

Factors shaping vaginal microbiota long-term community dynamics in young adult women

Tsukushi Kamiya^{1,*}, Nicolas Tessandier¹, Baptiste Elie^{1,2}, Claire Bernat^{2,3}, Vanina Boué², Sophie Grasset², Soraya Groc^{2,4}, Massilva Rahmoun², Christian Selinger^{2,5}, Michael S. Humphrys⁶, Marine Bonneau⁷, Christelle Graf⁷, Vincent Foulongne⁴, Jacques Reynes⁸, Vincent Tribout⁸, Michel Segondy⁴, Nathalie Boulle⁴, Jacques Ravel⁶, Carmen Lía Murall^{2,9}, and Samuel Alizon^{1,2,*}

¹Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, Université PSL, Paris, France

²MIVEGEC, CNRS, IRD, Université de Montpellier, France

³Institut de Génomique Fonctionnelle, Université de Montpellier, CNRS, INSERM, Montpellier, France.

⁴PCCEI, Univ. Montpellier, Inserm, EFS, Montpellier, France

⁵Current address: Swiss Tropical and Public Health Institute, Basel, Switzerland

⁶Institute for Genomic Sciences, University of Baltimore, USA

⁷Department of Obstetrics and Gynaecology, Centre Hospitalier Universitaire de Montpellier, Montpellier, France

⁸Department of Infectious and Tropical Diseases, Centre Hospitalier Universitaire de Montpellier, Montpellier, France

⁹National Microbiology Laboratory (NML), Montreal 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 known to affect women’s health. Yet, there is a notable paucity of high-resolution follow-up studies lasting several months, which would be required to interrogate the long-term dynamics and associations with demographic and behavioural covariates. Here, we present a high-resolution longitudinal cohort of 125 women followed for a median duration of 8.6 months, providing 11 samples per woman. Using a hierarchical Bayesian Markov model, we characterised the patterns of vaginal microbiota community persistence and transition, simultaneously estimated the impact of 16 covariates and quantified individual variability among women. We showed that ‘optimal’ (Community State Type (CST) I, II, and V) and ‘sub-optimal’ (CST III) communities are more stable over time than ‘non-optimal’ (CST IV) ones. Furthermore, we found that some covariates — most notably alcohol consumption — impacted the probability of shifting from one CST to another. We performed counterfactual simulations to confirm that alterations of key covariates, such as alcohol consumption, could shape the prevalence of different microbiota communities in the population. Finally, our analyses indicated that there is a relatively canalised pathway leading to the deterioration of vaginal microbiota communities, whereas the paths to recovery can be highly individualised among women. In addition to providing one of the first insights into vaginal microbiota dynamics over a year, our study showcases a novel application of a hierarchical Bayesian Markov model to clinical cohort data with many covariates. Our findings pave the way for an improved mechanistic understanding of

microbial dynamics in the vaginal environment and the development of novel preventative 21

and therapeutic strategies to improve vaginal health. 22

Introduction

23

Epithelia of the human body are host to a diverse array of microorganisms. These mi- 24
croorganisms are collectively referred to as microbiota and their compositions are tightly 25
associated with human health. In the human vaginal environment, the description of the 26
microbiota dates back to Albert Döderlein in 1892. Its composition has been demonstrated 27
to impact the acquisition risk of several sexually transmitted infections (STIs) [1], fertility 28
(especially in medically-assisted procreation procedures) [2], and general well-being [3]. 29

Vaginal microbiota communities comprise hundreds of species. To facilitate understand- 30
ing, the variation in community composition is usually reduced to a handful of categories 31
that capture key compositional signatures, such as the dominance of certain species or 32
species evenness. This dimensionality reduction filters out noise in the data and facilitates 33
the identification and visualisation of key patterns and relationships. 34

Potential drawbacks of reducing continuous variation include the risk of losing sub- 35
tle but meaningful signals within the microbiota, as less dominant or rare taxa may be 36
excluded despite their potential importance. Compared to the gut microbiota, however, 37
vaginal microbiota communities tend to be highly structured and are often dominated by 38
a small handful of species whose functional ecology is well-documented [4]. This contrasts 39
with the highly diverse gut microbiota, where defining discrete community types, such as 40
“enterotypes,” remains contentious [5]. The high diversity and evenness in gut microbiota 41
introduce continuous variations that can be oversimplified by strict categorical clustering. 42

In contrast, vaginal microbiota composition aligns more naturally with categorical clustering, providing a robust understanding of key microbial patterns without significantly sacrificing interpretability.

One dimensionality reduction framework, i.e., community state types (CSTs), introduced by Ravel et al. [6], categorises vaginal microbial communities into five discrete state types. The CSTs considered ‘optimal’ for health are dominated by *Lactobacillus* species; *Lactobacillus crispatus*, *L. gasseri*, and *L. jensenii* for CST I, II, and V, respectively. Lactobacilli produce lactic acid and hydrogen peroxide, which create an acidic environment that helps to inhibit the growth of harmful pathogens [7]. On the other end of the spectrum, CST IV is the primary microbial context of bacterial vaginosis (BV), which elevates the risk of STI acquisition and spontaneous preterm birth, and is associated with symptoms such as malodor, discharge, and itching [4, 8]. This community is characterised by a diverse assemblage of anaerobic bacterial species from the *Gardnerella*, *Prevotella*, and *Fannyhessea* genera: recent classifications include sub-categories within CST IV (i.e., IV-A, IV-B, IV-C), each with a distinct microbial profile [9]. Finally, CST III, characterised by a dominance of *L. iners*, is considered ‘sub-optimal’ for women’s health. While *L. iners* is a member of the *Lactobacillus* genus, it is less effective at producing lactic acid and hydrogen peroxide. As such, women with CST III tend to exhibit higher vaginal pH than those with CST I and are more prone to experiencing adverse health consequences, including vaginal infections [10].

The CST classification represents a snapshot of the microbiota community at the time of

sampling that facilitates the examination of clinically relevant microbiota variations across 64
time and women. The development of the modern pipeline — through meta-barcoding 65
sequencing of 16S DNA and clustering algorithms [9] — allows for CST-typing with en- 66
hanced efficiency and reduced observer bias compared to conventional microscopy-based 67
methods of vaginal microbiota community typing (e.g., Nugent score). 68

The composition of vaginal microbiota is characteristically variable over both short 69
and long timescales [11]. For instance, vaginal microbiota shifts throughout a woman’s life, 70
with prepubescent girls and postmenopausal women exhibiting lower levels of *Lactobacillus* 71
dominance compared to women of reproductive age, though their bacterial communities 72
are distinct from the CST IV typically seen during reproductive years [4]. On a short 73
timescale, daily CST fluctuations are observed in some women of reproductive age, while 74
others remain remarkably stable across menstrual cycles, suggesting that diverse factors 75
influence the dynamics of vaginal microbiota communities [12]. For example, menstruation 76
is a key driver of monthly dynamics, while clinical interventions such as antibiotics and 77
probiotics can cause temporary perturbations [4]. 78

A notable gap in the existing literature remains in the understanding of the long-term 79
dynamics of vaginal microbiota in reproductive-aged women across several months. While 80
some studies do follow this timespan, they focus on pregnancy-specific dynamics [13, 14], 81
have large intervals between samples (often exceeding three months) [15], or involve modest 82
sample sizes [16]. These limitations hinder our ability to fully understand the long-term 83
patterns of CST stability and transitions in the general population of reproductive-aged 84

women, and the influence of clinically relevant factors such as demography, lifestyle, sexual 85
practices, and medication. 86

In this study, we introduce an original follow-up cohort of 125 women in Montpellier, 87
France. Our cohort presents a high-resolution longitudinal follow-up study with 2,103 88
microbial samples, spanning a median duration of over 8.6 months and a median of 11 89
samples per woman. We devise a hierarchical Bayesian Markov model to estimate transition 90
probabilities between CSTs, associations between the transitions and 16 relevant covariates, 91
and individual variability among women. 92

Materials and Methods 93

Longitudinal clinical data 94

The samples originated from the PAPCLEAR monocentric longitudinal cohort study, which 95
followed 189 women longitudinally between 2016 and 2020. The participants were recruited 96
through posters and leaflets circulated at the main sexually transmitted infection detection 97
centre (CeGIDD) at the University Hospital of Montpellier (CHU) and at and around 98
university campuses in the city. The inclusion criteria were to be between 18 and 25 years 99
old, to be living in the area of Montpellier, France, to be in good health (no chronic disease), 100
not to have a history of human papillomavirus (HPV) infection (e.g., genital warts or high- 101
grade cervical lesion), and to report at least one new sexual partner over the last 12 months. 102
Additional details about the protocol can be found elsewhere [17]. The longitudinal data 103

analysed in the present study are available at <https://doi.org/10.57745/FHQR9Z>. 104

The inclusion visit was performed by a gynaecologist or a midwife at the CeGIDD 105
outside operating hours. After an interview, several samples were collected, including 106
vaginal swabs with eSwabs (Coppan) in Amies preservation medium from which microbiota 107
barcoding was later performed. The samples were aliquoted right after the visit and stored 108
at -20°C, before being transferred to -70°C within a month. The participants also filled in 109
a detailed questionnaire, which formed the basis of epidemiological covariates analysed in 110
this study. 111

Subsequent on-site visits were scheduled every two or four months, depending on the 112
HPV status. In between on-site visits, women were asked to perform eight self-samples at 113
home with eSwabs in Amies medium and to keep them in their freezer. The self-samples 114
were brought back in an isotherm bag at the next visit. These were then stored with the 115
swab at -70°C until processing. 116

Microbiota metabarcoding and quantification 117

The microbiota metabarcoding was performed on 200µL of vaginal swabs specimen stored 118
at -70° in Amies medium. The DNA extraction was performed using the MagAttract 119
PowerMicrobiome DNA/RNA kit (Qiagen). Next-generation sequencing of the V3-V4 120
region of the 16S gene [18] was performed on an Illumina HiSeq 4000 platform (150 base 121
pairs paired-end mode) at the Genomic Resource Center at the University of Maryland 122
School of Medicine. 123

The taxonomic assignment was performed using the software package SpeciateIT (<https://github.com/Ravel-Laboratory/speciateIT>) and the CSTs were determined using the VALENCIA software package [9]. To examine longitudinal patterns, the present study included participants who contributed at least three samples: 125 women met the inclusion criterion, giving 2,103 samples in total.

Covariates

In the PAPCLEAR study, a questionnaire was given to each participant to record patient-level meta-data. We initially considered the following covariates based on previously proposed roles in influencing the vaginal milieu:

1st menstr. Number of years since the first menstruation: The morphology of the human vagina changes throughout life and the onset of puberty marks a key event that triggers cascading changes [19].

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 [20].

Antibio. Application of antibiotics during the study, either systemic (*Antibio. (Systemic)*) or genital (*Antibio. (Genital)*): The bacterial composition responds rapidly and transiently to antibiotic treatments that target bacteria either broadly or with a narrow taxonomic scale [21].

