Title:

Vaginal microbiota changes during a natural menstrual cycle in the Women4Health cohort highlights specific features of the Italian population.

Authors

Elena Vinerbi[1,*], Fabio Chillotti[1,*], Andrea Maschio[1,*], Stefania Lenarduzzi[2,*], Silvia Camarda[3], Francesca Crobu[1], Daria V. Zhernakova[1], Valeria Lo Faro[1], Giulia Beltrame Vriz[3], Simona Incollu[1], Johanne Spreckels[4], Nataliia Kuzub[4], Alma Kadric[5], Ranko Gacesa[6], Alexandra Zhernakova[4], Francesco De Seta[2,3,7], Daniela Mazzà[2], Fabio Busonero[1], Maria Laura Ferrando[1], Giorgia Girotto[2,3,#], Serena Sanna[1,2,#]

Affiliations

1 Institute of Genetic and Biomedical Research (IRGB), National Research Council (CNR), Monserrato, Italy

2 Institute for Maternal and Child Health—IRCCS "Burlo Garofolo", Trieste, Italy

3 Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy

4 Department of Genetics, University Medical Center Groningen, Groningen, the Netherlands

5 Department of Life Science, University of Trieste, Trieste, Italy

6 Department of Gastroenterology and Hepatology, University Medical Center Groningen, Groningen, the Netherlands

7 Department of Obstetrics and Gynecology, IRCCS San Raffaele Scientific Institute, University Vita and Salute, Milan, Italy.

*** first shared author**

 # last shared author

Corresponding Authors

Serena Sanna, PhD

Institute of Genetic and Biomedical Research (IRGB), National Research Council (CNR), Monserrato, Italy

Email: serena.sanna@cnr.it

Tel: +39 0707654593

Giorgia Girotto, PhD

Institute for Maternal and Child Health—IRCCS "Burlo Garofolo", Trieste, Italy

Email: giorgia.girotto@burlo.trieste.it

Tel: +39 0403785539

ABSTRACT

Large sex hormonal fluctuations are thought to influence vaginal microbiota, but little is known about the impact of small, physiological variations. Here we tracked changes in vaginal microbiota during four key menstrual cycle phases in 61 healthy Italian women from the Women4Health cohort. The microbiota was primarily composed of *Lactobacillus*, with *L. iners* being the most abundant. Noteworthy, the high abundance of *L. iners* contrasts with previous studies in European populations, challenging its proposed pathogenic role, and suggesting distinct microbiota profiles within Europe. Individual microbiotas were generally stable, but beta diversity was higher during the follicular phase. Only 11 women exhibited compositional shifts and those mostly occurred between follicular and ovulatory phases. Finally, among the hormones evaluated, 17-beta estradiol had the largest impact on taxa abundance variations. Our study highlights specific features of the Italian population and points to the resilience of the vaginal microbiota to physiological hormonal changes.

INTRODUCTION

The vaginal microbiota refers to the community of microorganisms that colonize the vagina. Unlike microbiota from other mucosal sites (i.e., the gastrointestinal tract), it has lower diversity, and it is dominated by a few species, primarily *Lactobacillus spp*, which play a key role in regulating genital health. Producing lactic acid and antimicrobial molecules, these bacteria reduce the risk of colonization by other microorganisms, including potential pathogens involved in infections and diseases of the vaginal tract [1]; [2]; [3]; [4]. In general, the prevalence of *Lactobacillus* species in healthy women increases with puberty and stays stable until menopause[5].

The most common species of *Lactobacillus* are *L. iners*, *L. crispatus*, *L. gasseri*, and *L. jensenii*. Earliest studies showed that a vaginal microbiota dominated by *L. crispatus* was associated with a healthy status, whereas *L. iners* was often isolated during infections and therefore a microbiota dominated by it is often considered "dysbiotic". However, this hypothesis originated by studies with small sample size and a direct pathogenic role for *L. iners* has not been proved [6]; [7]; [8]. Likewise, a vaginal microbiota dominated by other anaerobic bacteria, such as *Gardnerella vaginalis* (recently reclassified as *Bifidobacterium vaginalis), Prevotella*, and *Fannyeshea*, have been associated with pathological conditions of the female genital tract due to the virulence potential of some strains [9]; [3]; [10].

The pathogenetic role of *L. iners* remains under debate since its prevalence can vary substantially among different populations [2]; [11]; [12]; [13]; [14]. The vaginal microbiota of African and Hispanic women during the reproductive age is dominated by *L. iners*, while is much rarer in Northern European women with *L. crispatus* being the most abundant and prevalent *Lactobacillus* species. The reasons for this north-to-south gradient in the prevalence of *L. crispatus* and *L. iners* are still unknown. It is worth noting that these two species can co-exist in the same microbiota, therefore it's unlikely they are antagonists.

The vaginal microbiota has been considered an overall stable ecosystem, although two recent studies conducted on short-term (42 days) and medium-term (18 months) longitudinal follow-up have shown that variations may occur in a small fraction of women [6]; [15]. These studies evaluated changes in the so-called Community State Types (CST), a method that classifies the vaginal microbiota into 5 groups depending on the composition and the most abundance species [2]; [16]. In the shorter 42-days study, only microbiotas classified as CST-I (dominated by *L. crispatus*) and CST-IV (dominated by anaerobic bacteria) remained stable, while variations were observed in other CST

types. In the longer 18 months longitudinal study, vaginal microbiota of CST-I and CST-III (dominated by *L. iners*) remained stable, while CST-IV microbiota displayed lower temporal stability (approximately 6 months). The underlying causes of these variations, how they occur in populations with different *Lactobacillus spp.* prevalences, and which species may be responsible for these changes in CSTs are not yet known. Worth mentioning, none of these studies directly characterized sex-hormones, being unable to link compositional changes to specific sex hormones fluctuations [3]. Here we present vaginal microbiome composition and its dynamics in of 61 healthy women from the Italian population cohort Women4Health (W4H), an observational longitudinal study centered on one menstrual cycle set up to explore the interactions between sex hormones, vaginal and intestinal microbiome, and lipid and glucose metabolism in women [17]. We characterized vaginal microbiome composition and evaluated dynamics at four time points pivotal for the menstrual cycle. We also evaluated host factors associated with diversity and identified sex-hormones that are mostly associated with the observed dynamic changes.

MATERIAL AND METHODS

The Women4Health cohort

The Women4Health (W4H) cohort is an ongoing short-term longitudinal study aiming to recruit up to 300 healthy women and to follow them up through a natural menstrual cycle, collecting various biological samples and metadata via questionnaires [17]. Briefly, women of European origin aged 18 to 45 years old, who were not using hormonal contraceptives and who have not a diagnosis of gynecological, metabolic, or gastrointestinal diseases, were enrolled in the study at the Institute for Maternal and Child Health – IRCCS Burlo Garofolo in Trieste, Italy. Upon inclusion and signature of an informed consent form, electronic questionnaires were used to gather data on family history, clinical and anthropometric measurements, lifestyle and diet. They were then scheduled for 4 weekly follow-up appointments to hand in self-collected vaginal swabs, stool samples, saliva (only at first appointment), to undergo a blood withdrawal and to collect an additional questionnaire related to their health and lifestyle in days preceding the appointment. Weekly follow-up visits were scheduled according to pivotal phases of the menstrual cycle: the 1st at the follicular phase (between days 6 and 9), the 2nd at the ovulatory phase (between days 12 and 16), the 3rd at the early luteal phase (between days 18 and 22), and the 4th at late luteal phase (between days 24 and

27). Blood samples were processed within 1 hour after collection. On serum, we measured 5 sex hormones (beta17-estradiol, luteinizing hormone, follicle stimulating hormone, progesterone and prolactin). During the first visit, we also measured testosterone and thyroid parameters (thyroid stimulating hormone and thyroxine (T_4)) to exclude endocrine disorders. Hormones were measured using 1.5 ml serum on the ADVIA Centaur CP Immunoassay System instrument from Siemens Healthineers.

The study protocol - including informed consent forms - was approved by the Friuli-Venezia Giulia Ethical Committee on 09/2021 with prot. N. 0034184/P/GEN/ARCS and modifications amended on 09/2023 with prot. N. 0033477/P/GEN/ARCS.

Here, we analyzed the vaginal swabs, metadata, and sex-hormone levels in the first 61 enrolled women. Descriptive statistics of the sample are provided in **Table S1.**

Vaginal swabs collection, processing, and microbial DNA extraction

Vaginal swabs were self-collected by volunteers before each follow-up visit (the same day or the day before) using the OMNIgene-VAGINAL kits (OMR-130, DNA Genotek), stored at room temperature and handed at IRCCS Burlo Garofolo within 24 hours, where kits were promptly stored, as recommended by the manufacturer, and shipped to the Institute for Genetic and Biomedical Research - National Research Council (IRGB-CNR) within 21 days. DNA was extracted from swabs with QIAamp PowerFecal Pro DNA kits (QIAgen) using a custom semiautomated protocol on a QIAcube platform. Prior to DNA extraction, a bead-beating step on a TissueLyser II (QIAgen) at 7.5 Hz for 5 min was performed to improve lysis. The composition of vaginal microbiota was established by sequencing amplicons of the V3-V4 and V7-V9 regions of the 16S rRNA gene, along with ITS1 gene of fungi. Library construction was performed with a 2-step PCR protocol and dual-barcoding strategy (QIAseq 16S/ITS Region Panel kits, QIAgen). Sequencing was carried out with the MiSeq sequencing system (Illumina) targeting an average of 100K 250-bp paired-end reads per sample. Samples were processed in three different batches, each including one QIAseq 16S/ITS Smart Control (QIAgen), a synthetic DNA construct used both as positive control for library preparation and as control for the identification of contaminants, and a PCR negative control (distilled sterile water). Data obtained were demultiplexed using Qiagen's GeneGlobe software.

Quality Control (QC) process

All demultiplexed *FastQ* underwent quality control screening before bioinformatic analysis. *FastQC* v.0.12.1 (see **URLs**) and *MultiQC* v.1.22.2 [18] (see **URLs**) tools were used to detect outliers based on read quality. The *trimmomatic* tool v.0.39 [19] (see **URLs**) was used for trimming low quality paired-end reads. We used the options *SLIDINGWINDOW:4:25* to evaluate whether the average quality score was at least 25 per four bases; if not, this part of the read was trimmed. The option MILEN:50 was used to remove reads shorter than 50 bp. After the trimming process, samples were re-analyzed with *FastQC* and *MultiQC* to evaluate their post-process quality.

