Vaginotypes of the human vaginal microbiome

environmental microbiology
environmental microbiology reports

# Vaginotypes of the human vaginal microbiome.

| | |
|---|---|
| Journal: | *Environmental Microbiology and Environmental Microbiology Reports* |
| Manuscript ID | EMI-2021-0080.R1 |
| Journal: | Environmental Microbiology |
| Manuscript Type: | EMI - Research article |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | Mancabelli, Leonardo; University of Parma, Life Sciences<br>Tarracchini, Chiara; University of Parma, Department of Chemistry, Life Sciences and Environmental Sustainability<br>Milani, Christian<br>Lugli, Gabriele Andrea; University of Parma, Department of Chemistry, Life Sciences and Environmental Sustainability<br>Fontana, Federico; Università degli Studi di Parma, Department of Chemistry, Life Sciences and Environmental Sustainability<br>Turroni, Francesca<br>van Sinderen, Douwe; National University of Ireland , Department of Microbiology<br>Ventura, Marco; University of Parma, Life Sciences |
| Keywords: | metagenomics, shotgun metagenomics, community state types |
| | |

# Vaginotypes of the human vaginal microbiome.

Leonardo Mancabelli[1], Chiara Tarracchini[1], Christian Milani[1,2], Gabriele Andrea Lugli[1], Federico Fontana[1], Francesca Turroni[1,2], Douwe van Sinderen[3] and Marco Ventura[1,2]

Laboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy[1]; Interdepartmental Research Centre "Microbiome Research Hub", University of Parma, 43124 Parma, Italy[2]; APC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland[3]

Running title: Meta-analysis of healthy vaginal microbiota

Correspondence. Mailing address for Marco Ventura Department of Life Sciences, University of Parma, Parco Area delle Scienze 11a, 43124 Parma, Italy. Phone: ++39-521-905666. Fax: ++39-521-905604. E-mail: marco.ventura@unipr.it

**Summary**

The human vaginal environment harbors a community of bacteria that plays an important role in maintaining vaginal health and in protecting this environment from various urogenital infections. This bacterial population, also known as vaginal microbiota, has been demonstrated to be dominated by members of the *Lactobacillus* genus. Several studies employing 16S rRNA gene-based amplicon sequencing have classified the vaginal microbiota into five distinct Community State Types (CSTs) or vaginotypes. To deepen our understanding of the vaginal microbiota we performed an in-depth meta-analysis of 1312 publicly available data sets concerning healthy vaginal microbiome information obtained by metagenomics sequencing. The analysis confirmed the predominance of taxa belonging to the *Lactobacillus* genus, followed by members of the genera *Gardnerella*, *Vibrio* and *Atopobium*. Moreover, the statistical robustness offered by this meta-analysis allowed us to disentangle the species-level composition of dominant and accessory taxa constituting each vaginotype and to revisit and refine the previously proposed CST classification. In addition, a functional characterization of the metagenomic datasets revealed particular genetic features associated with each assigned vaginotype.

## Introduction

The human body harbors thousands of microorganisms living in a mutualistic relationship with their host. The presence or absence of particular microbial species is determined by environmental conditions and host factors and thus will vary from site to site (Costello et al., 2009). In this context, the human vagina and its resident bacterial communities represent an example of a finely balanced mutual association (Ma et al., 2012). Remarkably, the human vaginal microbiota appears to play an important role in preventing several urogenital diseases, such as urinary tract infections, bacterial vaginosis (BV), yeast infections, human papillomavirus (HPV) and other sexually transmitted infections (STIs) (Taha et al., 1998; Donders et al., 2000; Wiesenfeld et al., 2003; Lai et al., 2009; De Seta et al., 2019). From a taxonomic perspective, this peculiar bacterial community is characterized by a relatively low microbial diversity and a predominance of *Lactobacillus* species, which appear to prevent colonization by (opportunistic) pathogens (Borges et al., 2014; Aldunate et al., 2015). *Lactobacillus* species are known to elicit their protective role by lowering the environmental pH through lactic acid production, by competitive exclusion, or through the production of particular bacteriostatic and/or bactericidal compounds (Boskey et al., 2001; Voravuthikunchai et al., 2006). In 2011, a cross-sectional study encompassing 394 healthy women of reproductive age allowed the classification of the human vaginal microbiota into five Community State Types (CSTs), also referred to vaginotypes (Ravel et al., 2011). In detail, CSTs I, II, III and V are characterized by the predominance of *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus iners*, and *Lactobacillus jensenii*, respectively, while CST IV is not associated with a particular dominant species. Interestingly, CST IV was initially further classified into type CST IV-A and IV-B, which are characterized by moderate proportions of *Lactobacillus* spp. concurrent with low abundance of several species of strictly anaerobic bacteria or higher relative abundance of the genus *Atopobium*, respectively (Gajer et al., 2012). Subsequently, Albert et al. (Albert et al., 2015) identified two further subgroups of CST IV, i.e. CST IV-C and CST IV-D, which are delineated by a predominance of species belonging to the *Gardnerella* genus or by a heterogeneous group of bacteria, such as

*Bifidobacterium*, *Lactobacillus, Alloscardovia, Gardnerella* and *Atopobium*, respectively. Furthermore, recent studies revealed that pregnancy induces reduced biodiversity and increases taxonomic composition stability of the vaginal microbiota (Freitas et al., 2017; Gupta et al., 2020).

The majority of studies aimed at dissecting the taxonomic composition of the human vaginal microbiota have been performed employing 16S rRNA gene-based amplicon sequencing (Ravel et al., 2011; Gajer et al., 2012; Virtanen et al., 2017; Cobo et al., 2019; Vargas-Robles et al., 2020). Such studies lead to the identification of correlations between taxonomic profiles and health condition, such as the beneficial effects of *L. crispatus* or the role of *G. vaginalis* in bacterial vaginosis (Pleckaityte et al., 2012; Chen et al., 2018; Cobo et al., 2019; Pramanick et al., 2019; Zwittink et al., 2020). Despite the widespread use of the 16S rRNA gene-based amplicon sequencing approach, this analysis is prone to technical biases, such as the efficiency of the DNA extraction method and performance of the primer pair used for PCR amplification, that may prevent accurate prediction of bacterial taxonomic ranks present in a sample, especially when aiming at species-level resolution (Yarza et al., 2014; Hillmann et al., 2018). Furthermore, 16S rRNA gene amplicon sequencing is unable to provide a functional overview of the genetic potential encoded by a microbial community. In this context, whole-metagenome shotgun (WMS) sequencing is now rapidly replacing 16S rRNA gene microbial profiling as the gold standard for the investigation of complex microbial communities thanks to a reduction in sequencing costs accompanied by development of associated bioinformatic software for data analysis of the generated sequence data sets. Compared to 16S rRNA gene-based microbial profiling, WMS represents a major step forward in terms of taxonomic profiling accuracy and functional investigation of the microbiome, while at the same time reducing the risk of technical biases (Jovel et al., 2016; Hillmann et al., 2018).

In order to provide a complete overview of the taxonomic composition of the human vaginal microbiota down to species level and to gain insights into the genetic potential harbored by the vaginal microbiome, we performed an in depth meta-analysis of seven publicly available shotgun metagenomics datasets corresponding to 1312 vaginal samples from healthy women.

## Results and discussion

**Selection of publicly available datasets.** An extensive literature search was performed in order to retrieve all publicly available data pertaining to studies involving shotgun metagenomics of a sufficient number of vaginal samples to reach robust statistical power, in accordance with previous studies focusing on human-associated microbiota. In detail, the literature survey allowed us to retrieve vaginal microbiota data from seven publicly available datasets (Lloyd-Price et al., 2017; Goltsman et al., 2018; Oliver et al., 2020; Yang et al., 2020) covering three different countries (Table 1 and Supplemental Table S1). Overall, the multi-population cohort meta-analysis performed in this study encompasses datasets corresponding to a total of 1312 vaginal samples from healthy adult women (average age 31 ± 6), including 333 pregnant individuals (Table 1 and Supplemental Table S2).

**Meta-analysis of healthy non-pregnant women microbiota.** A total of 979 publicly available samples from four cohorts encompassing vaginal samples of healthy non-pregnant women were retrieved (Lloyd-Price et al., 2017; Yang et al., 2020) (Table 1). Quality filtering resulted in a total of 2,795,011 Mbp with an average of 2,130 Mbp per sample (Supplemental Table S1). In accordance with previous studies (Duvallet et al., 2017; Bisanz et al., 2019; Greathouse et al., 2019; Mancabelli et al., 2020), we employed this large number of data sets and the possibilities offered by the shotgun metagenomic approach to accurately profile bacteria at species level through re-analysis with the METAnnotatorX platform (Milani et al., 2018).

The shotgun metagenomic meta-analysis allowed an in depth investigation into the biodiversity of the collected healthy non-pregnant vaginal samples. In detail, analysis of the species richness revealed an average number of species of 24 ± 18 (Figure S1a), confirming the previously proposed notion that the healthy vaginal microbiota is characterized by a rather low microbial biodiversity when compared to other human body sites (Wessels et al., 2017).

