy females and males using a sophisticated multi-omics approach. The study directly compared thirty-one non-coffee drinkers (NCD) with thirty-one coffee drinkers (CD) using a cross-sectional comparison. This comparison was followed by a two-week coffee abstinence wash-out period in the CD’s only, at which time they were block randomized into either a caffeinated or decaffeinated coffee consumption group in a double-blinded, parallel design for an additional three weeks. All participants underwent comprehensive collection for biological blood samples, urinary and faecal metabolomics, faecal gut microbiome, cognitive, impulsivity, emotion reactivity, urgency, memory, as well as physiological and psychological stress testing, and dietary intake. The authors conclude that coffee drinkers exhibit higher impulsivity and emotional reactivity, improved immune function, reduced neuroactive compounds, and increased alpha diversity of gut microbiome compared to non-coffee drinkers. Further, attention and vigilance increased whereas impulsivity and emotional reactivity decreased upon withdrawal of caffeine. Coffee reintroduction following abstinence, whether caffeinated or decaffeinated, decreased perceived stress and impulsivity and self-reported depression. Only reintroduction of caffeinated coffee reduced anxiety. The authors also conclude periods of coffee withdrawal impact mood, immune markers, and microbial diversity due to caffeine and/or other coffee containing compounds. The study uses cutting-edge methodologies to explore unique impacts of coffee on physiological, psychological, immunological, and microbial composition, which is novel and of great value. However, the basis of the comparison between NCD and CD participants has

significant limitations and this influences the interpretation of all the different analyses.

We thank the Reviewer for their considered and very constructive comments and for highlighting the novelty and cutting-edge methodologies used. We address their concerns below.

Major Comment:

The major limitation with the study is the lack of appropriate control of caffeine prior to the baseline testing day. This compromises the integrity of the remaining analyses performed in all outcome measures. Specifically, the NCD's were consuming almost 70mg of caffeine/day prior to baseline testing and abstained from nearly all of this prior to the baseline visit. This suggests they experienced a drastic change in their caffeine intake during baseline were essentially in caffeine withdrawal. Whereas, CD's were consuming 4x this amount on a daily basis and only abstained 1/10th (28mg) of their usual intake. Thus, the baseline data is comparing caffeine withdrawal in NCD's and caffeine habituation in CD's. The authors need to explain this within the context of their overall findings prior to any further consideration.

We thank the Reviewer for this important observation. We agree that caffeine withdrawal may have occurred in the non-coffee drinkers (NCD) group at baseline, given the near-complete abstinence from their usual low levels of caffeine intake (~70 mg/day). This is a valid concern and indeed, we considered this possibility when designing the broader study phases. However, we would like to clarify several points:

1. **Caffeine Intake in NCDs was minimal: While the average intake was ~70 mg/day (approximately one standard cup of tea), this value reflects occasional consumption of non-coffee caffeinated products (e.g., tea, chocolate, or soda). The absolute level of habitual intake in NCDs was therefore low and unlikely to generate strong physiological dependence, more than 100mg a day has been usually considered the lower bound threshold for withdrawal symptoms to occur,** (Juliano and Griffiths 2004).
2. **Withdrawal Symptoms Were Monitored: During the baseline assessment, we specifically monitored withdrawal symptoms (e.g., drowsiness, headache, low mood, fatigue) using validated questionnaires. At baseline (Visit 2), NCDs did not report elevated symptoms compared to CDs (Supplementary Fig. 9), and no significant increases were observed in fatigue or irritability. This suggests that, while withdrawal cannot be fully excluded, its impact was likely mild and not sufficient to explain between-group differences in cognition or mood.**
3. **Contrast With CD Habituation Effects: We agree that CDs were in a state of caffeine habituation at baseline. Importantly, the design of the study includes a subsequent washout and reintroduction phase, allowing us to evaluate within-subject changes in this group. The observed reduction in impulsivity and emotional reactivity after coffee withdrawal in CDs, as well as the re-emergence of certain effects upon reintroduction of both caffeinated and decaffeinated coffee, provide crucial context and reduce the reliance on the baseline comparison alone.**
4. **Analyses Not Solely Based on Baseline Comparison: While the baseline comparison between NCD and CD provides one axis of interpretation, our main conclusions are derived from the longitudinal within-subject changes observed in the CD group. This includes effects after abstinence and following randomised reintroduction of caffeinated vs. decaffeinated coffee. These phases were tightly controlled, double-blinded, and enable**

clearer causal inference about the role of coffee and caffeine on microbiome, mood, and cognition.

