Replies to the editor and reviewers

Dear Dr. Muñoz-Tamayo,

We sincerely appreciate the Recommender and three Reviewers for their thorough evaluation of our manuscript. We have carefully considered and addressed each of their comments, which we believe have significantly enhanced the clarity and quality of our work.
Regarding the sharing of the data, we increased the level of detail and now provide the relative abundances of all the taxa detected in the 16S data, in addition to the CSTs used in the model (which was already shared).

Please find below our responses to individual comments.

Tsukushi Kamiya and co-authors

Recommender

Your manuscript has been evaluated by three reviewers. The reviewers and I agree on the high value of your contribution. Two reviewers provided constructive comments to improve the clarity of the manuscript. I invite you to revise your article and address the remarks of the reviewers:
Please also consider my following remarks:

Q 0.1 *To comply with PCI requirements, please provide the microbiota data and the scripts for taxonomic assignment and for the determination of the community state type*

Reply: The scripts used for the taxonomic assignment and for the community state type have been published and are cited in the manuscript (SpeciateIT and VALENCIA, respectively). Regarding the data sharing, we now provide the detailed bacteria composition in the form of operational taxonomic unit for each sample. However, we currently do not have the right to share the raw files due to the clinical nature and associated data-sharing restrictions.

Q 0.2 *L276-277: please adjust the precision of the percentages of the CST to have the sum = 100*

Reply: Thank you for your suggestion. However, rounding errors are unavoidable in this context. We appreciate your understanding of this minor limitation.

Q 0.3 *Figure 1 and manuscript: the labelling I(II,V) might be misleading. Please mention explicitly in the manuscript that I(II,V) refers to the pool of optimal communities I, II, V*

Reply: We added the precision (i.e., CSTs I, II, and V combined) in the caption of Figure 2 (i.e., Figure 1 in the original submission).

Q 0.4 *L740-742: I think these findings should be placed as well in the results section as indication of the robustness of the modelling and identification approach*

Reply: The figure the recommender refers to concerns posterior contraction and posterior-z score, which collectively assess the accuracy, precision and identifiability of estimated parameters. We agree that this information is crucial in demonstrating the rigour of our findings. However, we believe this information is better situated in the Supplementary Information because a) it requires the rather technical context provided in Supplementary Information S2, b) such technicality of Bayesian inference reduces the accessibility of our work to biologists and epidemiologists, who are our core target audience, and c) methodologically, this is not the unique contribution that we would like to showcase.

Reviewer 1

This is a beautiful paper—both well-written and methodologically sound. I particularly like the hierarchical Bayesian Markov model, which seems to be working very well. I appreciate the authors for their detailed review and comparison with the existing literature, as well as for sharing the code to ensure reproducibility.

Reply: Thank you for your kind words.

Q 1.1 *I have no criticisms, but I do have a quick question: Was each MCMC chain run on a single CPU, and how long was the runtime?*

Reply: As seen in the Stan code provided, the model was parallelised per participant and MCMC fitting was carried out on a 48-core machine: the runtime was approximately 5 hours.

Q 1.2 *Additionally, do you know the scaling law for how runtime increases with the number of covariates?*

Reply: We did not carry out such exploration. However, the answer is likely contingent upon, not only the number of covariates, but the strength of information, both in the prior and data.

Reviewer 2

The paper "Factors shaping vaginal microbiota long-term community dynamics in young adult women" by Kamiya et al. analyses a new dataset of time-series of vaginal microbial community state types (CSTs) of 125 young women from the PAPCLEAR cohort. Analyses were carried out with a hierarchical bayesian model, which allowed to quantify the impact of 16 co-variables on the CST dynamics and inter-individual variability. The paper is well written and easy to follow. The study identified alcohol consumption as an important driver of transitions from optimal to non-optimal CST. It also showed up that dynamics towards non-optimal states are less variable than recovery, which is a very interesting pattern from a microbial ecology point of view. The analyzes are rigorously conducted and the results are correctly grounded by the analyzes. I think that this study is without any doubt of interest for the community of microbial ecologists. My main concern about this study is that it is built upon CSTs (i.e. the break down of complex metabarcoding data into 3 discrete types only) without introducing the CST computation method, nor discussing it (see below for extensive comments on this topic).