Focusing on the bacterial composition, the vaginal samples showed an overall predominance of taxa belonging to the *Lactobacillus* genus (average abundance of 68.35 % ± 38.09%), followed by

members of the genera *Gardnerella*, *Vibrio* and *Atopobium* (average abundance of 7.42 % ± 17.53 %, 3.10 % ± 10.51 % and 2.99 % ± 14.43 %, respectively). Moreover, only the species *L. crispatus* (average abundance of 41.52 % ± 42.63 %), *L. jensenii* (average abundance of 4.09 % ± 11.58 %), *L. iners* (average abundance of 13.87 % ± 27.21 %) and *L. gasseri* (average abundance of 4.73 % ± 15.80 %) revealed a prevalence of > 40%, confirming the predominance of *Lactobacillus* species in the vaginal environment (Mancabelli et al., 2020).

**Prediction of vaginotypes.** The collected data sets were used to confirm the existence of vaginal Community State Types (CSTs), i.e. common taxonomic profiles patterns also referred to as vaginotypes, and identify possible novel CSTs or sub-CSTs. Screening for vaginotypes was performed by cluster analysis through Hierarchical CLustering (HCL) involving the microbial taxonomic profiles at species level of healthy non-pregnant women (Figure 1a) and were confirmed by 3D Bray Curtis PCoA (Figure 1b). The identified clusters had to be represented by at least 10 samples to be defined as putative CSTs, as previously highlighted (Mancabelli et al., 2020) (Figure 1c and Supplemental Table S2). Clustering occurrences observed in the PCoA representation were statistically validated through PERMANOVA (p-value < 0.05, $R^2$ = 0.22). Moreover, a PCoA analysis based on the host geographical origin (Figure S1b) revealed that there was no correlation between host geographical origin and the vaginal microbiota profile (PERMANOVA p-value > 0.05, R2 = 0.02). The metagenomic analysis of healthy non-pregnant vaginal samples allowed us to confirm four CSTs, i.e. I, II, III and V, previously identify by Ravel et al. (Ravel et al., 2011) (Figure 1). In detail, our analysis revealed that 49.03 % of the samples are classified as CST I, followed by CST III, CST II and CST V, which exhibited a prevalence across the analysed samples of 16.65 %, 7.05 % and 2.86 %, respectively (Figure 1c). Interestingly, each proposed vaginal CST was typified by the presence of a dominant *Lactobacillus* species with an average abundance >20 % and a prevalence >90 % (Figure 1c and Supplemental Table S2) as previously reported by Ravel et al. (Ravel et al., 2011). In detail, CSTs, i.e. I, II, III and V showed dominance of *L. crispatus*, *L. gasseri*, *L. iners* and

*L. jensenii*, respectively. Furthermore, 17.47 % of the samples seems to correspond to CST IV, which has been described as the most heterogeneous and controversial vaginotype. In fact, CST IV was initially defined to be characterized to have no dominant species (Ravel et al., 2011), although subgroups dominated by non-*Lactobacillus* species, such as *Gardnerella* or *Atopobium*, were subsequently included (Gajer et al., 2012; Albert et al., 2015). In this context, based on the outcome from the current meta-analysis, 13.38 % of the samples appeared to be characterized by the dominant presence of the *Gardernella* genus, i.e. *G. vaginalis* and Unclassified *Gardenella* species, and is referred here as CST-*G*, which encompasses CST IV-B, -C and -D as previously defined in literature (Albert et al., 2015; Freitas et al., 2017). Moreover, the remaining 4.09 % of the total pool of samples seems to represent the previously reported CST IV-A (Gajer et al., 2012), characterized by a heterogeneous group of bacteria mainly represented by members of the genera *Bacteroides* and *Prevotella* (Figure 1c and Supplemental Table S2). Interestingly, our meta-analysis revealed the presence of two additional and possibly novel CSTs, which are characterized by dominance of *Kocuria rosea/Klebsiella pneumoniae* (named here CST-*KK*) and *Vibrio harveyi* (designated CST-*Vh*) with a prevalence among all vaginal swab samples of 3.06 % and 3.88 %, respectively (Figure 1c). Although the collected vaginal samples were cataloged as being derived from healthy subjects, these two putative novel CSTs appear to be characterized by presence of opportunistic pathogens of the urinary tract, i.e. *Kocuria*, *Klebsiella*, and *Vibrio* (Kandi et al., 2016; Cristea et al., 2017; Defoirdt et al., 2017), and as such may represent biomarkers for a shift from a healthy to a diseased status of the vaginal environment. Unfortunately, the absence of longitudinal data for these samples prevents validation of this hypothesis.

The taxonomic profiles obtained in this study were used to evaluate the most prevalent taxa which typified each assigned vaginotypes, i.e. those species that have been detected in >80 % of samples classified as a CST with a relative abundance of >0.05 % (Figure 2). As shown in Figure 2, we observed that no single species appears to be ubiquitous between CSTs, thus indicating the absence of a species-level core vaginal microbiota. Moreover, the presence of a total of 34 taxa seems to be

linked to specific CSTs, which indicates that the low biodiversity charactering the vaginal environment is accompanied by taxonomic variability that is strictly correlated with the established vaginotypes (Figure 2).

**Identification of sub-CSTs.** The high number of vaginal samples collected and our in depth shotgun metagenomic analysis allowed to highlight the presence of putative sub-CSTs. The sub-CSTs of healthy non-pregnant vaginal samples were identified by HCL analysis of taxonomic profiles at species level (Figure 1a) and 3D Bray Curtis PCoA (Figure 1b), and had to be supported by at least 10 samples. In detail, the HCL clustering analysis showed clear subgroups for CST I, II, III and CST-*G* (Figure 1c), confirmed by PERMANOVA of the PCoA representation (p-value < 0.05, $R^2 = 0.85$). Interestingly, CST I was shown to include a main subgroup, designated here as CST Ia, which is characterized by the predominance of *L. crispatus* (average abundance of 89.66 % ± 10.99 %) and three additional subgroups, i.e. CST Ib, CST Ic and CST Id, characterized by a high abundance of *L. crispatus* (>53.72 % ± 16.40 % in all cases) accompanied by *Klebsiella quasipneumoniae* (average abundance of 18.90 % ± 5.51 %) or *L. iners* (average abundance of 22.77 % ± 8.72 %) or *L. jensenii* (average abundance of 18.60 % ± 2.56 %), respectively (Figure 1c).

Furthermore, HCL analysis of CST II revealed two subclusters, i.e. CST IIa, which is characterized by dominance of *L. gasseri* (average abundance of 68.41% ± 27.46 %) and subgroup CST IIb in which the high abundance of *L. gasseri* (average abundance of 51.81 % ± 10.36 %) is accompanied by *Bifidobacterium scardovii* (average abundance of 28.35 % ± 2.72 %) (Figure 1c). CST III can be subdivided in subgroup CST IIIa, dominated by *L. iners* (average abundance of 72.44 % ± 17.99 %), and CST IIIb, which mainly constitutes *L. iners* (average abundance of 56.12 % ± 9.08 %) and *L. jensenii* (average abundance of 32.72 % ± 10.39 %) (Figure 1c).

Moreover, CST-*G* showed the presence of subgroups, i.e. CST-*G*a mainly represented by species of *Gardnerella* (average abundance of 54.51 % ± 14.72 %) and CST-*G*b characterized by species

belonging to *Gardnerella* genus and *Atopobium vaginae* (average abundance of 40.75 % ± 10.49 % and 36.77 % ± 11.45 %, respectively) (Figure 1c).

**Covariances between vaginotypes members and the role of dominant species in defining the taxonomic composition of the vaginal microbiota.** In order to identify if the main dominant taxa that characterize each vaginotype are implied in defining the overall taxonomic composition of the vaginal microbiota, we performed a covariance analysis through Spearman's rho coefficient (Supplemental Table S3). For this purpose, we correlated the relative abundance observed for all taxa which exhibit a total average abundance greater than 0.05 % and which are present in at least one sample with an abundance greater than 5 % (Supplemental Table S3). Interestingly, this analysis showed that *L. crispatus* and *L. jensenii* elicit the highest ability to negatively impact on the presence of other bacteria, as highlighted by negative correlations (p-value < 0.05) with more than 55 % of the taxa included in the analysis (Supplemental Table S3). Furthermore, *L. crispatus* revealed the lowest number of positive correlations, i.e. 9.38 %, compared to the other dominant taxa characterizing vaginotypes (Supplemental Table S3). These results support the notion that the CST I plays a key role in countering colonization by other bacteria and in maintaining low biodiversity in the vaginal environment, a condition considered to be associated with vaginal health (Ravel et al., 2011; Human Microbiome Project, 2012; Vargas-Robles et al., 2020). In contrast, *L. iners* and *G. vaginalis*, representative species of CST III and CST-*G* respectively, positively correlate with each other and appear to promote the presence of some pathogenic bacteria of the urinary tract, such as *Atopobium vaginae* (Burton et al., 2004; Burton et al., 2005), *Prevotella bivia* (Gilbert et al., 2019) and unknown species belonging to the genus *Megasphaera* (Fredricks et al., 2009), as indicated by positive correlations (Supplemental Table S3). Therefore, despite its rather high prevalence among women (16.65 %), CST III seems to facilitate vaginal infections (Jakobsson and Forsum, 2007; Petrova et al., 2017; Zheng et al., 2019). Intriguingly, *K. pneumoniae*/*K. rosea* and *V. harveyi*, species characteristic of putative CST-*KK* and CST-*Vh* respectively, positively correlate with species belonging to the

Proteobacteria phylum, such as *Raoultella planticola* and *Haemophilus parainfluenzae*, and negatively correlate with *Lactobacillus* species, as well as *Lactobacillus jensenii* (Supplemental Table S3), highlighting the ability of these taxa to alter the homeostasis of the vaginal microbiota.