5. **<u>Addressed in Manuscript Text:</u> Considering the Reviewer's comment, we have now added a more detailed explanation of this limitation and its implications in the Discussion section [Discussion, Page 16, paragraph 1], acknowledging that caffeine withdrawal in NCDs at baseline may have influenced group differences, while highlighting the strength of the crossover design in mitigating this issue.**

Minor Comments:
line 24 this assumes the gut aspects are different from the gut microbiome aspects?
We agree with the reviewer that this sentence lacked clarity, and we have amended the sentence to provide clarification.

44 the introduction is focused primarily on coffee, independent of any mention of caffeine. Given the age-related pharmokinetics, metabolic, and cardiovascular effects of caffeine, how might this affect the outcomes? This should be discussed more thoroughly in the introduction for perspective. Further, it appears the authors transition, unintentionally between coffee and caffeine throughout this section. Is the emphasis coffee or is it caffeine?
We are grateful for the Reviewer's comments on the Introduction and have redrafted it to delineate the specific roles off coffee and caffeine. The primary focus of the paper is coffee, while not forgetting about the role caffeine itself may have.

66-72 This appears to both address coffee's potential as a prebiotic by different mechanisms. The authors should state these differences, as such.
We have amended this sentence to make it clearer.
Coffee consumption may increase serum SCFA levels by delivering fibre-like compounds such as melanoidins to the gut, which act as prebiotics and stimulate the growth of SCFA-producing bacteria.

87 Was heart rate, blood pressure, fasting blood glucose, body weight/composition, etc. measured to determine physical wellbeing?
At each study visit, heart rate, blood pressure, body weight / composition was measured (Supplementary Tables 3, 4, 5). Fasting blood glucose was not measured at any study visit, anyone with diabetes would have not been admitted to the study due to the exclusion criteria defined in the methods section. As detailed in Figure S17, there were no participant withdrawal following the randomisation, all withdrawals were before this point. Exclusions were due to no longer meeting inclusion/exclusion such as going on an antibiotic or withdrawal for personal reasons such as they no longer had time to participate. Just to clarify none were withdrawn for researcher-led reasons. There were no exclusions due to adverse events during the intervention.

94 This seems like a major limitation with the study. NCD's were consuming almost 70mg of caffeine/day and abstained from nearly all of this prior to baseline visit. This suggests they experienced a drastic change in their caffeine intake during baseline. Whereas, CD's were consuming 4x this amount on a daily basis and only abstained 1/10th (28mg) of their usual intake. Thus, the baseline data is comparing caffeine withdrawal in NCD's and caffeine habituation in CD's. The authors need to explain this within the context of their overall findings prior to any further consideration.

Please see our response to this point on page 20 of this document.

121 interesting that cognitive performance improved following coffee washout period. If it was due to task repetition, performance would be improved following the coffee intervention in both groups. Was it?
We have amended this specific section of the results to clarify the fact that task repetition was not solely responsible for the differences seen in the results of the PASAT [Page 7].

144 NCD had improved sleep and physical activity. This seems like a significant difference and benefit.
We have amended this sentence to better explain the significance of the findings.
The STAI total score decreased in caffeinated drinkers, while HSCL, sleep quality (PSQI), and physical activity (IPAQ) scores improved in decaffeinated drinkers (Supplementary Table 6) suggesting that caffeinated and decaffeinated coffee may have distinct psychological and physiological benefits, potentially influencing recommendations for their consumption based on individual health goals.

148 Here again, NCD improved memory compared to CD, which is notable. If it was a practice effect why did only NCD improve and not CD?
We appreciate the reviewers' comments on this matter and have revised our explanation accordingly.
The decaffeinated group demonstrated significant improvements in several ModRey test components, whereas the caffeinated group did not. These findings may be attributable to the beneficial impact of improved sleep quality and increased physical activity, both of which are well-documented to support episodic memory performance (Berres & Erdfelder, 2021; Moore & Loprinzi, 2021)

185 There seems to be little doubt that NCD performs better on most important health outcomes than CD. This should be explained further.
We agree with the reviewers point and have amended this section of the results and will also discuss these results more in the Discussion.