We first processed 185 samples, 4 smart controls and 3 negative controls in two plates and decided to remove **i)** all samples that had less than 80K reads as input before the trimming process and/or less than 60K reads survived post-trimming (N=10samples), **ii)** one sample with very high number of ITS reads mapping on *Sporobolomyces spp*, a known environmental contaminant, and **iii)** samples whose libraries were prepared together with a negative control that showed contamination (>500 reads)(N=10). These 21 samples were repeated in a third plate, containing additional 28 samples, 1 smart control and 1 negative control. The same quality control filters were applied and only 1 sample was removed from plate 3. Finally, one sample that showed 1.5 million reads and 1.3 million post-trimming reads as input, due to an error in library equimolar pooling, was subjected to the subsampling process by randomly selecting 187K reads (the average value of reads survived post trimming for Plate 3), through the R package *ShortRead* v1.62.0 (see **URLs**).

After all these QC steps, a total of 212 samples were left for downstream analyses with the median number of reads per sample being 138,531 (127,518, 141,450 and 190,072 for plates 1, 2 and 3, respectively).

Characterization of the vaginal microbiota

To characterize the bacterial composition of the samples, the *DADA2* v3.19.0 package [20] (see **URLs**) was used in the R (v 4.4.1) environment. Using the *"filterAndTrim()"* function, only sequences with a minimum length of 160 bp were selected. With the "*removeBimeraDenovo()"* function, all chimeras were removed.

The microorganisms were classified up to the taxonomic level of species with the *"assignTaxonomy()"* and *"addSpecies()"* function Genome Taxonomy DataBase (GTDB) v.r95 [21] was used to profile a phylogenetically consistent microbial taxonomy (see **URLs**). Hypervariable regions V3V4 and V7V9 were processed together. Rare amplicon sequence variants (ASVs) resulting from DADA2 were removed; only ASVs that had at least 5 reads in at least 2 samples were maintained for

181 subsequent analyses. To increase taxonomic resolution for the genus *Lactobacillus,* we followed a
182 previously described protocol [11]. Using their code and data files provided (see **URLs**), the genus
183 *Lactobacillus* was reclassified into different subgenera, allowing the identification of more
184 *Lactobacillus* species. Finally, *Lactobacillus* species names were replaced by the respective subgenus
185 identified (**Table S2**). Next, we selected ambiguous ASVs – i.e. those ASVs that were assigned by the
186 algorithm as *Lactobacillus* species but belonged to other genera - and uploaded the corresponding
187 sequences on BLAST to identify the most similar species. Since ambiguity remained, we decided to
188 assign these ASVs the label "unclassified" (**Table S3).**
189 To characterize the fungi composition of the samples, the ITS fastQ files were processed with the
190 *DADA2* v3.19.0 [20] (see **URLs**) package on R (v.4.4.1). To this end, "*filterAndTrim()*" function was set
191 as "*minLen = 50*" and "*maxLen = 200*". We profiled samples with the Unite database v.10.0 [22] (see
192 **URLs).**
193 We derived global metrics of bacterial microbiome composition such as alpha and beta diversity
194 using the *vegan* package v.2.6-4 [23] (see **URLs**). Alpha diversity was calculated based on taxa counts
195 according to the Shannon-Weiner (H) and Simpson (D) indices. Beta diversity was calculated
196 according to the Bray-Curtis dissimilarity index using taxa relative abundances and according to
197 Euclidean distance using taxa relative abundances normalized using the Centered log ratio (CLR)
198 transformation. The Euclidean distance was used in addition to Bray-Curtis because it allowed us to
199 use normalized and corrected data (**Table S4**). To perform CLR transformation we used our own
200 function that calculates the transformation in parallel among taxa, using the *Parallel* package v.4.4.1
201 on R (see **URLs**).
202 In addition, we classified our samples according to the Community State Types (CSTs) definition of
203 the VALENCIA algorithm [16] (see **URLs**). Toward this end, it was necessary to manually modify the
204 taxonomy names according to those required by VALENCIA (**Table S5)**. To identify which ASV could
205 be attributed to *Gardenella Vaginalis*, given the recent change in nomenclature, we used DADA2
206 with the database SILVA v.138.1 where the nomenclatures of species align to those used in the
207 VALENCIA algorithm.
208 Finally, we used the t-SNE algorithm [24] to visualize the vaginal microbiome composition in a two-
209 dimensional space and used color-coding labels to describe the observed variation in terms of taxa
210 abundance as well as CSTs. For this analysis, we employed the t-SNE algorithm to CLR-transformed
211 data using the R package *Rtsne* v.0.17 [25] (see **URLs**).

212 Finally, we profiled microbiotas using the same pipeline but subsampling 30K (and 60K) survived
213 reads per sample using the R package *ShortRead* (see **URLs**).
214

215 **Identification of factors associated with beta diversity**

216 We evaluated the impact on beta diversity of technical confounders, characteristics of the volunteers
217 (as recorded in questionnaires) and of sample collection. Using the *vegan* package [23] (see **URLs**),
218 we first investigated whether the first five Bray-Curtis beta diversity principal coordinate analyses
219 (PCoA) axes were correlated with technical variables, such as microbial DNA concentration, library
220 concentration, and total reads counts (survived reads). Each PCoA axis was linearly adjusted using
221 only the covariates with which it showed significant correlation. As the dataset included longitudinal
222 measurements, independence tests were conducted separately for each visit to ensure
223 independence between observations. Namely, we used the Kruskal-Wallis independence test in R
224 through the function *kruskal.test()* of *stats* package v.4.4.1 [26]; (see **URLs).**

225 Variables from the questionnaires encompassed both categorical and continuous data related to
226 demographics (e.g., age), anthropometric measures (e.g., body mass index [BMI], waist-to-hip ratio
227 [WHR]), lifestyle factors (e.g., smoking, hormonal contraceptive use), clinical history related to
228 women's health (e.g., number of pregnancies), and conditions surrounding vaginal swab collection
229 (e.g., medication or supplement use prior to collection, sexual intercourse within two days of
230 collection, swab timing, and stool consistency assessed using the Bristol stool scale). Continuous
231 variables were categorized to allow application of the independence test. All tested variables are
232 summarized in **Table S6**.
233 We also run PERMANOVA analyses for the significant associations resulting from independence test,
234 using *adonis2()* function from *vegan* R package [23] (see **URLs).**
235

236 **Statistical analysis to evaluate vaginal microbiota dynamics**
237 To evaluate changes in alpha diversity, beta diversity and taxa relative abundance across the four
238 phases of the menstrual cycle, we used the non-parametric Friedman test. To evaluate changes
239 between two phases (pairwise comparisons) we used the non-parametric Wilcoxon pairwise test for
240 the untransformed data, and the parametric pairwise t-test for the CLR-transformed data. All these
241 comparative tests were run using the *stat* package v.4.4.1 [26] (see **URLs**). All pairwise comparisons
242 were corrected using the Bonferroni approach.

To evaluate beta diversity changes (both Bray-Curtis and Euclidean distance, adjusted for technical covariates) we used two approaches. First, we calculated mean and median beta diversity across all women within each menstrual cycle phase and compared these across the different phases, using the paired Wilcoxon test (**Table S7**).

In a second step, we calculated mean and median beta diversity for each woman among her different collections (up to 4), and compared the distribution obtained when comparing her first collection to 3 collections from different women from the other 3 phases, using again the unpaired Wilcoxon test. All comparisons were corrected using Bonferroni.

Linear Models

We fitted three groups of linear models to: **i)** investigate whether bacterial abundances vary across the menstrual cycle; **ii)** explore the relationship between hormones and bacterial abundances variations between two consecutive phases and **iii)** explore the overall relationship between hormones and bacteria relative abundances.

For all linear models, taxa were filtered using a prevalence > 20% and a mean relative abundance threshold of 0.10 % **(Table S8).**

In the first group of models, we fitted the following gaussian mixed models:

$$\text{Taxa_adj} \sim \text{Visit_number} + (1|\text{Woman}) \qquad \textbf{(1)}$$

where visit_number is a numeric variable coded as 1, 2, 3, 4 according to the four phases of the menstrual cycle, and Taxa_adj are the residuals of the linear models:

$$\text{Taxa_CLR-transformed} \sim \text{covariates}$$

fitted using the function "*LinearRegression()*" of *sklearn* package v.1.5.2 in Python v.3.10.12 [27] (see **URLs**). The placeholder "covariates" here means that we considered several sets of covariates based on the results of independence tests and technical variables to evaluate if these changes could be confounded by a specific covariate. We opted for stepwise models with increasing number of covariates, rather than running a model with the full set of covariates, given the limited sample size of the study. A total of 10 models were considered **(Table S9).**

From model (1) we used the regression coefficients and the p-values for the variable Visit_number to identify taxa with significant changes. We used permutation to derive an empirical p-value and evaluate robustness of significant results ($p<0.05$). We permuted the column containing taxa abundances and re-run the model 1000 times. The proportion of "fake" p-values less than or equal

to the original p-value was calculated to provide the empirical p-value. In addition, false discovery rate (FDR) was instead used to adjust p-values for multiple testing (original pvalues, not empirical pvalues), with the function *"p.adjust()"* of *stats* package and method 'BH' [28].

In addition, we calculated the interclass correlation coefficient (ICC) for the model where bacteria were adjusted for the technical covariates (model **i)** above). This parameter quantifies the proportion of total variance attributable to between-subject variability, reflecting the stability of bacterial abundance across different time points within subjects. The ICC was calculated using the function "*icc()"* from the package *performance v 0.12.4* [28] (see **URLs**).

In the second group of models, we again used linear mixed models to investigate if changes in bacteria relative abundance between two phases were associated with changes in sex hormones.

Firstly, we removed outlier values in the sex hormone variables exceeding 4 standard deviations from the mean. Secondly, we calculated the differences between two consecutive phases of both sex hormones and bacteria relative abundances and then we normalized these differences using the function "*ordernorm()*" of the R package *bestNormalize v 1.9.1* [29](see **URLs**).

The following gaussian mixed models were fitted:

$$(\text{Hormone}_{t+1} - \text{Hormone}_{t}) \sim (\text{Taxa_adj}_{t+1} - \text{Taxa_adj}_{t}) + (1|\text{Women}) \qquad \textbf{(2)}$$

where t represents the menstrual cycle phase, and taxa_adj is the CLR-transformed abundance adjusted by technical covariates, calculated as the residuals of the models:

$$\text{Taxa_CLR_transformed} \sim \text{Qubit_DNA} + \text{Qubit_Library} + \text{Total_counts} \qquad \textbf{(3)}$$

where Qubit_DNA is the DNA concentration, Qubit_Library is the library concentration, and Total_counts is the total number of survived reads.