BMI Body mass index (BMI): Obesity has been implicated in elevating vaginal microbiota 143
diversity and promoting *Prevotella* associated with BV [22]. 144

Caucasian Identity as Caucasian ethnicity or other: Ethnicity has been linked to variation in 145
vaginal microbiota compositions in several studies [6]. However, causal mechanisms 146
remain an open question. 147

Cigarettes Cigarette smoking: Smoking has been implicated in the development of BV due to its 148
anti-estrogenic effects and the presence of harmful substances such as benzo[a]pyrene 149
diol epoxide (BPDE) [23]. 150

Horm. contra. Use of hormonal contraception during the study: The vaginal hormonal landscape is 151
affected by the use of hormonal contraceptives [24]. 152

Lubricant Use of lubricant during the study: Personal lubricants contain various chemicals that 153
differentially impact the growth of vaginal microbes in-vitro [25]. 154

Menstr. cup Use of menstrual cups during the study: The vaginal microenvironment may be 155
altered by the use of menstrual cups both physically and chemically. An elevated 156
risk of fungal infections has been reported [26]. 157

Partners Cumulative number of sexual partners: The genital microbiome can be transferred 158
between sexual partners [27]. Such an external input could destabilise the resident 159
community. 160

Red meat Average number of meals that include red meat consumption per week: Diet alters 161
the vaginal environment for microbes. An unhealthy diet, linked to a high proportion 162
of red meat consumption, has been linked to an elevated risk of BV [28]. 163

Regular condom Regular use of condoms during sexual intercourse: Condom use can modify the vagi- 164
nal microenvironment by altering the exchange of microbes between partners [29]. 165

Regular sport Engaging in regular sporting activities, over 50% of the time: Physical activities 166
influence immune responses, with leisure-time physical activity associated with a 167
reduced risk of suspected bacterial infections compared to sedentary behaviour [30]. 168

Stress Average stress level reported from 0 (min) to 3 (max): Stress hormones may disrupt 169
vaginal flora, for instance, by inhibiting glycogen production, which is the primary 170
fuel for lactobacilli [31]. 171

Tampon Use of tampons during the study: The use of internal menstrual health products 172
like tampons directly alters the vaginal environment, although negative effects from 173
tampon use are seldom reported [32]. 174

Vag. product Use of vaginal cream/tablet/capsule/gel/wipe during the study: Women frequently 175
use over-the-counter vulvovaginal treatments that contain a variety of chemical com- 176
ponents. However, the clinical effectiveness of these products in preventing BV is 177
seldom systematically evaluated [33]. 178

Chlamydia Tested positive for chlamydia. 179

Female/male affinity Affinity to female/male partner: Genital microbiome transfers during sexual activity 180

are anticipated to vary based on the genders of the partners [34]. 181

Pregnancy History of pregnancy: Pregnancy significantly changes the cervicovaginal environ- 182

ment, with increased estrogen from the ovaries and placenta leading to higher vaginal 183

glycogen. This supports the growth of *Lactobacillus* species [35]. 184

Spermicide Use of spermicide during the study: Spermicides are chemicals that prevent sperm 185

from reaching an egg, but their use can change the vaginal microflora, potentially 186

increasing the risk of genitourinary infections [36]. 187

Vag. douching Use of vaginal douching during study: Vaginal douching, the practice of washing in- 188

side the vagina with a liquid solution, has been shown to increase the risk of disturbing 189

the natural balance of vaginal flora [37]. 190

Out of the covariates initially considered above, we excluded six (*Chlamydia*, *Female* 191
affinity, *Male affinity*, *Pregnancy*, *Spermicide* and *Vag. douching*) as data were severely 192
skewed towards the most common value (> 90% of data). During the study, any use of 193
antibiotics was recorded with the date and we distinguished systemic (*Antibio. (Systemic)*) 194
and genital topical (*Antibio. (Genital)*) applications, corresponding to ‘Gynecological anti- 195
infectives and antiseptics’ (‘G01’ ATC codes), which consisted of metronidazole treatments, 196
and ‘Antibacterials for systemic use’ (‘J01’ ATC codes), which were more diverse. Since 197
the exact dates of treatment were recorded, *Antibio. (Systemic)* and *Antibio. (Genital)* 198
were included as time-inhomogenous covariates in the model. All other covariates were 199

considered time-homogeneous meaning that the variation is among women, and static 200
through time because the precise timing of changes in the covariate values was unknown. 201

To facilitate the comparison of covariate effects, we centred and scaled continuous vari- 202
ables [38] and deviation-coded binary variables. These transformations ensure that all 203
covariates are modelled in a comparable scale and the intercept is located at a “representa- 204
tive reference value” of the modelled population: i.e., the population mean for continuous 205
and the theoretical mid-point for binary values. Four continuous covariates (i.e., *Alcohol*, 206
BMI, *Partners*, and *Red meat*) were log-transformed before scaling due to their right- 207
skewed distribution. We found no strong correlations among the covariates included in the 208
analysis (Supplementary Information S1). 209

Modelling 210

Markov model 211

Markov models are statistical models used to represent systems that transition between 212
discrete states over time. These models are ‘memoryless’, meaning that the probability of 213
transition to another state depends on the current state, but not its historical path. In 214
clinical research, these models are often used to predict the transitions among health states 215
(e.g., health, illness and remission), and the propensity to transition between these states is 216
estimated from longitudinal follow-up data. Clinical follow-up data are typically modelled 217
using the continuous-time Markov model [39], in which the probability of transition over 218

a given interval depends on the instantaneous transition intensity and the amount of time 219
spent in the current state. 220

Vaginal microbiota state transitions are classically studied using continuous-time Markov 221
models [13–15, 40, 41]. Our application of the continuous-time Markov model differs from 222
those of the existing literature in its hierarchical Bayesian formulation, which allowed us 223
to quantify individual variability among women (as unobserved heterogeneity, or random 224
effects) and to estimate many covariate effects simultaneously (through the use of weakly 225
informative priors). 226

Transition intensities 227

Transition intensities, q , refer to the instantaneous rate of moving from state i to state j
in a participant p (e.g., CST I to CST IV), 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:

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

where $\mu_{p,i,j}$ is the intercept and $\beta_{i,j}$ is the coefficient expressing the impact of a covari-
ate(s). This intercept is further defined by the equation,

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

where $\bar{\mu}$ and μ_{sd} are the prior mean and standard deviation of the intercept such that 228
 $\hat{\mu}_{i,j} \cdot \mu_{sd} + \bar{\mu}$ constitutes the non-centred parameterisation of the population-level intercept, 229
 $\mu_{i,j}$ and is assumed to be normally distributed, i.e., $\hat{\mu}_{i,j} \sim \mathcal{N}(0, 1)$. 230

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

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

We sampled from the corresponding weakly informative priors, namely $sd_s \sim t_4(0, 1)$, 231
 $L_s \sim \text{LKJCorrCholesky}(2)$ (which slightly favours correlations among unobserved hetero- 232
geneity closer to zero, reducing the likelihood of extreme positive or negative correlations), 233
and $z_s \sim \mathcal{N}(0, 1)$, as recommended by the Stan development community [42, 43]. The 234
multivariate normal density and the LKJ prior require the matrix parameters to be de- 235
composed, which can be computationally intensive if done repeatedly. To ensure compu- 236
tational efficiency and numerical stability, the model was directly parameterised using the 237
Cholesky factors of correlation matrices. This approach uses a multivariate version of the 238
non-centred parameterisation. 239

For regression coefficients, the Student-t distributions with degrees of freedom 4 to 7 are 240
recommended as generic, weakly informative, priors [43]: we sampled β from $\beta \sim t_4(0, 1)$, 241
which places a comparatively wide tail within the recommendation. As all of our covariates 242
have been proposed to impact vaginal microbiota communities *a priori* (see above), we did 243
not strongly regularise the priors, for example, through the use of horseshoe priors [44]. 244

We note that all covariates were modelled simultaneously, such that the interpretation of 245
each coefficient is conditional upon other covariates included and accounts for the influence 246
of other factors. We assumed that the covariates affect the transitions symmetrically (i.e., 247
 $\beta_{j,i} = -\beta_{i,j}$), meaning that the influence of a covariate on the affinity (or aversion) towards 248
a particular CST is consistent, regardless of the direction of the transition. 249

Collectively, the transition intensities form the matrix, Q_p , in which the sum of inten-
sities 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 [39]:

$$q_{p,i,i} = - \sum_{j \neq i} q_{p,i,j}. \quad (4)$$

Transition probabilities and likelihood 250

Taking the matrix exponential of the Q_p matrix for each participant, p , we compute the
matrix P_p such that:

$$P_p = \text{Exp}((t_{k+1} - t_k) Q_p), \quad (5)$$

where k represents the sample identity for a given individual. The P_p matrix contains 251
the transition probabilities between two observations (at k and $k+1$) and $t_{k+1} - t_k$ indicates 252
the elapsed time between two observations. 253

Finally, the probability of observing a given state at the next sampling event (i.e., at

$k + 1$) is modelled by the categorical distribution:

$$y_{k+1} \sim \text{Categorical}(P_p[y_k,]) \quad (6)$$

where $P_p[y_k,]$ is the y_k^{th} row of the P_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. In total, the model consisted of 57 parameters and 12 hyper-parameters. Our model was written in Stan 2.26.1 and fitted through the RStan interface 2.32.3 [45]. The Stan programme is available at <https://doi.org/10.57745/FHQR9Z>.

One participant lacked information on the years since their initial menstruation. We imputed missing values using the mice package [46] and generated 10 imputed datasets to be fitted separately. For each imputed dataset, we fitted the model in parallel using four independent chains, each with 10,000 sampled iterations and 1,000 warm-up iterations. The MCMC samples from separate runs (i.e., based on differently imputed data) were subsequently combined for inference.