312 what would cause the increase in velonella parvula in CD similar to increased levels in NCD?
341 what would account for the increased velonella in NCD then?
The increase in Velonella parvula in CDs to NCD-like levels may be due to a few factors. Firstly, it is possibly due to disruption of coffee-associated microbiota during the washout phase or the removal or reduction of competing microbes. It could be a delayed or indirect response to the reintroduction of coffee (particularly non-caffeine compounds).

Coffee may be utilising coffee derived compounds, Velonella species, including V. parvula, are known to metabolize lactate and short-chain organic acids, Coffee (even decaffeinated) contains polyphenols (e.g. chlorogenic acids) dietary fibre and fermentable oligosaccharides. These compounds can be metabolized by other gut microbes into lactate, which Velonella then uses as an energy source. So, coffee may indirectly support V. parvula by boosting lactate-producing bacteria (Scheiman et al. 2019)**.**

Overall, while Velonella is abundant in the oral cavity and the upper gastrointestinal tract, it of relatively low abundance in the faecal microbiome (The Human Microbiome Project Consortium 2012; MetaHIT Consortium et al. 2010)**. Thus, the importance of Velonella parvula in the context of health and wellbeing may not be of consequence, though it is responsive to the presence and reintroduction of coffee.**

342 delete either "associated" or "correlated".
We have amended this sentence.

361 How do you explain CD having higher impulsivity and emotional reactivity, yet greater stress tolerance and anti-inflammatory responses?
This is a good question, and one we have considered at length. Firstly, impulsivity and emotional reactivity (as measured by UPPS-P and ERS) are more stable personality traits. Stress tolerance and inflammation, however, are often state-dependent, influenced by acute physiological factors like caffeine's effects on neurotransmitters, hormones, and immune signalling.

CDs had distinct gut microbiota profiles, including species like Velonella that may support neurotransmitter synthesis, energy metabolism, and immune modulation. Changes in gut-derived metabolites were associated with improved cognition and mood. These microbiota shifts might play a compensatory role, offsetting trait vulnerabilities (like impulsivity) by enhancing resilience and anti-inflammatory responses.

Regular caffeine consumption leads to physiological adaptation, e.g. desensitization of adenosine receptors and tolerance to stress-related hormonal spikes (like cortisol). CDs may have adapted to regular arousal, meaning they do not overreact to stressors, even if their baseline trait profile (e.g. impulsivity) is higher.

406 What explains the lower GABA in CD, while reduced anxiety? It would suggest these should respond in opposition to each other.
We agree with the reviewer that at first this may seem at odds with the expected relationship between GABA and anxiety. There are several possibilities that may explain this, while fully acknowledging that our data is not conclusive in this regard.
Firstly, lower GABA in faeces may indicate greater uptake of GABA in the gut rather than lower production. This could mean that coffee drinking enhances the absorption of GABA into the bloodstream or increases its utilization by enterocytes or neurons. GABA produced in the gut can act on the enteric nervous system or be transmitted via the gut-brain axis to affect brain function. Secondly, Peripheral GABA levels may not directly reflect central GABA activity. Finally, coffee consumption might shift microbial populations to favour those that either: use GABA more efficiently or suppress overproduction, avoiding excess GABA being excreted (Belelli et al. 2025)**.**

410 This sentence is unclear. Please reword.
We have rewritten this sentence to improve its clarity.
"Targeted metabolomics of caffeine-related compounds revealed a clear link between coffee consumption and metabolic patterns, showing decreased

levels of 3-hydroxybenzoic acid in the NCD group, while 4-hydroxybenzoic acid levels increased. Notably, abstinence from coffee and its reintroduction had a marked impact on these metabolite concentrations."