In the third group of models we investigate relationships between taxa relative abundances and sex hormones changes across the entire menstrual cycle by fitting the following:

$$\text{Hormone} \sim \text{Visit_number} + \text{Taxa_adj} + (1|\text{Woman}) + \text{Pregnancy_category} \qquad \textbf{(4)}$$

In both models **(2)** and **(4)** we used the same strategies described above for models (1) to derive empirical p-values and to correct p-values for multiple testing. To fit all linear mixed models in (1), (2) and (4) we used the *lme4 v 1.1.35.5* [30] *(see* **URLs***)* and the *lmerTest v.3.1.3* [28] R packages (see **URLs**).

Whole-genome sequencing

Ten samples with different microbiome composition were selected for whole-genome sequencing (WGS). We used the beta diversity matrix to maximize dissimilarity among selected samples, as follows. The first three selected samples were those showing a high relative abundance of *L. iners, L. crispatus* and *L. gasseri,* respectively, and had the maximum beta diversity between them. Then, we selected seven samples showing maximum distance from the first three samples. Sequencing was performed at Prebiomics srl on the Illumina NovaSeq 6000 sequencing system, and microbiome profiling was carried out using *MetaPhlan 4* [31]. An average of 3 million bacterial reads per sample was obtained **(Table S10a)**. Relative abundances from the two data sets (16S and WGS) were compared for taxa with abundances > 0 in at least one of the samples. Spearman's correlation test was used to check the correlations between relative abundances of taxa detected by both approaches and for which at least 3 samples had no-zero value **(Table S10b).**

RESULTS

Vaginal Microbiome Composition

A total of 61 women contributed 213 vaginal swabs to this study, of which 212 passed quality control (QC) as described in **Methods**. Bacterial composition was profiled using previously described protocols (see **Methods**). At the genus level, *Lactobacillus* was the most common genus, followed by *Bifidobacterium* and *Streptococcus*. The two most abundant *Lactobacillus spp* were *L. iners* and *L. crispatus* (**Figure S1**) and were identified in almost all samples (prevalence > 90%). When restricting to samples with a relative abundance > 60%, *L. iners* was detected in 92 samples (from 38 women) while *L. crispatus* was detected in 74 samples (from 34 women). In addition to *Lactobacillus spp*, we identified other species, such as *Bifidobacterium vaginale* and *Fannyhessea vaginae* (**Figure S2** and **Table S8**).

Overall, vaginal microbiota remained stable throughout the menstrual cycle. There were no significant differences in alpha diversity across the phases, as measured by the Shannon (H) and Simpson (D) indices (**Figure S3** and **Table S11**). The average beta diversity was instead higher in the follicular phase compared to ovulatory phase (O) ($p_{wilcoxon}$ = 0.009 and $p_{t\text{-}test}$ = 0.007) and early luteal phase (EL) ($p_{wilcoxon}$ = 0.04 and $p_{t\text{-}test}$ = 0.03) (**Figure 1A** and **Table S7).** Similar results were obtained when using median instead of average (**Figure S4** and **Table S7**), or the overall distribution **(Figure S5** and **Table S12)**. Nevertheless, these changes were small and did not lead women's microbiota to

334 align to a similar composition during a particular phase. In fact, the average beta diversity of a
335 woman's microbiota across all her samples was lower than the average obtained when using random
336 samples from other women (p value << 0.05 for all comparisons), indicating that microbiota is largely
337 determined by individual specificity, rather than by menstrual phase (**Figure 1B**).
338 Analyses of dynamic changes could not be performed for the mycobiome, due to the limited number
339 of samples for which mycobiome could be profiled. Although we sequenced the ribosomal internal
340 transcribed spacer (ITS) region in all 213 samples, only 25 samples had at least 200 reads, cutoff
341 deemed necessary to obtain fungi classification (**Figure S6** and **Table S13)**. Twenty-five fungal species
342 were identified, the most prevalent being *Candida albicans*, a fungus associated with vaginal
343 infections (candidiasis) but also frequently found as part of the normal vaginal microbiota. The
344 relatively low number of reads overall detected in the ITS region may be due to the use of a protocol
345 and kit optimized for bacterial rather than fungal DNA extraction, as reported in previous studies
346 [32].
347
348 **The impact of host phenotypes**
349 We then investigated the impact on vaginal microbiota of host phenotypes, lifestyle and information
350 on sample collection, by correlating genus-based and species-based beta diversity PCoA axes with
351 available metadata. Among the 12 variables tested, the variable "having had at least one pregnancy"
352 had the higher impact at both genus and species level, followed by "having used oral hormonal
353 contraceptives in the past" (**Table S14**). Other variables that were significant albeit with less impact
354 include age, waist-to-hip ratio, smoking habits, intercourse in the days preceding the swab, time of
355 collection, and use of medication or supplements **(Figure S7)**. Intriguingly, using Permanova test we
356 observed that the microbiota of women who had at least one pregnancy before participating in the
357 study had a lower likelihood to be dominated by *L. crispatus* (p=0.005), in line with observations
358 from the ISALA study [11], suggesting the existence of a potential lifelong pregnancy signature
359 (**Figure S8)**. Likewise, using Permanova we observed that the microbiota of older women is less likely
360 dominated by *Lactobacillus spp.* (p=0.003), in line with the observed decline in their abundance
361 during menopause (**Figure S9**)[5]; [33]. Similarly, dominance of *Lactobacillus spp*. is less frequent
362 when swabs are collected after feces (p=0.02) **(Figure S10),** and in those collected in the morning
363 (p=0.003) **(Figure S11).**
364
365 **Frequency of *L. iners* and *L. crispatus* along the menstrual cycle**

366 In contrast to findings from cohort studies on Northern European women, our study found a greater
367 abundance of *L. iners* than *L. crispatus*, a microbiome composition typically observed instead in
368 Hispanic and African populations [11]; [6]; [12]; [14]; [2] (**Table 1**). We ruled out the possibility that
369 this result is driven by a potential bias from the 16S rRNA gene profiling method, since comparison
370 of the relative abundances of these two *Lactobacillus spp* in 10 samples showed very high
371 concordance with profiles obtained using shot-gun whole-genome sequencing (WGS) method
372 (**Figure S12** and **Table S10b**). Likewise, we also rule out the possibility that this result is driven by the
373 higher number of reads compared to protocols used in previous studies. In fact, taxa relative
374 abundances were almost identical when using a subsampling approach on our 16S rRNA data to
375 restrict analyses to maximum 30K or 60K reads per sample (**Figure S13**).
376 In our cohort, *L. iners* and *L. crispatus* have a similar prevalence in the follicular phase (98% and 95%,
377 respectively), but the average relative abundance was higher for *L.* iners (average relative abundance
378 40% and 25% for *L. iners* and *L. crispatus*, respectively). The difference in average relative abundance
379 between these two species was negligible in the other 3 phases (absolute difference between
380 percentages: 2.03, -1.27, 2.21 in the ovulatory, early and late luteal phases, respectively).
381 Interestingly, all microbiotas dominated by *L. crispatus* in the follicular phase were stable during the
382 menstrual cycle, while a small fraction (20%) of those dominated by *L. iners* at the follicular phase
383 switched to *L. crispatus*-dominant - or to other species - in the ovulatory or in the late luteal phase
384 (**Figure 2**). In addition, the microbiota of two women dominated by low prevalent species in the
385 follicular phase became *L. crispatus*-dominant in the ovulatory phase. Inter-class correlation (ICC)
386 analyses indicated that these two species explained the largest variance of total variability between
387 women (**Table S15**). While these results suggest a trend for women's microbiota to switch to *L.*
388 *crispatus* during the ovulatory phase, *L. iners* still remains the most abundant species in the Italian
389 population across all time points.
390
391 **Vaginal microbiota dynamics**
392 Given these clear-cut changes observed in main *Lactobacillus* spp, we sought to thoroughly
393 investigate microbiota dynamics along the menstrual cycle. First, we assessed the composition and
394 variations along the menstrual cycle using the CST classification (see **Methods**). Overall, we found
395 that CST-V (*L. Jensenii* dominated) was the rarest among all samples and phases (2.5%), followed by
396 CST-II (*L. gasseri* dominated) (10.2%) and CST-IV (defined by moderate to high abundance of different
397 anaerobic species, such as *Prevotella spp* and *G. vaginalis,* and low abundance of *Lactobacillus spp*

398 (9.2%) (**Figure S14A**). In line with relative abundances, the CST-III (*L. iners* dominated) was the most
399 common (42.9%) and with higher prevalence in the follicular phase, while the second most common
400 CST was CST-I (*L. crispatus* dominated) (**Figure S14B**), whose prevalence increased during ovulatory
401 phase (**Figure S14B**)**.** Most women maintained a stable CST throughout the menstrual cycle: only 11
402 out of 61 women changed CST at any time point (18 if we consider changes in subCST) (**Figure 3A**).
403 Interestingly, 72% (N=8) of these changes occurred from follicular to ovulatory phases (**Figure 3B**),
404 and of which 6 were changes from CST-III or CST-IV (N=2) to CST-I (N=4). The remaining changes
405 occurred from early to luteal phase, with 3 women's microbiota switching to CST-IV.
406 None of the women's microbiotas classified as CST-I exhibited CST shifts, indicating the high stability
407 of communities dominated by *L. crispatus*. Of the women who undergo CST shifts, only one changed
408 CST types more than once (**Figure S15**). We were unable to detect association with CST/subCST
409 changes and presence of fungi, given the limited amount of information (only 4 out of 11 women
410 exhibiting shits in CST had fungi information at the relevant time points). Likewise, we did not detect
411 association with CST/subCST changes with having had sex intercourse in the two days prior sample
412 collection ($p>0.05$).
413 Secondly, we examined the longitudinal changes of taxa relative abundances using linear mixed
414 models and controlling for potential confounding variables, as described in **Methods**. We found that
415 the relative abundance of a few bacteria species varies significantly across phases, with the majority
416 remaining stable when different covariates were adjusted. For the genus *Lactobacillus*, *L. iners, L.*
417 *gasseri* and *L. jensenii* showed significant variation across mostly all corrections applied, while *L.*
418 *crispatus* was not significant in any of the models (**Figure 4**). Likewise, for the genus *Streptococcus,*
419 two taxa, namely *S. unclassified* and *S. agalactiae* varied significantly across phases regardless of
420 covariates used. Finally, another bacterial species, *Dialister B micraerophilus* showed significant
421 variation in 3 of the 10 models (**Figure 4).**

422

423 **Longitudinal changes of bacteria abundance associate with sex hormone levels**
424 We investigated whether changes in *Lactobacillus* and *Streptococcus* species between two
425 consecutive phases were associated with corresponding changes in sex hormone levels. Changes in
426 *L. iners* between ovulatory and early luteal phases were associated with changes in luteinizing
427 hormone (LH) levels ($p=0.04$), and those between early and late luteal phases in *L. jensenii* with
428 prolactin levels ($p=0.04$) (**Table S16** and **Figure 5A**). These results were however only nominally
429 significant, therefore larger samples size and/or additional cohorts are needed to be confirmed.