We confirmed over 1,000 effective samples per imputed dataset and ensured convergence of independent chains ($\hat{R} < 1.01$) for all parameters [47]. We carried out a posterior predictive check by comparing the observed and predicted CST frequency. We also quanti-

fied the posterior z -score and posterior contraction to examine the accuracy and precision 270
of posterior distributions and the relative strength of data to prior information [48] (Sup- 271
plementary Information S2). 272

Counterfactual predictions 273

We took advantage of the parameterised model to simulate the population-level outcomes 274
of each covariate, assuming that all covariates, but a focal one, are at the representative 275
reference value (as described above) and then varying the focal parameter within the range 276
of values observed in the studied cohort. The model predictions were generated by ran- 277
domly drawing 100 samples from the posterior distributions and simulating the Markov 278
model for each sampled parameter set. We focused on the CST frequency as the outcome 279
of interest. 280

Results and Discussion 281

CSTs in the cohort 282

As is typical of vaginal microbiota communities, the microbial compositions sampled in 283
PAPCLEAR were highly structured, and characterised by a relatively small number of 284
operational taxonomic units (OTUs). The dominant species within these communities 285
aligned closely with specific community state types (CSTs) as defined by Ravel et al. [6]. 286
For example, CST I was primarily associated with *L. crispatus* and CST III with *L. iners*. 287

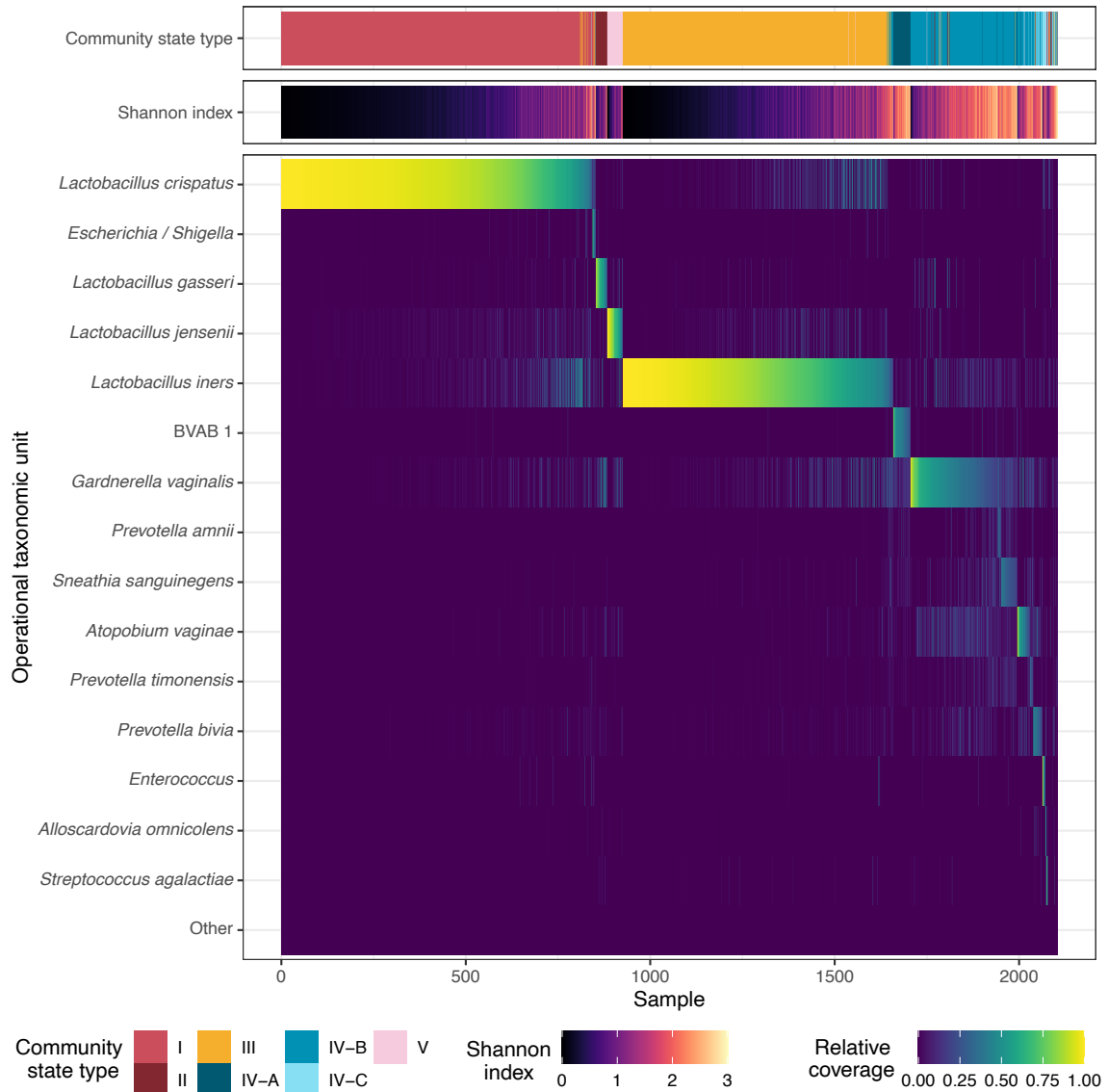


Figure 1: **Vaginal community state types (CSTs), diversity (Shannon Index), and relative coverage of the 15 most common taxonomic operational units (OTU) of 2,103 samples from the PAPCEAR cohort.** In over 98.5% of samples, a single of these 15 OTUs represented the most common OTU.

In contrast, and as expected, CST IV communities exhibited a higher degree of microbial 288
diversity compared to CSTs dominated by lactobacilli, reflecting a broader range of species 289
typical of this community type (Fig. 1). 290

Our longitudinal dataset from the PAPCLEAR cohort represents one of the largest 291
analysed to date in the context of the vaginal microbiota. Detailed participant character- 292
istics are presented in Table 1. Briefly, the participants were between 18 and 25 years old 293
and the majority of the 2,103 samples (73.7%) were self-collected at home, the rest being 294
collected during on-site visits (Fig. 2a). The median follow-up duration was 8.64 months 295
and the most common intervals between analysed samples were seven and 28 days (Fig. 2a 296
& b). On average, each of the 125 participants contributed 11 samples (Fig. 2c). 297

The metabarcoding analysis on 16S RNA with the VALENCIA algorithm [9] was used 298
to assign each sample to a CST. The vaginal microbiota communities were variable across 299
women and over time (Fig. 2d). As CSTs I, II, and V are all dominated by lactobacilli 300
and considered ‘optimal’ in terms of health, yet the latter two are rare (~4% of all sam- 301
ples combined), we pooled the three optimal communities for further investigation. Over- 302
all, optimal communities were the most frequent, representing 44.5% of samples, followed 303
by ‘sub-optimal’ (CST III) at 35.2% and ‘non-optimal’ communities (CST IV) at 20.4% 304
(Fig. 2e and Table 1). 305

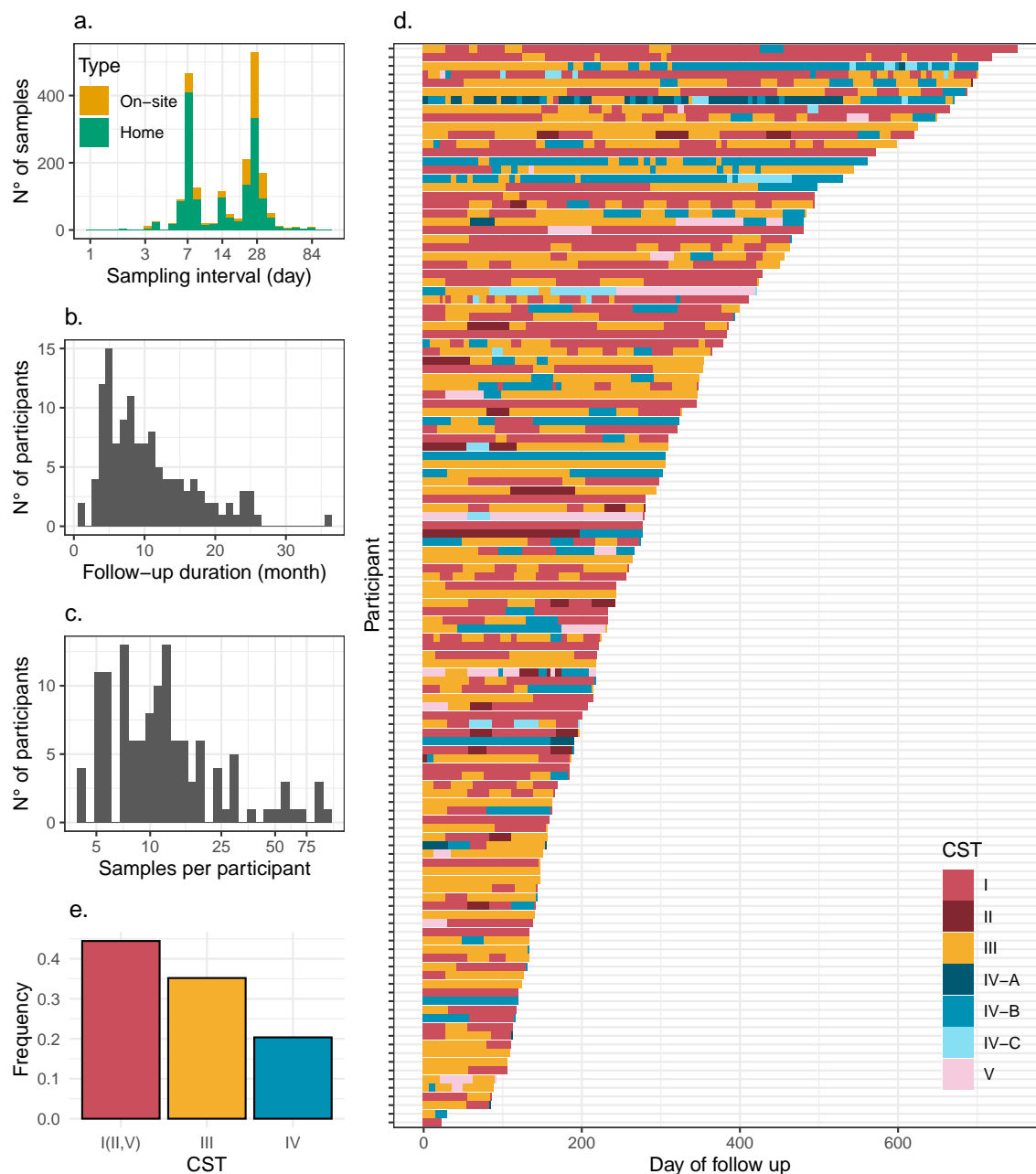


Figure 2: **Summary of vaginal microbiota samples analysed in the PAPCLEAR study.** a) Intervals between sampling events for clinical (i.e., on-site) and home samples. b) Follow-up duration per participant. c) Number of samples analysed per participant. d) Vaginal microbiota Community State Types (CST) over time in 125 participants. For visualisation, data are truncated at 750 days for a single individual whose duration exceeds this threshold. e) Frequency of the optimal (i.e., CSTs I, II, and V combined), sub-optimal (CST III) and non-optimal (CST IV) communities in all samples.

Probabilities of CST persistence

306

We implemented a continuous-time Markov model to capture the CST dynamics. Simulations based on the estimated parameters of our model (i.e., posterior predictive check) confirmed that it accurately captures the observed CST prevalence (Fig. 3a). The optimal,

Table 1: Summary profile of vaginal 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 Materials and Methods for the covariate definitions.

