Factors shaping vaginal microbiota community long-term dynamics

Tsukushi Kamiya[1,*], Nicolas Tessandier[1], Baptiste Elie[1,3], Claire Bernat[3,4], Vanina Boué[2], Sophie Grasset[2], Soraya Groc[2,3], Massilva Rahmoun[3], Christian Selinger[2,5], Michael S. Humphrys[6], Marine Bonneau[7], Christelle Graf[7], Vincent Foulongne[3], Jacques Reynes[8], Vincent Tribout[8], Michel Segondy[3], Nathalie Boulle[3], Jacques Ravel[6], Carmen Lía Murall[2,9], and Samuel Alizon[1,2,*]

[1]Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, Université PSL, Paris, France
[2]MIVEGEC, CNRS, IRD, Université de Montpellier, France
[3]Institut de Génomique Fonctionnelle, Université de Montpellier, CNRS, INSERM, Montpellier, France.
[4]PCCEI, Univ. Montpellier, Inserm, EFS, Montpellier, France
[5]Swiss Tropical and Public Health Institute, Basel, Switzerland
[6]Institute for Genomic Sciences, University of Baltimore, USA
[7]Department of Obstetrics and Gynaecology, Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[8]Department of Infectious and Tropical Diseases, Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[9]National Microbiology Laboratory (NML), Public Health Agency of Canada (PHAC), Canada
[*]Corresponding authors: tsukushi.kamiya@college-de-france.fr, samuel.alizon@college-de-france.fr

Abstract

The vaginal microbiota is structured into five main community state types (CST) that are known to affect women's health. CST shifts can occur in less than a day, but there is a lack of follow-ups lasting more than a few months; a gap in the knowledge that hampers our understanding of long-term dynamics. Analysing a longitudinal cohort of 125 women followed for a median duration of 10 months, we show that 'optimal' (CST I, II, and V) or 'sub-optimal' CST III are more stable in time than the 'non-optimal' CST IV. We also find that some probabilities of shifting from one CST to another are associated with covariates such reported number of sexual partners or alcohol consumption. Finally, we simulate population-level consequences of variations in behaviours. In addition to providing one of the first insights on vaginal microbiota dynamics over a year, along with a robust methodological analysis, these results open new perspectives to improve our mechanistic understanding of microbial interactions in the vaginal environment and develop new therapeutic strategies.

Introduction

Epithelia of the human body are host to diverse arrays of microorganisms, referred to collectively as microbiota, that are tightly associated with our health. This is particularly true for the vaginal microbiota because its composition is strongly linked to the acquisition risk of many sexually transmitted infections (STIs) [1], fertility, especially in the context of medically-assisted procreation procedures [2], and general well-being [3].

Over the last two decades, meta-barcoding sequencing of 16S DNA revealed that the variations in vaginal microbiota, which traces back to Albert Döderlein in 1892, can be explained through five main community state types, or CSTs [4]. Three of these are referred to as 'optimal' and are dominated by a lactobacillus species (*L. crispatus* for CST I, *L. gasseri* for CST II, and *L. jensenii* for CST III). A fourth one, CST III, can be seen as 'sub-optimal' because it is dominated by *L. iners*, which is metabolically distinct from other lactobacilli and can be associated with pathologies [5]. Finally, CST IV is characterised by a diverse assemblage of other anaerobic bacterial species from the *Gardnerella*, *Prevotella*, or *Fannyhessea* genera. This latter CST is the one usually associated with diseases such as bacterial vaginosis [6]. Metagenomics and transcriptomics techniques have allowed us to refine the compositional structure of these CSTs and to better understand their functioning [7], yet a dynamical perspective is still starting to emerge.

The vaginal microbiota composition evolves through life [8] and daily variations have been scrutinised over a couple of months, with pioneering work analysing daily samples

during 16 weeks in 32 women [9]. While longer follow-ups exist, they either specifically focus on pregnancy [10, 11], the samples are taken at large intervals (typically more than three months) [12], or sample sizes are modest (< 100) [13]. These circumstances limit the scope to generally understand the transient dynamics of vaginal microbiota in reproductive-aged women. Methodologically, many longitudinal studies aim to estimate the patterns of transitions and stability among distinct community types [10, 11, 14]. However, the lack of statistical power prevents these studies from examining how key risk factors modulate community transitions. Therefore, despite the inherently dynamic nature of vaginal microbiota communities, the current understanding of the impacts of key covariates — including demography, lifestyle factors, sexual practices, or medication — relies almost entirely on cross-sectional studies [15]. As such, there exists a knowledge gap in factors responsible for stabilising and destabilising vaginal microbiota communities.

We explore vaginal microbiota long-term dynamics through a unique collection of 2,103 samples, which provide us with a high-resolution longitudinal follow-up of 125 women over more than 10 months for each on average. We implement Bayesian Markov models to estimate transition probabilities between CSTs, simultaneously estimating associations between the transitions and relevant covariates. Our study offers a new insight into the stability and variability of the vaginal microbiota over a year and identifies important covariates that can explain the variation we witness in human populations.

Results

CSTs in the cohort

With 2,103 samples spanning over 1,323 total months of follow-up for 125 women, our longitudinal dataset from the PAPCLEAR cohort represents one of the longest analysed to date in the context of the vaginal microbiota (Figure 1). Study participants were between 18 and 25 years old at inclusion and additional characteristics are shown in Table 1. The majority of the samples (73.7%) were self-collected at home, the rest being collected upon on-site clinical visits, and the most common interval between analysed samples were seven and 14 days (Fig. 1a). The median follow-up duration was 8.64 (5.36, 14.0) months (Fig. 1b). On average, each participant contributed 11 samples (Fig. 1c).

The metabarcoding analysis on 16S RNA allowed us to assign each sample to a CST (see the Methods). These were variable across women and over time (Fig. 1d). CST I, II, and V are considered 'optimal' but the latter two are rare, so we pooled all three in further investigation and refer to them as 'CST I(II, V)'. Overall, optimal communities were the most frequent, representing 44.5% of samples, followed by 'sub-optimal' (CST III) at 35.2% and 'non-optimal' communities (CST IV) at 20.4% (Fig. 1e).

Probabilities of CST persistence

Given the longitudinal nature of our data, we used a continuous-time Markov model to model CST variations over time (see the Methods). Simulations based on the estimated