430 Furthermore, we identified several bacteria whose longitudinal fluctuations along the entire
431 menstrual cycle were significantly associated with levels of sex hormones – for a total of 14
432 associations (**Figure 5B** and **Figure S16** and **Table S17)**. Most of the associations were observed with
433 beta-17 estradiol (BES17). Some were in common between BES17 and LH, including *Prevotella* genus
434 and species *P. timonensis_A*, and genera *Finegoldia* and *Anaerococcus* genera. Only one association
435 was detected with prolactin (PRL), namely with abundance of *Bifidobacterium vaginale*. All these
436 associations were negative, indicating negative correlations between sex hormones levels and the
437 relative abundance of these bacteria (**Figure 5B** and **Figure S16**). When considering permuted p-
438 values to evaluate robustness of results to outliers, the number of significant associations decreased
439 to 9 out of 14. Notably, 4 of these 9 associations, all linked to BES17, also meet a stricter significance
440 threshold applied after correcting for multiple testing, suggesting that these are likely genuine.
441

442 # DISCUSSION

443 We carried out the first short-term longitudinal study on the vaginal microbiota of non-pregnant,
444 healthy women from an Italian population. In the W4H cohort we observed a higher abundance of
445 *L. iners* compared to *L. crispatus*, in contrast from earlier studies on cohorts from Northern Europe
446 [6]; [11] (**Table 1**). Of note, a recent study on young French women [15] reported again a higher
447 prevalence of *L. crispatus*-dominated CST (CST-I) over the *L. iners*-dominated *CST* (CST-III), although
448 the difference in frequency between the two CSTs was less pronounced (40.5 vs 38.1%, while our
449 study has obtained 35% for CST-I and 43% for CST-III) (Figure S14) compared to observations from
450 cohorts based in Belgium (ISALA study), Danmark and Sweden (MiMens study) [11]; [6]. In this
451 cohort, 70% of women were enrolled in Bordeaux (south of France). Hence, considered these results
452 and the higher prevalence of *L. iners* in African and Hispanic cohorts [12]; [2]; [13], our study of the
453 Italian population provides evidence for a north-to-south gradient of increased *L. iners* abundance
454 even within Europe.
455 The reasons for these geographical differences in microbiota composition remain unclear but results
456 from our study and those from the ISALA study offer insights for speculation. For example, in our
457 study we found that women who previously had children were less likely to have a microbiota
458 dominated by *L. crispatus*. Likewise, in the ISALA study having children was negatively correlated
459 with *L. crispatus* and positively with *L. iners*; this variable was the strongest host factor associated -
460 with opposite direction - with the abundance of these two species [11]. Geographic differences may
461 thus be related to the delivery mode, or to post-pregnancy health care practices. Another

contributing factor might be related to different cultural and hygiene practices, such as the soap type used, the frequency of washing, the use of intimate wipes and even the type of pads, tampons or other periods supplies [34]; [4]. While we do not have this information in our cohort, results from the ISALA study indicate that the use of menstrual pads is negatively correlated with *L. crispatus* abundance, while the use menstrual cups was positively correlated [11]. Intriguingly, the ISALA study also reported a positive correlation with *L. Crispatus* abundance with use of hormonal contraceptive methods (combination pill, vaginal ring or patch). Previously, also Tuddenham S. [35] showed that the proportion of CST I (*L. crispatus*-dominated) was higher among users of oral hormonal contraceptive than the proportion of CST I among non-users, albeit this result was significant only in White and not African American. No association was instead seen in these two studies between hormonal contraceptives use and *L.iners*.

All our volunteers had a natural menstrual cycle and did not use hormonal contraceptives for at least the past 30 days. This, together with the overall limited use of hormonal contraceptive in the Italian population (19.1% of females during reproductive age versus ~30% in Northern and Western European countries, according to WHO [36] could explain the observed low abundance of *L. Crispatus*. Finally, differences in abundance of the two *Lactobacillus* species may be attributed to variations in the host-immune system, in the amount and quality of vaginal secretion, in genetic variants of gene encoding receptors expressed on the surface of the vaginal epithelial cells, and other genetically determined factors specific to each host.

The very high prevalence and abundance of *L. iners* in our cohort of healthy women is striking and challenges the idea that *L. iners* dominated microbiotas are 'unhealthy' or ‘dysbiotic’. In fact, W4H volunteers have reported no current symptomatic vaginal infections, so we can rule out the abundance of *L. iners* in our samples due to either vaginal pathologies or infections. We speculate that the presence of different strains could be the most plausible explanation of geographic and clinical differences. Unfortunately, with the 16S sequencing approach we (and others) are unable to identify which strains of *L. iners* are present in the studied vaginal microbiota. However, given its high genetic variability, is reasonable to assume that certain strains of *L. iners* may promote healthy conditions, while others may increase risk of infection, thus explaining its presence as species abundance in both healthy and dysbiotic vaginal conditions [37]; [8]. In addition, a woman can be colonized by multiple coexisting strains with different genetic characteristics and metabolic functions, enabling the survival of this species in a variety of vaginal microenvironments [38]. A deep

493 genetic study on *L. iners* strains is needed to better understand the ecology of this bacterial species
494 and their differential impact on infections and other gynecological diseases.
495 Our results on dynamic changes along menstrual cycle phases agree with previous studies reporting
496 that vaginal microbiota is generally stable over time and support current knowledge on the strong
497 stability of microbiomes dominated by *L. crispatus* [3]; [15]; [6]. Furthermore, we were able to
498 pinpoint to species responsible for the dynamic changes occurring in a minority of women, namely
499 *L. iners*, *L. gasseri, L. jensenii* and *Streptococcus spp*. The decreased stability conferred by these
500 species could be explained by the production of two distinct isoforms of lactic acid by these vaginal
501 microorganisms. Unlike *L. Crispatus, L. iners*, *Bifidobacterium spp*. and *Streptoccosus spp*. produce
502 the L- enantiomer of lactic acid that allows easier colonization by other microbes than D- enantiomer
503 [39]; [4]. Furthermore, *L. iners* possesses additional genes compared to *L. crispatus*, encoding stress-
504 tolerance proteins, iron–sulfur proteins, exogenous L-cysteine transport, and exhibits superior
505 metabolic adaptation to the changing carbohydrate sources in the vaginal environment [40]; [1].

506 Intriguingly, we observed that even if only a few women's microbiotas change along the menstrual
507 cycle, most changes occur from follicular to ovulatory phase, with most of these aligning to an *L.*
508 *crispatus* profile. Beta diversity was also lower during the ovulatory and early luteal phases
509 compared to follicular phase, which could be due to a protective mechanism to support a potentially
510 fertilized ovum. It is possible that the fluctuation of sex hormones plays a key role in vaginal
511 microbiota changes. Our analyses showed that BES17 has the major impact on longitudinal variation
512 in taxa abundances, although none of the hormone-taxa association involved *Lactobacillus* species
513 and thus they cannot fully explain changes we observed during ovulatory phase. Nonetheless, these
514 associations with BES17 are fascinating. In fact, we detected significant negative relationships with
515 *Prevotella spp*., *Finegoldia* and *Corynebacterium* genera, taxa that are often found at high
516 abundances in women after menopause with menopause-associated symptoms, such as vulvo-
517 vaginal atrophy and genito-urinary symptoms [41]. The decline in estrogen levels during menopause
518 may play a prominent role in driving microbiota change, as reduced estrogen leads to a decrease in
519 glycogen by vaginal cells. This causes an elevated vaginal pH, creating conditions that favor the
520 colonization of multiple microbial species adapted to a less acidic environment.

521 Our study has several strengths and novelties, including the high depth 16S rRNA sequencing, the
522 enrollment of naturally menstruating women, the direct measurement of sex hormones, and the
523 investigation of a south European healthy population. While our results have not a direct clinical

implication, knowledge of vaginal microbiota dynamics under physiological sex hormones variations in all populations is an essential route for investigating the potential role of microbes also in diagnosis and prevention of diseases conditions such infertility and endometriosis.

We also recognize the limitations of the study. First, we did not collect samples during menses, a stage where important changes in bacterial composition may occur. Moreover, we have focused on a single menstrual cycle and therefore we were unable to assess if observed changes were cyclic, i.e. the microbiota of that minority of women experiencing shift returns to the original profile after menstruation. Furthermore, we have no measurement of vaginal pH, a crucial factor in the regulation of the microbiota. Finally, while we have employed a very deep 16S sequencing approach, we are limited to the relative abundance of bacteria and do not have information on bacterial function or strains, features that metagenomic sequencing could provide.

In conclusion, our study highlights the importance to extend investigations on vaginal microbiota in healthy women from diverse populations, also from the same ethnicity, and the need to assess stability and dynamics with sex hormone changes, including those occurring during a natural menstrual cycle.

DATA AVAILABILITY STATEMENT

All our code used to analyze the data and instructions are available at: https://github.com/Sanna-s-LAB/Women4Health

The human-depleted 16S rRNA and ITS sequences are deposited at the Sequence Read Archive (SRA) repository under accession number PRJNA1222832 [which will be released upon acceptance of the manuscript]. Participant-level personally identifiable data cannot be shared in respect to participants' informed consent and are protected under the Italian Personal Data Protection Code and European Regulation 2016/679 of the European Parliament and of the Council (GDPR) that prohibit distribution even in pseudo-anonymized form.

All data necessary to support the conclusions drawn in this study are available in the Supplementary Tables.