424 How is it that coffee abstinence does not influence diversity, yet species level changes occur after abstinence? Please explain.
We appreciate the reviewer's insightful question. Firstly, we believe that diversity quantifies how many different species are present and how evenly they are distributed while species levels refer to which microbial species are more, or less abundant. Thus, diversity can remain stable, while the proportion of certain species can change, indeed species that disappear may be replaced by other species. It is also possible that coffee, which contains polyphenols, antioxidants and caffeine, may act as a prebiotic that are still present following coffee abstinence (Mena and Bresciani 2020; "Prebiotic Potential of Coffee and Coffee By-Products" 2025)**. Thus, species that rely on coffee or it's metabolites may disappear while species that do not will replace them.**

427 Is this Day 2 of the Intervention phase? Please specify
That is correct, it is Day 2 of the intervention or Day 16 overall, we have amended the sentence to provide more clarity.

436 It's assumed the authors are referring to Cryptobacterium curtum here, correct?
We thank the Reviewer for their comments. Unfortunately, there is a mismatch between the line numbers referred to and the manuscript, we hope we have addressed all the concerns accordingly.

438 The authors should consider organizing the Discussion similar to how it's presented in the figure.
We have amended the discussion to reflect the reviewer's suggestion.

441 This is a very long sentence and rather confusing. Please rephrase.
We have amended this sentence to make it clearer.
Targeted metabolomics of caffeine-related compounds revealed a clear link between coffee consumption and metabolic patterns, showing decreased levels of 3-hydroxybenzoic acid in the NCD group, while 4-hydroxybenzoic acid levels increased. Notably, abstinence from coffee and its reintroduction had a marked impact on these metabolite concentrations.

446 This statement appears redundant as the authors state caffeine metabolites change which are associated with microbial species and cognitive/behavioral outcomes in lines 441-445..
We have removed this sentence to avoid repetition.

461 The authors state their data agrees with the hypothesis that the gut microbiome may influence the variability in phenolic metabolism. Yet, they later state this connection is far more connected with the urinary metabolome than the gut microbiome. This discrepancy needs addressing.
We agree with the Reviewer that this is confusing and have amended the relevant text to reflect this.

1109 While groups differed in total caffeine intake, the NCD group still consumed the

equivalent of one cup of coffee per day. Thus, they were not caffeine naive. How does this impact the results and conclusions?
We appreciate the Reviewer's response and have responded to this point earlier in the rebuttal.

1110 Given pharmokinetic differences in caffeine matabolism between ethnicities and sexes and the fact there were 10 times more "other" ethnicities between groups and more women, how might this influence the outcomes?
We agree with the reviewer, that it is important to consider ethnic origin and sex as potential confounders in our analysis. Sex was indeed used as a confounder in all applicable analysis [Page 30, Statistical Analysis], we believe that this would account for any potential gender specific pharmacokinetic differences in our data. Regarding ethnicities, we acknowledge that not considering it as a confounder may be a limitation, but of the 10 ethnicities in the NCD group, there are 5 different ethnic groups represented, thus, we feel that we would not have the statistical power to derive any meaningful data. We will also include this as a limitation in the study. Indeed, there are also many cultural nuances involved in the drinking of coffee across different cultures that we have also not considered but should be considered for further work in this area.

References for Rebuttal

Arnarson, ÞórÐur Örn, Daníel Þór Ólason, Jakob Smári, and Jón FriÐrik SigurÐsson. 2008. "The Beck Depression Inventory Second Edition (BDI-II): Psychometric Properties in Icelandic Student and Patient Populations." *Nordic Journal of Psychiatry* 62 (5): 360–65. https://doi.org/10.1080/08039480801962681.

Belelli, Delia, Jeremy J Lambert, Murphy Lam Yim Wan, Ana Rita Monteiro, David J Nutt, and Jerome D Swinny. 2025. "From Bugs to Brain: Unravelling the GABA Signalling Networks in the Brain–Gut–Microbiome Axis." *Brain* 148 (5): 1479–506. https://doi.org/10.1093/brain/awae413.

Bengtsson, Mariette, Oskar Hammar, Thomas Mandl, and Bodil Ohlsson. 2011. "Evaluation of Gastrointestinal Symptoms in Different Patient Groups Using the Visual Analogue Scale for Irritable Bowel Syndrome (VAS-IBS)." *BMC Gastroenterology* 11 (1). https://doi.org/10.1186/1471-230x-11-122.