Major issues:

Q 2.1 *Description and discussion of the CST construction method should be added. In the whole paper, CST are described and introduced as natural concepts well accepted by the community of researchers interested in vaginal*

microbiota, and CST computation is presented as a state-of-the-art black box. I however think that CST computation deserves at least a small introduction and discussion. Please find hereafter suggestions to illustrate my point.
1.a: CST computation is a hard clustering method that may bring potential drawback on Markov model interpretation. The original paper about VALENCIA, the CST computation software, indicates that CSTs result from a nearest-centroid clustering method. An issue with this type of hard clustering method is that CST changes, that are always interpreted in the markov model as large ecological shifts, could only be the result of a small microbial variation across the boundary between two CSTs, if the community state is close enough to the boundary. Such a CST change would have small ecological significance. Is there a way to take this into account in a Markov model ? (e.g. by modeling the evolution of the distance to the centroid rather than switches between discrete class). Could it be discussed?
1.b: hard clustering may introduce 'false' inter-individual variability. You mentioned two variability sources: the lack of covariable or time-varying parameters. But a potentially important source of variability could come from the CST clustering method, that hide inter-individual microbial variability. The initial distance of the community to the class centroid could be a reason why an individual (falsely) appears more sensitive to a covariable: if initially close to a boundary, a small effect will induce a label change but if initially far from the boundary, a strong effect will be needed for the same CST switch.
1.c: discussion about hard vs soft clustering. In the gut microbiota community, the state-of-the-art discrete community types, the enterotypes, have been recently challenged by alternative concepts, such as enterobranches (10.1038/s41467-023-38558-7) or enterosignatures (10.1016/j.chom.2023.05.024), that provide more continuous descriptions of community changes. Enterosignatures are based on soft clustering methods. Is this kind of concept of interest for vaginal microbial ecology?

Reply: Upon reading the comments, we realised that our introduction of the CSTs was indeed too hasty, especially for a broad community. To address this, we added a paragraph expanding on the context of the discrete categorisation of vaginal microbiota including its drawbacks (L35-45).

We now also include a new figure (new Figure 1), illustrating the relative abundance of the top 15 operational taxonomic units in all of our samples. The dominance of a relatively small number of taxa (compared to the gut counterpart) is a hallmark of vaginal microbiota communities and we believe that this new figure highlights the striking concordance between a handful of dominant taxa, CSTs, and microbial diversity profiles. We also now contrast this high degree of clustering with the difficulty of defining clear communities, or 'enterotypes', in the gut microbiota.

As illustrated by our new figure, some CST assignments are indeed more closely aligned with the dominance of a specific taxon than others. However, we are unsure that this would affect our inference. We expect these assignments to average out over the entire dataset of over 2000 samples (e.g., a sample in between CST III and CST IV would sometimes be categorised as the former, sometimes as the latter). Regarding the suggestion to model the dynamics of a multi-dimensional quantitative trait, i.e., the distance to three (or more) CSTs, this led to many discussions in the team. Indeed, we agree that such an approach would represent meaningful progress in improving the precision of our understanding of the vaginal microbial community. However, given the state-of-the-art, we believe that it would be a significant methodological leap, beyond the scope of our present study. We now discuss the potential merits and possibility of going beyond CSTs (L491-509).
Beyond the scope of our study, a recent possibility to more accurately define the community composition is to use metagenomics to define what Holm et al. (2023) define as metagenomic community state types (mgCSTs). They show that instead of the 5 main CST, or 12 sub-CST, it is possible to distinguish between 27 MgCSTs. We unfortunately do not have the funding to perform metagenomics sequencing on our 2,000 samples but we now cite this study in our article (L505-509).

Q 2.2 *I did not find access to the metabarcoding data (only CST time series are provided). I do not know if it is mandatory for PCI, but it would be for sure of interest for the community.*

Reply: We now provide the detailed taxa abundance resulting from SpeciateIT, in addition to the detailed CST inference by Valencia. As indicated above, for patient confidentiality protection, we are currently not allowed to share the raw data, these could be accessed but with a specific application.

Q 2.3 *The bayesian approach would gain to be better described.*
3.a - equation terms in eq. (3) should be explained. A rationale of this decomposition should be provided.
3.b - the choice of priors should be better explained in eq. (3)
3.c - what is the Q matrix? (line 231). Is it patient dependent ? (in such case, please be consistent: use Q_p notation.). Is it the average over p of $q_{p,i,j}$? (in such case, please explain).
3.d - In eq. (4) : what is $q_{i,j}$? I assume that it is actually $q_{p,i,j}$ and that the conservation equation is patient-dependent.
3.e - Could you provide somewhere the number of parameters, so that we could have an idea of the ratio between data and parameters?