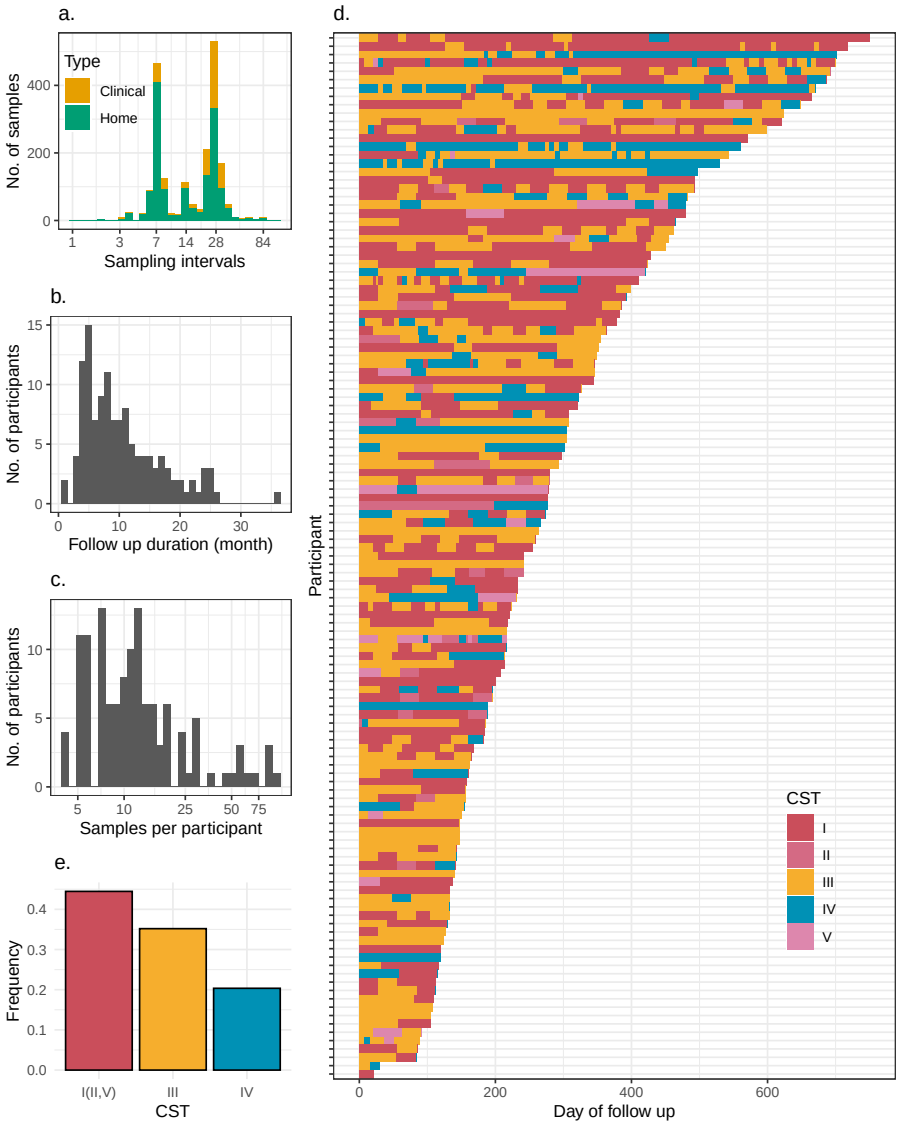

Figure 1: **Summary of vaginal microbiota samples analysed in the PAPCLEAR study.** a) Intervals between sampling events for on-site clinical and home samples. b) Follow-up duration per participant. c) Number of samples analysed per participant d) Vaginal microbiota Community State Types (CST) over time in 125 participants. e) Frequency of the optimal (CST I, II, and V), sub-optimal (CST III) and non-optimal (CST IV) communities in all samples.

parameters of our model (i.e., a posterior predictive check), confirmed that it accurately captures the observed CST prevalence (Fig. 2a). The optimal, CST I(II, V), and sub-optimal, CST III, communities showed a high degree of stability, with weekly probabilities to remain in the current state estimated at 87% (95% credibility interval, CrI, of 78 to 93%) and 81% (95% CrI of 68 to 90%), respectively. In contrast, the weekly persistence

Table 1: Summary profile of microbiota samples and covariates in the PAPCLEAR study. Q1 and Q3 refer to first (25%) and third (75%) quantiles. Level = 1 indicates the presence of a binary condition. See Appendix S1 for the covariate definitions.

	Level	Summary
Samples (Participants)		2103 (125)
CST (%)	I(II,V)	935 (44.5)
	III	740 (35.2)
	IV	428 (20.4)
Sample type (%)	Clinical	553 (26.3)
	Home	1550 (73.7)
Sampling interval (median (Q1,Q3))		21 (7, 28)
Follow-up duration (median (Q1,Q3))		8.64 (5.36, 14.0)
Samples per subject (median (Q1,Q3))		11 (7, 16)
Covariates		
Identifying as 'Caucasian' (%)	1	102 (81.6)
BMI (median (Q1,Q3))		21.19 (19.78, 23.46)
Alcohol (median (Q1,Q3))		3.14 (1.40, 5.07)
Smoker (%)	1	36 (28.8)
Stress level (from 0 to 3, median (Q1,Q3))		1.41 (1.00, 1.75)
Regular sport practice (%)	1	61 (48.8)
Red meat consumption (times per week, median (Q1,Q3))		0.50 (0.16, 1.00)
Years since 1st menstruations (median (Q1,Q3))		9 (7, 10)
Hormonal contraception (%)	1	32 (25.6)
Menstrual cup user (%)	1	46 (36.8)
Vaginal product user (%)	1	73 (58.4)
Tampon user (%)	1	89 (71.2)
Lifetime number of partners (median (Q1,Q3))		5 (3, 11)
Lubricant use (%)	1	58 (46.4)
Regular condom use by partner (%)	1	23 (18.4)
Male affinity (%)	1	124 (99.2)
Chlamydia infection at inclusion (%)	1	7 (5.6)
Pregnancy during follow-up (%)	1	4 (3.2)
Vaginal douching (%)	1	4 (3.2)
Spermicide user (%)	1	1 (0.8)
Female affinity (%)	1	10 (8.0)
Systemic antibiotic treatment during the study (%)	1	65 (52.0)
Genital antibiotic treatment during the study (%)	1	30 (24.0)

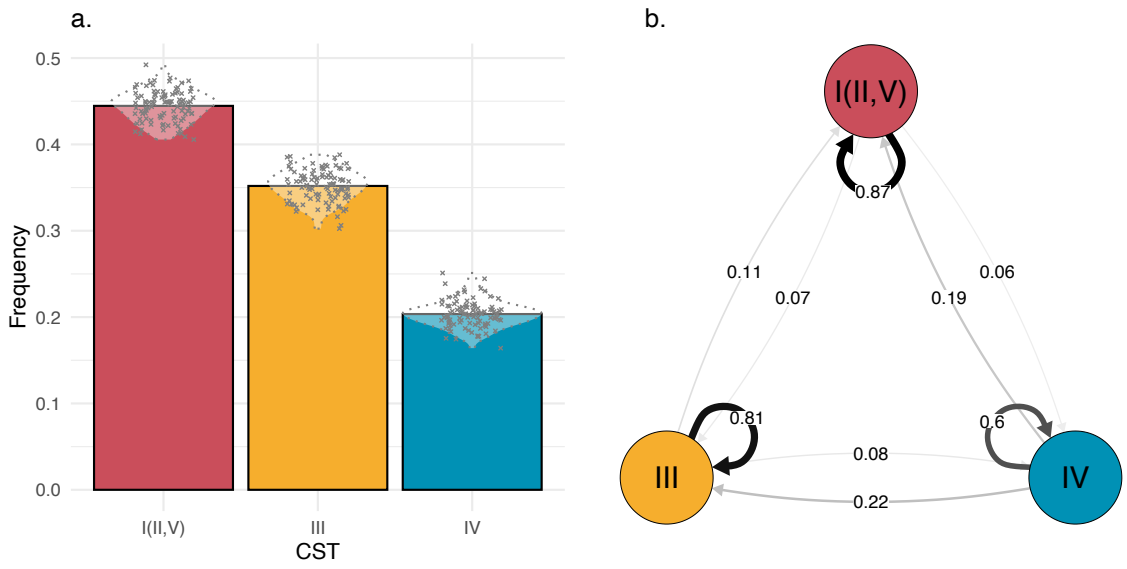

Figure 2: **Prevalence and transition probabilities among vaginal microbiota community state types (CSTs).** a. Observed (bars) and predicted prevalence (crosses) of CSTs I (II, V), III and IV. The model predictions were generated by drawing 100 random samplings from the posterior distributions and simulating the Markov model for each sampled parameter set. b) Mean estimated weekly transition probabilities of CSTs I (II, V), III and IV. The arrow thickness indicates the persistence or transition probability.

probability of the non-optimal CST IV was only 60% (95% CrI of 32 to 80%, Fig. 2b).

Covariates effect on transitions

To gain a mechanistic understanding of the CST shifts, we estimate the effect of covariates jointly with the transition probabilities. Focusing on 16 covariates of interest and assuming that they have a symmetrical effect on CST transitions, we identified alcohol consumption as the most consistent driver of vaginal microbiota communities as it favoured the sub-optimal over the optimal community (with 98% certitude, Fig. 3). This means that alcohol

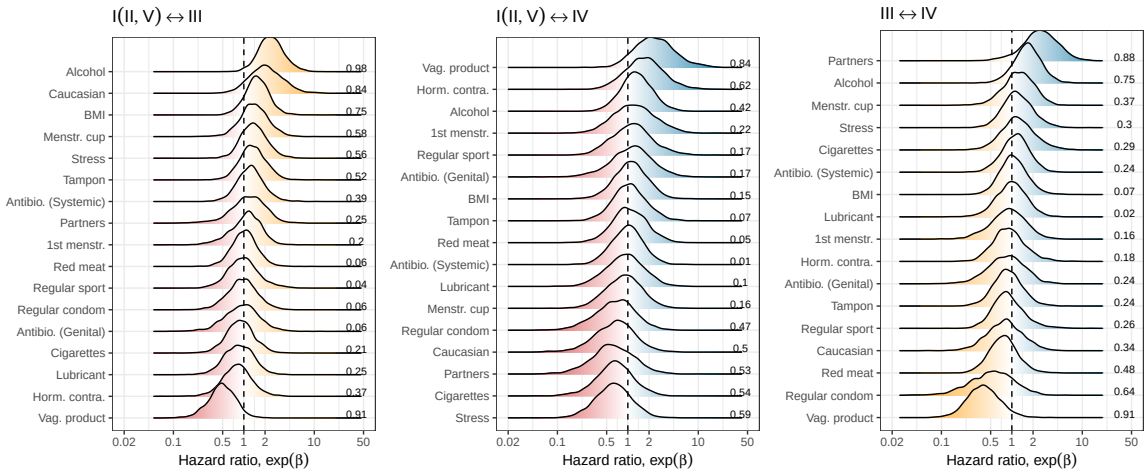

Figure 3: **Estimated covariate effects on community transition rates.** With the symmetry assumption, there are only three main types of transitions. The impact of covariates on community transition rates was estimated for a given set of community states as the log hazard ratio, β. The figure shows the posterior distributions of $\exp(\beta)$, the hazard ratio for the three sets of transition sets, and the corresponding 16 covariates. The numbers on the right-hand side of each panel indicate the probability that the estimated effect is different from the hazard ratio of 1 (i.e., the proportion of posterior distributions sampled on the dominant side of the effect).

consumption increases the pace of transition from CST I(II, V) to CST III or reduces that in the opposite direction by the same magnitude (our symmetry assumption does not allow us to distinguish between the two). In addition, alcohol consumption tended to favour CST IV over CST III (with 75% certitude, Fig. 3).