**Meta-analysis of healthy pregnant women microbiota.** During pregnancy, hormonal changes leads to immune modulation and physico-chemical changes in the mucosa of the genital tract (Gupta et al., 2020). These changes can affect the composition and the function of the vaginal microbiota, making it distinctive from non-pregnant women. In order to identify possible variations between healthy non-pregnant and pregnant women, we analyzed the metagenomics data of a total of 333 vaginal samples obtained from publicly available studies and encompassing pregnant women. Quality filtering of shotgun metagenomic data resulted in a total of 36,574 Mbp with an average of ~109 Mbp per sample (Supplemental Table S2). Analysis of the species richness revealed a statistically significant simplification of the vaginal microbiota of pregnant women (average species richness of 17 ± 14) when compared to non-pregnant samples (average species richness of 24 ± 18) (t-test $p$-value < 0.01) (Figure S1a), confirming the shift of the vaginal microbiota towards to low biodiversity during pregnancy (Romero et al., 2014; Freitas et al., 2017; Gupta et al., 2020). Moreover, HCL and PCoA analysis allowed the identification of associations between vaginotypes and pregnancy (Figure 3). In detail, the meta-analysis allowed us to observe that the CST I and CST III each had a prevalence of > 30 % amongst pregnant women, thus being the most common CSTs, while CST II, CST-*G*, CST IV and CSTV had a prevalence of < 13 % (Figure 3C). Nevertheless, the prevalence of CST I is lower compared to that of healthy non-pregnant women (36.94 % in pregnant vs 49.03 % in non-pregnant women). Moreover, pregnancy seems to be correlated with a higher prevalence of CST III and CST V (Figure 3c), thereby indicating an overall destabilization of vaginal microbiota homeostasis. This result corroborates the suggestion that bacterial communities in pregnancy do shift from one vaginotype dominated by *Lactobacillus* spp. to another CST dominated by *Lactobacillus* spp., but rarely to CST-*G* or CST-IV (Romero et al., 2014). Notably, the putative CST-*KK* and CST-*Vh* seem

to be absent in pregnant women probably due to the simplification of the vaginal microbiota that characterize pregnant women, as previously reported (Romero et al., 2014; Freitas et al., 2017; Gupta et al., 2020). In this context, the lower number of samples analyzed compared to non-pregnant samples may prevent identification of vaginotypes at low prevalence such as CST-*KK* and CST-*Vh*, and therefore these findings require further validation.

In order to identify possible specific sub-CSTs in vaginal samples from pregnant women, cluster analyses were performed. The HCL and PCoA analyses allowed the identification of subgroups CST Ia (prevalence 33.63 %) and CST Ic (prevalence 3.30 %), previously identified in the non-pregnant samples. Moreover, CST V could be subdivided in the subgroup CST Va, dominated by *L. jensenii* (average abundance of 57.36% ± 25.52 %), and a new subgroup CST Vb characterized by high abundance of *L. jensenii* (average abundance of 66.78 % ± 13.36 %) and *L. iners* (average abundance of 25.88 % ± 13.33%) (Figure 3C). Notably, CST Vb was not identified by analysis of non-pregnant women due to insufficient (<10 women) prevalence. In fact, only the non-pregnant samples SRR513792 can be assigned to this sub-CST. The data suggest that future integration of this comparative analysis with novel samples may result in the identification of additional sub-CSTs.

In order to identify possible differences in the taxonomic profiling of the vaginal microbiota between non-pregnant and pregnant women, we performed a *t*-test between the average taxonomic composition observed for the two groups. Focusing on bacterial taxa showing an average relative abundance of > 0.1% in at least one of the two groups. The analysis allowed the identification of statistically significant differences for 46 microbial taxa (Supplemental Table S3). Remarkably, samples collected from pregnant women showed higher relative abundance of *L. iners* (+211 % compared to samples from non-pregnant women) and *Gardnerella* species (+60 % compared to the non-pregnant group) (p-value < 0.05) (Supplemental Table S3). Notably, these changes are indicative of disruption of vaginal microbiota homeostasis. This idea is reinforced by average relative abundance increases in opportunistic pathogens such as *Ralstonia pickettii* and *Ureaplasma parvum* (+14185 % and +1242 %, when compared to samples from the non-pregnant group) (Shurin et al.,

1975; Ryan et al., 2006; Normann et al., 2009; Ryan et al., 2011; Ryan and Adley, 2014; Combaz-Sohnchen and Kuhn, 2017; de Goffau et al., 2019).

**Functional capabilities of vaginotypes.** The different taxonomic profiles associated with each predicted vaginotypes are assumed to correspond to specific microbiomes, each with their particular genetic repertoires. In order to explore the genetic features characterizing each of the identified CST, a total of 133 shotgun metagenomics samples were classified. We focused on a comparison between various identified CSTs and CST I, i.e. the CST dominated by *L. crispatus*, due to its generally accepted positive role in supporting vaginal health (Nardini et al., 2016; Wang et al., 2017; Chee et al., 2020).

Screening for genes related to bacteriocins showed that, on average, CST I encodes a 10-fold higher abundance of bacteriocins when compared to all other CSTs (average of 0.03 % and 0.003 % of the whole metagenomic dataset, respectively) (ANOVA p-value < 0.01) (Figure 4a). In detail, this difference is caused by a higher abundance of Class III bacteriocins in CST I when compared to other CSTs (ANOVA p-value < 0.01) (Figure 4a). Intriguingly, Class III bacteriocins represent peptides that cause bacterial cells death by cell wall degradation (Class IIIa) and peptides that dissipate the cytoplasmic membrane potential and cause cell death without cell lysis. The higher abundance of bacteriocins in CST I may explain the low biodiversity associated with this CST and the ability of *L. crispatus* to dominate the vaginal microbiota while inhibiting colonization of opportunistic pathogens (Nardini et al., 2016; Wang et al., 2017; Atassi et al., 2019; Chee et al., 2020).

Moreover, we performed a screening of the genetic repertoire involved in catabolic pathways based on the MetaCyc database (Caspi et al., 2018) (Figure 4b). In dept evaluation of changes in the relative abundance of each biosynthetic and degradative pathway profiles allowed the identification of 44 pathways whose abundance in CST I is higher than all 6 other CSTs included in the analysis (Figure 4b). Among these, N-acetylneuraminate and N-acetylmannosamine degradation II, mannitol degradation I and D-arabitol revealed a statistical significance compared to all other CSTs (ANOVA

304 $p$-value < 0.05, Tukey post-hoc test $p$-value < 0.05) (Figure 4b). Notably, vaginal mucus secretions
305 are rich in sialic acids and the higher abundance of genes involved in the degradation of N-
306 acetylneuraminate, the most common form of sialic acid, indicates a greater adaptability of the CST
307 I-associated microbes to the vaginal environment (Haines-Menges et al., 2015). Furthermore, the
308 competition in the degradation of sialic acids could disfavor colonization by BV-associated bacterium
309 *Gardnerella vaginalis* (Lewis et al., 2013), promoting the stability of the vaginal environment (Lewis
310 et al., 2013). Similarly, mannitol has been suggested to support *L. crispatus* in adhering to the
311 epithelial layer and inhibit the colonization of other microbes, potentially by altering the mucin
312 structure (Wu et al., 2015), further highlighting the possible beneficial role of CST I.
313

## CONCLUSIONS

The human vaginal environment is characterized by bacteria, i.e. vaginal microbiota, inhabiting the human vaginal tract and presumed to play a key role in supporting a healthy host status. In this study, we performed an in-depth meta-analysis based on a total of seven publicly available shotgun metagenomics datasets of 1312 vaginal samples from healthy women. The performed meta-analysis confirmed the existence of vaginotypes defined as Community State Types (CSTs), i.e. CST I, CST II, CST III and CST V, and allowed a detailed dissection of the controversial CST IV. Based on our findings, we propose the new vaginotype CST-*G* which appears to be typified by the dominant presence of members of the *Gardnerella* genus. Furthermore, covariance analyses between vaginotypes and the taxonomic composition of the vaginal microbiota supported the positive role of the CST I in maintaining vaginal health, preventing the colonization of other bacteria, in particular vaginal pathogens, and preserving low biodiversity. In contrast, CST III and CST-*G* seem to promote the establishment of putative pathogenic bacteria in the urogenital tract. Furthermore, analysis of the vaginal microbiota of pregnant women revealed a significant reduction in species richness when compared to non-pregnant samples and showed high prevalence of CST III and CST V, thus suggesting an overall destabilization of vaginal microbiota homeostasis. In addition, CST I was predicted to encode a higher abundance of Class III bacteriocins when compared to all other CSTs and appears to encompass a genetic repertoire that plays a beneficial role by promoting the stability of the vaginal environment.