URLs

Code for the reclassification of *Lactobacillus sp.:*

https://github.com/LebeerLab/Citizen-science-map-of-the-vaginal-microbiome

Code and files for VALENCIA algorithm

https://github.com/ravel-lab/VALENCIA

Tools

FastQC: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Trimmomatic: https://github.com/timflutre/trimmomatic?tab=readme-ov-file

MultiQC: https://github.com/MultiQC/MultiQC

R packages:

vegan: https://github.com/vegandevs/vegan/

lme4: 10.32614/CRAN.package.lme4

lmerTest: 10.32614/CRAN.package.lmerTest

Rtsne: 10.32614/CRAN.package.Rtsne

Parallel: https://www.rdocumentation.org/packages/parallel/versions/3.6.2

STAT: https://doi.org/10.32614/CRAN.package.STAT

DADA2: https://www.bioconductor.org/packages/release/bioc/html/dada2.html

ShortRead: https://kasperdanielhansen.github.io/genbioconductor/html/ShortRead.html

Performance: 10.32614/CRAN.package.performance

sklearn: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Databases:

GTDB: https://gtdb.ecogenomic.org/stats/r95

UNITE DB: https://doi.org/10.15156/BIO/2959330

AUTHOR'S CONTRIBUTION

Conceptualization: F.D.S., G.G., S.S.

Writing - original draft Preparation: E.V., F.C., A.M., S.S.

Writing – review and editing: F.B., M.L.F.

Project supervision: S.S.

Funding acquisition: G.G., S.S.

Statistical and Bioinformatic analyses: E.V., F.C., S.S.

Critical support for bioinformatic analyses: D.V.Z., V. LF., R.G., J.S., N.K., A.Z., S.S.

Samples and data collection: S.L., S.C., G.V.B., A.K., F.D.S., D.M, G.G.

Vaginal samples processing: A.M., F.Cr., S.I., F.B., M.L.F.

All authors read and revised the manuscript and approved its final version.

ACKNOWLEDGMENTS

We thank all the volunteers who participated in the study for their time and commitment. We thank the physicians of the University of Trieste for their contribution to the recruitment phase, including but not limited to, Francesco Cracco, Roberta Maria Gentile, Elena Stefani and Michele Stracquadaini, the Directors of I.R.C.C.S. Burlo-Garofolo and of IRGB-CNR for logistic support, the Agenzia Regionale Sardegna Ricerche for providing access to laboratories, technologists Davide Murrau, Marco Masala and Michele Marongiu from the IRGB-CNR for IT support. Finally, we thank Dr. Alessandra Meloni and Dr. Ferdinando Coghe from the Azienda Ospedaliero Universitaria di Cagliari and Prof. Stefano Guerriero from the University of Cagliari for having joined the Women4Health project and initiating a second recruiting center in Cagliari starting in October 2024.

FUNDING STATEMENT

This study was co-funded by the Italian Ministry of Health through the contribution given to the Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste - Italy (SD 02/21 to G.G.), by the European Union (ERC Stg 2022 to S.S., acronym SEMICYCLE, GA n.101075624), by the Next Generation EU, in the context of the National Recovery and Resilience Plan, Investment PE8 – Project Age-It: “Ageing Well in an Ageing Society, ” [DM 1557 11.10.2022 to S.S.] and by NutrAGE grant (CNR Project FOE-2021 DBA.AD005.225). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. In addition, S.S. also received a PRIN2022 grant by the Next Generation EU funds (DSB.PN004.021 2022PMZKEC_LS2_PRIN2022 SANNA, CUP B53D23008300006).

USE OF AI STATEMENT

The authors declare that they have used generative artificial intelligence, specifically Quillbot, ChatGPT and Reverso online platforms only to check spelling and grammar. The authors declare that ChatGPT was also used to improve R scripts for handling data formatting and generation of figures.

618 **Table 1. Prevalence and relative abundance of *L. iners* and *L. crispatus* in our study and other European cohorts.**

619 The table reports the statistics from previously published studies on European cohorts and those from our study, for *L. iners* and *L. crispatus*. The
620 column Time point indicates weather statistics refer to samples collected in a specific menstrual cycle phase or at any time during menstrual cycle
621 (or menopause).

							Frequency (%)		**Prevalence (%)**		**Average Relative Abundance (%)**	
Study	**Country**	**Sequencing Method**	**N reads (K)**	**N samples**	**Age range**	**Time point**	***CST-I***	***CST-III***	***L.iners***	***L.crispatus***	***L.iners***	***L.crispatus***
ISALA	Belgium	16S (V4)	25	3000	18-73	any	43	28	72	90	24	38
MiMens	Denmark, Sweden	16S (V3V4) WGS	--	49	20-28	Follicular phase	--	--	63	76	12	59
						Ovulatory phase	--	--	66	83	14	67
i-Predict	France	16S (V3V4)	34	241	18-25	any	41	38	--	--	38	45
W4H	Italy	16S (V3V4, V7V9)	155	61	18-45	Follicular phase	22	31	98	95	40	25
						Ovulatory phase	31	27	93	93	37	35

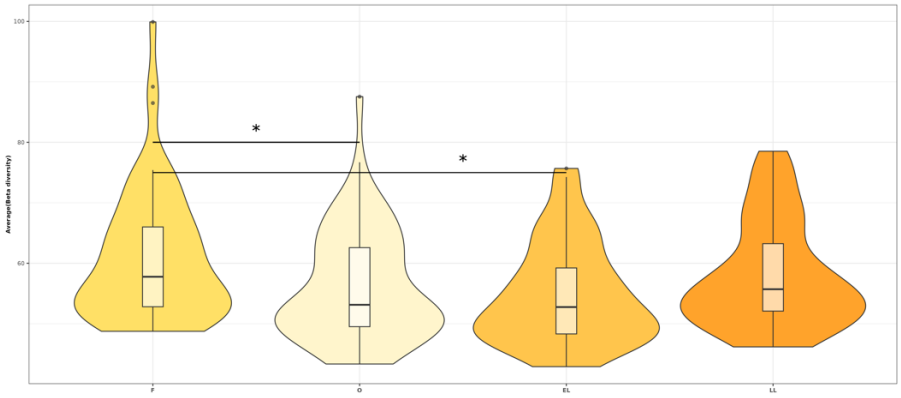
622

Figures

Figure 1. Average beta diversity across menstrual cycle phases.

In **(A)** Average Beta diversity across the four menstrual cycle phases: Follicular phase (F), Ovulatory phase (O), Early Luteal phase (EL), and Late Luteal phase (LL). The black line inside the boxplot represents the median, with the lowest and highest values within the 1.5 interquartile range represented by the whiskers. The black lines between violins indicate significant differences (*p< 0.01). P-values were obtained using paired T-test. In **(B)** Average beta diversity within the same woman across phases (Self) and average of beta diversity among different women (Random 1 to Random 10). In both panels, beta diversity was calculated using Euclidean distances on transformed taxa adjusted for technical covariates (see **Material and methods)**.

A)



B)

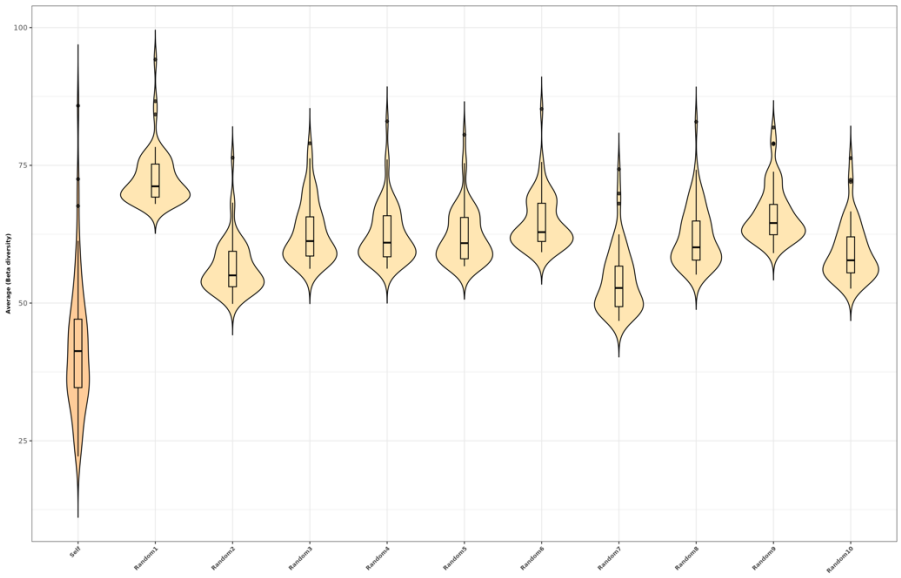
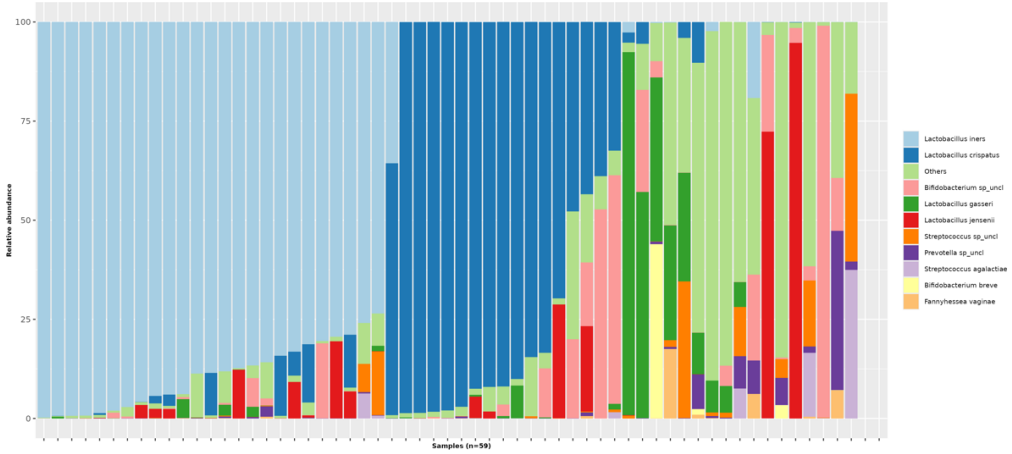
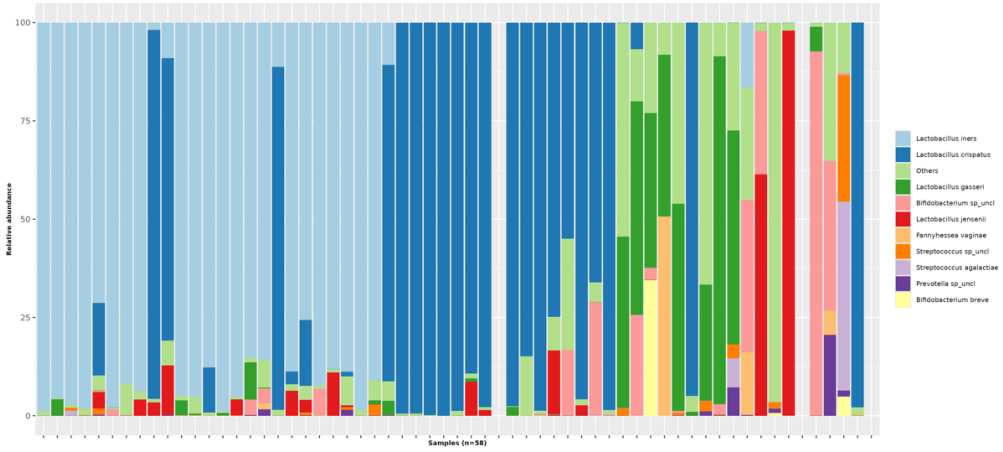


Figure 2: Distribution of the most abundant taxa across menstrual cycle phases.
Relative abundance of the 10 most abundant species in the four different phases. In **(A)** samples are sorted according to the abundance of *L. iners*, *L. crispatus*, *L. gasseri*, respectively. Samples in **(B)**, **(C)**, and **(D)** are ordered as in **(A)**. The grey bars show missing information for those women in the specific week. In the x-axis title we indicate for how many samples microbiome profile was available.