Other factors of possible effects on transitions (i.e., with more than 75% certitude) included self-reported Caucasian identity, number of sexual partners, use of vaginal intimate hygiene products, and body mass index (BMI). More precisely, identifying oneself as a 'Caucasian' and having a larger BMI both appeared to favour CST III over CST I(II, V)

with 84 and 75% certitude, respectively. The number of sexual partners was also found to potentially favour CST IV over CST III, increasing the risk of maintaining (or transitioning to) CST IV with 88% certitude.

Finally, the use of vaginal hygiene products appeared to have multifaceted effects. Between CST I(II, V) and CST III, their use was positively linked to maintaining or transitioning to CST I(II, V) with 91% certitude. For the CST I(II, V) and CST IV pair, this product use tended to favour a shift towards CST IV, with an 84% certitude. Between CST III and CST IV, their use was more likely to support the persistence or move towards CST III, also with a 91% probability. These findings indicate that further mechanistic understanding of individual products marketed for 'vaginal intimate hygiene' is warranted.

Population-level consequence of covariate effects

Analysing transition rates at the individual level can help guide more mechanistic studies at the within-host level, but they offer limited insights about population-level outcomes. This is particularly true for covariates that are strongly involved in more than one type of CST. Therefore, we performed simulations in which we varied a covariate while setting all other factors to their representative reference value. Thanks to our estimated hazard ratios (Fig. 3), we could predict the expected proportion of each CST in this simulated population. This allowed us to see that covariates such as the reported use of vaginal hygiene products, which are involved in many shits, appear to have a limited population effect, with only a slightly decreasing frequency of CST IV with more frequent usage.

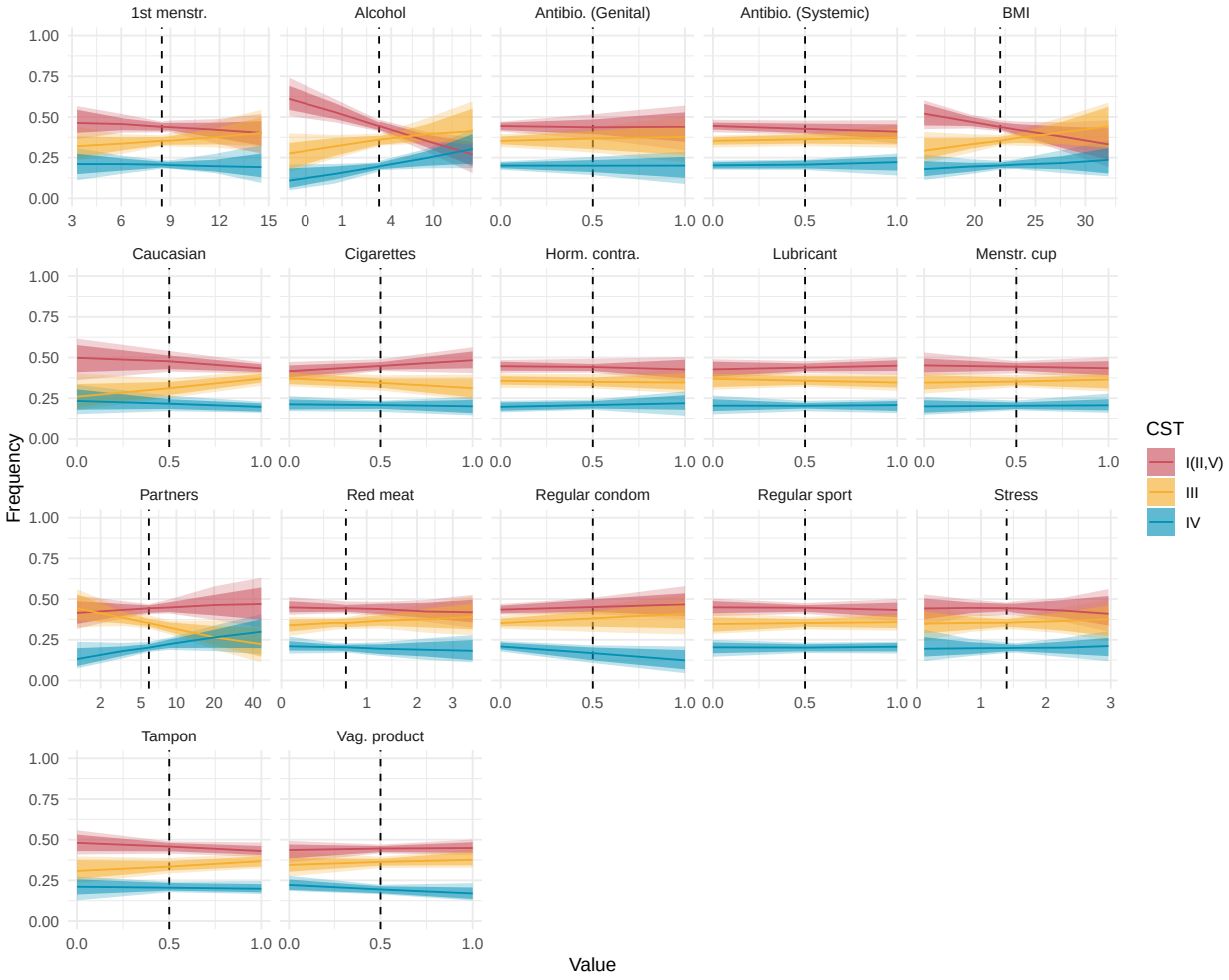

Figure 4: **Prevalence simulations predict population-level consequences of covariates.** Based on estimated hazard ratios (Fig. 3), the population-level impact was simulated for each covariate. The vertical dashed lines indicate the intercept used in estimation: i.e., the population mean for continuous and 0.5 for binary variables. For continuous variables, the range of values explored was determined by the minimum and maximum values reported in the PAPCLEAR study .

Conversely, these simulations allowed us to confirm that the effect of alcohol on community transition rates translates into a marked decline in CST I prevalence in favour of CST

III at the population level in our simulations (Fig. 4). Compared to the average alcohol consumer in this cohort, the predicted prevalence of the optimal community (CST I, II and V) was 17% higher and lower for non-drinkers and the heaviest drinkers in the cohort, respectively (>99% certitude). The alcohol-induced downfall of the optimal communities is accompanied by an increase in both the sub- and non-optimal communities, although the increasing trends are less certain separately. Nonetheless, the non-optimal CST IV tends to be 11% (>88% certitude) higher among average drinkers than non-drinkers.

Finally, as expected from the transition rates, population-level simulations predict that an increasing number of sexual partners tends to reduce the prevalence of the sub-optimal CST III with up to 92% certitude in favour of non-optimal CST IV (with up to 84% certitude). Perceived ethnic identity also hinted at trends at the population level as participants who identified as 'Caucasian' were estimated to show an 11% higher occurrence of CST III (with 84% certitude). We also find that CST IV (non-optimal) communities tend to be less frequent in regular condom users (8% lower prevalence with 88% certitude), which is consistent with previous association studies [16–18].

Unobserved individual variability in community transition

While we incorporated 16 covariates into our model, some variations among women remain unaccounted for. To quantify these, we estimated the extent of individual variability (i.e., unobserved heterogeneity) in community transitions for each transition pair.