## Materials and Methods

**Database selection.** In this meta-analysis, we retrieved seven publicly available data sets from studies involving the taxonomic determination of the vaginal microbiota. In order to reduce the variability in the input data, we selected shotgun metagenomics datasets obtained by an Illumina sequencing platform. In detail, we selected shotgun metagenomics data sets from 1312 vaginal samples of women covering four geographic regions, ensuring that vaginal samples corresponded to healthy subjects only, while also including samples from 333 pregnant subjects (Table 1 and Supplemental Tables S1 and S2).

**Taxonomic classification of sequence reads.** Taxonomic profiling of sequenced reads was performed employing the METAnnotatorX bioinformatics platform (Milani et al., 2018). Taxonomic classification of up to 100,000 reads was achieved by means of megablast (Chen et al., 2015) employing a manually curated and pre-processed database of genomes retrieved from the National Center for Biotechnology Information (NCBI).

**Functional prediction.** Functional profiling of sequenced reads was performed with the METAnnotatorX bioinformatics platform (Milani et al., 2018). Functional classification of reads was performed to reveal metabolic pathways based on the MetaCyc database (Caspi et al., 2016). Identification and functional assignment of genes related to bacteriocin biosynthesis and immunity was performed using the BAGEL4 tool (de Jong et al., 2006)

**Vaginal Community State Type (VCST) prediction.** The hierarchical clustering (HCL) of samples was obtained using bacterial composition at species level and was calculated through TMeV 4.8.1 software using Pearson correlation as a distance metric based on information at species level. The data obtained was represented by a cladogram.

359 **Statistical analysis**. ORIGIN 2021 (https://www.originlab.com/2021) and SPSS software
360 (www.ibm.com/software/it/analytics/spss/) were used to compute statistical analyses.
361 PERMANOVA analyses were performed using 1,000 permutations to estimate p-values for
362 differences among populations in PCoA analyses. Furthermore, differential abundance of bacterial
363 genera was tested by t-test analysis. Moreover, we also calculated ANOVA and the post hoc analysis
364 Tukey's HSD (Honestly Significant Difference) test for multiple comparison. Covariance analyses
365 were calculated through Spearman's rho coefficient correlation.
366

**Acknowledgements**

Part of this research is conducted using the High Performance Computing (HPC) facility of the University of Parma. We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. DvS is a member of APC Microbiome Ireland funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (Grant no's. SFI/12/RC/2273-P1 and SFI/12/RC/2273-P2).

**Author Contributions**

LM processed the metagenomic data, conducted the analyses and wrote the manuscript. CT contributed to the metagenomic analyses. CM participated in the design of the study and contributed to the manuscript preparation. GAL contributed to the metagenomic analyses. FF contributed to the statistical analyses. FT participated in the design of the study. DvS participated and supervised the study. MV conceived the study, participated in its design and coordination and contributed to the manuscript preparation. All authors have read and approved the final manuscript.

Declaration of interest: none.

## References


Albert, A.Y., Chaban, B., Wagner, E.C., Schellenberg, J.J., Links, M.G., van Schalkwyk, J. et al. (2015) A Study of the Vaginal Microbiome in Healthy Canadian Women Utilizing cpn60-Based Molecular Profiling Reveals Distinct Gardnerella Subgroup Community State Types. *PLoS One* **10**: e0135620.

Aldunate, M., Srbinovski, D., Hearps, A.C., Latham, C.F., Ramsland, P.A., Gugasyan, R. et al. (2015) Antimicrobial and immune modulatory effects of lactic acid and short chain fatty acids produced by vaginal microbiota associated with eubiosis and bacterial vaginosis. *Front Physiol* **6**: 164.

Atassi, F., Pho Viet Ahn, D.L., and Lievin-Le Moal, V. (2019) Diverse Expression of Antimicrobial Activities Against Bacterial Vaginosis and Urinary Tract Infection Pathogens by Cervicovaginal Microbiota Strains of Lactobacillus gasseri and Lactobacillus crispatus. *Front Microbiol* **10**: 2900.

Bisanz, J.E., Upadhyay, V., Turnbaugh, J.A., Ly, K., and Turnbaugh, P.J. (2019) Meta-Analysis Reveals Reproducible Gut Microbiome Alterations in Response to a High-Fat Diet. *Cell Host Microbe* **26**: 265-272 e264.

Borges, S., Silva, J., and Teixeira, P. (2014) The role of lactobacilli and probiotics in maintaining vaginal health. *Arch Gynecol Obstet* **289**: 479-489.

Boskey, E.R., Cone, R.A., Whaley, K.J., and Moench, T.R. (2001) Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. *Hum Reprod* **16**: 1809-1813.

Burton, J.P., Devillard, E., Cadieux, P.A., Hammond, J.A., and Reid, G. (2004) Detection of Atopobium vaginae in postmenopausal women by cultivation-independent methods warrants further investigation. *J Clin Microbiol* **42**: 1829-1831.

Burton, J.P., Chilcott, C.N., Al-Qumber, M., Brooks, H.J., Wilson, D., Tagg, J.R., and Devenish, C. (2005) A preliminary survey of Atopobium vaginae in women attending the Dunedin gynaecology out-patients clinic: is the contribution of the hard-to-culture microbiota overlooked in gynaecological disorders? *Aust N Z J Obstet Gynaecol* **45**: 450-452.

Caspi, R., Billington, R., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M. et al. (2018) The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res* **46**: D633-D639.

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M. et al. (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **44**: D471-480.

Chee, W.J.Y., Chew, S.Y., and Than, L.T.L. (2020) Vaginal microbiota and the potential of Lactobacillus derivatives in maintaining vaginal health. *Microb Cell Fact* **19**: 203.

Chen, H.M., Chang, T.H., Lin, F.M., Liang, C., Chiu, C.M., Yang, T.L. et al. (2018) Vaginal microbiome variances in sample groups categorized by clinical criteria of bacterial vaginosis. *BMC Genomics* **19**: 876.

Chen, Y., Ye, W., Zhang, Y., and Xu, Y. (2015) High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res* **43**: 7762-7768.

Cobo, T., Vergara, A., Collado, M.C., Casals-Pascual, C., Herreros, E., Bosch, J. et al. (2019) Characterization of vaginal microbiota in women with preterm labor with intra-amniotic inflammation. *Sci Rep* **9**: 18963.

Combaz-Sohnchen, N., and Kuhn, A. (2017) A Systematic Review of Mycoplasma and Ureaplasma in Urogynaecology. *Geburtshilfe Frauenheilkd* **77**: 1299-1303.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009) Bacterial community variation in human body habitats across space and time. *Science* **326**: 1694-1697.

Cristea, O.M., Avramescu, C.S., Balasoiu, M., Popescu, F.D., Popescu, F., and Amzoiu, M.O. (2017) Urinary tract infection with Klebsiella pneumoniae in Patients with Chronic Kidney Disease. *Curr Health Sci J* **43**: 137-148.

de Goffau, M.C., Lager, S., Sovio, U., Gaccioli, F., Cook, E., Peacock, S.J. et al. (2019) Human placenta has no microbiome but can contain potential pathogens. *Nature* **572**: 329-334.