Reply: We addressed all of your queries. More specifically:

3.a & b We added an explanation for the Cholesky decomposition and LKJ distribution parameterisation (L231-239).

3.c & d As defined in L249, the Q matrix consists of the state transition intensities. This is a formulation typical of the continuous-time Markov model and we now provide a reference to a relevant resource on the topic. We have also added the subscripts p where appropriate to emphasize that the quantities are participant-specific.

3.e In hierarchical models with partial pooling, it is not possible to define the exact number of parameters because each group's parameter is only partially independent. The model doesn't allocate fully separate parameters for each group, nor does it completely share a single parameter across all groups. Instead, individual group parameters are "shrunk" toward a shared group-level distribution, blending individual and shared information. With this caveat withstanding, we now provide the number of parameters and hyper-parameters (L258-259).

Q 2.4 *CST categories should be consistent all over the paper. CST IV-A, IV-B and IV-C appear in Figure 1 but are never introduced. They do not appear again after that. Is it necessary to introduce them? If you think so, please describe them.*

Reply: We now introduce the notion of sub-CSTs (L56-57).

Minor issues:

Q 2.5 *figure 1.b and 1.d : the outlier in figure 1.b (more than 35 months of follow-up) does not seem to appear in figure 1.d. Please indicate if (and why) you removed it.*

Reply: Thank you for noticing this. We have added in the caption: "For visualisation, data are truncated at 750 days for a single individual whose duration exceeds this threshold."

Q 2.6 *lines 287 to 290 : you are computing here predicted sojourn times. How do they compare to observed sojourn times?*

Reply: As the sampling intervals are uneven in our study, the observed "sojourn time" does not provide reliable information on how long a state persisted. The continuous-time Markov model alleviates this issue by incorporating the sampling intervals into the estimation of the sojourn time.

Q 2.7 *Could you please provide an example of interpretation of one distribution in figure 3 legend to facilitate the reading?*

Reply: We have added in the caption the following: "For example, alcohol consumption was estimated to favour CST III over CST I (II,V) at a credibility level of 97%."

Q 2.8 *line 384 to 393: could you indicate somewhere that you are dealing with figure 3 here?*

Reply: Thank you for the suggestion. The figure is now cited in the paragraph.

Q 2.9 *line 404 : "trend was less clear for individual CSTs". What is individual CST here? Optimal CST?*

Reply: We see that the phrasing here lacked clarity. We rephrased as "CST I (II,V) and CST IV, individually".

Q 2.10 *please double check line 431. I have the feeling that it should be "from optimal to non-optimal" here. Am I correct?*

Reply: Thank you for noticing this. The error is now corrected.

Q 2.11 *final line (line 486) : "development of preventive and therapeutic strategies to improve vaginal health". Few lines could be added before in the discussion to illustrate what could be the use of your approach in a clinical context, (or the work that remains to do between now and then).*

Reply: The phrasing "*the development of preventive and therapeutic strategy*" was perhaps too vague and ambitious. We have rephrased the sentence as "*Our work paves the way for an improved ecological understanding of microbial dynamics within the vaginal environment and indicates lifestyle alterations (such as reduced alcohol consumption) that may promote vaginal health.*" (L521-524):

Q 2.12 *line 747. 'for all but one covariate effect'. Which one?*

Reply: We realise that the original phrasing was imprecise and partially incorrect (as the lowest contraction was at 75%). We rephrased the sentence as follows: "*We found that the posterior distributions for covariate coefficients, β, contracted by 86% on average, and at least 75%, compared to the prior distribution...*"

Q 2.13 *figure S2. If you only interpret the absolute value of the z-score, why not display the absolute value on the graph?*

Reply: This is a fair point, but we followed the plotting convention outlined by Betancourt M. Towards a principled Bayesian workflow; 2020: https://betanalpha.github.io/assets/case_studies/principled_663bayesian_workflow.html. Although the direction of the Z-score is not a particular concern, it could indicate the direction of bias if any were present: bias was not detected in our study.

Reviewer 3

While I am not an expert in vaginal microbiomes and Bayesian approaches, I have read with great interest the manuscript entitled "Factors shaping vaginal microbiota long-term community dynamics in young adult women" by Kamiya and colleagues.
Using a hierarchical Bayesian Markov model, the authors elucidate the drivers of state transitions in the vaginal microbiome within a large cohort of 125 women, followed over a median of 8 months with an average of 11 samples per participant. This study is impressive in scope and innovative in its application of advanced computational techniques, providing critical insights into long-term vaginal microbiota dynamics and covariate associations.

