Microbiota effects and predictors of *Lactobacillus crispatus* colonization after treatment with a vaginal live biotherapeutic: results from a randomized, double-blinded, placebo-controlled trial

Authors & affiliations

Seth M. Bloom*[1,2,3], Laura Symul*[4], Joseph Elsherbini[1], Jiawu Xu[1], Salina Hussain[1,5], Johnathan Shih[1,6], Ashley Sango[1,7], Caroline M. Mitchell[1,3,8,9], Anke Hemmerling[10], Thomas P. Parks[11], Aditi Kannan[1,12], Fatima A. Hussain[1,3,13], Craig R. Cohen[10], Susan P. Holmes[14], Douglas S. Kwon[1,2,3]

1. Ragon Institute of Mass General Brigham, MIT, and Harvard, Cambridge, Massachusetts, USA
2. Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts, USA
3. Harvard Medical School, Boston, Massachusetts, USA
4. Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA), Louvain Institute of Data Analysis and Modelling (LIDAM), UCLouvain, Belgium
5. Current affiliation: The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
6. Current affiliation: Graduate Program in Quantitative Biosciences, Georgia Institute of Technology, Atlanta, Georgia, USA
7. Current affiliation: African Women Leaders Network
8. Department of Obstetrics & Gynecology, Massachusetts General Hospital, Boston, Massachusetts, USA
9. Vincent Center for Reproductive Biology, Massachusetts General Hospital, Boston, Massachusetts, USA
10. Department of Obstetrics, Gynecology, & Reproductive Sciences, University of California San Francisco, San Francisco, California, USA
11. Osel, Inc., Mountain View, California, USA
12. Current affiliation: Chemical Biology Graduate Program, Harvard University, Cambridge, Massachusetts, USA

25 13. Current affiliation: Department of Biology, Tufts University, Medford, Massachusetts, USA

26 14. Department of Statistics, Stanford University, Palo Alto, California, USA

27 *Contributed equally

28 Correspondence: Douglas S. Kwon, dkwon@mgh.harvard.edu

Abstract

Bacterial vaginosis (BV) affects >25% of women worldwide and often recurs after standard-of-care metronidazole (MTZ) treatment. LACTIN-V, a live biotherapeutic product (LBP) containing *Lactobacillus crispatus* strain CTV-05, reduced recurrent BV in a Phase 2b clinical trial, but efficacy was incomplete. We characterized microbiota and immune effects and correlates of treatment success in trial samples. By week 12, *L. crispatus*-dominant microbiota was achieved in 30% of LBP recipients compared to 9% of placebo (benefit ratio: 3.31; $p<0.005$). This effect was mostly due to CTV-05, but native *L. crispatus* strains were also present and increased over time. Inflammatory cytokines decreased in both arms after MTZ, but returned to baseline in placebo recipients. *L. crispatus* colonization was associated with pre-MTZ microbiota, baseline cytokine profiles, post-MTZ bacterial load, and clinical and behavioral variables. These findings elucidate LBP microbiota effects and identify predictors of treatment success, informing improved intervention strategies to advance women's health.

41 **Introduction**

42 Bacterial vaginosis (BV), a vaginal syndrome characterized by *Lactobacillus*-deficient vaginal microbiota,
43 affects 23-29% of women of reproductive age globally, with an estimated annual economic burden of
44 $4.8 billion worldwide (Peebles et al., 2019; Workowski et al., 2021). BV symptoms include vaginal
45 discharge, malodor, pain, and itching, leading to significant impairments in quality of life, self-esteem,
46 and sexual function (Bilardi et al., 2013; Brusselmans et al., 2023). In addition, BV is linked to mucosal
47 inflammation and increased risk for numerous adverse health outcomes, such as HIV acquisition,
48 sexually transmitted infections, preterm birth, human papillomavirus (HPV) infection, and cervical cancer
49 (Anahtar et al., 2015; Atashili et al., 2008; Brusselaers et al., 2019; Gosmann et al., 2017; Hillier et al.,
50 1988; Leitich and Kiss, 2007). While BV occurs globally, it disproportionately affects those with lower
51 socioeconomic status and members of racial or ethnic minority groups across diverse settings (Allsworth
52 and Peipert, 2007; Kenyon et al., 2013; Marconi et al., 2015; Peebles et al., 2019). Development of
53 effective BV treatments is therefore a key objective to improve women's health worldwide (Bradshaw and
54 Sobel, 2016).