de Jong, A., van Hijum, S.A., Bijlsma, J.J., Kok, J., and Kuipers, O.P. (2006) BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res* **34**: W273-279.
De Seta, F., Campisciano, G., Zanotta, N., Ricci, G., and Comar, M. (2019) The Vaginal Community State Types Microbiome-Immune Network as Key Factor for Bacterial Vaginosis and Aerobic Vaginitis. *Front Microbiol* **10**: 2451.
Defoirdt, T., Vlaeminck, S.E., Sun, X., Boon, N., and Clauwaert, P. (2017) Ureolytic Activity and Its Regulation in Vibrio campbellii and Vibrio harveyi in Relation to Nitrogen Recovery from Human Urine. *Environ Sci Technol* **51**: 13335-13343.
Donders, G.G., Bosmans, E., Dekeersmaecker, A., Vereecken, A., Van Bulck, B., and Spitz, B. (2000) Pathogenesis of abnormal vaginal bacterial flora. *Am J Obstet Gynecol* **182**: 872-878.
Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J. (2017) Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* **8**: 1784.
Fredricks, D.N., Fiedler, T.L., Thomas, K.K., Mitchell, C.M., and Marrazzo, J.M. (2009) Changes in vaginal bacterial concentrations with intravaginal metronidazole therapy for bacterial vaginosis as assessed by quantitative PCR. *J Clin Microbiol* **47**: 721-726.
Freitas, A.C., Chaban, B., Bocking, A., Rocco, M., Yang, S., Hill, J.E. et al. (2017) The vaginal microbiome of pregnant women is less rich and diverse, with lower prevalence of Mollicutes, compared to non-pregnant women. *Sci Rep* **7**: 9212.
Gajer, P., Brotman, R.M., Bai, G., Sakamoto, J., Schutte, U.M., Zhong, X. et al. (2012) Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* **4**: 132ra152.
Gilbert, N.M., Lewis, W.G., Li, G., Sojka, D.K., Lubin, J.B., and Lewis, A.L. (2019) Gardnerella vaginalis and Prevotella bivia Trigger Distinct and Overlapping Phenotypes in a Mouse Model of Bacterial Vaginosis. *J Infect Dis* **220**: 1099-1108.
Goltsman, D.S.A., Sun, C.L., Proctor, D.M., DiGiulio, D.B., Robaczewska, A., Thomas, B.C. et al. (2018) Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome. *Genome Res* **28**: 1467-1480.
Greathouse, K.L., White, J.R., Padgett, R.N., Perrotta, B.G., Jenkins, G.D., Chia, N., and Chen, J. (2019) Gut microbiome meta-analysis reveals dysbiosis is independent of body mass index in predicting risk of obesity-associated CRC. *BMJ Open Gastroenterol* **6**: e000247.
Gupta, P., Singh, M.P., and Goyal, K. (2020) Diversity of Vaginal Microbiome in Pregnancy: Deciphering the Obscurity. *Front Public Health* **8**: 326.
Haines-Menges, B.L., Whitaker, W.B., Lubin, J.B., and Boyd, E.F. (2015) Host Sialic Acids: A Delicacy for the Pathogen with Discerning Taste. *Microbiol Spectr* **3**.
Hillmann, B., Al-Ghalith, G.A., Shields-Cutler, R.R., Zhu, Q., Gohl, D.M., Beckman, K.B. et al. (2018) Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems* **3**.
Human Microbiome Project, C. (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207-214.
Jakobsson, T., and Forsum, U. (2007) Lactobacillus iners: a marker of changes in the vaginal flora? *J Clin Microbiol* **45**: 3145.
Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T. et al. (2016) Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Front Microbiol* **7**: 459.
Kandi, V., Palange, P., Vaish, R., Bhatti, A.B., Kale, V., Kandi, M.R., and Bhoomagiri, M.R. (2016) Emerging Bacterial Infection: Identification and Clinical Significance of Kocuria Species. *Cureus* **8**: e731.
Lai, S.K., Hida, K., Shukair, S., Wang, Y.Y., Figueiredo, A., Cone, R. et al. (2009) Human immunodeficiency virus type 1 is trapped by acidic but not by neutralized human cervicovaginal mucus. *J Virol* **83**: 11196-11200.
Lewis, W.G., Robinson, L.S., Gilbert, N.M., Perry, J.C., and Lewis, A.L. (2013) Degradation, foraging, and depletion of mucus sialoglycans by the vagina-adapted Actinobacterium Gardnerella vaginalis. *J Biol Chem* **288**: 12067-12079.

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A.B. et al. (2017) Erratum: Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **551**: 256.
Ma, B., Forney, L.J., and Ravel, J. (2012) Vaginal microbiome: rethinking health and disease. *Annu Rev Microbiol* **66**: 371-389.
Mancabelli, L., Tarracchini, C., Milani, C., Lugli, G.A., Fontana, F., Turroni, F. et al. (2020) Multi-population cohort meta-analysis of human intestinal microbiota in early life reveals the existence of infant community state types (ICSTs). *Comput Struct Biotechnol J* **18**: 2480-2493.
Milani, C., Casey, E., Lugli, G.A., Moore, R., Kaczorowska, J., Feehily, C. et al. (2018) Tracing mother-infant transmission of bacteriophages by means of a novel analytical tool for shotgun metagenomic datasets: METAnnotatorX. *Microbiome* **6**: 145.
Nardini, P., Nahui Palomino, R.A., Parolin, C., Laghi, L., Foschi, C., Cevenini, R. et al. (2016) Lactobacillus crispatus inhibits the infectivity of Chlamydia trachomatis elementary bodies, in vitro study. *Sci Rep* **6**: 29024.
Normann, E., Lacaze-Masmonteil, T., Eaton, F., Schwendimann, L., Gressens, P., and Thebaud, B. (2009) A novel mouse model of Ureaplasma-induced perinatal inflammation: effects on lung and brain injury. *Pediatr Res* **65**: 430-436.
Oliver, A., LaMere, B., Weihe, C., Wandro, S., Lindsay, K.L., Wadhwa, P.D. et al. (2020) Cervicovaginal Microbiome Composition Is Associated with Metabolic Profiles in Healthy Pregnancy. *mBio* **11**.
Petrova, M.I., Reid, G., Vaneechoutte, M., and Lebeer, S. (2017) Lactobacillus iners: Friend or Foe? *Trends Microbiol* **25**: 182-191.
Pleckaityte, M., Zilnyte, M., and Zvirbliene, A. (2012) Insights into the CRISPR/Cas system of Gardnerella vaginalis. *BMC Microbiol* **12**: 301.
Pramanick, R., Mayadeo, N., Warke, H., Begum, S., Aich, P., and Aranha, C. (2019) Vaginal microbiota of asymptomatic bacterial vaginosis and vulvovaginal candidiasis: Are they different from normal microbiota? *Microb Pathog* **134**: 103599.
Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S., McCulle, S.L. et al. (2011) Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* **108 Suppl 1**: 4680-4687.
Romero, R., Hassan, S.S., Gajer, P., Tarca, A.L., Fadrosh, D.W., Nikita, L. et al. (2014) The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome* **2**: 4.
Ryan, M.P., and Adley, C.C. (2014) Ralstonia spp.: emerging global opportunistic pathogens. *Eur J Clin Microbiol Infect Dis* **33**: 291-304.
Ryan, M.P., Pembroke, J.T., and Adley, C.C. (2006) Ralstonia pickettii: a persistent gram-negative nosocomial infectious organism. *J Hosp Infect* **62**: 278-284.
Ryan, M.P., Pembroke, J.T., and Adley, C.C. (2011) Genotypic and phenotypic diversity of Ralstonia pickettii and Ralstonia insidiosa isolates from clinical and environmental sources including High-purity Water. Diversity in Ralstonia pickettii. *BMC Microbiol* **11**: 194.
Shurin, P.A., Alpert, S., Bernard Rosner, B.A., Driscoll, S.G., and Lee, Y.H. (1975) Chorioamnionitis and colonization of the newborn infant with genital mycoplasmas. *N Engl J Med* **293**: 5-8.
Taha, T.E., Hoover, D.R., Dallabetta, G.A., Kumwenda, N.I., Mtimavalye, L.A., Yang, L.P. et al. (1998) Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. *AIDS* **12**: 1699-1706.
Vargas-Robles, D., Morales, N., Rodriguez, I., Nieves, T., Godoy-Vitorino, F., Alcaraz, L.D. et al. (2020) Changes in the vaginal microbiota across a gradient of urbanization. *Sci Rep* **10**: 12487.
Virtanen, S., Kalliala, I., Nieminen, P., and Salonen, A. (2017) Comparative analysis of vaginal microbiota sampling using 16S rRNA gene analysis. *PLoS One* **12**: e0181477.
Voravuthikunchai, S.P., Bilasoi, S., and Supamala, O. (2006) Antagonistic activity against pathogenic bacteria by human vaginal lactobacilli. *Anaerobe* **12**: 221-226.

Wang, S., Wang, Q., Yang, E., Yan, L., Li, T., and Zhuang, H. (2017) Antimicrobial Compounds Produced by Vaginal Lactobacillus crispatus Are Able to Strongly Inhibit Candida albicans Growth, Hyphal Formation and Regulate Virulence-related Gene Expressions. *Front Microbiol* **8**: 564.

Wessels, J.M., Lajoie, J., Vitali, D., Omollo, K., Kimani, J., Oyugi, J. et al. (2017) Association of high-risk sexual behaviour with diversity of the vaginal microbiota and abundance of Lactobacillus. *PLoS One* **12**: e0187612.

Wiesenfeld, H.C., Hillier, S.L., Krohn, M.A., Landers, D.V., and Sweet, R.L. (2003) Bacterial vaginosis is a strong predictor of Neisseria gonorrhoeae and Chlamydia trachomatis infection. *Clin Infect Dis* **36**: 663-668.

Wu, N., Zhang, X., Li, F., Zhang, T., Gan, Y., and Li, J. (2015) Spray-dried powders enhance vaginal siRNA delivery by potentially modulating the mucus molecular sieve structure. *Int J Nanomedicine* **10**: 5383-5396.

Yang, Q., Wang, Y., Wei, X., Zhu, J., Wang, X., Xie, X., and Lu, W. (2020) The Alterations of Vaginal Microbiome in HPV16 Infection as Identified by Shotgun Metagenomic Sequencing. *Front Cell Infect Microbiol* **10**: 286.

Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F.O., Ludwig, W., Schleifer, K.H. et al. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* **12**: 635-645.

Zheng, N., Guo, R., Yao, Y., Jin, M., Cheng, Y., and Ling, Z. (2019) Lactobacillus iners Is Associated with Vaginal Dysbiosis in Healthy Pregnant Women: A Preliminary Study. *Biomed Res Int* **2019**: 6079734.

Zwittink, R.D., van den Munckhof, E.H.A., Leverstein-van Hall, M.A., Boers, K., Molijn, A., Knetsch, C.W., and Kuijper, E.J. (2020) The vaginal microbiota in the course of bacterial vaginosis treatment. *Eur J Clin Microbiol Infect Dis*.

559 **Table 1.** Metadata of samples included in the meta-analysis.

| Bioproject | PMID/DOI | Females status | Nation | Age | n° of samples |
|---|---|---|---|---|---|
| PRJEB38528 | - | non-pregnant | Sweden | 34 ± 6 | 74 |
| PRJNA275349 | 29022944 | non-pregnant | USA (HMP) | 26 ± 5 | 29 |
| PRJNA48479 | | non-pregnant | USA (HMP) | | 841 |
| PRJNA576566 | 32656096 | non-pregnant | China | 35 ± 6 | 35 |
| PRJNA288562 | 30232199 | pregnant | USA | 31 ± 6 | 101 |
| PRJNA612083 | 32843557 | pregnant | USA | 28 ± 5 | 35 |
| PRJNA639592 | doi: https://doi.org/10.1101/2020.06.26.173922 | pregnant | - | 29 ± 4 | 197 |

560

561

562 **Figure legends**

563 **Figure 1.** Identification of vaginotypes. Panel a shows a circular cladogram of the healthy vaginal
564 samples obtained by means of hierarchical clustering (HCL) analysis. The cladogram highlighted the
565 different CSTs identified by through HCL analysis.
566 Panel b reports the principal coordinate analysis (PCoA) of the healthy vaginal samples, subdivided
567 by vaginotypes.
568 Panel c displays the average abundance and prevalence of bacteria that correspond to an identified
569 CST and sub-CST.
570
571 **Figure 2.** Evaluation of the most prevalent taxa characterizing each predicted vaginotypes. In detail,
572 we selected those species that have been detected in >80 % of samples classified as a CST with a
573 relative abundance >0.05 %.
574
575 **Figure 3.** Identification of pregnancy vaginotypes. Panel a shows a circular cladogram of the healthy
576 vaginal samples of pregnancy women, obtained by means of hierarchical clustering (HCL) analysis.
577 The cladogram highlighted the different CSTs identified by through HCL analysis.
578 Panel b reports the principal coordinate analysis (PCoA) of the healthy vaginal samples of pregnancy
579 women, subdivided by vaginotypes.
580 Panel c displays the average abundance and prevalence of bacteria that correspond to an identified
581 CST and sub-CST.
582
583 **Figure 4.** Functional capabilities of vaginotypes. Panel a shows the abundance of bacteriocins class,
584 i.e. class I, II and III, in different CSTs.
585 Panel b reveals the relative abundance of each biosynthetic and degradative pathway of the 44
586 pathways whose abundance in CST I is higher than all 6 other CSTs included in the analysis.

Significant Tukey post-hoc analysis between CST I and the other CSTs are highlighted with a violet outline.

**Additional files**

**Figure S1.** Evaluation of the species richness and evaluation of possible correlation between vaginotypes and host geographical origin. Panel a reports the Whiskers plot representing the species richness identified from non-pregnant and pregnant women. The x axis represents the different groups, while the y axis indicates the number of species. The boxes are determined by the 25$^{th}$ and 75$^{th}$ percentiles. The whiskers are determined by standard deviation. The line in the boxes represented the average, while the circle represents the median.

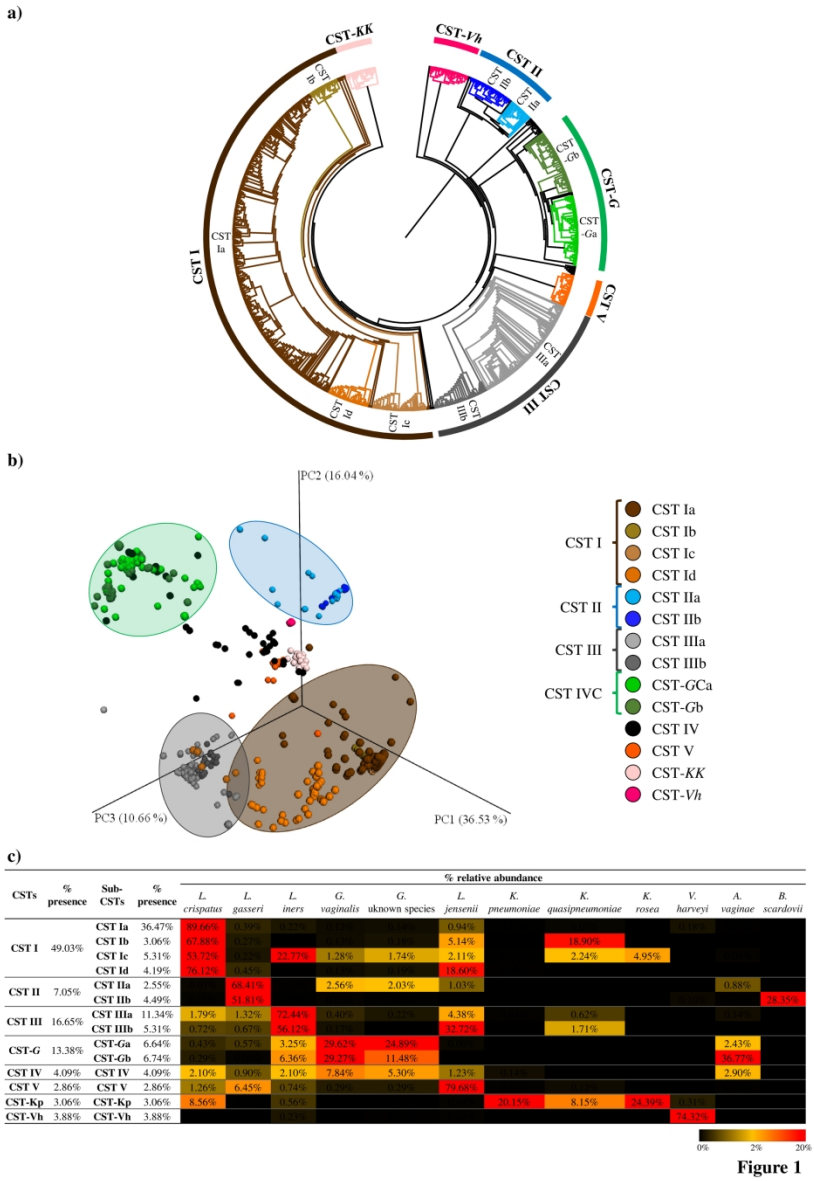
Panel b shows the PCoA depicting the beta diversity of samples in relation to geographical origin.

**Supplementary tables S1.** Studies included in this meta-analysis covering non-pregnant women.

**Supplementary tables S2.** Studies included in this meta-analysis covering pregnant women.

**Supplementary tables S3.** Covariance analysis based on the retrieved taxonomic profiles.

**Supplementary tables S4.** Species whose relative abundance differs in non-pregnant versus pregnat women with t-test p-value <0.05.




| CSTs | % presence | Sub-CSTs | % presence | *L. crispatus* | *L. gasseri* | *L. iners* | *G. vaginalis* | *G.* uknown species | *L. jensenii* | *K. pneumoniae* | *K. quasipneumoniae* | *K. rosea* | *V. harveyi* | *A. vaginae* | *B. scardovii* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CST I | 49.03% | CST Ia | 36.47% | 89.66% | 0.59% | 0.25% | 0.25% | 0.25% | 0.94% | | | | 0.25% | | |
| | | CST Ib | 3.06% | 67.88% | 0.28% | | | | 5.14% | | 18.90% | | | | |
| | | CST Ic | 5.31% | 53.72% | 0.28% | 22.77% | 1.28% | 1.74% | 2.11% | | 2.24% | 4.95% | | | |
| | | CST Id | 4.19% | 76.12% | 0.45% | | | 0.25% | 18.60% | | | | | | |
| CST II | 7.05% | CST IIa | 2.55% | | 68.41% | | 2.56% | 2.03% | 1.03% | | | | | 0.88% | |
| | | CST IIb | 4.49% | | 51.81% | | | | | | | | 0.25% | | 28.35% |
| CST III | 16.65% | CST IIIa | 11.34% | 1.79% | 1.32% | 72.44% | 0.40% | 0.25% | 4.38% | | 0.62% | | | | |
| | | CST IIIb | 5.31% | 0.72% | 0.67% | 56.12% | 0.25% | | 32.72% | | 1.71% | | | | |
| CST-*G* | 13.38% | CST-Ga | 6.64% | 0.43% | 0.57% | 3.25% | 29.62% | 24.89% | | | | | | 2.43% | |
| | | CST-Gb | 6.74% | 0.25% | | 6.36% | 29.27% | 11.48% | | | | | | 36.77% | |
| CST IV | 4.09% | CST IV | 4.09% | 2.10% | 0.90% | 2.10% | 7.84% | 5.30% | 1.23% | 0.13% | | | | 2.90% | |
| CST V | 2.86% | CST V | 2.86% | 1.26% | 6.45% | 0.74% | 0.29% | 0.29% | 79.68% | | | | | | |
| CST-Kp | 3.06% | CST-Kp | 3.06% | 8.56% | | 0.56% | | | | 20.15% | 8.15% | 24.39% | 30.31% | | |
| CST-Vh | 3.88% | CST-Vh | 3.88% | | | 0.25% | | | | | | | 74.32% | | |

Figure 1

Figure 1. Identification of vaginotypes. Panel a shows a circular cladogram of the healthy vaginal samples obtained by means of hierarchical clustering (HCL) analysis. The cladogram highlighted the different CSTs identified by through HCL analysis.
Panel b reports the principal coordinate analysis (PCoA) of the healthy vaginal samples, subdivided by vaginotypes.
Panel c displays the average abundance and prevalence of bacteria that correspond to an identified CST and sub-CST.