I commend the authors for their comprehensive data collection and novel modeling approach. However, I have several comments and suggestions that could further enhance the study's clarity, interpretation, and overall impact.

Reply: Thank you for the detailed summary and the appreciation of our work.

Q 3.1 *The concept of microbiome "states" remains somewhat controversial, particularly about the assumption that they are discrete and stable over time. The study relies on VALENCIA for CST classification. Yet, recent work by Lebeer et al. (2023, Nature Microbiology) suggests that vaginal microbiome composition may be better represented as a continuum, especially when considering subdominant genera. In this study, the authors highlight the presence of transitional states between CSTs, which casts doubt on the binary stability of certain microbiota categories. This continuum-based framework could offer valuable insights into the dynamics the authors are modeling. Given your longitudinal data, I suggest leveraging this aspect to characterize state stability and resilience better. Can the authors explore whether tipping points between CSTs (e.g., transitions between dominant and subdominant taxa) exhibit bimodal distribution? This could help identify more nuanced community structures, as observed by Lebeer et al.*

Reply: As indicated in our reply to Reviewer #2, we agree that our introduction of the CSTs was too hasty. As the majority of our readers are likely to be non-experts of the vaginal microbiota, we now pay close attention to introducing the concept (L35-62) and highlighting the striking level of structuring observed in vaginal microbiota samples using our original data (i.e., new Figure 1).

Regarding the viewing of CSTs as a continuum, such a view was explored in the work by Gajer *et al.* (2012, *Science Transl Med*), in particular in their Figure 3. However, as indicated above, we are not aware of quantitative, multinomial approaches for modelling longitudinal vaginal microbiota data and we believe such a study would be a meaningful contribution to the field. We added this point to the Discussion in the revised manuscript (L491-509).

However, we believe it is inaccurate to state that the use of CST and the application of Markov models is contentious in the vaginal microbiota literature (at least to the extent seen in the gut microbiota literature). While we acknowledge that the gut microbiota exhibits higher dimensionality and hence the concept of "enterotypes" is contested by some, vaginal microbiota communities are more amenable to dimensionality reduction due to their relative simplicity (as highlighted by the new Figure 1).

For its relative simplicity, we argue that the vaginal microbiota is precisely an ideal system to circumvent this continuity issue (or the "curse of dimensionality" as generally referred to in statistics) which arguably has hindered the understanding of other microbiota. Indeed, this microbiota is highly and consistently structured. This is visible in Lebeer *et al.* with what they call 'modules' and a handful of taxa they discover associated with health outcomes and lifestyle covariates are the same constituent taxa of CSTs classified by VALENCIA. Thus, the approach Lebeer *et al.* is also an exercise in dimensionality reduction similar to CSTs. We use the CSTs for several reasons. The first reason is practicality: there exists an established algorithm to compute CSTs, which is not the case for the Lebeer et al. modules. Second, CSTs have been widely used for the last decade, and adopting this classification allows interpreting our results in light of existing literature. In fact, an advantage of the CSTs is that they are computed using a reference and performing clustering on our dataset would have internal relevance only. Third, CSTs are not arbitrary but derived from clustering approaches where the number of clusters is not assumed a priori but is estimated based on the data. In other words, in Ravel et al. (2011, *PNAS*) and the follow-up work by France et al. (2020, *Microbiome*), the structuring into five groups was informed by data. The addition of new samples confirmed this main structuring in five groups (or CSTs) with the ability to distinguish 12 sub-groups (or sub-CSTs) (France et al. 2020, *Microbiome*).

Finally, we are unsure about which type of 'bimodal distribution' the reviewer is referring to because CSTs are computed based on at least 5 distances, thus the axis of variation is not binary.

We now present data on taxonomic relative abundance to illustrate the high degree of structuring in vaginal microbiota communities (new Figure 1). Finally, we also explain that finer-grain classification is possible, yet at the expense of interpretability as is the case in any high-dimensional data (L491-509).