Booth, Michael. 2000. "Assessment of Physical Activity: An International Perspective." *Research Quarterly for Exercise and Sport* 71 (sup2): 114–20. https://doi.org/10.1080/02701367.2000.11082794.

Buysse, Daniel J., Charles F. Reynolds, Timothy H. Monk, Susan R. Berman, and David J. Kupfer. 1989. "The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research." *Psychiatry Research* 28 (2): 193–213. https://doi.org/10.1016/0165-1781(89)90047-4.

Cade, Janet, Rachel Thompson, Victoria Burley, and Daniel Warm. 2002. "Development, Validation and Utilisation of Food-Frequency Questionnaires – a Review." *Public Health Nutrition* 5 (4): 567–87. https://doi.org/10.1079/phn2001318.

Cohen, Sheldon, Tom Kamarck, and Robin Mermelstein. 1983. "A Global Measure of Perceived Stress." *Journal of Health and Social Behavior* 24 (4): 385. https://doi.org/10.2307/2136404.

Cotillard, Aurélie, Agnès Cartier-Meheust, Nicole S Litwin, et al. 2022. "A Posteriori Dietary Patterns Better Explain Variations of the Gut Microbiome than Individual Markers in the American Gut Project." *The American Journal of Clinical Nutrition* 115 (2): 432–43. https://doi.org/10.1093/ajcn/nqab332.

Cyders, Melissa A., Andrew K. Littlefield, Scott Coffey, and Kenny A. Karyadi. 2014. "Examination of a Short English Version of the UPPS-P Impulsive Behavior Scale." *Addictive Behaviors* 39 (9): 1372–76. https://doi.org/10.1016/j.addbeh.2014.02.013.

Degen, L P, and S F Phillips. 1996. "How Well Does Stool Form Reflect Colonic Transit?" *Gut* 39 (1): 109–13. https://doi.org/10.1136/gut.39.1.109.

Ferreira, R., and J. Murray. 1983. “Spielberger’s State-Trait Anxiety Inventory: Measuring Anxiety with and without an Audience during Performance on a Stabilometer.” *Perceptual and Motor Skills* 57 (1): 15–18. https://doi.org/10.2466/pms.1983.57.1.15.

Grubbs, Frank E. 1950. “Sample Criteria for Testing Outlying Observations.” *The Annals of Mathematical Statistics* 21 (1): 27–58. https://doi.org/10.1214/aoms/1177729885.

Heaton, K W, J Radvan, H Cripps, R A Mountford, F E Braddon, and A O Hughes. 1992. “Defecation Frequency and Timing, and Stool Form in the General Population: A Prospective Study.” *Gut* 33 (6): 818–24. https://doi.org/10.1136/gut.33.6.818.

Juliano, Laura M., and Roland R. Griffiths. 2004. “A Critical Review of Caffeine Withdrawal: Empirical Validation of Symptoms and Signs, Incidence, Severity, and Associated Features.” *Psychopharmacology* 176 (1): 1–29. https://doi.org/10.1007/s00213-004-2000-x.

Koo, Terry K., and Mae Y. Li. 2016. “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.” *Journal of Chiropractic Medicine* 15 (2): 155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

Lewis, S. J., and K. W. Heaton. 1997. “Stool Form Scale as a Useful Guide to Intestinal Transit Time.” *Scandinavian Journal of Gastroenterology* 32 (9): 920–24. https://doi.org/10.3109/00365529709011203.

Linseisen, J, A A Welch, M Ocké, et al. 2009. “Dietary Fat Intake in the European Prospective Investigation into Cancer and Nutrition: Results from the 24-h Dietary Recalls.” *European Journal of Clinical Nutrition* 63 (S4): S61–80. https://doi.org/10.1038/ejcn.2009.75.

Lipman, Ronald S., Lino Covi, and Arthur K. Shapiro. 1979. “The Hopkins Symptom Checklist (HSCL).” *Journal of Affective Disorders* 1 (1): 9–24. https://doi.org/10.1016/0165-0327(79)90021-1.