	Level	Summary
Samples (Participants)		2103 (125)
CST (%)	I	847 (40.3)
	II	39 (1.85)
	III	740 (35.2)
	IV-A	54 (2.57)
	IV-B	342 (16.3)
	IV-C	32 (1.52)
	V	49 (2.33)
Sample type (%)	On-site	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)
<i>Covariates</i>		
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 menstruation (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 (%)	1	65 (52.0)
Genital antibiotic treatment (%)	1	30 (24.0)

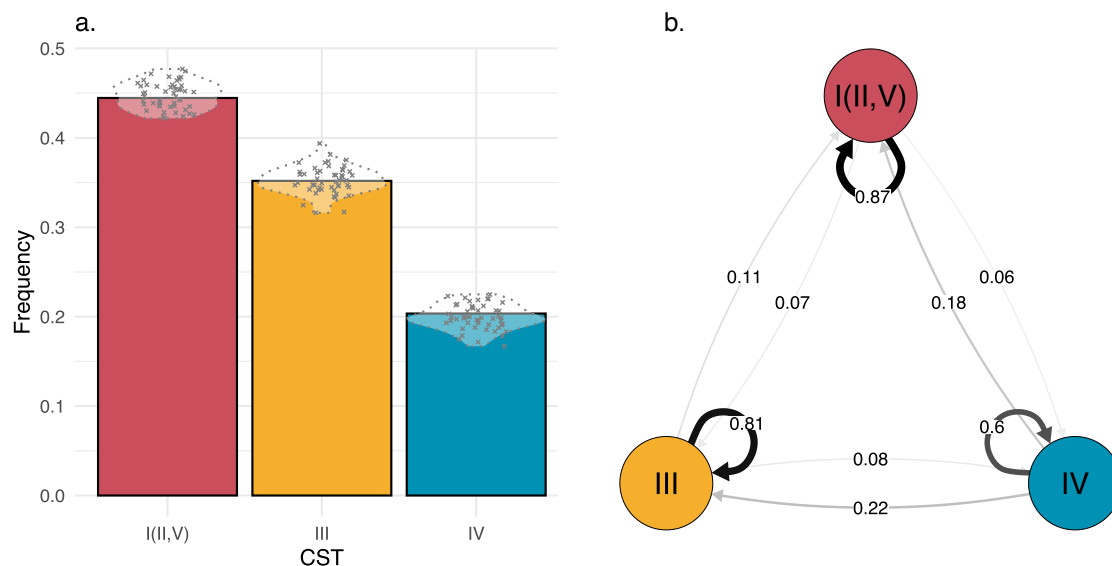


Figure 3: Prevalence and transition probabilities among vaginal microbiota community state types (CSTs). a) Observed (bars) and predicted prevalence (crosses) of CSTs I (II, V), III and IV. The model predictions were generated by drawing 100 random samples from the posterior distributions and simulating 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.

CST I (II, V), and sub-optimal, CST III, communities showed a high degree of stability, 310
with weekly probabilities to remain in the current state estimated at 87% (95% credibility 311
interval (95CrI): 78 - 93%) and 81% (95CrI: 68 - 90%), respectively (Fig. 3b). In contrast, 312
the weekly persistence probability of the non-optimal CST IV was 60% (95CrI: 35 - 80%, 313
Fig. 3b). These transition probabilities translate into sojourn times (i.e., the expected time 314
spent in a given state before moving to another) in CST I (II, V), III and IV of 6.9 days 315
(95CrI: 2.9 - 13.6 days), 4.23 days (95CrI: 1.8 - 8.4 days) and 1.6 days (95CrI: 0.58 - 3.8 316
days), respectively. 317

The reported persistence and transition probabilities in the literature vary widely based 318
on the cohort characteristics. For example, focusing on women during pregnancy, DiGiulio 319
et al. [13] estimated that the four *Lactobacillus*-dominated CSTs (CSTs I, II, III, and 320
V) were more stable than CST IV. Notably, both CST I and II showed 98% probability 321
of weekly persistence. The enhanced persistence of *Lactobacillus*-dominated communities 322
during pregnancy owes itself to specific vaginal conditions during pregnancy including the 323
up-regulation of oestrogen and progesterone that facilitates lactobacilli [13, 49]. 324

In addition, the temporal dynamics of vaginal microbiota are notably different in women 325
with BV. In contrast to pregnant women, those experiencing symptomatic BV generally 326
exhibit less stable vaginal microbiota communities. In the cohort of Ravel et al. [50], 327
which focused on women with symptomatic BV, Brooks et al. [40] found significantly lower 328
stability across all CSTs. The probability of these CSTs persisting ranged from 38% to 329
48%, with CST I persisting only 46% of the time over a week. 330

Among studies that focused on non-pregnant, healthy young women — with no partic- 331
ular emphasis on BV — the analysis by Brooks et al. [40] of the Chaban et al. cohort [16] 332
(N = 27; Canada) estimated weekly persistence probabilities of 75% for CST I, 78% for III, 333
60% for IV-A, and 88% for V. In the Gajer et al. dataset [12] (N = 32; USA), analysed again 334
by Brooks et al., [40], CST I, II and III demonstrated 72%, 84% and 77% weekly persis- 335
tence probabilities, respectively. In this dataset, CST IV sub-categories showed markedly 336
different stability with CST IV-A with weekly persistence of 38% and CST IV-B with 337
persistence of 82%. A third study, Munoz et al. [15] (N = 88; South Africa), reported 338

the stability of vaginal microbiota in women in a three-month time frame using a differ- 339
ent microbiota classification system consisting of four categories predominantly associated 340
with: *L. crispatus* (similar to CST I), *L. iners* (similar to CST III), *G. vaginalis* (similar 341
to CST IV), or *Prevotella* spp. (similar to CST IV). They found similar persistence for 342
CST I and CST IV-like communities ranging from 51 to 53% over three months while the 343
CST III-like community was more stable at 62% over the same period. Recasting in the 344
three-month time scale, our estimates show the same extent of stability for CST I(II, V) at 345
51% (95% CrI: 29-72%) while CST III (38%, 95% CrI: 19-61%) and CST IV (15%, 95% CrI: 346
5-34%) were less stable. Taken together, our estimates of vaginal microbiota community 347
stability are within the range of values reported in other cohorts. However, the dynamics 348
of vaginal microbiota communities are likely geographically variable even among healthy 349
young women. 350

Covariate effects on transitions

 351

The Bayesian approach, which can accommodate vaguely informative priors on the co- 352
variate effects, allows for the simultaneous inclusion of many covariates (as hazard ratios; 353
Eq. 1) which would otherwise prove difficult in Markov models [39]. We identified 16 co- 354
variates based on previously proposed roles in influencing the vaginal milieu and assumed 355
that covariates have a symmetrical effect on CST transitions: e.g., the magnitude of a given 356
covariate effect on the transition from CST I to III is identical to that on the transition 357
from CST III to I. We identified alcohol consumption as the strongest and most consistent 358

effect while several other covariates were identified as possible drivers of CST transitions. 359

Alcohol consumption 360

The estimated hazard ratios on community transitions indicate that alcohol consumption 361
favoured the sub-optimal (CST III) community over optimal (CST I(II, V)) with 97% 362
probability (Fig. 4). Because of our symmetry assumption, this can mean that alcohol 363
consumption increases the pace of transition from CST I(II, V) to CST III or reduces that 364
in the opposite direction by the same magnitude. Alcohol consumption also tended to 365
favour CST IV over CST III, although with a lower credibility level (with 73% probability 366
of the hazard ratio $\neq 1$, Fig. 4). 367

To examine how these effects translate to the population level, we carried out counter- 368
factual simulations in which all participant characteristics were set to the representative 369
value observed in the studied cohort, except for alcohol consumption, which ranged from 370
non-drinking to the level of the heaviest drinking observed in our cohort (19 drinks per 371
week). The simulations demonstrated that the expected prevalence of the optimal (CST I 372
(II, V)) community was 18% (95% CrI of 9 to 27%) higher in a hypothetical population 373
of non-drinkers compared to that of average-level drinkers who consumed three drinks per 374
week (Fig. 5; Supplementary Information S3). In turn, the prevalence of the optimal com- 375
munity was 19% (95% CrI of 10 to 29%) higher in the population of average-level drinkers 376
than in the heaviest drinkers. As the optimal community declined with alcohol consump- 377
tion, the prevalence of the non-optimal (CST IV) community was found to be 9% (95% CrI 378

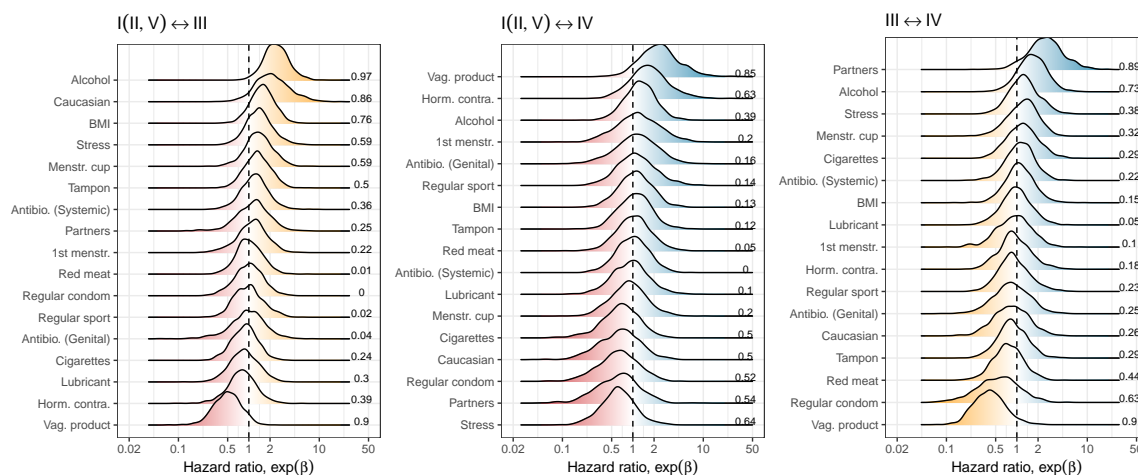


Figure 4: **Estimated covariate effects on community transition rates.** With the symmetry assumption, there are three sets of covariate effects on transitions. The impact of covariates on community transition rates was estimated for a given set of community states as the log hazard ratio, β . The figure 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). For example, alcohol consumption was estimated to favour CST III over CST I (II,V) at a credibility level of 97%.

of 2 to 15%) higher among average drinkers compared to non-drinkers. Therefore, while the 379
strongest impact of alcohol on community transitions appears to be between the optimal 380
(CST I (II, V)) and sub-optimal (CST III) communities, an additional, non-zero impact 381
on the sub-optimal to non-optimal (CST IV) transition means that alcohol consumption 382
ultimately promotes non-optimal communities at the expense of optimal ones. As the ef- 383
fects of covariates are estimated simultaneously, potential confounding factors, including 384
the number of partners, condom use and smoking, are controlled for in our findings. 385

Alcohol consumption may impact the vaginal microbiota through a variety of mecha- 386