Q 3.2 *The transition dynamics between CSTs are well addressed. Yet, it would strengthen the manuscript to explore whether specific covariates correlate with CST transitions and shifts at the species or genus level. For instance, alcohol consumption has been implicated as a factor influencing shifts in microbiome composition. Still, it would be beneficial to see whether this and other covariates drive a gradual or critical transition in the community structure (i.e., whether there are tipping points in bacterial abundance). Can the authors show how these covariates relate to specific bacterial species or genera and determine if transitions occur along a continuum or through critical shifts? This analysis would align with classical numerical ecology approaches, often incorporating alpha and beta diversity metrics.*

Reply: As we now illustrate in the new Figure 1, there is a high degree of correspondence between species abundance and CST classification. We would indeed prefer to have a greater variety of 'states', and even tried initially to fit the model with 5 CSTs. Unfortunately, we do not have the statistical power to do so (as some states are rare), which is why we pool CSTs I, II, and V.
Regarding the idea of following diversity metrics, this is another dimensionality reduction approach applied to microbiota systems and we have now included the Shannon diversity index in the new Figure 1. However, this does not address the technical issue, which would be to model the dynamics of such a trait.

Q 3.3 *While the hierarchical Bayesian Markov model is novel and powerful, the manuscript needs more discussion of basic ecological diversity measures such as alpha and beta diversity. These indices would provide valuable context for comparing the observed vaginal microbiome dynamics to those of other studies, particularly to CST classifications. Including classical diversity measures (alpha, beta) and comparing them across CSTs could add depth to the ecological understanding of microbial dynamics. This would allow a clearer comparison with other longitudinal or cross-sectional studies on vaginal microbiota, such as those by Ravel et al. or Lebeer et al.*

Reply: We agree with Reviewer #3 that diversity measures provide a useful angle in inspecting bacterial communities. As such, the revised manuscript now includes a figure describing the detailed composition of the CSTs including the diversity metrics (new Figure 1).

Q 3.4 *The authors describe their hierarchical Bayesian Markov model as "novel." However, it would be beneficial for readers to understand how this approach is superior or different from more commonly used methods such as ALDeX2, ANCOM-BC, DESeq, limma, Maaslin2, or linear regression. What advantages does the hierarchical Bayesian framework offer, particularly in CST dynamics? How does it improve upon or complement these other methods? A brief comparison of the Bayesian Markov model to these standard microbiome analysis methods would clarify the advantages and justify the novelty of your approach. Specifically, what insights were gained through the Markovian assumptions about state transitions that alternative statistical methods might not capture?*

Reply: A key aspect of our data and focus of analysis is that it is longitudinal. The approaches developed for cross-sectional analyses mentioned by the Reviewer are not designed to incorporate the strong temporal auto-correlation inherent in longitudinal data. Markov models are one of the most widely used methods to address this issue. The key methodological innovation of our study is that we here estimate the effect of many covariates while estimating the transition probabilities in the Markov model, which would be challenging outside the Bayesian framework. Furthermore, we introduce random effects to account for individual differences and implement the whole model in a Bayesian setting.
Tools like ALDeX2 or ANCOM-BC typically do not include such a feature. The MaAsLin software package appears to be very strong but their handling of longitudinal data is perhaps suboptimal, as they seem to only include a random effect per individual to correct for this, thereby neglecting the temporal structure itself.

Overall, our methodological novelty is in relation to the literature on the continuous-time Markov model, which has previously been applied to study vaginal microbiota. We emphasize this point in L221-226. Thus, cross-sectional analysis methods are not the most appropriate point of comparison for our study.

Q 3.5 *The study by Lebeer et al. (2023, Nature Microbiology) presents a citizen-science-based approach to mapping the vaginal microbiome, emphasizing the importance of life course, lifestyle, and environmental factors. This work could provide a useful reference for expanding the discussion on covariates, such as hormonal contraceptive use and lifestyle factors, which were also shown to influence vaginal microbiota in their study. The authors of this manuscript could benefit from discussing their findings in the context of such large-scale studies, which consider the continuous nature of microbial shifts and community co-occurrence networks.*

Reply: We agree with the Reviewer that the citizen-science-based approach of Lebeer et al., holds promise for extending our work in the future and we now mention their approach in Discussion (L470-472).
Nonetheless, we note a fundamental difference in the perspective between the two studies. By estimating the effects of covariates from longitudinal data using a continuous-time Markov model, our study focuses on covariate effects on the process of transitions (e.g., what factors affect the chance of moving from an optimal to a sub-optimal state). In contrast, any statistical inference based on cross-sectional data would be agnostic on the process of community transitions. Instead, their perspective is inherently static, e.g., what factors influence the chance of “being" in a particular state. Thus, the dynamical perspective of longitudinal data offers a mechanistic insight grounded in microbial community ecology that is absent in cross-sectional studies.