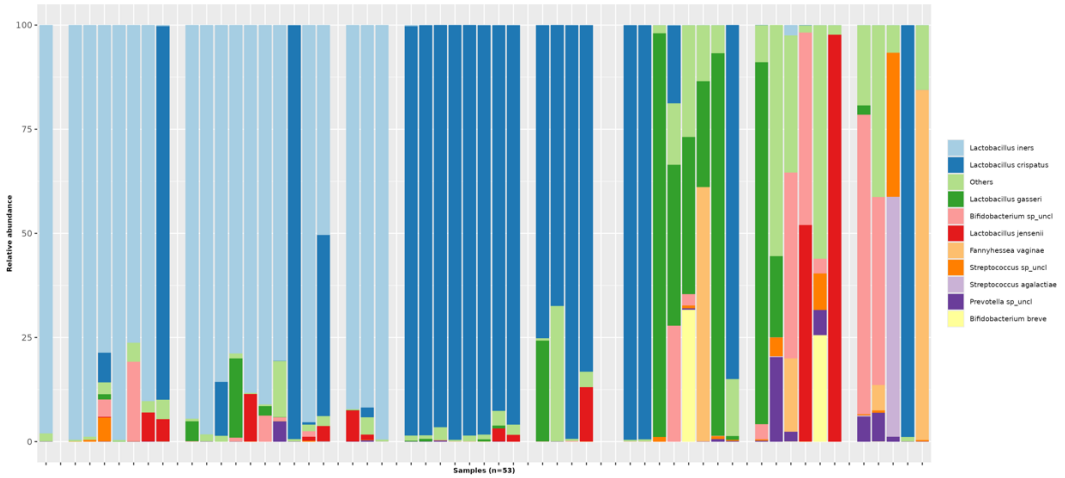
A)



B)



C)


100
75
50
25
0
Relative abundance
Samples (n=53)
Lactobacillus iners
Lactobacillus crispatus
Others
Lactobacillus gasseri
Bifidobacterium sp_uncl
Lactobacillus jensenii
Fannyhessea vaginae
Streptococcus sp_uncl
Streptococcus agalactiae
Prevotella sp_uncl
Bifidobacterium breve

D)

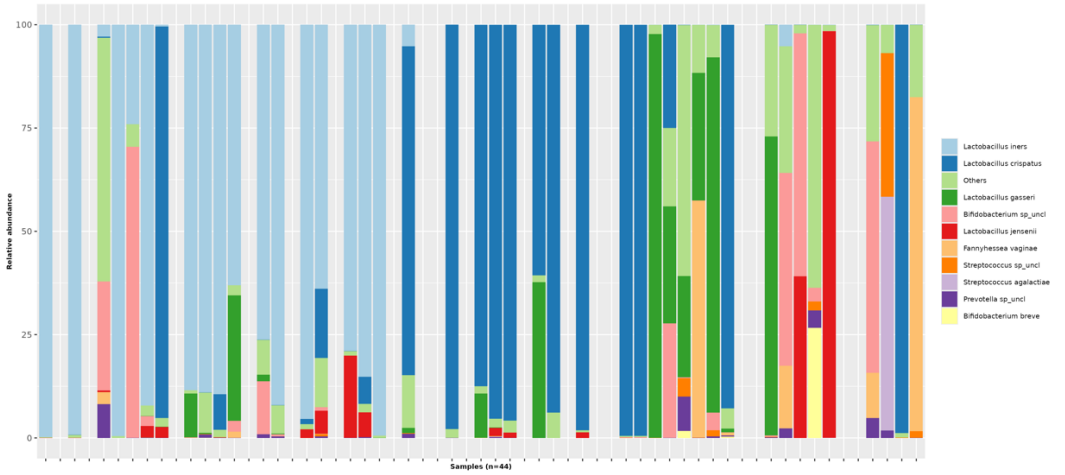
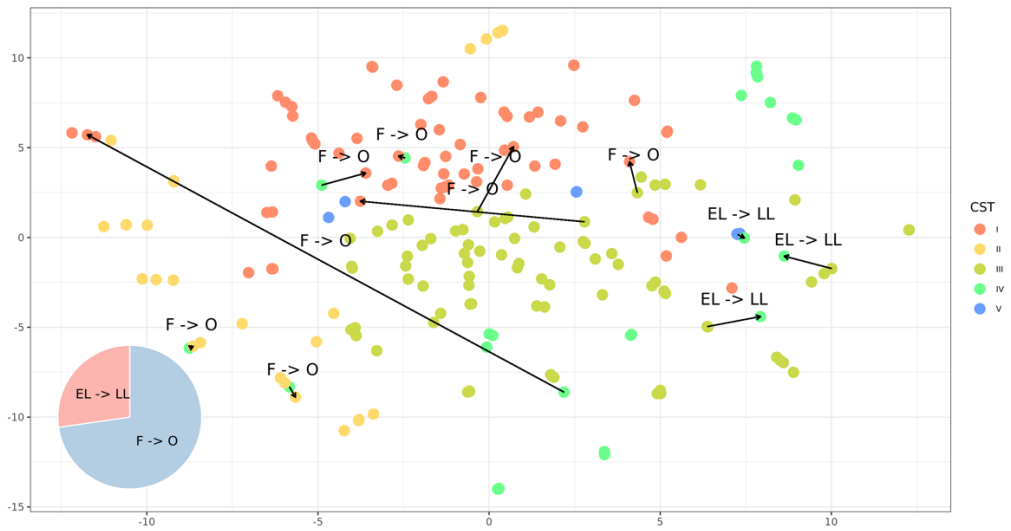

100
75
50
25
0
Relative abundance
Samples (n=44)
Lactobacillus iners
Lactobacillus crispatus
Others
Lactobacillus gasseri
Bifidobacterium sp_uncl
Lactobacillus jensenii
Fannyhessea vaginae
Streptococcus sp_uncl
Streptococcus agalactiae
Prevotella sp_uncl
Bifidobacterium breve

Figure 3: Community State Type (CST) across menstrual cycle phases

In **(A)** t-SNE representation of the vaginal microbiome composition, with samples colored according to their CST classification. Arrows highlight instances where a woman changes CST between consecutive phases, with the notation start_phase → final_phase displayed above each arrow. In the bottom-left corner, a pie chart shows the proportions of phases where CST changes occur. The legend provides the color coding for each CST. In **(B)** Alluvial plot showing Community State Type (subCST) changes between phases. The four columns represent the Follicular (F), Ovulatory (O), Early Luteal (EL), and Late Luteal (LL) phases, respectively. In legend of the plot: CH: a change occurred in at least one phase, complete data available (15/61); CH + NA: a change occurred in at least one phase, missing samples in some phases (3/61); NO CH: No changes occurred, complete data available (26/61); NO CH + NA: No changes occurred, but missing samples in some phases (17/61).

A)



B)

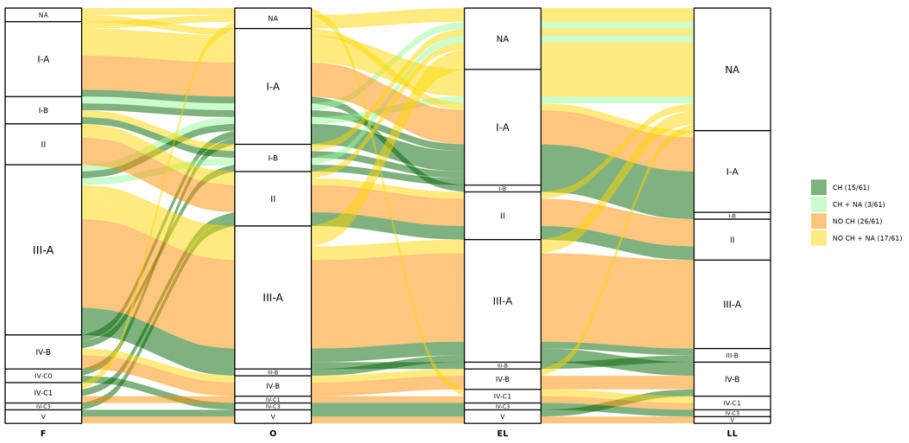


Figure 4: Longitudinal changes in taxa relative abundance across phases

This figure shows the results of a linear model investigating whether bacterial abundances change across menstrual cycle phases. In each panel, the model coefficients for each taxon are shows with points representing the beta coefficient of the variable time and the bars the 95% confidence interval of the estimate. The actual p-values are shown above each point, and the permuted p-values are displayed in blue below each point for actual p-values <0.05, or NA otherwise. Colour coding indicates the significance of the associations according to original p-values.

Each panel corresponds to the same model using different correction for the taxa, with the addition of an increasing number of covariates. In particular: model 1 is the basic model with CLR transformed bacteria, model 2 adds to previous model the technical covariates, model 3 adds Age, model 4 adds sex_less_than_two_days, model 5 adds Pregnancy_category, model 6 adds Pill_use, model 7 adds Swab_after_feces, model 8 adds Bristol_stool_scale, model adds Swab_morning_or_not and model 10 adds WHR_ranges.