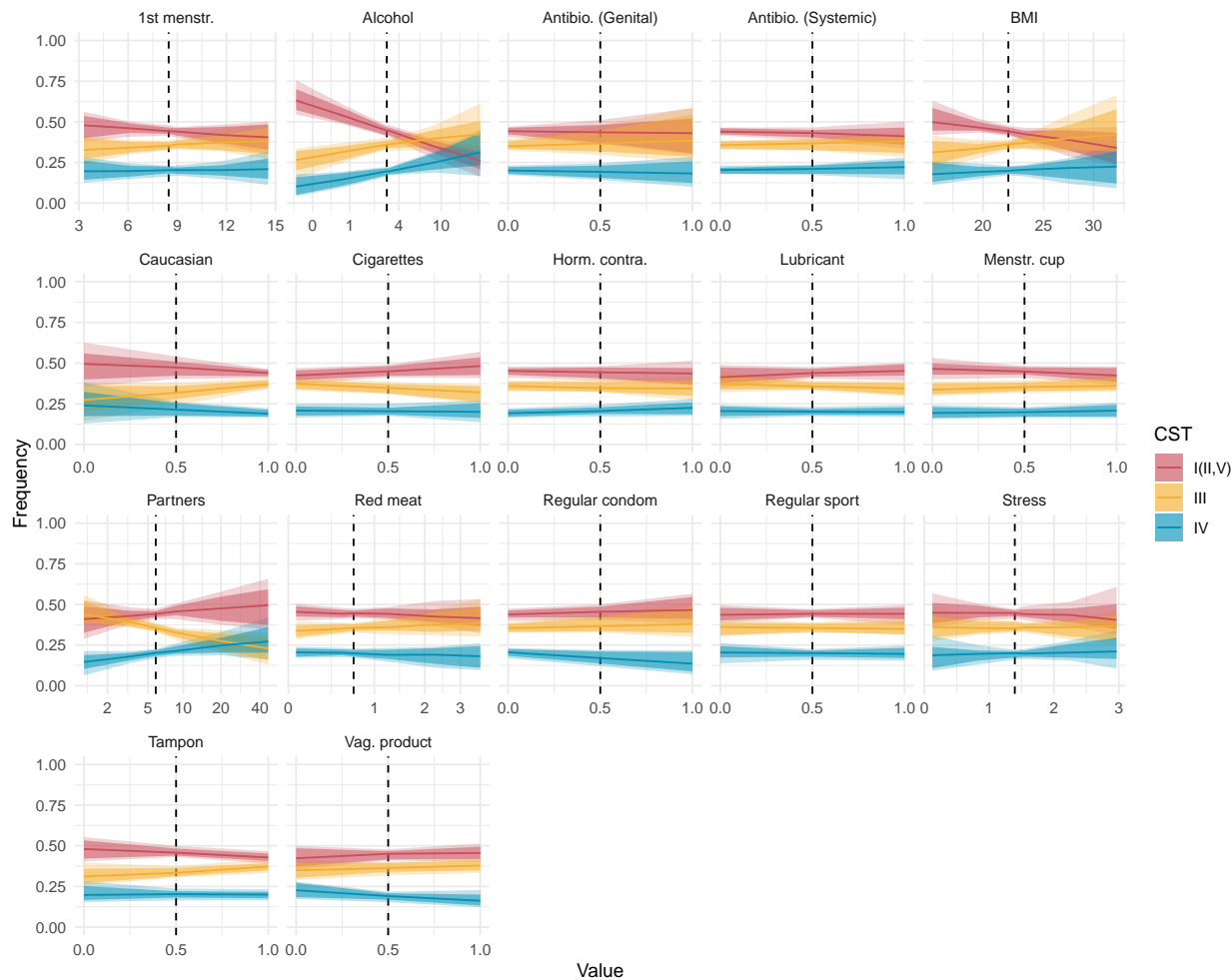


Figure 5: Counter-factual simulations predict population-level consequences of covariates. Based on estimated hazard ratios (Fig. 4), the population-level impact (i.e., the prevalence of CST I (II,V), III and IV) was simulated for each covariate. The vertical dashed lines indicate the intercept used in estimation: i.e., the population mean for continuous and midpoint for binary variables. For continuous variables, the range of values explored was determined by the minimum and maximum values reported in the PAPER CLEAR study.

nisms. Physiologically, the chronic presence of alcohol in the genital environment has been 387
 linked to a shift in immune and microbiological conditions [20]. In addition, alcohol is a 388

known modifier of sexual behaviour, which in turn has been demonstrated to increase the 389
risk of BV, linked to CST IV [51]. Finally, alcohol alters the microbial profile in other 390
body parts, which in turn could cross over to the vaginal milieu. For example, *Prevotella*, 391
a genus commonly found in CST IV communities, is enriched in the oral microbiota of 392
drinkers [52]. Similarly, others postulate the effect of alcohol on the gut microbiota may 393
have a concurrent influence on the vaginal microbiota [53]. 394

While there remains a lack of consensus among existing studies (briefly reviewed by 395
Froehle et al. [53]), cohort and cross-sectional studies from diverse geographical contexts 396
(namely, Australia, Denmark, Sweden, Thailand, Tanzania, Uganda and USA) have pre- 397
viously reported an association between alcohol consumption and BV [53–61]. In addition 398
to corroborating these findings, our Markov model offers a novel insight into the ecology of 399
microbial communities underlying these observations: alcohol consumption destabilises the 400
optimal (CST I (II, V)) communities towards sub-optimal (CST III), which opens the gate 401
for the deterioration towards non-optimal (CST IV), associated with BV. To the authors’ 402
knowledge, there have been no alcohol cessation studies reporting its impact on vaginal 403
microbiota. Such studies are necessary to establish causal links, similar to those conducted 404
on the effects of smoking [23], douching [62], and antibiotics [21] on vaginal microbiota 405
compositions. 406

Potential effects of other covariates

407

Other factors with possible effects on transitions (i.e., with more than 80% probability of 408
hazard ratio $\neq 1$) included the use of vaginal intimate hygiene products, number of sexual 409
partners and self-reported ‘Caucasian’ identity. 410

Vaginal hygiene products: The use of vaginal hygiene products, defined broadly 411
here to include vaginal cream, tablet, capsule, gel and wipe, appeared to have multifaceted 412
effects. Between CST I (II, V) and CST III, their use was positively linked to maintaining 413
or transitioning to CST I (II, V) with 90% probability (Fig. 4). For the CST I (II, V) and 414
CST IV pair, it tended to favour a shift towards CST IV, with 85% probability. Finally, 415
between CST III and CST IV, their use was more likely to support the persistence or a 416
shift towards CST III, also with 90% probability. The circular effects suggest that women 417
may experience different effects of the products marketed for ‘vaginal intimate hygiene’ 418
depending on the predisposition with certain CSTs. Nonetheless, the circular effects on 419
community transitions meant that there was no noticeable impact at the population level 420
in our counterfactual simulations (Fig. 5). 421

Number of sexual partners: A higher number of sexual partners was also found 422
to potentially favour CST IV over CST III, increasing the risk of maintaining (or transi- 423
tioning to) CST IV with 89% probability of the hazard ratio $\neq 1$. The association between 424
CST IV and the lifetime number of partners is consistent with the hypothesis that external 425
importation of microbes could alter the dynamics of vaginal microbiota and is in line with 426

earlier work [63, 64]. Population-level simulations predict that an increasing number of 427
sexual partners tends to reduce the prevalence of the sub-optimal (CST III) community. 428
For example, CST III was 13% (95% CrI of 2 to 21%) less common in a hypothetical pop- 429
ulation with the highest number of partners than one conforming to the average number. 430
The decrease was accompanied by a tendency for the other CSTs to increase, although the 431
trend was less clear for CST I(II,V) and CST IV, individually (Fig. 5). 432

Causasian identity: It is worth noting that our cohort was not designed to achieve 433
comprehensive coverage of self-reported ethnic identity, with over 80% identifying as Cau- 434
casians (Table 1). Nonetheless, identifying oneself as a ‘Caucasian’ tended to favour CST III 435
over CST I(II, V) with 86% probability. European studies focusing on the role of ethnicity 436
are rare. However, a North American study has observed a qualitatively opposite trend: 437
CST III communities are comparably rare in women who identify themselves as Caucasian 438
compared to those identifying as Asian, Black and Hispanic (26.8 versus 42.7, 31.4 and 439
36.1%, respectively [6]). While previous studies have revealed differences in vaginal mi- 440
crobiota compositions among ethnic groups, the relative importance of biological, societal, 441
and environmental factors remains an open question [6, 65–67]. 442

Antibiotics: Notably, we found little association between antibiotic consumption and 443
CST transitions, neither for local treatment for BV (genital application of metronidazole) 444
nor systemic treatment (antibiotic treatment via oral intake). Such a lack of effect in 445
our study may be because the changes in the vaginal microbiota compositions following an 446
antibiotic treatment take place in a shorter time scale than our sampling intervals: the most 447

common sampling intervals were either 7 or 28 days (Table 1). In comparison, Brooks et al. [40] found rapid CST transition following BV medication in the cohort of Ravel et al. [50], which involved daily sampling. On a longer time scale, the re-emergence of BV-associated communities following treatments is a well-documented clinical challenge [68–70].

Unobserved individual variability in community transition

While we incorporated 16 covariates into our Markov model, some variations among women remain unaccounted for. To quantify these, we estimated the extent of individual variability (i.e., unobserved heterogeneity, or random effects) in community transitions for each transition pair using a hierarchical Bayesian approach (Eq.2 & 3).

The highest variability was observed among women in the transitions involving ‘recovery’ to an optimal (CST I (II,V)) from a non-optimal (CST IV) state (Fig. 6). On the other hand, inverse transitions from optimal to non-optimal 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. These findings suggest that there are relatively limited pathways leading to the deterioration of vaginal microbiota communities, whereas the routes to recovery can be more individualised and the source of this variation remains to be fully elucidated.

The presence of individual-level random effects indicates that a considerable part of the variability remains unaccounted for by the 16 covariates in this study. One possible cause is that our study left out key drivers of the vaginal milieu. For example, while menstrual cycles

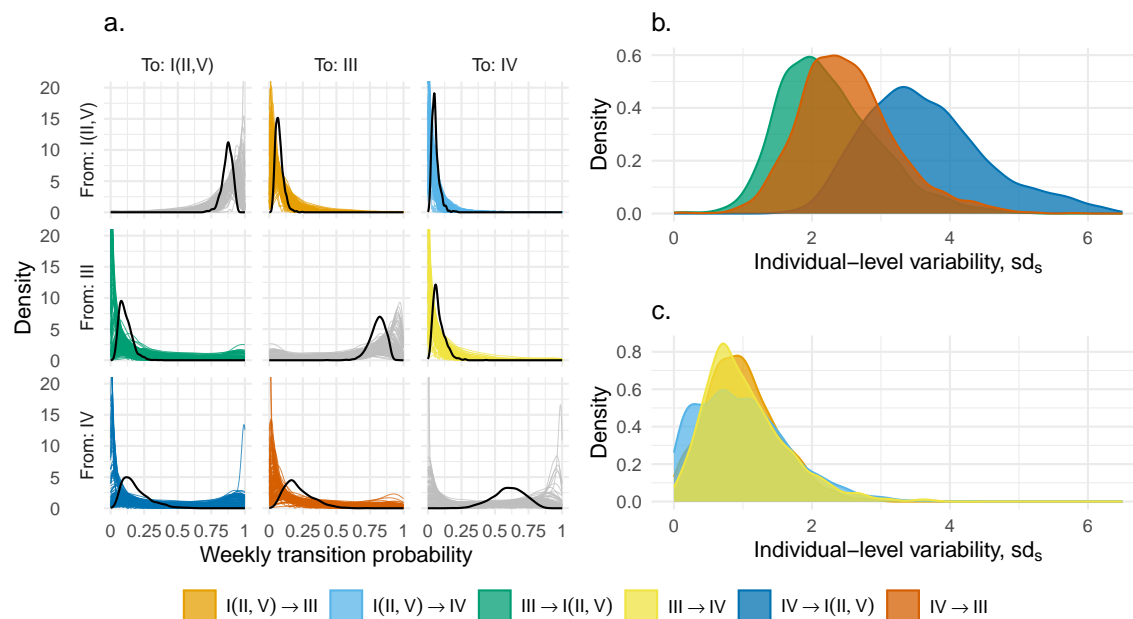


Figure 6: **Individual-level variability in vaginal microbiota community state type (CST) transitions.** a) The population average (thick black) and individual (thinner colours) weekly transition probabilities. Between-women individual variation for transitions to b) a more optimal and c) a less optimal state. Colours indicate the type of transition between CSTs.