The highest variability was observed among women in the transitions involving 'recov-

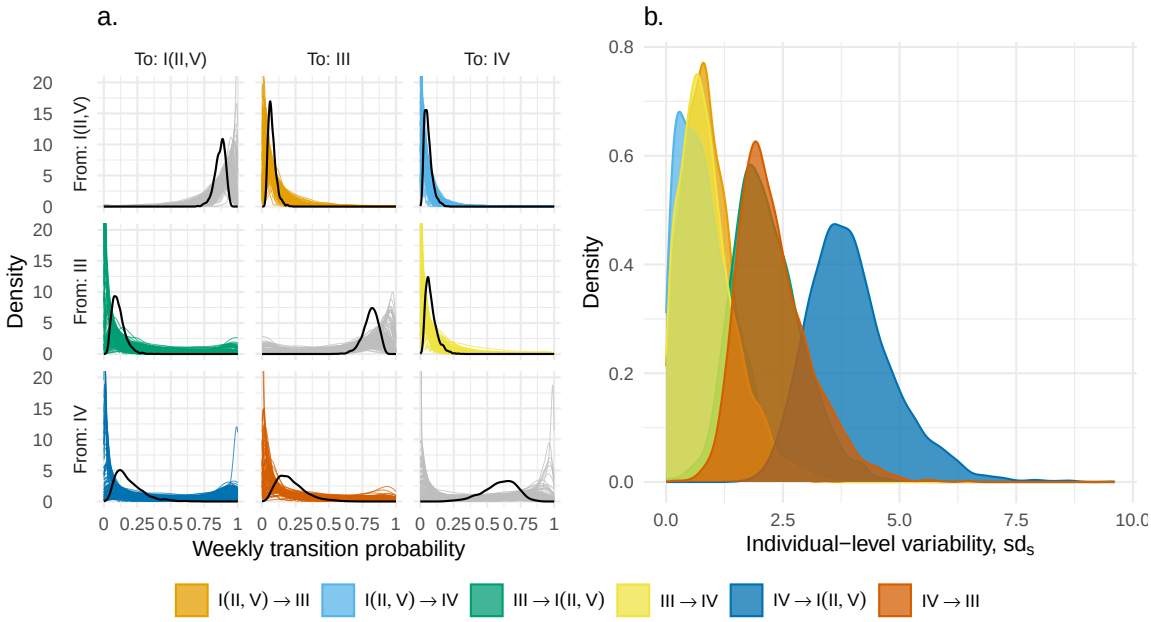

Figure 5: **Individual-level variability in vaginal microbiota community state type transitions.** a) The population average (thick black) and individual (thinner colours) weekly transition probabilities. b) Between-women individual variation for different transition combinations. Colours indicate the type of transition between CSTs.

ery' to an optimal CST from CST IV (Fig. 5). On the other hand, inverse transitions exhibited some of the lowest individual variability. The same is true, although to a lesser extent, for the shifts from sub-optimal CST III to optimal CSTs.

These results suggest that there are limited pathways leading to the deterioration of vaginal microbiota communities, whereas the routes to recovery can be highly individualised. As these variations remained unexplained by the 16 covariates we investigated, further research attention is warranted to better understand the diversity of microbiota recovery trajectory.

Discussion

Vaginal microbiota is a key component of women's health. Pioneer work has studied its daily variations for up to 16 weeks [9], but longer-term dynamics remain largely unknown. By analysing follow-ups in 125 women with unprecedented length (more than 10 months on average and more than 16 samples per participant), we estimate the probabilities to shift from one CST to another. These are consistent with the ones from earlier studies, especially with a model that included the data from four cohorts [14]. Interestingly, in a cohort following pregnant women, CST I was found to be more stable than in ours, which could be due to hormonal changes [10].

We also succeeded in testing the effect of relevant covariates on these transition rates, while estimating them, which represents a strong methodological advance. We identify several covariates with strong effects, some of which are echo earlier studies. For example, the association between CST IV and the number of partners is consistent with the hypothesis that external importation of microbes could alter the dynamics of vaginal microbiota and in line with earlier work []. The same is true for our result that (non-optimal) CST IV tends to be less frequent in regular condom users, which is consistent with previous association studies [16–18].

Our study represents a rare European long-term cohort, yet it does not support a balanced design concerning ethnicity with over 80% of participants representing the Caucasian identity (Table 1). We do find that participants who do not identify themselves as

Caucasian are less found in CST I is consistent with many studies showing associations between self-perceived ethnicity and CST [4, 19, 20]. As the relative importance of biological, societal and environmental factors remains an open question [18], further studies are warranted to investigate in Europe.

The reported use of products for vaginal hygiene (e.g. intimate soap) appears to have an effect on several transitions, which makes it difficult to link to a CST in particular. Finally, one of the strongest associations we find is between the amount of self-reported alcohol consumption and CST III. There are reports that go in this direction but, generally, there is little data on the link between diet and CST.

Strikingly, antibiotic consumption had little effect on CST transitions. This was true for local treatment (genital application of metronidazole) or system treatment (antibiotic treatment via oral intake). The strongest signal was between systemic use and CST III, which is consistent with existing literature. An interpretation of this result is that CST dynamics can be fast, as shown by daily monitoring. Therefore, since our sampling is typically independent from treatment, CSTs are likely to have returned to a more basal state when sampled.

These results have limitations. One of these is that, for our model to achieve convergence, we had to assume symmetry in the transition risks. This means that for alcohol consumption, for example, we do not know whether it increases the risk to shift away from an optimal CST or decreases the likelihood of leaving a sub-optimal CST. Another limitation inherent to all these studies is that even if we included many covariates, these

patterns could be explained by 'hidden variables' or participants' general lifestyles.

Our model outputs, especially the inspection of the random effects, indicate that a large part of the variability remains to be explained. One possibility could be that we miss important covariates. Another could be that the CSTs classification level is too coarse. Future studies could analyse sub-CSTs, but this would greatly increase the number of possible transitions, thereby raising an acute statistical challenge (here we already had to collapse CSTs I, II, and V to introduce covariates into the model). In the same vein, analysing metagenomics data could provide additional valuable insights on vaginal microbiota long-term dynamics [21]. For example, for women who appear to have stable CSTs, we could identify the proportion of lineage replacements. This would also allow us to investigate the effect of antibiotic treatments on the prevalence of resistance genes in the metagenome.

Finally, a promising avenue for future studies would be to jointly analyse CST dynamics and sexually transmitted infections, especially human papillomaviruses (HPV). Earlier studies have found a weak association between CST IV and the risk of HPV detection [22] but the CST effect was tested after estimating transition rates and, more importantly, this study pooled all high-risk HPV and all low-risk HPVs, making it difficult to identify coinfections or reinfections. The PAPCLEAR cohort having genotype-specific follow-ups [23], it could provide new insights into the link between CST and HPV infection, while even identifying causal relationships.

Material and methods

Longitudinal clinical data

Samples originate from the PAPCLEAR monocentric longitudinal cohort study, which followed $N = 149$ women longitudinally between 2016 and 2020. Its inclusion criteria were to be between 18 and 25 years old, to live in the area of Montpellier, France, to be in good health (no chronic disease), not to have a history of HPV infection (e.g., genital warts or high-grade cervical lesion), and to report at least one new sexual partner over the last 12 months. Additional details about the protocol can be found elsewhere [24].

Participants were enrolled by putting up posters and handing out leaflets at the main sexually transmitted infection detection centre (CeGIDD) within the University Hospital of Montpellier (CHU) and in the Universities of the city. To increase enrolment, posters were also hung at bus stops near the CHU.

The inclusion visit was performed by a gynaecologist or a midwife at the CeGIDD outside operating hours. After an interview, several samples were collected, including vaginal swabs with eSwabs (Coppan) in Amies preservation medium. The samples were aliquoted right after the visit and stored at -20°C, before being transferred at -70°C within a month. Participants also filled in a detailed questionnaire, which formed the basis of epidemiological covariates analysed in this study.

Subsequent clinical visits were scheduled every 2 or 4 months, depending on the HPV status. Between two visits, women were asked to perform 8 self-samples at home with

eSwabs in Amies medium and to keep them in their freezer. Samples were brought back in an isotherm bag at the next visit. These were stored with the swab at -70°C until processing.

Microbiota metabarcoding and quantification

The microbiota metabarcoding was performed on 200μL of vaginal swabs specimen stored at -70° in Amies medium. The DNA extraction was performed using the MagAttract PowerMicrobiome DNA/RNA Kit (Qiagen). Next-generation sequencing of the V3-V4 region of the 16S gene [25] was performed on an Illumina HiSeq 4000 platform (150 bp paired-end mode) at the Genomic Resource Center at the University of Maryland School of Medicine.

The taxonomic assignment was performed using the internal software package SpeciateIT (`https://github.com/Ravel-Laboratory/speciateIT`) and the community state type was determined using the VALENCIA software package [26].

Covariates

In the PAPCLEAR study, questionnaires were given to each participant to record patient-level meta-data. In Supplementary Methods, we outline these variables with a variable label (in italics), brief description and relevance for the vaginal environment.