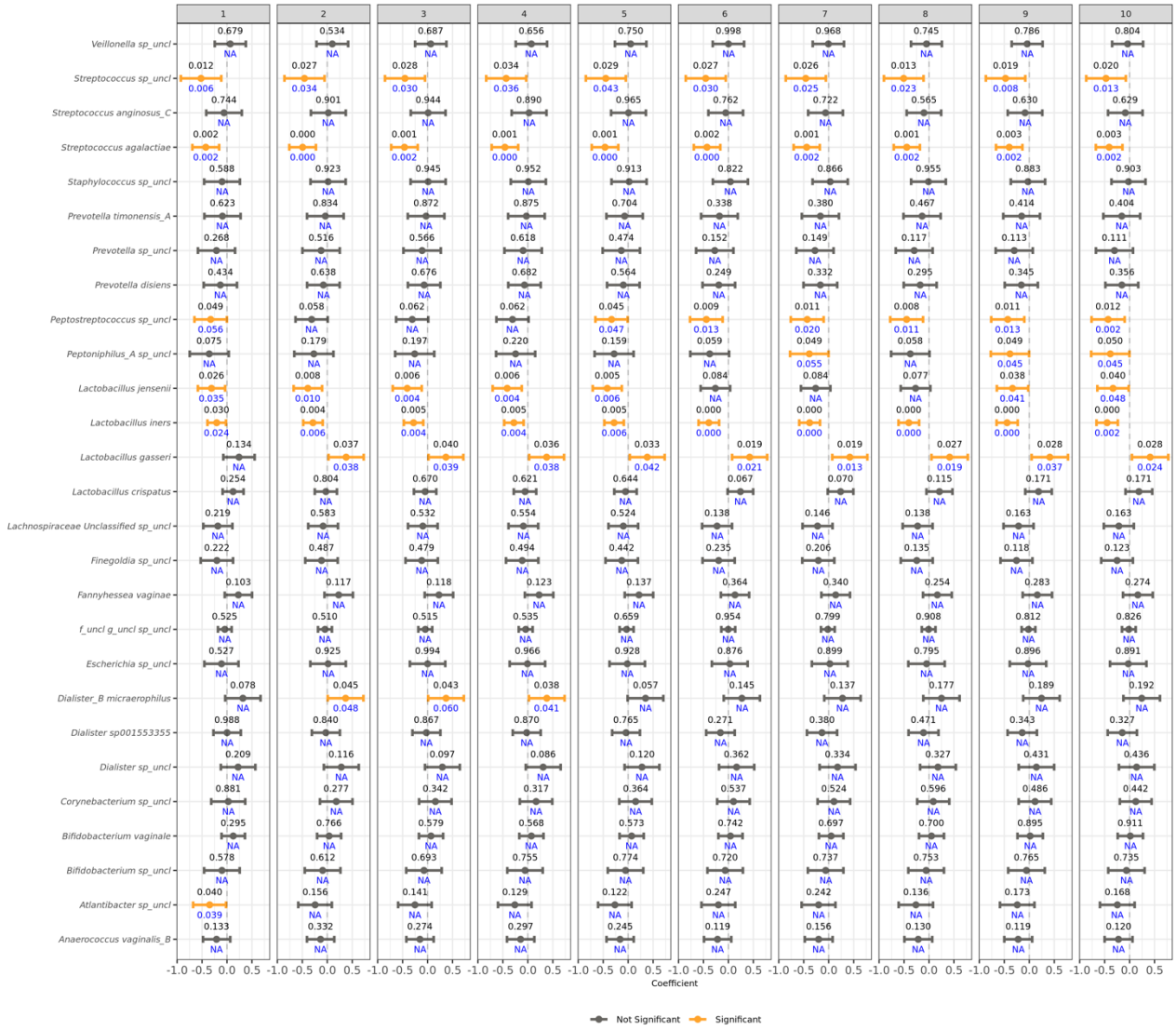
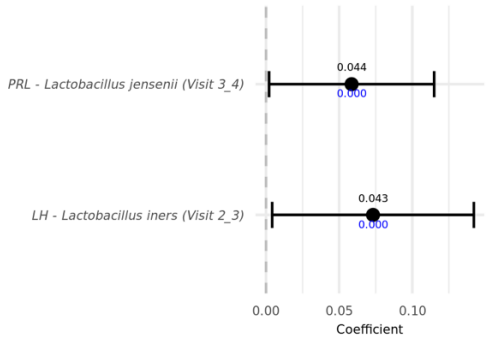


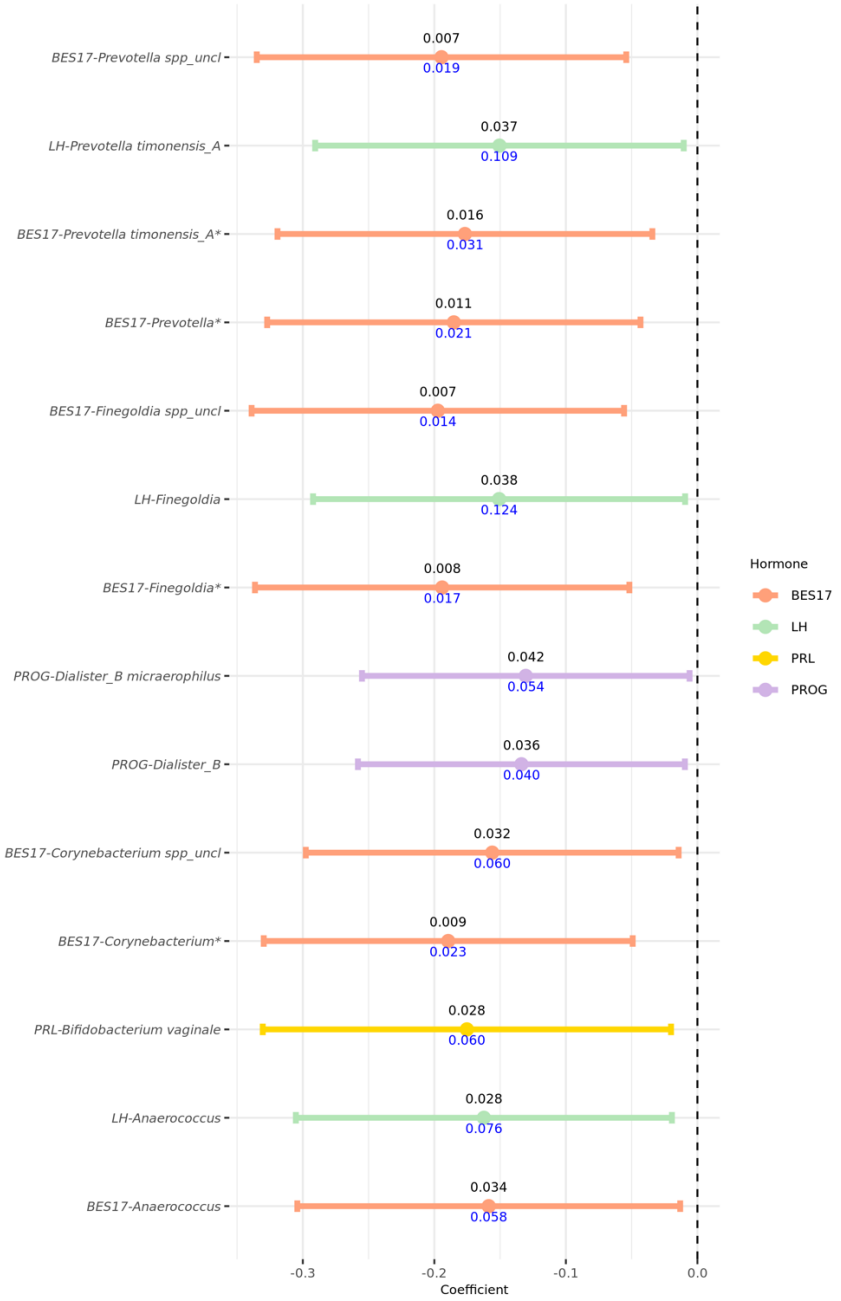
Figure 5: Relationship between sex hormones and bacteria

These forest plots display the significant associations between sex hormones and bacterial taxa when considering two consecutive phases **(A)** or in the overall menustral cycle **(B)**. The x-axis represents the model coefficients, indicating the strength and direction of each association, while the y-axis lists bacterial taxa. Each line illustrates the 95% confidence interval for the coefficient, with colors representing the sex hormone associated with each taxon. The p-value for each model is shown above the corresponding point, and the permuted p-value (in blue) is shown below. Associations that remain significant after FDR correction, with a threshold of 0.2, are highlighted with an asterisk on the y-axis. In the legend, the color coding for the hormones is as follows: BES17=beta-estradiol, FSH=follicle-stimulating hormone, LH=luteinizing hormone, PRL=prolactin, and PROG=progesterone.

A)



B)


BES17-Prevotella spp_uncl
0.007
0.019
LH-Prevotella timonensis_A
0.037
0.109
BES17-Prevotella timonensis_A*
0.016
0.031
BES17-Prevotella*
0.011
0.021
BES17-Finegoldia spp_uncl
0.007
0.014
LH-Finegoldia
0.038
0.124
BES17-Finegoldia*
0.008
0.017
PROG-Dialister_B micraerophilus
0.042
0.054
PROG-Dialister_B
0.036
0.040
BES17-Corynebacterium spp_uncl
0.032
0.060
BES17-Corynebacterium*
0.009
0.023
PRL-Bifidobacterium vaginale
0.028
0.060
LH-Anaerococcus
0.028
0.076
BES17-Anaerococcus
0.034
0.058
-0.3
-0.2
-0.1
0.0
Coefficient
Hormone
BES17
LH
PRL
PROG

REFERENCE

[1] M. T. France, H. Mendes-Soares, e L. J. Forney, «Genomic Comparisons of Lactobacillus crispatus and Lactobacillus iners Reveal Potential Ecological Drivers of Community Composition in the Vagina», *Appl. Environ. Microbiol.*, vol. 82, fasc. 24, Art. fasc. 24, dic. 2016, doi: 10.1128/AEM.02385-16.

[2] J. Ravel *et al.*, «Vaginal microbiome of reproductive-age women», *Proc. Natl. Acad. Sci.*, vol. 108, fasc. supplement_1, pp. 4680–4687, mar. 2011, doi: 10.1073/pnas.1002611107.