have been demonstrated to influence daily and weekly transitions [12], they were omitted 468
 from our analysis because the timing of menstruation was ambiguous in the PAPCLEAR 469
 study. Furthermore, while large-scale longitudinal studies present logistical challenges, a 470
 citizen-science-based approach offers the potential for expanding the cohort size, thereby 471
 enhancing the statistical power needed to examine additional covariates [71]. Secondly, 472
 further resolution on individual variability may be gained by incorporating time-varying 473
 covariates, which could accommodate changes in participant behaviours during the follow- 474
 up. In continuous-time Markov models, time-varying covariates are assumed piecewise 475

constant, meaning they are constant between sampling events [39]. Such an assumption is 476
convenient as covariate values are rarely known between sampling events. Without precise 477
knowledge of the timing of the covariate changes, however, it is unclear whether the previous 478
covariate value (at $t - 1$) or the new covariate value (at t) should influence the transition. 479
Consequently, our analysis focused on static covariates, with the exception of antibiotic 480
treatments for which the exact application dates were known. Aggregating participant 481
behaviours as static covariates eliminated the uncertainty of covariate dynamics, albeit at 482
a potentially lost opportunity for further precision. 483

Limitations and opportunities 484

A potential limitation from a clinical methodological perspective is that the majority of 485
samples were collected at home during the PAPCLEAR study. While home sampling could 486
introduce variability, the participants were provided with detailed instructions to minimise 487
the difference in swabbing techniques between on-site and home samples, and we verified 488
consistency in sampling dates by having participants fill out online questionnaires during 489
sampling. 490

Another possible limitation of our study is the resolution of microbiota community 491
classification. We focused on three CST groups with varying health implications: optimal 492
(CST I (II,V)), sub-optimal (III), and non-optimal (IV). This decision stemmed from the 493
fact that detailed classifications in a Markov model would increase the number of possible 494
transitions, and it would be difficult to estimate transitions between rare types. However, 495

significant functional differences may exist within these CSTs. For instance, the VALEN- 496
CIA algorithm classifies subcategories within some CSTs [9], and Brooks et al. demon- 497
strated that CST IV-B is more stable than CST IV-A [40]. We also note that there are 498
several clustering algorithms of microbial communities besides the CST framework [71,72], 499
which may offer differing insights on community transitions. Furthermore, the centroid 500
distance computed by VALENCIA for CST assignment may also be leveraged to develop a 501
quantitative, multi-dimensional perspective of the vaginal microbiota communities. Such 502
a quantitative perspective may enhance our understanding of within-CST variabilities — al- 503
though we are unaware of an existing approach that accommodates the temporal patterns 504
in such data. Finally, the metagenomics approach holds the promise to uncover within- 505
species diversities: e.g., metagenomics CSTs (MgCSTs) have identified with 25 distinct 506
communities [73]. Such an approach helps to identify lineage replacements in women with 507
stable CSTs and investigating the impact of antibiotic treatments on the prevalence of 508
resistance genes could yield insights into the within-species dynamics of vaginal microbes. 509

A promising direction for future research is the joint analysis of CST dynamics and 510
sexually transmitted infections such as HPV. Previous studies have found a weak associ- 511
ation between CST IV and HPV detection risk [74]. However, these studies tested the 512
CST effect after estimating transition rates and pooled all high-risk and low-risk HPVs, 513
making it difficult to identify coinfections or reinfections. The PAPCLEAR cohort, with 514
genotype-specific follow-ups, could provide new insights into the link between CST and 515
HPV infection, potentially identifying causal relationships. 516

Conclusion

517

We showcased a novel application of a hierarchical Bayesian Markov model to original 518
clinical cohort data of vaginal microbiota dynamics. Our approach facilitated the simulta- 519
neous estimation of several covariate effects on community transitions and the identification 520
of unobserved variability in these transitions. Our work paves the way for an improved 521
ecological understanding of microbial dynamics within the vaginal environment and indi- 522
cates lifestyle alterations (such as reduced alcohol consumption) that may promote vaginal 523
health. 524

Ethics

525

This study has been approved by the Comité de Protection des Personnes (CPP) Sud 526
Méditerranée I (reference number 2016-A00712-49); by the Comité Consultatif sur le Traite- 527
ment de l'Information en matière de Recherche dans le domaine de la Santé (reference 528
number 16.504); by the Commission Nationale Informatique et Libertés (reference num- 529
ber MMS/ABD/ AR1612278, decision number DR-2016-488), by the Agence Nationale 530
de Sécurité du Médicament et des Produits de Santé (reference 20160072000007), and is 531
registered at ClinicalTrials.gov under the ID NCT02946346. 532

Funding

533

This project has received funding from the European Research Council (ERC) under the 534
European Union’s Horizon 2020 research and innovation programme (grant agreement No 535
648963, to SA). The authors acknowledge further support from the Fondation pour la 536
Recherche Medicale (to TK), the Agence Nationale de la Recherche contre le SIDA (ANRS- 537
MIE, to NT), and the MemoLife Labex (to BE). 538

Acknowledgements

539

The authors thank Olivier Supplisson for his helpful feedback. We acknowledge the ISO 540
9001 certified IRD i-Trop HPC (member of the South Green Platform) at IRD Montpel- 541
lier for providing HPC resources that have contributed to the research results reported 542
within this article (bioinfo.ird.fr and www.southgreen.fr). DNA extracts were (partly) per- 543
formed through the genotyping and sequencing facilities of ISEM (Institut des Sciences de 544
l’Evolution-Montpellier) and Labex Centre Méditerranéen Environnement Biodiversité. 545

References

546

- [1] van de Wijert JHHM. The vaginal microbiome and sexually transmitted infec- 547
tions are interlinked: Consequences for treatment and prevention. PLOS Medicine. 548
2017;14(12):e1002478. 549

- [2] Haahr T, Zacho J, Bräuner M, Shathmigha K, Skov Jensen J, Humaidan P. Repro- 550
ductive outcome of patients undergoing in vitro fertilisation treatment and diagnosed 551
with bacterial vaginosis or abnormal vaginal microbiota: a systematic PRISMA review 552
and meta-analysis. *BJOG*. 2019;126(2):200–207. 553
- [3] Bilardi JE, Walker S, Temple-Smith M, McNair R, Mooney-Somers J, Bellhouse C, 554
et al. The burden of bacterial vaginosis: women’s experience of the physical, emotional, 555
sexual and social impact of living with recurrent bacterial vaginosis. *PLOS ONE*. 556
2013;8(9):e74378. 557
- [4] France M, Alizadeh M, Brown S, Ma B, Ravel J. Towards a deeper understanding of 558
the vaginal microbiota. *Nature Microbiology*. 2022;7(3):367–378. 559
- [5] Cheng M, Ning K. Stereotypes about enterotype: the old and new ideas. *Genomics, 560
Proteomics and Bioinformatics*. 2019;17(1):4–12. 561
- [6] Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal 562
microbiome of reproductive-age women. *Proceedings of the National Academy of 563
Sciences*. 2011;108:4680–4687. 564
- [7] France MT, Fu L, Rutt L, Yang H, Humphrys MS, Narina S, et al. Insight into the 565
ecology of vaginal bacteria through integrative analyses of metagenomic and meta- 566
transcriptomic data. *Genome Biology*. 2022;23(1):66. 567

- [8] Coudray MS, Madhivanan P. Bacterial vaginosis—A brief synopsis of the literature. *European Journal of Obstetrics & Gynecology and Reproductive Biology*. 2020 Feb;245:143–148. 568
569
- [9] 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. 571
572
573
- [10] Petrova MI, Reid G, Vaneechoutte M, Lebeer S. *Lactobacillus iners*: Friend or Foe? *Trends in Microbiology*. 2017;25(3):182–191. 574
575
- [11] 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. 576
577
578
- [12] Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UM, Zhong X, et al. Temporal dynamics of the human vaginal microbiota. *Science Translational Medicine*. 2012;4(132):132ra52–132ra52. 579
580
581
- [13] 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. 582
583
584

- [14] Serrano MG, Parikh HI, Brooks JP, Edwards DJ, Arodz TJ, Edupuganti L, et al. 585
Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. 586
Nature Medicine. 2019;25(6):1001–1011. 587
- [15] Munoz A, Hayward MR, Bloom SM, Rocafort M, Ngcapu S, Mafunda NA, et al. 588
Modeling the temporal dynamics of cervicovaginal microbiota identifies targets that 589
may promote reproductive health. Microbiome. 2021;9(1):1–12. 590
- [16] Chaban B, Links MG, Jayaprakash TP, Wagner EC, Bourque DK, Lohn Z, et al. 591
Characterization of the vaginal microbiota of healthy Canadian women through the 592
menstrual cycle. Microbiome. 2014;2:1–12. 593
- [17] Murall CL, Rahmoun M, Selinger C, Baldellou M, Bernat C, Bonneau M, et al. 594
Natural history, dynamics, and ecology of human papillomaviruses in genital in- 595
fections of young women: protocol of the PAPCLEAR cohort study. BMJ Open. 596
2019;9(6):e025129. 597
- [18] Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. Critical eval- 598
uation of two primers commonly used for amplification of bacterial 16S rRNA genes. 599
Applied and Environmental Microbiology. 2008;74(8):2461. 600
- [19] Farage M, Maibach H. Lifetime changes in the vulva and vagina. Archives of Gyne- 601
cology and Obstetrics. 2006;273:195–202. 602

- [20] Loganantharaj N, Nichols WA, Bagby GJ, Volaufova J, Dufour J, Martin DH, et al. 603
The effects of chronic binge alcohol on the genital microenvironment of simian im- 604
munodeficiency virus-infected female rhesus macaques. *AIDS Research and Human* 605
Retroviruses. 2014;30(8):783–791. 606
- [21] Mayer BT, Srinivasan S, Fiedler TL, Marrazzo JM, Fredricks DN, Schiffer JT. Rapid 607
and profound shifts in the vaginal microbiota following antibiotic treatment for bac- 608
terial vaginosis. *Journal of Infectious Diseases*. 2015;212(5):793–802. 609
- [22] Si J, You HJ, Yu J, Sung J, Ko G. *Prevotella* as a hub for vaginal microbiota under 610
the influence of host genetics and their association with obesity. *Cell Host & Microbe*. 611
2017;21(1):97–105. 612
- [23] Brotman RM, He X, Gajer P, Fadrosch D, Sharma E, Mongodin EF, et al. Association 613
between cigarette smoking and the vaginal microbiota: a pilot study. *BMC Infectious* 614
Diseases. 2014;14(1):1–11. 615
- [24] Achilles SL, Austin MN, Meyn LA, Mhlanga F, Chirenje ZM, Hillier SL. Impact of 616
contraceptive initiation on vaginal microbiota. *American Journal of Obstetrics and* 617
Gynecology. 2018;218(6):622–e1. 618
- [25] Laniewski P, Owen KA, Khnanisho M, Brotman RM, Herbst-Kralovetz MM. Clinical 619
and personal lubricants impact the growth of vaginal lactobacillus species and colo- 620