215x279mm (300 x 300 DPI)

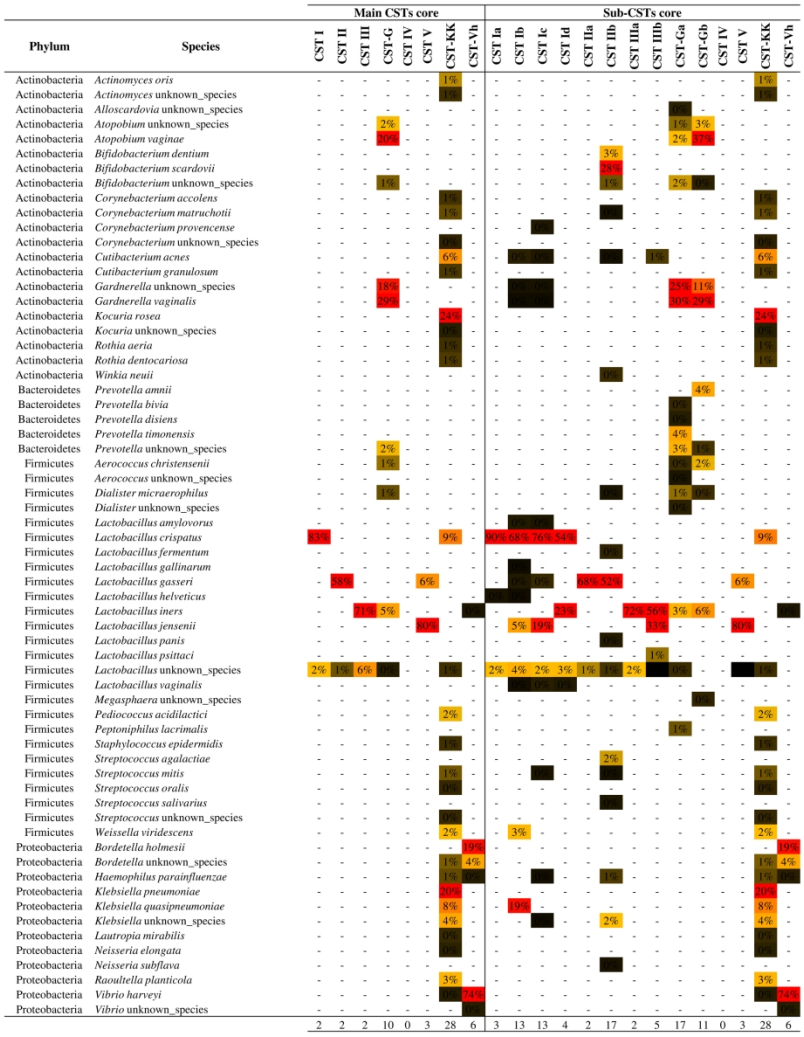
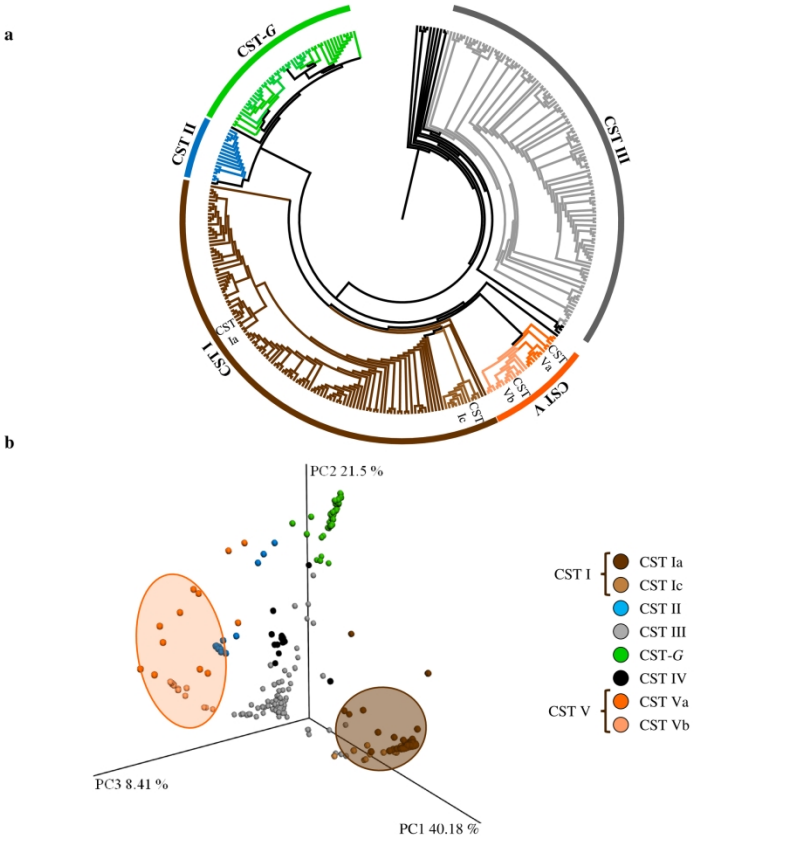



**Figure 2**

Figure 2. Evaluation of the most prevalent taxa characterizing each predicted vaginotypes. In detail, we selected those species that have been detected in >80 % of samples classified as a CST with a relative abundance >0.05 %.

215x279mm (300 x 300 DPI)




| CSTs | % presence | Sub-CSTs | % presence | % relative abundance | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | *L. crispatus* | *L. gasseri* | *L. iners* | *G. vaginalis* | *G.* uknown species | *L. jensenii* |
| CST I | 36.94% | CST Ia | 33.63% | 85.82% | 0.39% | 0.60% | 0.37% | 0.26% | 0.81% |
| | | CST Ic | 3.30% | 65.24% | 0.18% | 20.49% | 0.18% | 0.17% | 3.70% |
| CST II | 4.80% | CST II | 4.80% | 0.26% | 72.74% | | 2.91% | 3.29% | |
| CST III | 32.13% | CST III | 32.13% | 2.29% | 1.28% | 74.95% | 1.99% | 1.30% | 2.41% |
| CST-*G* | 13.81% | CST-*G* | 13.81% | 0.23% | 1.19% | 3.96% | 49.40% | 24.89% | 0.25% |
| CST IV | 4.50% | CST IV | 4.50% | 1.37% | 1.19% | 1.56% | 1.03% | 0.28% | 0.35% |
| CST V | 7.81% | CST Va | 3.60% | 0.11% | 2.12% | 0.18% | 10.22% | 4.98% | 57.36% |
| | | CST Vb | 4.20% | | 0.24% | 25.88% | | | 66.78% |

**Figure 3**

Figure 3. Identification of pregnancy vaginotypes. Panel a shows a circular cladogram of the healthy vaginal samples of pregnancy women, obtained by means of hierarchical clustering (HCL) analysis. The cladogram highlighted the different CSTs identified by through HCL analysis.
Panel b reports the principal coordinate analysis (PCoA) of the healthy vaginal samples of pregnancy women, subdivided by vaginotypes.
Panel c displays the average abundance and prevalence of bacteria that correspond to an identified CST and sub-CST.