[3] P. Gajer *et al.*, «Temporal Dynamics of the Human Vaginal Microbiota», *Sci. Transl. Med.*, vol. 4, fasc. 132, pp. 132ra52-132ra52, mag. 2012, doi: 10.1126/scitranslmed.3003605.

[4] J. B. Holm *et al.*, «Integrating compositional and functional content to describe vaginal microbiomes in health and disease», *Microbiome*, vol. 11, fasc. 1, p. 259, nov. 2023, doi: 10.1186/s40168-023-01692-x.

[5] P. Łaniewski e M. M. Herbst-Kralovetz, «Connecting microbiome and menopause for healthy ageing», *Nat. Microbiol.*, vol. 7, fasc. 3, pp. 354–358, mar. 2022, doi: 10.1038/s41564-022-01071-6.

[6] L. W. Hugerth *et al.*, «Defining Vaginal Community Dynamics: daily microbiome transitions, the role of menstruation, bacteriophages, and bacterial genes», *Microbiome*, vol. 12, fasc. 1, Art. fasc. 1, ago. 2024, doi: 10.1186/s40168-024-01870-5.

[7] J. Ravel *et al.*, «Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis», *Microbiome*, vol. 1, fasc. 1, Art. fasc. 1, dic. 2013, doi: 10.1186/2049-2618-1-29.

[8] M. I. Petrova, G. Reid, M. Vaneechoutte, e S. Lebeer, «*Lactobacillus iners*: Friend or Foe?», *Trends Microbiol.*, vol. 25, fasc. 3, pp. 182–191, mar. 2017, doi: 10.1016/j.tim.2016.11.007.

[9] S. M. Bloom *et al.*, «Cysteine dependence of Lactobacillus iners is a potential therapeutic target for vaginal microbiota modulation», *Nat. Microbiol.*, vol. 7, fasc. 3, Art. fasc. 3, mar. 2022, doi: 10.1038/s41564-022-01070-7.

[10] C. A. Muzny, P. Łaniewski, J. R. Schwebke, e M. M. Herbst-Kralovetz, «Host–vaginal microbiota interactions in the pathogenesis of bacterial vaginosis», *Curr. Opin. Infect. Dis.*, vol. 33, fasc. 1, p. 59, feb. 2020, doi: 10.1097/QCO.0000000000000620.

[11] S. Lebeer *et al.*, «A citizen-science-enabled catalogue of the vaginal microbiome and associated factors», *Nat. Microbiol.*, vol. 8, fasc. 11, Art. fasc. 11, nov. 2023, doi: 10.1038/s41564-023-01500-0.

[12] M. G. Serrano *et al.*, «Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy», *Nat. Med.*, vol. 25, fasc. 6, Art. fasc. 6, giu. 2019, doi: 10.1038/s41591-019-0465-8.

[13] X. Wei *et al.*, «Vaginal microbiomes show ethnic evolutionary dynamics and positive selection of *Lactobacillus* adhesins driven by a long-term niche-specific process», *Cell Rep.*, vol. 43, fasc. 4, p. 114078, apr. 2024, doi: 10.1016/j.celrep.2024.114078.

[14] O. S. E. Roachford, A. T. Alleyne, e K. E. Nelson, «Insights into the vaginal microbiome in a diverse group of women of African, Asian and European ancestries», *PeerJ*, vol. 10, p. e14449, nov. 2022, doi: 10.7717/peerj.14449.

[15] J. Tamarelle *et al.*, «Vaginal microbiota stability over 18 months in young student women in France», *Eur. J. Clin. Microbiol. Infect. Dis.*, set. 2024, doi: 10.1007/s10096-024-04943-3.

[16] M. T. France *et al.*, «VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition», *Microbiome*, vol. 8, fasc. 1, Art. fasc. 1, nov. 2020, doi: 10.1186/s40168-020-00934-6.
[17] F. Busonero *et al.*, «The Women4Health cohort: a unique cohort to study women-specific mechanisms of cardio-metabolic regulation», *Eur. Heart J. Open*, vol. 4, fasc. 2, Art. fasc. 2, mar. 2024, doi: 10.1093/ehjopen/oeae012.
[18] P. Ewels, M. Magnusson, S. Lundin, e M. Käller, «MultiQC: summarize analysis results for multiple tools and samples in a single report», *Bioinforma. Oxf. Engl.*, vol. 32, fasc. 19, Art. fasc. 19, ott. 2016, doi: 10.1093/bioinformatics/btw354.
[19] A. M. Bolger, M. Lohse, e B. Usadel, «Trimmomatic: a flexible trimmer for Illumina sequence data», *Bioinforma. Oxf. Engl.*, vol. 30, fasc. 15, Art. fasc. 15, ago. 2014, doi: 10.1093/bioinformatics/btu170.
[20] B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, e S. P. Holmes, «DADA2: High resolution sample inference from Illumina amplicon data», *Nat. Methods*, vol. 13, fasc. 7, Art. fasc. 7, lug. 2016, doi: 10.1038/nmeth.3869.
[21] D. H. Parks, M. Chuvochina, P.-A. Chaumeil, C. Rinke, A. J. Mussig, e P. Hugenholtz, «A complete domain-to-species taxonomy for Bacteria and Archaea», *Nat. Biotechnol.*, vol. 38, fasc. 9, Art. fasc. 9, set. 2020, doi: 10.1038/s41587-020-0501-8.
[22] K. Abarenkov *et al.*, «The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered», *Nucleic Acids Res.*, vol. 52, fasc. D1, Art. fasc. D1, gen. 2024, doi: 10.1093/nar/gkad1039.
[23] J. Oksanen *et al.*, «Vegan: Community Ecology Package. R package version 2.0-2», gen. 2012.
[24] L. van der Maaten e G. Hinton, «Visualizing Data using t-SNE», *J. Mach. Learn. Res.*, vol. 9, fasc. 86, pp. 2579–2605, 2008.
[25] J. Krijthe e L. van der M. (Author of original C. code), *Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation*. (7 dicembre 2023). Consultato: 16 dicembre 2024. [Online]. Disponibile su: https://cran.r-project.org/web/packages/Rtsne/index.html
[26] K. Bolar, *STAT: Interactive Document for Working with Basic Statistical Analysis*. (1 aprile 2019). Consultato: 16 dicembre 2024. [Online]. Disponibile su: https://cran.r-project.org/web/packages/STAT/index.html
[27] F. Pedregosa *et al.*, «Scikit-learn: Machine Learning in Python», *J Mach Learn Res*, vol. 12, fasc. null, pp. 2825–2830, nov. 2011.
[28] A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, e S. P. Jensen, *lmerTest: Tests in Linear Mixed Effects Models*. (23 ottobre 2020). Consultato: 16 dicembre 2024. [Online]. Disponibile su: https://cran.r-project.org/web/packages/lmerTest/index.html
[29] R. A. Peterson, «Finding Optimal Normalizing Transformations via bestNormalize», *R J.*, vol. 13, fasc. 1, pp. 294–313, giu. 2021.
[30] D. Bates *et al.*, *lme4: Linear Mixed-Effects Models using «Eigen» and S4*. (11 gennaio 2025). Consultato: 14 gennaio 2025. [Online]. Disponibile su: https://cran.r-project.org/web/packages/lme4/index.html
[31] A. Blanco-Míguez *et al.*, «Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4», *Nat. Biotechnol.*, vol. 41, fasc. 11, Art. fasc. 11, nov. 2023, doi: 10.1038/s41587-023-01688-w.

[32] S. Ahannach *et al.*, «Microbial enrichment and storage for metagenomics of vaginal, skin, and saliva samples», *iScience*, vol. 24, fasc. 11, p. 103306, nov. 2021, doi: 10.1016/j.isci.2021.103306.
[33] A. L. Muhleisen e M. M. Herbst-Kralovetz, «Menopause and the vaginal microbiome», *Maturitas*, vol. 91, pp. 42–50, set. 2016, doi: 10.1016/j.maturitas.2016.05.015.
[34] A. M. Holdcroft, D. J. Ireland, e M. S. Payne, «The Vaginal Microbiome in Health and Disease-What Role Do Common Intimate Hygiene Practices Play?», *Microorganisms*, vol. 11, fasc. 2, p. 298, gen. 2023, doi: 10.3390/microorganisms11020298.
[35] S. Tuddenham *et al.*, «*Lactobacillus*-dominance and rapid stabilization of vaginal microbiota in combined oral contraceptive pill users examined through a longitudinal cohort study with frequent vaginal sampling over two years», *eBioMedicine*, vol. 87, p. 104407, gen. 2023, doi: 10.1016/j.ebiom.2022.104407.
[36] United Nations, *Contraceptive Use by Method 2019: Data Booklet*. United Nations, 2019. doi: 10.18356/1bd58a10-en.
[37] W. Kwak *et al.*, «Complete Genome of Lactobacillus iners KY Using Flongle Provides Insight Into the Genetic Background of Optimal Adaption to Vaginal Econiche», *Front. Microbiol.*, vol. 11, mag. 2020, doi: 10.3389/fmicb.2020.01048.
[38] J. B. Holm, K. A. Carter, J. Ravel, e R. M. Brotman, «Lactobacillus iners and Genital Health: Molecular Clues to an Enigmatic Vaginal Species», *Curr. Infect. Dis. Rep.*, vol. 25, fasc. 4, pp. 67–75, apr. 2023, doi: 10.1007/s11908-023-00798-5.
[39] S. S. Witkin, H. Mendes-Soares, I. M. Linhares, A. Jayaram, W. J. Ledger, e L. J. Forney, «Influence of Vaginal Bacteria and d- and l-Lactic Acid Isomers on Vaginal Extracellular Matrix Metalloproteinase Inducer: Implications for Protection against Upper Genital Tract Infections», *mBio*, vol. 4, fasc. 4, p. 10.1128/mbio.00460-13, ago. 2013, doi: 10.1128/mbio.00460-13.
[40] J. M. Macklaim, G. B. Gloor, K. C. Anukam, S. Cribby, e G. Reid, «At the crossroads of vaginal health and disease, the genome sequence of Lactobacillus iners AB-1», *Proc. Natl. Acad. Sci.*, vol. 108, fasc. supplement_1, pp. 4688–4695, mar. 2011, doi: 10.1073/pnas.1000086107.
[41] M. N. Anahtar *et al.*, «Cervicovaginal Bacteria Are a Major Modulator of Host Inflammatory Responses in the Female Genital Tract», *Immunity*, vol. 42, fasc. 5, pp. 965–976, mag. 2015, doi: 10.1016/j.immuni.2015.04.019.