- nization of vaginal epithelial cells: an in vitro study. *Sexually Transmitted Diseases*. 621
2021;48(1):63–70. 622
- [26] Tessandier N, Uysal IB, Elie B, Selinger C, Bernat C, Boué V, et al. Does exposure 623
to different menstrual products affect the vaginal environment? *Molecular Ecology*. 624
2023;32(10):2592–2601. 625
- [27] Vodstrcil LA, Twin J, Garland SM, Fairley CK, Hocking JS, Law MG, et al. The 626
influence of sexual activity on the vaginal microbiota and *Gardnerella vaginalis* clade 627
diversity in young women. *PLOS ONE*. 2017;12(2):e0171856. 628
- [28] Noormohammadi M, Eslamian G, Kazemi SN, Rashidkhani B. Association between 629
dietary patterns and bacterial vaginosis: a case–control study. *Scientific Reports*. 630
2022;12(1):12199. 631
- [29] Hutchinson KB, Kip KE, Ness RB. Condom use and its association with bacterial 632
vaginosis and bacterial vaginosis-associated vaginal microflora. *Epidemiology*. 2007;p. 633
702–708. 634
- [30] Pape K, Ryttergaard L, Rotevatn TA, Nielsen BJ, Torp-Pedersen C, Overgaard C, 635
et al. Leisure-time physical activity and the risk of suspected bacterial infections. 636
Medicine and Science in Sports and Exercise. 2016;48(9):1737–1744. 637
- [31] Amabebe E, Anumba DO. Psychosocial stress, cortisol levels, and maintenance of 638
vaginal health. *Frontiers in Endocrinology*. 2018;p. 568. 639

- [32] Klebanoff MA, Nansel TR, Brotman RM, Zhang J, Yu KF, Schwebke JR, et al. 640
Personal hygienic behaviors and bacterial vaginosis. *Sexually Transmitted Diseases*. 641
2010;37(2):94. 642
- [33] Van Kessel K, Assefi N, Marrazzo J, Eckert L. Common complementary and al- 643
ternative therapies for yeast vaginitis and bacterial vaginosis: a systematic review. 644
Obstetrical & Gynecological Survey. 2003;58(5):351–358. 645
- [34] Ma ZS. Microbiome transmission during sexual intercourse appears stochastic and 646
supports the red queen hypothesis. *Frontiers in Microbiology*. 2022;12:789983. 647
- [35] Juliana NC, Peters RP, Al-Nasiry S, Budding AE, Morr e SA, Ambrosino E. Composi- 648
tion of the vaginal microbiota during pregnancy in women living in sub-Saharan Africa: 649
a PRISMA-compliant review. *BMC Pregnancy and Childbirth*. 2021;21(1):1–15. 650
- [36] Gupta K, Hillier SL, Hooton TM, Roberts PL, Stamm WE. Effects of contraceptive 651
method on the vaginal microbial flora: a prospective evaluation. *Journal of Infectious* 652
Diseases. 2000;181(2):595–601. 653
- [37] Brotman RM, Klebanoff MA, Nansel TR, Andrews WW, Schwebke JR, Zhang J, 654
et al. A longitudinal study of vaginal douching and bacterial vaginosis—a marginal 655
structural modeling analysis. *American Journal of Epidemiology*. 2008;168(2):188–196. 656
- [38] Gelman A. Scaling regression inputs by dividing by two standard deviations. *Statistics* 657
in Medicine. 2008;27(15):2865–2873. 658

- [39] Christopher H Jackson. Multi-State Models for Panel Data: The msm Package for R. 659
Journal of Statistical Software. 2011;38(8):1–29. 660
- [40] Brooks JP, Buck GA, Chen G, Diao L, Edwards DJ, Fettweis JM, et al. Changes 661
in vaginal community state types reflect major shifts in the microbiome. Microbial 662
Ecology in Health and Disease. 2017;28(1):1303265. 663
- [41] Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, et al. The 664
vaginal microbiome and preterm birth. Nature Medicine. 2019;25(6):1012–1021. 665
- [42] Team SD. Stan Functions Reference; 2024. Accessed: 2024-02-05. [https://mc-stan.](https://mc-stan.org/docs/functions-reference) 666
[org/docs/functions-reference.](https://mc-stan.org/docs/functions-reference) 667
- [43] Stan Development Team. Prior Choice Recommendations; 2023. Accessed: 2024-02- 668
15. [https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations.](https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations) 669
- [44] Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and 670
other shrinkage priors. Electronic Journal of Statistics. 2017;11(2):5018–5051. 671
- [45] Stan Development Team. RStan: the R interface to Stan; 2023. R package version 672
2.32.3. Available from: <https://mc-stan.org/>. 673
- [46] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained 674
equations in R. Journal of Statistical Software. 2011;45(3):1–67. 675

- [47] Stan Development Team. The Stan Core Library; 2018. Version 2.18.0. Available 676
from: <http://mc-stan.org/17>. 677
- [48] Betancourt M. Towards a principled Bayesian workflow; 2020. Avail- 678
able from: [https://betanalpha.github.io/assets/case_studies/principled_](https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html) 679
[bayesian_workflow.html](https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html). 680
- [49] Odogwu NM, Onebunne CA, Chen J, Ayeni FA, Walther-Antonio MR, Olayemi OO, 681
et al. *Lactobacillus crispatus* thrives in pregnancy hormonal milieu in a Nigerian 682
patient cohort. Scientific Reports. 2021;11(1):18152. 683
- [50] Ravel J, Brotman RM, Gajer P, Ma B, Nandy M, Fadrosch DW, et al. Daily temporal 684
dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. 685
Microbiome. 2013;1(1):1–6. 686
- [51] Fethers KA, Fairley CK, Hocking JS, Gurrin LC, Bradshaw CS. Sexual risk factors 687
and bacterial vaginosis: a systematic review and meta-analysis. Clinical Infectious 688
Diseases. 2008;47(11):1426–1435. 689
- [52] Liao Y, Tong XT, Jia YJ, Liu QY, Wu YX, Xue WQ, et al. The effects of alco- 690
hol drinking on oral microbiota in the Chinese population. International Journal of 691
Environmental Research and Public Health. 2022;19(9):5729. 692

- [53] Froehle L, Ghanem KG, Page K, Hutton HE, Chander G, Hamill MM, et al. Bacterial 693
vaginosis and alcohol consumption: a cross-sectional retrospective study in Baltimore, 694
Maryland. *Sexually Transmitted Diseases*. 2021;48(12):986–990. 695
- [54] Smart S, Singal A, Mindel A. Social and sexual risk factors for bacterial vaginosis. 696
Sexually Transmitted Infections. 2004;80(1):58–62. 697
- [55] Thorsen P, Vogel I, Molsted K, Jacobsson B, Arpi M, Møller BR, et al. Risk factors for 698
bacterial vaginosis in pregnancy: a population-based study on Danish women. *Acta* 699
Obstetrica et Gynecologica Scandinavica. 2006;85(8):906–911. 700
- [56] Shoubnikova M, Hellberg D, Nilsson S, Mårdh PA. Contraceptive use in women with 701
bacterial vaginosis. *Contraception*. 1997;55(6):355–358. 702
- [57] Rugpao S, Sriplienchan S, Rungruengthanakit K, Lamlertkittikul S, Pinjareon S, Wer- 703
awatakul Y, et al. Risk factors for bacterial vaginosis incidence in young adult Thai 704
women. *Sexually Transmitted Diseases*. 2008;35(7):643–648. 705
- [58] Baisley K, Chagalucha J, Weiss HA, Mugeye K, Everett D, Hambleton I, et al. 706
Bacterial vaginosis in female facility workers in north-western Tanzania: prevalence 707
and risk factors. *Sexually Transmitted Infections*. 2009;85(5):370–375. 708
- [59] Francis SC, Looker C, Vandepitte J, Bukonya J, Mayanja Y, Nakubulwa S, et al. Bac- 709
terial vaginosis among women at high risk for HIV in Uganda: high rate of recurrent 710
diagnosis despite treatment. *Sexually Transmitted Infections*. 2016;92(2):142–148. 711

- [60] Cu-Uvin S, Hogan JW, Warren D, Klein RS, Peipert J, Schuman P, et al. 712
Prevalence of lower genital tract infections among human immunodeficiency virus 713
(HIV)—seropositive and high-risk HIV-seronegative women. *Clinical Infectious Dis-* 714
eases. 1999;29(5):1145–1150. 715
- [61] French AL, Adeyemi OM, Agniel DM, Evans CT, Yin MT, Anastos K, et al. The 716
association of HIV status with bacterial vaginosis and vitamin D in the United States. 717
Journal of Women’s Health. 2011;20(10):1497–1503. 718
- [62] Brotman RM, Ghanem KG, Klebanoff MA, Taha TE, Scharfstein DO, Zenilman JM. 719
The effect of vaginal douching cessation on bacterial vaginosis: a pilot study. *American* 720
Journal of Obstetrics and Gynecology. 2008 Jun;198(6):628.e1–628.e7. 721
- [63] Sobel JD, Sobel R. Current and emerging pharmacotherapy for recurrent bacterial 722
vaginosis. *Expert Opinion on Pharmacotherapy.* 2021;22(12):1593–1600. 723
- [64] Morsli M, Gimenez E, Magnan C, Salipante F, Huberlant S, Letouzey V, et al. 724
The association between lifestyle factors and the composition of the vaginal micro- 725
biota: a review. *European Journal of Clinical Microbiology & Infectious Diseases.* 726
2024;43(10):1869–1881. 727
- [65] Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P, et al. Differences 728
in the composition of vaginal microbial communities found in healthy Caucasian and 729
black women. *ISME J.* 2007;1(2):121–133. 730

- [66] 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. 731-733
- [67] 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. 734-736
- [68] Lambert JA, John S, Sobel JD, Akins RA. Longitudinal analysis of vaginal microbiome dynamics in women with recurrent bacterial vaginosis: recognition of the conversion process. *PLOS ONE*. 2013;8(12):e82599. 737-739
- [69] Srinivasan S, Liu C, Mitchell CM, Fiedler TL, Thomas KK, Agnew KJ, et al. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PLOS ONE*. 2010;5(4):e10197. 740-742
- [70] Armstrong E, Hemmerling A, Joag V, Huibner S, Kulikova M, Crawford E, et al. Treatment success following standard antibiotic treatment for bacterial vaginosis is not associated with pre-treatment genital immune or microbial parameters. *Open Forum Infectious Diseases*. 2023 Jan;p. ofad007. 743-746
- [71] Lebeer S, Ahannach S, Gehrman T, Wittouck S, Eilers T, Oerlemans E, et al. A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. *Nature Microbiology*. 2023;8(11):2183–2195. 747-749