Out of the 22 covariates initially considered, we excluded six (*Chlamydia*, *Female affinity*, *Male affinity*, *Pregnancy*, *Spermicide* and *Vag. douching*) as data were severely

skewed towards the most common value ($> 90\%$ of data). All these covariates are time-homogeneous meaning that the variation is among women, and static through time. To facilitate the comparison of covariate effects, we centred and scaled continuous variables [27] and deviation-coded binary variables. These transformations ensure that all covariates are modelled in a comparable scale and the intercept is located at a "representative reference value" of the modelled population. Four continuous covariates (i.e., *Alcohol*, *BMI*, *Partners*, and *Red meat*) were log-transformed before scaling due to their right-skewed distribution.

During their follow-ups, any use of medication was listed by the study staff. Within these, we identified antibiotic treatments and separated the ones corresponding to 'Gynecological anti-infectives and antiseptics' ('G01' ATC codes), which all consisted in metronidazole treatments, to those corresponding to 'Antibacterials for systemic use' ('J01' ATC codes), which were more diverse in terms of molecules. Since we had the exact dates of treatment, these could be included as time-varying covariates into the model.

The model

Markov modelling

Markov models are statistical models used to represent systems that transition between discrete states over time. These models are said to be 'memoryless', meaning that the probability of transition to another state depends on the current state, but not its historical

path. In clinical research, these models are often used to predict the transitions among health states (e.g., health, illness and remission), and the propensity to transition between these states is estimated from longitudinal follow-up data. Because the exact timing of these transitions is rarely known, clinical follow-up data are typically modelled using the continuous-time Markov model [28], in which the probability of transition over a given interval depends on the instantaneous transition intensity and the amount of time spent in the current state.

Transition intensities

Transition intensities refer to the instantaneous rate of moving from state i to state j, a process that may be affected by a vector of covariates, X. Taking the form of a proportional hazards model, these rates can be expressed as:

$$q_{p,i,j} = \mathrm{Exp}(\mu_{p,i,j} + \beta_{i,j}\ X), \tag{1}$$

where $\mu_{p,i,j}$ is the log-intercept, i.e. the baseline when all covariates are 0, and $\beta_{i,j}$ is the log-coefficient expressing the impact of a covariate(s). This intercept is further defined by the equation

$$\mu_{p,i,j} = (\hat{\mu}_{i,j} + s_{p,i,j}) \cdot \mu_{sd} + \bar{\mu}, \tag{2}$$

where $\bar{\mu}$ and μ_{sd} are the prior mean and standard deviation of the log-intercept such that $\hat{\mu}_{i,j} \cdot \mu_{sd} + \bar{\mu}$ constitutes the non-centred parameterisation of the population-level intercept, $\mu_{i,j}$ and is assumed to be normally distributed, i.e., $\hat{\mu}_{i,j} \sim \mathcal{N}(0, 1)$.

Additionally, we allow for unobserved heterogeneity in μ, i.e., $s_{p,i,j}$, where

$$\mathbf{s} = \text{diag}(\mathbf{sd_s}) \cdot \mathbf{L_s} \cdot \mathbf{z_s}. \tag{3}$$

We sample from the corresponding weakly informative priors, namely $sd_s \sim t_4(0, 1)$, $L_s \sim \text{LKJCorrCholesky}(2)$, and $z_s \sim \mathcal{N}(0, 1)$, as recommended by the Stan development community [29, 30].

For regression coefficients, the student-t distributions with degrees of freedom 4 to 7 are recommended as generic, weakly informative, priors [30]. We sampled β from $\beta \sim t_4(0, 1)$, which places a comparatively wide tail within the recommendation. As all of our covariates have been proposed to impact vaginal microbiota communities *a priori* (see above), we did not strongly regularise the priors, for example, through the use of the horseshoe priors [31]. We note that all covariates were modelled simultaneously such that the interpretation of

each coefficient is conditional upon other covariates included and accounts for the influence of other factors. We assumed that the covariates symmetrically affect the transitions (i.e., $\beta_{j,i} = -\beta_{i,j}$), implying that the influence of any particular factor on moving from one community state to another remains consistent, regardless of the direction the transition.

Collectively, the transition intensities form the Q matrix, in which the sum of intensities across a row, i.e., all transitions from a particular state, is defined to be zero, such that we have the following equation for the diagonal entries:

$$q_{i,i} = -\sum_{j \neq i} q_{i,j}. \tag{4}$$

Transition probabilities and likelihood

Taking the matrix exponential of the Q matrix for each participant, we compute the matrix P such that:

$$P = \mathrm{Exp}\big((t_{k+1} - t_k)\ Q\big), \tag{5}$$

where k represents the sample identity for a given individual. The P matrix contains the transition probabilities between two observations (at k and $k+1$) and $t_{k+1} - t_k$ indicates the elapsed time between two observations.

Finally, the probability of observing a given state at the next sampling event (i.e. at $k+1$) is modelled by the categorical distribution:

$$y_{k+1} \sim \text{Categorical}\big(P[y_k,]\big) \tag{6}$$

where $P[y_k,]$ is the $y_k{}^{\text{th}}$ row of the P matrix containing the probabilities of transition from the state observed at k.

Model fitting

We used a Bayesian approach to fit the above continuous-time Markov model to longitudinal data of vaginal microbiota CSTs. Our model was written in Stan 2.26.1 and fitted through the RStan interface 2.32.3 [32].

One participant lacked information on the years since their initial menstruation. We imputed missing values using the mice package [33] and generated 20 imputed datasets to be fitted separately. For each imputed dataset, we fitted the model in parallel using four independent chains, each with $2,000$ sampled iterations and $1,000$ warmup iterations. The MCMC samples from separate runs (i.e., based on differently imputed data) were subsequently combined for inference.

We confirmed over $4,000$ combined effective samples and ensured convergence of independent chains ($\hat{R} < 1.1$) for all parameters [34]. We carried out a posterior predictive

check by comparing the observed and predicted CST frequency. We also quantified the posterior z-score and posterior contraction to examine the accuracy and precision of posterior distributions and the relative strength of data to prior information [35] (Supporting Information S1).

Counterfactual predictions

We took advantage of the parameterised model to simulate the population-level outcomes of each covariate, assuming that all covariates, but a focal one, are at the representative reference value (as described above) and then varying the focal parameter within the range of values observed in the studied cohort. The model predictions were generated by randomly drawing 100 samplings from the posterior distributions and simulating the Markov model for each sampled parameter set. We focused on the CST frequency as the outcome of interest.

Acknowledgments

The authors thank Olivier Supplison for his helpful feedback. They acknowledge the ISO 9001 certified IRD i-Trop HPC (member of the South Green Platform) at IRD Montpellier for providing HPC resources that have contributed to the research results reported within this article (bioinfo.ird.fr and www.southgreen.fr).

DNA extracts were (partly) performed through the genotyping and sequencing facilities

of ISEM (Institut des Sciences de l'Evolution-Montpellier) and Labex Centre Méditerranéen Environnement Biodiversité.

The raw data and R scripts used will be shared upon publication.

References

[1] van de Wijgert JHHM. The vaginal microbiome and sexually transmitted infections are interlinked: Consequences for treatment and prevention. PLoS Medicine. 2017;14(12):e1002478.

[2] Haahr T, Zacho J, Bräuner M, Shathmigha K, Skov Jensen J, Humaidan P. Reproductive outcome of patients undergoing in vitro fertilisation treatment and diagnosed with bacterial vaginosis or abnormal vaginal microbiota: a systematic PRISMA review and meta-analysis. BJOG. 2019;126(2):200–207.

[3] Bilardi JE, Walker S, Temple-Smith M, McNair R, Mooney-Somers J, Bellhouse C, et al. The Burden of Bacterial Vaginosis: Women's Experience of the Physical, Emotional, Sexual and Social Impact of Living with Recurrent Bacterial Vaginosis. PLoS ONE. 2013;8(9):e74378.

[4] Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Nat Acad Sci USA. 2011;108(S1):4680–4687.

[5] France MT, Fu L, Rutt L, Yang H, Humphrys MS, Narina S, et al. Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data. Genome Biology. 2022;23(1):66.

[6] McKinnon LR, Achilles SL, Bradshaw CS, Burgener A, Crucitti T, Fredricks DN, et al. The Evolving Facets of Bacterial Vaginosis: Implications for HIV Transmission. AIDS Research and Human Retroviruses. 2019;35(3):219–228.

[7] Holm JB, France MT, Gajer P, Ma B, Brotman RM, Shardell M, et al. Integrating compositional and functional content to describe vaginal microbiomes in health and disease. Microbiome. 2023;11(1):259.

[8] Cancelo-Hidalgo MJ, Coello LB. Genitourinary Syndrome of the Menopause: Vaginal Health and Microbiota. In: Cano A, editor. Menopause: A Comprehensive Approach. Cham; 2017. p. 91–107.