55 Current first-line therapy for BV consists of oral or intra-vaginal antibiotics such as metronidazole (MTZ),
56 which target many species within the diverse anaerobic bacterial communities characteristic of BV
57 (Bradshaw and Sobel, 2016; Workowski et al., 2021). In most cases, MTZ reduces abundance of BV-
58 associated anaerobes and leads to emergence of bacterial communities dominated by *Lactobacillus*
59 species (which are intrinsically MTZ-resistant) (Bradshaw and Brotman, 2015), but BV frequently recurs
60 after treatment. An Australian study of women with symptomatic BV found that 58% experienced
61 recurrent BV (rBV) within 1 year of MTZ therapy, while a US trial that enrolled women with symptomatic
62 BV and a prior history of post-treatment recurrence found that >75-80% experienced rBV within 16 weeks
63 of treatment (Bradshaw et al., 2006; Schwebke et al., 2021). It is hypothesized that the incomplete
64 efficacy of MTZ often results from a failure to eradicate BV-associated bacterial communities and/or from
65 their post-MTZ replacement with *Lactobacillus* species prone to reverting to BV (Bradshaw and Brotman,

66 2015). Studies have shown that *Lactobacillus iners* is associated with an increased risk of transition to
67 BV-like states (DiGiulio et al., 2015; Munoz et al., 2021; Tamarelle et al., 2022) and with higher rates of
68 adverse health outcomes (Colbert et al., 2023; Gosmann et al., 2017; Kindinger et al., 2017; Norenhag
69 et al., 2020; van Houdt et al., 2018). MTZ treatment frequently results in establishment of vaginal
70 microbiota dominated by *L. iners* rather than *L. crispatus* (Ferris et al., 2007; Joag et al., 2019; Mitchell
71 et al., 2012; Ravel et al., 2013; Srinivasan et al., 2010; Verwijs et al., 2020), supporting the hypothesis
72 that BV therapies which preferentially promote *L. crispatus* over *L. iners* may improve treatment outcomes
73 and promote vaginal health (Bradshaw and Brotman, 2015; Nilsen et al., 2020; Zhu et al., 2024).

74 New strategies to enhance *L. crispatus* colonization during BV treatment are in various stages of
75 development, including vaginal microbiome transplants, adjunctive therapies that selectively inhibit *L.*
76 *iners* and/or promote *L. crispatus* growth, and *L. crispatus*-containing live biotherapeutic products (LBPs)
77 (Bloom et al., 2022; Chetty et al., 2025; Cohen et al., 2020; Lev-Sagie et al., 2019; Nilsen et al., 2020;
78 Ravel et al., 2025; Yockey et al., 2022; Zhu et al., 2024). The most advanced LBP in development is
79 LACTIN-V, a vaginally-administered LBP containing the *L. crispatus* strain CTV-05, and the only LBP
80 tested in large-scale clinical trials (Cohen et al., 2020; Hemmerling et al., 2009). Cohen and colleagues
81 recently reported results of a phase 2b multi-center, randomized, placebo-controlled, double-blinded
82 study showing that participants who received 11 weeks of intravaginal LACTIN-V after a 5-day course of
83 intravaginal MTZ developed rBV at lower rates than placebo recipients (Cohen et al., 2020). Benefits
84 from LACTIN-V were observed at both week 12 (approximately 1 week after completing therapy; relative
85 risk of rBV = 0.66) and week 24 (approximately 13 weeks after completing therapy; relative risk of rBV =
86 0.73). However, rBV rates remained high even in the LBP arm, with 30% of recipients experiencing rBV
87 by week 12 and 39% experiencing recurrence by week 24 in the intention-to-treat analysis. In prior
88 reports, rBV was characterized using clinical measures, but comprehensive molecular analysis of vaginal
89 microbiota composition and factors associated with colonization were not assessed.

90 Here we characterize the effects of LACTIN-V on vaginal microbiota composition and identify correlates
91 of successful colonization by analyzing vaginal microbiota composition, bacterial strain dynamics,
92 cytokine profiles, and demographic/behavioral parameters of LACTIN-V phase 2b trial participants
93 (Cohen et al., 2020). We find that the reduction in rBV among LBP recipients corresponded to >3-fold
94 higher rates of *L. crispatus*-dominant colonization, but only 30% of recipients achieved *L. crispatus*-
95 dominant vaginal microbiota at week 12, consistent with the observed incomplete clinical efficacy.
96 Microbiota trajectory and strain dynamics revealed CTV-05 accounted for the majority of *L. crispatus*
97 colonization among LBP recipients, though endogenous strains increased over time and sometimes
98 displaced CTV-05. We further describe the effects of LBP treatment and microbiota composition on
99 vaginal immune markers, and show that pre-MTZ microbiota composition, baseline inflammatory cytokine
100 profiles, post-MTZ total bacterial load, and selected clinical/behavioral variables are associated with LBP
101 treatment success. These results show how an *L. crispatus* LBP alters vaginal microbiota composition to
102 improve BV treatment efficacy while identifying factors linked to LBP success that can guide development
103 of future therapies to promote more optimal vaginal microbiota communities and improved women's
104 health outcomes.