215x279mm (300 x 300 DPI)

**a**

| | CSTI | CSTII | CSTIII | CST-*G* | CSTIV | CSTV | CST-Vh |
|---|---|---|---|---|---|---|---|
| **Class I** | 0.00018% | 0.00003% | 0.00186% | 0.00005% | 0.00580% | 0.00045% | 0.00002% |
| **Class II** | 0.00001% | 0.00003% | 0.00001% | 0.00122% | 0.00011% | 0.00000% | 0.00000% |
| **Class III** | 0.02717% | 0.00250% | 0.00109% | 0.00005% | 0.00163% | 0.00014% | 0.00002% |

0.02%
0%

**b**

| Pathways Class | Pathway | CST I | CST II | CST III | CST-G | CST IV | CST V | CST-Vh | ANOVA *p*-value |
|---|---|---|---|---|---|---|---|---|---|
| Alcohol Degradation | glycerol degradation I | 0.65% | 0.11% | 0.50% | 0.07% | 0.12% | 0.30% | 0.36% | 0.000 |
| Amine and Polyamine Biosynthesis | histamine biosynthesis | 0.03% | 0.00% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% | 0.654 |
| Amino Acid Biosynthesis | L-ornithine degradation I (L-proline biosynthesis) | 0.08% | 0.00% | 0.06% | 0.03% | 0.04% | 0.06% | 0.00% | 0.017 |
| Amino Acid Biosynthesis | L-arginine degradation VII (arginase 3 pathway) | 0.08% | 0.00% | 0.06% | 0.03% | 0.04% | 0.06% | 0.00% | 0.017 |
| Amino Acid Biosynthesis | L-arginine degradation (Stickland reaction) | 0.44% | 0.19% | 0.39% | 0.36% | 0.43% | 0.21% | 0.38% | 0.041 |
| Amino Acid Degradation | L-leucine degradation I | 0.12% | 0.08% | 0.12% | 0.04% | 0.11% | 0.06% | 0.00% | 0.422 |
| Amino Acid Degradation | L-lysine degradation XI (mammalian) | 0.02% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.884 |
| Aromatic Compound Biosynthesis | dipicolinate biosynthesis | 0.32% | 0.19% | 0.05% | 0.06% | 0.11% | 0.00% | 0.22% | 0.000 |
| Aromatic Compound Biosynthesis | 3-dehydroquinate biosynthesis II (archaea) | 2.16% | 1.03% | 1.55% | 0.12% | 0.16% | 1.40% | 0.88% | 0.000 |
| Aromatic Compound Biosynthesis | chorismate biosynthesis II (archaea) | 2.16% | 1.06% | 1.55% | 0.12% | 0.16% | 1.42% | 0.95% | 0.000 |
| Aromatic Compound Degradation | γ-resorcylate degradation I | 0.02% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.889 |
| Aromatic Compound Degradation | 4-ethylphenol degradation (anaerobic) | 0.02% | 0.01% | 0.00% | 0.00% | 0.01% | 0.00% | 0.00% | 0.886 |
| Carbohydrate Biosynthesis | protein N-glycosylation processing phase (yeast) | 0.06% | 0.00% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% | 0.904 |
| Carbohydrate Biosynthesis | trehalose biosynthesis IV | 0.11% | 0.02% | 0.06% | 0.03% | 0.11% | 0.10% | 0.00% | 0.703 |
| Carbohydrate Degradation | N-acetylneuraminate and N-acetylmannosamine degradation I | 0.25% | 0.03% | 0.02% | 0.03% | 0.02% | 0.00% | 0.08% | 0.000 |
| Carboxylate Degradation | oxalate degradation II | 0.40% | 0.37% | 0.02% | 0.01% | 0.00% | 0.00% | 0.00% | 0.000 |
| Carboxylate Degradation | pyruvate fermentation to acetate and (S)-lactate I | 1.46% | 0.72% | 1.27% | 0.54% | 0.69% | 0.36% | 0.87% | 0.000 |
| Carboxylate Degradation | pyruvate fermentation to acetate and lactate II | 1.46% | 0.74% | 1.27% | 0.54% | 0.69% | 0.36% | 0.95% | 0.000 |
| Carboxylate Degradation | glycerol and glycerophosphodiester degradation | 0.65% | 0.13% | 0.50% | 0.08% | 0.12% | 0.30% | 0.36% | 0.000 |
| Carboxylate Degradation | superpathway of acetate utilization and formation | 0.62% | 0.29% | 0.54% | 0.52% | 0.54% | 0.27% | 0.47% | 0.020 |
| Carboxylate Degradation | pyruvate fermentation to acetate II | 0.58% | 0.28% | 0.51% | 0.52% | 0.53% | 0.26% | 0.47% | 0.028 |
| Carboxylate Degradation | acetate and ATP formation from acetyl-CoA I | 0.62% | 0.28% | 0.54% | 0.60% | 0.57% | 0.27% | 0.47% | 0.010 |
| Carboxylate Degradation | pyruvate fermentation to acetate I | 0.58% | 0.31% | 0.51% | 0.54% | 0.57% | 0.26% | 0.55% | 0.031 |
| Carboxylate Degradation | malonate degradation I (biotin-independent) | 0.06% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.897 |
| Carboxylate Degradation | acetate conversion to acetyl-CoA | 0.02% | 0.01% | 0.02% | 0.00% | 0.00% | 0.00% | 0.00% | 0.821 |
| Cell Structure Biosynthesis | poly(ribitol phosphate) wall teichoic acid biosynthesis II (S. aureus) | 4.77% | 2.76% | 3.19% | 0.04% | 1.13% | 2.46% | 1.36% | 0.000 |
| Cell Structure Biosynthesis | poly(glycerol phosphate) wall teichoic acid biosynthesis | 4.92% | 2.77% | 3.36% | 0.08% | 1.11% | 2.58% | 1.07% | 0.000 |
| Cell Structure Biosynthesis | type IV lipoteichoic acid biosynthesis (S. pneumoniae) | 4.55% | 2.54% | 2.88% | 0.09% | 1.21% | 2.36% | 1.82% | 0.000 |
| Cell Structure Biosynthesis | poly(ribitol phosphate) wall teichoic acid biosynthesis I (B. subtilis) | 4.77% | 2.76% | 3.19% | 0.05% | 1.13% | 2.46% | 1.32% | 0.000 |
| Cell Structure Biosynthesis | type I lipoteichoic acid biosynthesis (S. aureus) | 4.84% | 2.85% | 3.37% | 0.10% | 1.18% | 3.13% | 0.99% | 0.000 |
| Cofactor, Carrier, and Vitamin Biosynthesis | γ-glutamyl cycle | 0.26% | 0.20% | 0.24% | 0.13% | 0.15% | 0.11% | 0.00% | 0.000 |
| Cofactor, Carrier, and Vitamin Biosynthesis | superpathway of heme b biosynthesis from glutamate | 1.20% | 0.74% | 1.11% | 0.19% | 0.27% | 0.41% | 0.80% | 0.000 |
| Cofactor, Carrier, and Vitamin Biosynthesis | NAD de novo biosynthesis I (from aspartate) | 0.43% | 0.25% | 0.39% | 0.02% | 0.04% | 0.20% | 0.15% | 0.000 |
| Cofactor, Carrier, and Vitamin Biosynthesis | NAD salvage pathway I (PNC VI cycle) | 0.43% | 0.23% | 0.38% | 0.02% | 0.02% | 0.20% | 0.07% | 0.000 |
| Cofactor, Carrier, and Vitamin Biosynthesis | NAD salvage pathway II (PNC IV cycle) | 0.43% | 0.23% | 0.37% | 0.01% | 0.01% | 0.20% | 0.07% | 0.000 |
| Degradation/Utilization/Assimilation | glutathione-mediated detoxification I | 0.26% | 0.20% | 0.25% | 0.15% | 0.15% | 0.13% | 0.00% | 0.001 |
| Detoxification | 4-hydroxy-2-nonenal detoxification | 0.26% | 0.20% | 0.24% | 0.13% | 0.15% | 0.11% | 0.00% | 0.000 |
| Generation of Precursor Metabolites and Energy | heterolactic fermentation | 1.16% | 0.56% | 1.03% | 0.34% | 0.34% | 0.30% | 0.60% | 0.000 |
| Generation of Precursor Metabolites and Energy | lactate biosynthesis (archaea) | 2.08% | 0.94% | 1.52% | 0.08% | 0.10% | 1.40% | 0.66% | 0.000 |
| Generation of Precursor Metabolites and Energy | L-alanine fermentation to propanoate and acetate | 0.11% | 0.03% | 0.02% | 0.09% | 0.09% | 0.01% | 0.09% | 0.007 |
| Secondary Metabolite Biosynthesis | achromobactin biosynthesis | 0.02% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.905 |
| Secondary Metabolite Degradation | mannitol degradation I | 0.21% | 0.01% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% | 0.000 |
| Secondary Metabolite Degradation | D-arabitol degradation | 0.21% | 0.01% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% | 0.000 |
| Tetrapyrrole Biosynthesis | tetrapyrrole biosynthesis I (from glutamate) | 1.16% | 0.73% | 1.11% | 0.19% | 0.23% | 0.37% | 0.49% | 0.000 |

0% 4.5%

**Figure 4**

Figure 4. Functional capabilities of vaginotypes. Panel a shows the abundance of bacteriocins class, i.e. class I, II and III, in different CSTs.
Panel b reveals the relative abundance of each biosynthetic and degradative pathway of the 44 pathways whose abundance in CST I is higher than all 6 other CSTs included in the analysis. Significant Tukey post-hoc analysis between CST I and the other CSTs are highlighted with a violet outline.

215x279mm (300 x 300 DPI)