[9] Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UME, Zhong X, et al. Temporal Dynamics of the Human Vaginal Microbiota. Science Translational Medicine. 2012;4(132):132ra52–132ra52.

[10] DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, et al. Temporal and spatial variation of the human microbiota during pregnancy. Proceedings of the National Academy of Sciences. 2015;112(35):11060–11065.

[11] Serrano MG, Parikh HI, Brooks JP, Edwards DJ, Arodz TJ, Edupuganti L, et al. Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. Nature medicine. 2019;25(6):1001–1011.

[12] Munoz A, Hayward MR, Bloom SM, Rocafort M, Ngcapu S, Mafunda NA, et al. Modeling the temporal dynamics of cervicovaginal microbiota identifies targets that may promote reproductive health. Microbiome. 2021;9(1):1–12.

[13] Chaban B, Links MG, Jayaprakash TP, Wagner EC, Bourque DK, Lohn Z, et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. Microbiome. 2014;2:1–12.

[14] Brooks JP, Buck GA, Chen G, Diao L, Edwards DJ, Fettweis JM, et al. Changes in vaginal community state types reflect major shifts in the microbiome. Microbial ecology in health and disease. 2017;28(1):1303265.

[15] Ma B, Forney LJ, Ravel J. Vaginal Microbiome: Rethinking Health and Disease. Annual Review of Microbiology. 2012 Oct;66(1):371–389. Available from: `http://www.annualreviews.org/doi/10.1146/annurev-micro-092611-150157`.

[16] Hutchinson KB, Kip KE, Ness RB. Condom use and its association with bacterial vaginosis and bacterial vaginosis-associated vaginal microflora. Epidemiology. 2007;p. 702–708.

[17] Yotebieng M, Turner AN, Hoke TH, Van Damme K, Rasolofomanana JR, Behets F. Effect of consistent condom use on 6-month prevalence of bacterial vaginosis varies by baseline BV status. Tropical Medicine & International Health. 2009;14(4):480–486.

[18] Borgdorff H, Veer Cvd, Houdt Rv, Alberts CJ, Vries HJd, Bruisten SM, et al. The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. PLoS ONE. 2017;12(7):e0181135.

[19] Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P, et al. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. ISME J. 2007;1(2):121–133.

[20] Fettweis JM, Brooks JP, Serrano MG, Sheth NU, Girerd PH, Edwards DJ, et al. Differences in vaginal microbiome in African American women versus women of European ancestry. Microbiology. 2014;160(Pt 10):2272.

[21] France M, Ma B, Ravel J. Persistence and In Vivo Evolution of Vaginal Bacterial Strains over a Multiyear Time Period. mSystems. 2022 Nov;0(0):e00893–22. Publisher: American Society for Microbiology. Available from: `https://journals.asm.org/doi/10.1128/msystems.00893-22`.

[22] Brotman RM, Ravel J, Bavoil PM, Gravitt PE, Ghanem KG. Microbiome, sex hormones, and immune responses in the reproductive tract: Challenges for vaccine development against sexually transmitted infections. Vaccine. 2014;32(14):1543–1552.

[23] Beneteau T, Groc S, Murall CL, Boué V, Elie B, Tessandier N, et al.. HPV detection patterns in young women from the PAPCLEAR longitudinal study: implications for HPV screening policies. medRxiv; 2023. Available from: `https://www.medrxiv.org/content/10.1101/2023.09.30.23296382v1`.

[24] Murall CL, Rahmoun M, Selinger C, Baldellou M, Bernat C, Bonneau M, et al. Natural history, dynamics, and ecology of human papillomaviruses in genital infections of young women: protocol of the PAPCLEAR cohort study. BMJ Open. 2019;9(6):e025129.

[25] Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. Critical Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S rRNA Genes. Appl Environ Microbiol. 2008;74(8):2461.

[26] France MT, Ma B, Gajer P, Brown S, Humphrys MS, Holm JB, et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. Microbiome. 2020;8(1):166.

[27] Gelman A. Scaling regression inputs by dividing by two standard deviations. Statistics in medicine. 2008;27(15):2865–2873.

[28] Christopher H Jackson. Multi-State Models for Panel Data: The msm Package for R. Journal of Statistical Software. 2011;38(8):1–29.

[29] Stan Development Team. Stan Functions Reference; 2023. Accessed: 2024-02-24. Available from: https://mc-stan.org/docs/functions-reference/index.html.

[30] Stan Development Team. Prior Choice Recommendations; 2023. Accessed: 2024-02-15. https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations.

[31] Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics. 2017;11(2):5018–5051.

[32] Stan Development Team. RStan: the R interface to Stan; 2023. R package version 2.32.3. Available from: https://mc-stan.org/.

[33] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45(3):1–67.

[34] Stan Development Team. The Stan Core Library; 2018. Version 2.18.0. Available from: http://mc-stan.org/17.

[35] Betancourt M. Towards a principled Bayesian workflow; 2020.

[36] Farage M, Maibach H. Lifetime changes in the vulva and vagina. Archives of gynecology and obstetrics. 2006;273:195–202.

[37] Loganantharaj N, Nichols WA, Bagby GJ, Volaufova J, Dufour J, Martin DH, et al. The effects of chronic binge alcohol on the genital microenvironment of simian im-

munodeficiency virus-infected female rhesus macaques. AIDS research and human retroviruses. 2014;30(8):783–791.

[38] Mayer BT, Srinivasan S, Fiedler TL, Marrazzo JM, Fredricks DN, Schiffer JT. Rapid and profound shifts in the vaginal microbiota following antibiotic treatment for bacterial vaginosis. The Journal of infectious diseases. 2015;212(5):793–802.

[39] Si J, You HJ, Yu J, Sung J, Ko G. Prevotella as a hub for vaginal microbiota under the influence of host genetics and their association with obesity. Cell host & microbe. 2017;21(1):97–105.

[40] Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proceedings of the National Academy of Sciences. 2011;108(supplement_1):4680–4687.

[41] Brotman RM, He X, Gajer P, Fadrosh D, Sharma E, Mongodin EF, et al. Association between cigarette smoking and the vaginal microbiota: a pilot study. BMC infectious diseases. 2014;14(1):1–11.

[42] Achilles SL, Austin MN, Meyn LA, Mhlanga F, Chirenje ZM, Hillier SL. Impact of contraceptive initiation on vaginal microbiota. American journal of obstetrics and gynecology. 2018;218(6):622–e1.

[43] Laniewski P, Owen KA, Khnanisho M, Brotman RM, Herbst-Kralovetz MM. Clinical and personal lubricants impact the growth of vaginal lactobacillus species and colo-

nization of vaginal epithelial cells: an in vitro study. Sexually transmitted diseases. 2021;48(1):63–70.

[44] Tessandier N, Uysal IB, Elie B, Selinger C, Bernat C, Boué V, et al. Does exposure to different menstrual products affect the vaginal environment? Molecular Ecology. 2023;32(10):2592–2601.

[45] Vodstrcil LA, Twin J, Garland SM, Fairley CK, Hocking JS, Law MG, et al. The influence of sexual activity on the vaginal microbiota and Gardnerella vaginalis clade diversity in young women. PLOS One. 2017;12(2):e0171856.

[46] Noormohammadi M, Eslamian G, Kazemi SN, Rashidkhani B. Association between dietary patterns and bacterial vaginosis: a case–control study. Scientific Reports. 2022;12(1):12199.

[47] Pape K, Ryttergaard L, Rotevatn TA, Nielsen BJ, Torp-Pedersen C, Overgaard C, et al. Leisure-time physical activity and the risk of suspected bacterial infections. Medicine and science in sports and exercise. 2016;48(9):1737–1744.

[48] Amabebe E, Anumba DO. Psychosocial stress, cortisol levels, and maintenance of vaginal health. Frontiers in endocrinology. 2018;p. 568.

[49] Klebanoff MA, Nansel TR, Brotman RM, Zhang J, Yu KF, Schwebke JR, et al. Personal hygienic behaviors and bacterial vaginosis. Sexually transmitted diseases. 2010;37(2):94.

[50] Van Kessel K, Assefi N, Marrazzo J, Eckert L. Common complementary and alternative therapies for yeast vaginitis and bacterial vaginosis: a systematic review. Obstetrical & gynecological survey. 2003;58(5):351–358.

[51] Ma ZS. Microbiome transmission during sexual intercourse appears stochastic and supports the red queen hypothesis. Frontiers in Microbiology. 2022;12:789983.

[52] Juliana NC, Peters RP, Al-Nasiry S, Budding AE, Morré SA, Ambrosino E. Composition of the vaginal microbiota during pregnancy in women living in sub-Saharan Africa: a PRISMA-compliant review. BMC pregnancy and childbirth. 2021;21(1):1–15.