Longstreth, George F., W. Grant Thompson, William D. Chey, Lesley A. Houghton, Fermin Mearin, and Robin C. Spiller. 2006. “Functional Bowel Disorders.” *Gastroenterology* 130 (5): 1480–91. https://doi.org/10.1053/j.gastro.2005.11.061.

Manghi, Paolo, Amrisha Bhosle, Kai Wang, et al. 2024. “Coffee Consumption Is Associated with Intestinal Lawsonibacter Asaccharolyticus Abundance and Prevalence across Multiple Cohorts.” *Nature Microbiology* 9 (12): 3120–34. https://doi.org/10.1038/s41564-024-01858-9.

Mena, Pedro, and Letizia Bresciani. 2020. “Dietary Fibre Modifies Gut Microbiota: What’s the Role of (Poly)Phenols?” *International Journal of Food Sciences and Nutrition* 71 (7): 783–84. https://doi.org/10.1080/09637486.2020.1826913.

MetaHIT Consortium, Junjie Qin, Ruiqiang Li, et al. 2010. “A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing.” *Nature* 464 (7285): 59–65. https://doi.org/10.1038/nature08821.

Nock, Matthew K., Michelle M. Wedig, Elizabeth B. Holmberg, and Jill M. Hooley. 2008. “The Emotion Reactivity Scale: Development, Evaluation, and Relation to Self-Injurious Thoughts and Behaviors.” *Behavior Therapy* 39 (2): 107–16. https://doi.org/10.1016/j.beth.2007.05.005.

O’Donnell, L J, J Virjee, and K W Heaton. 1990. “Detection of Pseudodiarrhoea by Simple Clinical Assessment of Intestinal Transit Rate.” *BMJ* 300 (6722): 439–40. https://doi.org/10.1136/bmj.300.6722.439.

Park, Yikyung, Kevin W Dodd, Victor Kipnis, et al. 2018. “Comparison of Self-Reported Dietary Intakes from the Automated Self-Administered 24-h Recall, 4-d Food Records, and Food-Frequency Questionnaires against Recovery Biomarkers.” *The American Journal of Clinical Nutrition* 107 (1): 80–93. https://doi.org/10.1093/ajcn/nqx002.

“Prebiotic Potential of Coffee and Coffee By-Products.” 2025. In *Coffee in Health and Disease Prevention*, by Marlene Machado, M. Beatriz P.P. Oliveira, Helena Ferreira, and Rita C. Alves. Elsevier. https://doi.org/10.1016/b978-0-443-13868-3.00075-2.

“Reproducibility and Validity of Food Frequency Questionnaires.” 2012. In *Nutritional Epidemiology*, by Walter Willett and Elizabeth Lenart. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199754038.003.0006.

Rétey, J V, M Adam, R Khatami, et al. 2007. “A Genetic Variation in the Adenosine A2A Receptor Gene (ADORA2A) Contributes to Individual Sensitivity to Caffeine Effects on Sleep.” *Clinical Pharmacology & Therapeutics* 81 (5): 692–98. https://doi.org/10.1038/sj.clpt.6100102.

Saad, Richard J, Satish Sc Rao, Kenneth L Koch, et al. 2010. “Do Stool Form and Frequency Correlate With Whole-Gut and Colonic Transit? Results From a Multicenter Study in

Constipated Individuals and Healthy Controls." *American Journal of Gastroenterology* 105 (2): 403–11. https://doi.org/10.1038/ajg.2009.612.
Salvini, Simonetta, David J Hunter, Laura Sampson, et al. 1989. "Food-Based Validation of a Dietary Questionnaire: The Effects of Week-to-Week Variation in Food Consumption." *International Journal of Epidemiology* 18 (4): 858–67. https://doi.org/10.1093/ije/18.4.858.
Scheiman, Jonathan, Jacob M. Luber, Theodore A. Chavkin, et al. 2019. "Meta-Omics Analysis of Elite Athletes Identifies a Performance-Enhancing Microbe That Functions via Lactate Metabolism." *Nature Medicine* 25 (7): 1104–9. https://doi.org/10.1038/s41591-019-0485-4.
The Human Microbiome Project Consortium. 2012. "Structure, Function and Diversity of the Healthy Human Microbiome." *Nature* 486 (7402): 207–14. https://doi.org/10.1038/nature11234.