- [72] Symul L, Jeganathan P, Costello EK, France M, Bloom SM, Kwon DS, et al. Sub- 750
communities of the vaginal microbiota in pregnant and non-pregnant women. Pro- 751
ceedings of the Royal Society B. 2023;290(2011):20231461. 752
- [73] Holm JB, France MT, Gajer P, Ma B, Brotman RM, Shardell M, et al. Integrating 753
compositional and functional content to describe vaginal microbiomes in health and 754
disease. *Microbiome*. 2023;11(1):259. 755
- [74] Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, et al. Interplay 756
between the temporal dynamics of the vaginal microbiota and human papillomavirus 757
detection. *Journal of Infectious Diseases*. 2014;210(11):1723–1733. 758

Supplementary Information 759

Competing interests 760

SAIizon is a recommender for PCI Evolutionary Biology and PCI Ecology. 761

JReynes reports personal fees from Gilead (payment or honoraria for lectures, presen- 762
tations, or educational events) Merck (payment or honoraria for lectures, presentations, 763
or educational events), Moderna (honoraria for lectures, or manuscript writing), Shionogi 764
(honoraria for presentation) and ViiV Healthcare (consulting and payment or honoraria for 765
lectures, presentations, or educational events) and support for attending meetings and/or 766
travel from Gilead and Pfizer, all outside of the submitted work. 767

JRavel is co-founder of LUCA Biologics, a biotechnology company focusing on translat- 768
ing microbiome research into live biotherapeutics drugs for women’s health. He is Editor- 769
in-Chief at *Microbiome*. 770

None of the other authors report any conflict of interest. 771

S1: Pairwise correlations between covariates 772

There were no strong correlations among covariates, with the strongest correlation found 773
between BMI and stress ($r = 0.41$; Fig. S1). 774

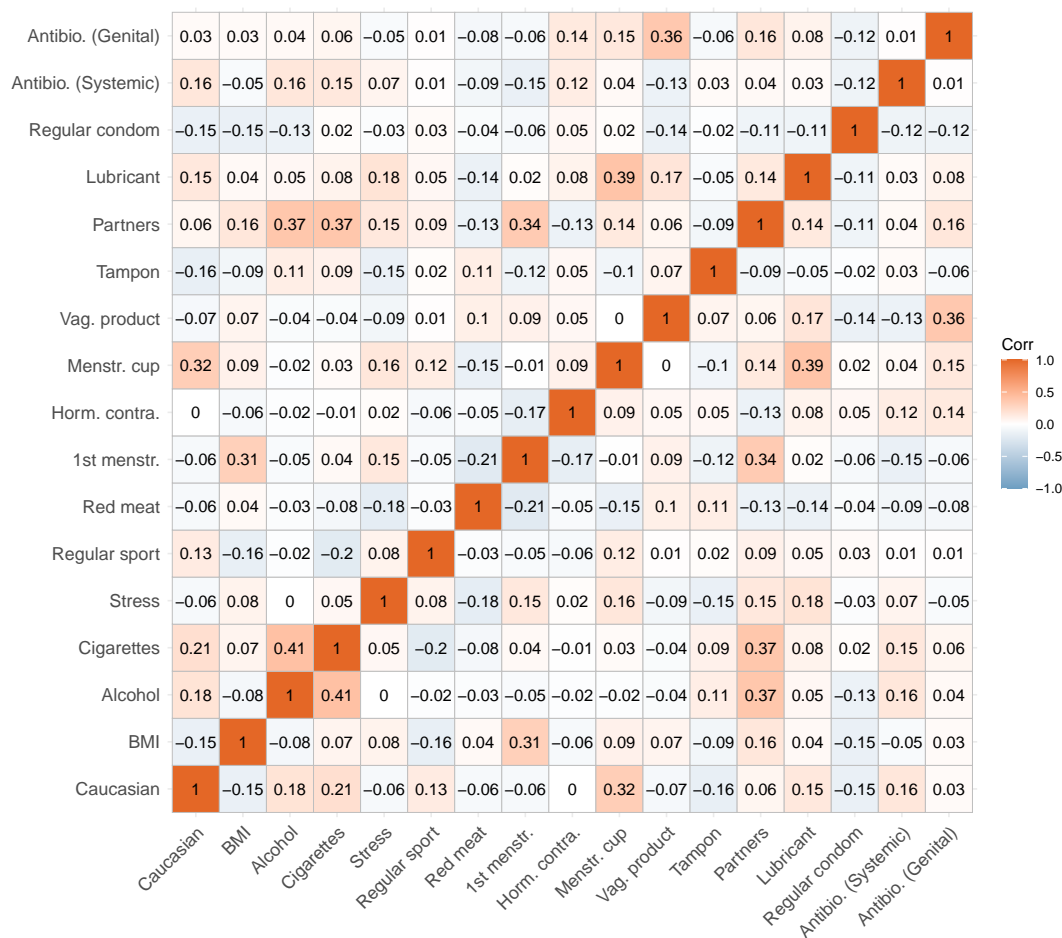


Figure S1: Correlation between covariates. Pairwise Pearson’s correlation coefficients between covariates. Parameter descriptions are found in Materials and Methods.

S2: Assessment of posterior accuracy, precision and prior contraction

775

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 [48]:

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

where \mathbb{E}_{post} denotes the posterior mean of the fit to the actual data that we consider the ‘true’ parameter. \mathbb{E}_{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 [48]. In contrast, large z-values may be indicative of overfitting and, or poor prior specifications [48].

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

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

where σ_{post}^2 and σ_{prior}^2 correspond to the variance of posterior and prior distributions, 781
respectively. The k values close to zero indicate that data contain little information (i.e., 782
rendering priors strongly informative). Conversely, values close to 1 indicate that data are 783
much more informative than the prior [48]. 784

We found that most of our model parameters and hyperparameters — were estimated 785
with accuracy, precision, and identifiability, with the absolute posterior z-scores below three 786
(Fig. S2). The posterior distributions for covariate coefficients, β , contracted by 86% on 787
average, and at least 75%, compared to the prior distribution, meaning that the covariate 788
coefficients were well-identified from data (Fig. S2). Although we used generic priors 789
recommended by Stan [42], the L_s parameters that define correlations among between- 790
woman variation showed limited posterior contraction (i.e., $\leq \sim 0.25$), indicating that these 791
parameters are poorly informed by data. As such, we refrain from making biological 792
inferences about these correlations. 793

S3: Predicted difference in community state type (CST) prevalence at 794 various counterfactual scenarios. 795

Our counterfactual simulations predicted that alcohol consumption and the number of 796
partners are factors that impact the population-level outcome in terms of the prevalence 797
of different community state types. The full list of comparisons is available in Fig. S3. 798

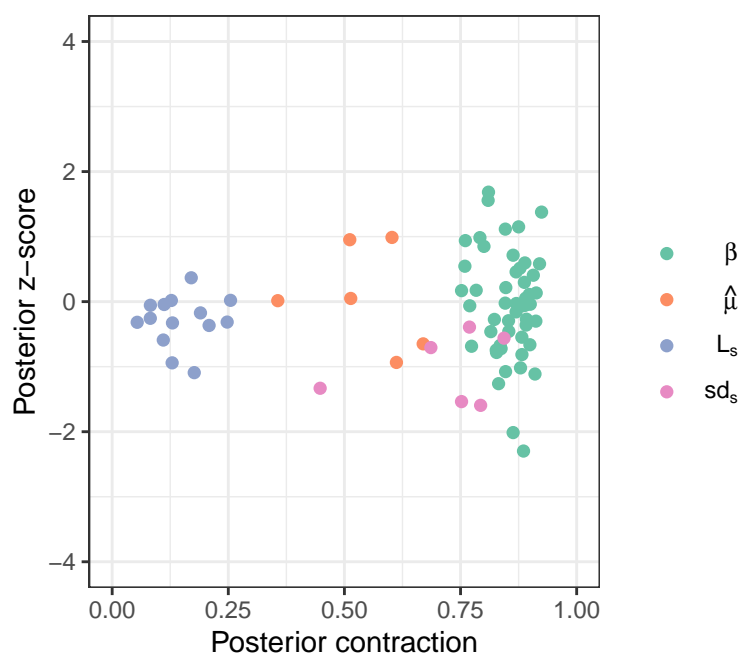


Figure S2: Accuracy, precision and identifiability of estimated parameters. Posterior z-score (y-axis) measures how closely the posterior recovers the parameters of the true data-generating process and posterior contraction (x-axis) evaluates the influence of the likelihood function over the prior, respectively. 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 [48]. 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.

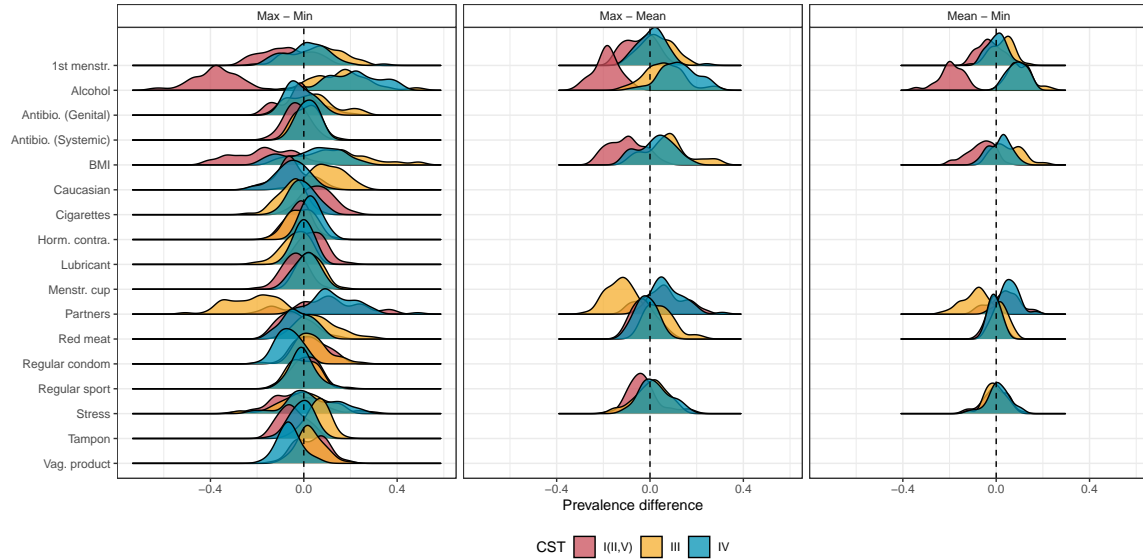


Figure S3: Difference in community state type (CST) prevalence at predicted various counterfactual scenarios. The differences were calculated from posterior samples simulated at 0 and 1 for binary variables and at the population maximum and minimum values recorded by the PAPCLEAR for continuous variables (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). Parameter descriptions are found in Materials and Methods.