[53] Gupta K, Hillier SL, Hooton TM, Roberts PL, Stamm WE. Effects of contraceptive method on the vaginal microbial flora: a prospective evaluation. The Journal of infectious diseases. 2000;181(2):595–601.

[54] Brotman RM, Klebanoff MA, Nansel TR, Andrews WW, Schwebke JR, Zhang J, et al. A longitudinal study of vaginal douching and bacterial vaginosis—a marginal structural modeling analysis. American journal of epidemiology. 2008;168(2):188–196.

[55] Team SD. Stan Functions Reference; 2024. Accessed: 2024-02-05. `https://mc-stan.org/docs/functions-reference`.

Supplementary methods

Competing interests

JReynes reports personal fees from Gilead (consulting and payment or honoraria for lectures, presentations, speaker's bureaus, manuscript writing, or educational events), Janssen (payment or honoraria for lectures, presentations, speaker's bureaus, manuscript writing, or educational events), Merck (payment or honoraria for lectures, presentations, speaker's bureaus, manuscript writing, or educational events), Theratechnologies (payment or honoraria for lectures, presentations, speaker's bureaus, manuscript writing, or educational events), and ViiV Healthcare (consulting and payment or honoraria for lectures, presentations, speaker's bureaus, manuscript writing, or educational events) and support for attending meetings and/or travel from Gilead and Pfizer, outside of the submitted work.

JRavel is co-founder of LUCA Biologics, a biotechnology company focusing on translating microbiome research into live biotherapeutics drugs for women's health. He is Editor-in-Chief at *Microbiome*.

All the other authors do not report any conflict of interest.

Ethics

The PAPCLEAR study was promoted by the CHU of Montpellier (France). This study has been approved by the Comité de Protection des Personnes (CPP) Sud Méditerranée I (reference number 2016-A00712-49); by the Comité Consultatif sur le Traitement de

l'Information en matière de Recherche dans le domaine de la Santé (reference number 16.504); by the Commission Nationale Informatique et Libertés (reference number MMS/ ABD/ AR1612278, decision number DR-2016–488), by the Agence Nationale de Sécurité du Médicament et des Produits de Santé (reference 20160072000007), and is registered at ClinicalTrials.gov under the ID NCT02946346.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648963, to SA). The authors acknowledge further support from the Fondation pour la Recherche Medicale (to TK), the Agence Nationale de la Recherche contre le SIDA (ANRS-MIE, to NT and OS), and the MemoLife Labex (to BE).

S1: Description of the study covariates

1st menstr. Number of years since the first menstruation: The morphology of the human vagina changes throughout the life of a woman and the onset of puberty marks a key event that triggers cascading changes [36].

Alcohol Average number of glasses of alcoholic drinks consumed per week: Chronic presence of alcohol in the genital environment has been linked to a shift in the immune and microbiological conditions [37].

Antibiotics Consumption of antibiotics during the study: The bacterial composition responds rapidly and transiently to antibiotics treatments that target bacteria either broadly or with a narrow taxonomic scale [38].

BMI Body mass index (BMI): Obesity has been implicated in elevating vaginal microbiota diversity and promoting *Prevotella* associated with bacterial vaginosis [39].

Caucasian Identity as Caucasian ethnicity: Ethnicity has been linked to variation in vaginal microbiota compositions in several studies [40]. However, causal mechanisms remain an open question.

Cigarettes Cigarette smoking: Smoking has been implicated in the development of BV due to its anti-estrogenic effects and the presence of harmful substances such as benzo[a]pyrene diol epoxide (BPDE). [41].

Horm. contra. Use of hormonal contraception during the study: The vaginal hormonal landscape is affected by the use of hormonal contractions [42].

Lubricant Use of lubricant during the study: Personal lubricants contain various chemicals that differentially impact the growth of vaginal microbes in-vitro [43].

Menstr. cup Use of menstrual cup during the study: The vaginal microenvironmental may be altered by the use of menstrual cups both physically and chemically. An elevated risk of fungal infections has been reported [44].

Partners Cumulative number of sexual partners: The genital microbiome can be transferred between sexual partners [45]. Such an external input could destabilise the resident community.

Red meat Average number of meals that include red meat consumption per week: Diet alters the vaginal environment for microbes. An unhealthy diet, linked to a high proportion of red meat consumption, has been linked to an elevated risk of BV [46].

Regular condom Regular use of condoms during sexual intercourse: Condom use can modify the vaginal microenvironment by altering the exchange of microbes between partners [16].

Regular sport Engaging in regular sporting activities, over 50% of the time: Physical activities influence immune responses, with leisure-time physical activity associated with a reduced risk of suspected bacterial infections compared to sedentary behaviour [47].

Stress Average stress level reported from 0 (min) to 3 (max): Stress hormones may disrupt vaginal flora, for instance, by inhibiting glycogen production, which is the primary fuel for lactobacilli [48].

Tampon Use of tampons during the study: The use of internal contraceptives like tampons directly alters the vaginal environment, although negative effects from tampon use are seldom reported [49].

Vag. product Use of vaginal cream/tablet/capsule/gel/wipe during the study: Women frequently use over-the-counter vulvovaginal treatments that contain a variety of chemical com-

ponents. However, the clinical effectiveness of these products in preventing bacterial vaginosis (BV) is seldom systematically evaluated [50].

Chlamydia Tested positive for chlamydia

Female affinity Affinity to female: Genital microbiome transfers during sexual activity are anticipated to vary based on the genders of the partners [51]

Male affinity Affinity to male: Genital microbiome transfers during sexual activity are anticipated to vary based on the genders of the partners [51]

Pregnancy History of pregnancy: Pregnancy significantly changes the cervicovaginal environment, with increased estrogen from the ovaries and placenta leading to higher vaginal glycogen. This supports the growth of Lactobacillus species [52].

Spermicide Use of spermicide during study: Spermicides use chemicals to prevent sperm from reaching an egg, but their use can change vaginal microflora, potentially increasing the risk of genitourinary infections [53].

Vag. douching Use of vaginal douching during study: Vaginal douching, the practice of washing inside the vagina with a liquid solution, has been shown to increase the risk of disturbing the natural balance of vaginal flora [54].

Supplementary Results

S2: Pairwise correlations between covariates

There were no strong correlations among covariates, with the strongest correlation found between BMI and stress ($r = 0.41$).

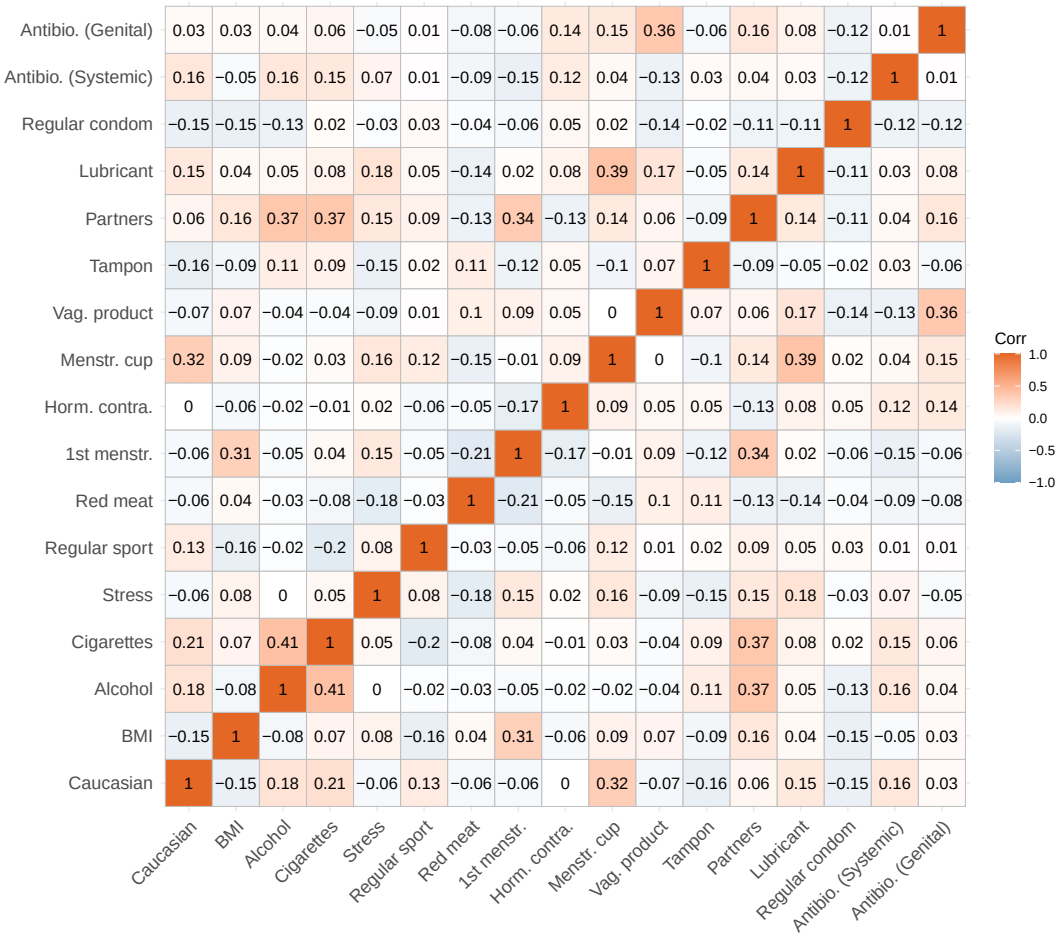

Figure S1: Correlation between covariates. Pairwise Pearson's correlation coefficients between covariates

S3: Assessment of posterior accuracy, precision and prior contraction

We leveraged the properties of posterior distributions to identify potential model fitting problems that might manifest from our model assumptions. To examine the accuracy and precision of posterior distributions, we first generated simulated observations based on the estimated posterior mean parameters. We then refitted our model to the simulated observations (i.e., secondary fitting) to compute the posterior z-score for each parameter, which measures how closely the posterior recovers the parameters of the data generating process [35]:

$$z = \frac{\mathbb{E}_{\text{sim}} - \mathbb{E}_{\text{post}}}{\sigma_{\text{sim}}},$$

where $\mathbb{E}_{\text{post}}$ denotes the posterior mean of the fit to the actual data that we consider the 'true' parameter. $\mathbb{E}_{\text{sim}}$ and σ_{sim} denote the mean and standard deviation of the posterior distribution of the secondary fitting. The smaller the z-score, the closer the bulk of the posterior is to the true parameter [35]. In contrast, large z-values may be indicative of overfitting and, or poor prior specifications [35].

To examine the influence of the likelihood function in relation to prior information, we computed the posterior contraction, k:

$$k = 1 - \frac{\sigma^2_{\text{post}}}{\sigma^2_{\text{prior}}}$$

where σ^2_{post} and σ^2_{prior} correspond to the variance of posterior and prior distributions, respectively. The k values close to zero indicate that data contain little information (i.e., rendering priors strongly informative). Conversely, values close to 1 indicate that data are much more informative than the prior [35].

We found that the most of our model parameters and hyperparameters — were estimated with accuracy and precision and identifiability, with the absolute posterior z-scores below three (Fig. S1). A small number of individual variation in transition rates, sd_s showed a tendency towards overfitting (the absolute posterior z-scores above three). Thus, caution might be warranted when interpreting the extent of between-women variation in CST transition rates, a small number of z-scores exceeding the absolute number of three is unlikely to be a cause of concern [35]. We found that the posterior distributions for covariate coefficients, β, contracted by over 75% compared to the prior distribution, for all but one covariate effect, meaning that the covariate coefficients were well-identified from data (Fig. S1). Although we used generic priors recommended by Stan [55], the L_s parameters that define correlations among between-woman variation showed limited posterior contraction (i.e., $\leq\sim 0.25$), indicating that these parameters are poorly informed by data. As such, we refrain from making biological inferences about these correlations.

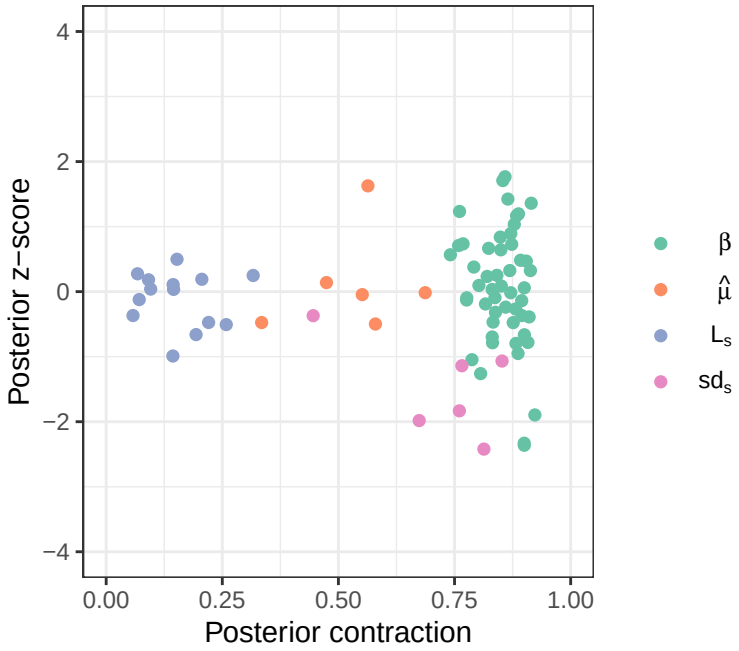

Figure S2: Accuracy, precision and identifiability of estimated parameters. Posterior z-score (y-axis) measures how closely the posterior recovers the parameters of the true data-generating process and posterior contraction (x-axis) evaluates the influence of the likelihood function over the prior, respectively. Smaller absolute posterior z-scores indicate that the posterior accurately recovers the parameters of the data-generating process: the absolute value beyond three to four may indicate substantial bias [35]. The posterior contraction values close to one indicate that data are much more informative than the prior. The estimated parameters are represented by a filled dot.

S4: Predicted difference in community state type (CST) prevalence at various counterfactual scenarios.

Our counterfactual simulations predicted that alcohol consumption and the number of partners are factors that impact the population-level outcome in terms of the prevalence of different community state types.

The prevalence of the optimal community (CST I (II, V)) in a counterfactual population comprising the heaviest drinkers was predicted to be 37% lower than that of non-drinkers (with >99% certitude) (Fig. S??). The prevalence difference of the sub-optimal and non-optimal communities in the same comparisons was an increase of 22% (98% certitude) and 16% (84% certitude), respectively. Similar comparisons between the populations of heaviest drinkers and average drinkers yielded a reduction of 20% in the optimal communities (>99% certitude) and the population of average drinkers versus non-drinkers predicted a 17% lower prevalence of the optimal communities.

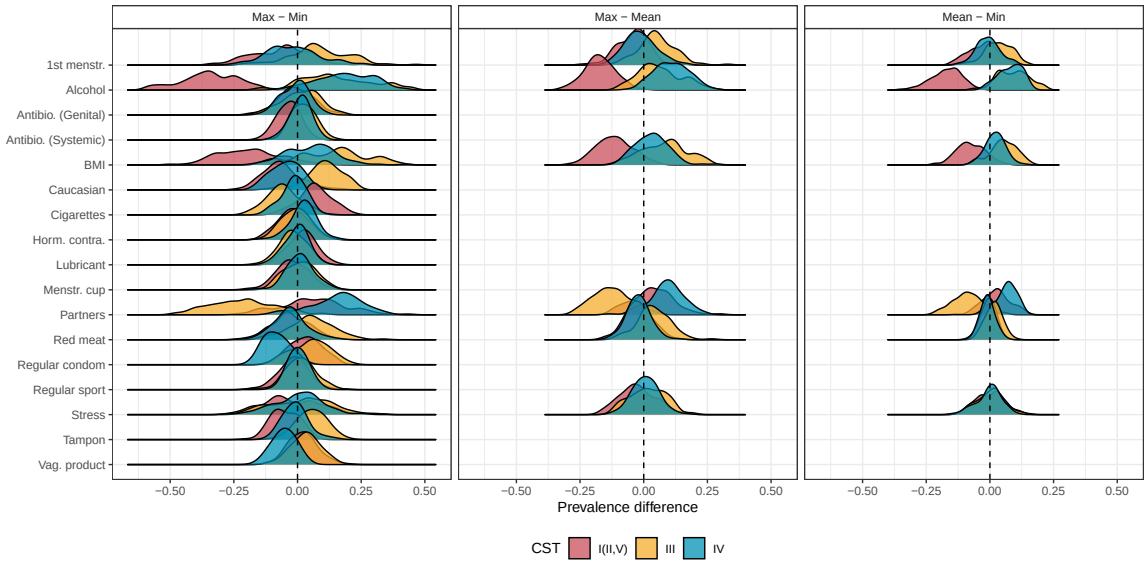

Figure S2: Difference in community state type (CST) prevalence at predicted various counterfactual scenarios. The differences were calculated from posterior samples simulated at 0 and 1 for binary variables and at the population maximum and minimum values recorded by the PAPCLEAR for continuous variables (left panel). Additional differences were computed between the population maximum and mean (middle panel) and the population mean and minimum for continuous variables (right panel).