105 **Results**

106 **Study design and participant characteristics**

107 Women aged 18 to 45 years with BV completed five days of standard intravaginal MTZ therapy and were
108 then randomized 2:1 to receive intravaginal LACTIN-V (Osel, Inc.) or placebo as previously described
109 (Cohen et al., 2020). Enrollment occurred at four U.S. sites. LACTIN-V is a powder formulation containing
110 the live *L. crispatus* strain CTV-05, administered via prefilled, single-use vaginal applicators at a dose of
111 $2x10^9$ colony-forming units (CFU). The placebo contained the same inactive ingredients without bacteria
112 and was visually indistinguishable. BV diagnosed at the baseline "pre-MTZ" screening visit required at
113 least three of four Amsel criteria and a Nugent score of ≥4 (Amsel et al., 1983; Nugent et al., 1991).

114 Participants received five days of intravaginal MTZ gel within 30 days of diagnosis, and then were seen
115 at a "post-MTZ" visit within 48 hours of completing antibiotics for randomization to LBP or placebo. The
116 first treatment dose was clinician-administered at the post-MTZ visit, then participants self-administered
117 doses daily for four consecutive days, followed by twice weekly dosing for 10 weeks (ending at week 11
118 post-randomization; **Figure 1A**). Clinician-collected vaginal swabs were obtained at the pre-MTZ and
119 post-MTZ visits prior to randomization, and at weeks 4, 8, 12, and 24 post-randomization. Detailed
120 demographic, clinical, and behavioral data showed no notable differences between arms (Cohen et al.,
121 2020). The trial enrolled and randomized 228 participants (152 in the LBP arm and 76 in the placebo
122 arm); samples from 213 of these participants (142 in the LBP arm and 71 in the placebo arm) were
123 available for this analysis, representing a total of 1,156 unique study visits.

124 **LBP Effects on *L. crispatus* and total *Lactobacillus* colonization**

125 Bacterial microbiota composition was determined by bacterial 16S V4 ribosomal RNA (rRNA) gene
126 sequencing. Four samples were excluded due to technical failures of PCR amplification and sequencing
127 ($<10^3$ processed reads per sample). Of the remaining 1,152 samples (**Table S1, Figure S1A**), the median
128 number of analyzable reads per sample was 31,177 (IQR 23,774-41,251). No significant differences in
129 microbiota composition were observed between study arms at either the pre-MTZ or post-MTZ visits
130 (PERMANOVA $p = 0.63$ and $p = 0.14$, respectively; **Figure S1B-C, SI**).

131 To assess microbiota impacts of LBP treatment, we defined two outcome parameters based on
132 microbiota proportional composition: ≥50% relative abundance of *L. crispatus* ("*L. crispatus*-dominant")
133 and ≥50% summed relative abundance of *Lactobacillus* genus ("*Lactobacillus*-dominant", including *L.*
134 *crispatus*). The primary microbiota endpoint for this analysis was establishment of an *L. crispatus*-
135 dominant vaginal community at week 12, the timepoint corresponding to the trial's primary clinical

136 endpoint (Cohen et al., 2020). Secondary microbiota endpoints were *L. crispatus*-dominance at week 24
137 and total *Lactobacillus*-dominance at weeks 12 or 24. Almost all participants in both treatment and
138 placebo arms had non-*Lactobacillus*-dominant microbiota (<50% *Lactobacillus*) at the pre-MTZ screening
139 visit, consistent with the fact that all had met criteria for clinical BV (**Figure 1B**). The majority in both arms
140 transitioned to *Lactobacillus*-dominance at the post-MTZ visit that was largely driven by high relative
141 abundance of *L. iners* (**Table S2**). Only one participant in each arm exhibited *L. crispatus*-dominant
142 colonization at the post-MTZ visit (**Figure 1B**). Microbiota composition of the two arms diverged after
143 randomization, with 30% of LBP recipients (n=37) compared to only 9% of placebo recipients (n=5)
144 achieving *L. crispatus*-dominant colonization at week 12 (the primary endpoint), resulting in a benefit ratio
145 of 3.4 (95% CI: 1.4 - 8.1; $p < 0.005$; **Table 1**, **Figure 1B-C**). At week 24, 35% of LBP recipients (n=39)
146 versus just 8% of placebo recipients (n=4) achieved *L. crispatus*-dominant colonization, for a benefit
147 ratio of 4.5 (95% CI: 1.7 - 11.9). The LBP did not increase rates of total *Lactobacillus*-dominant
148 colonization at week 12 compared to placebo (56% and 48% of LBP and placebo recipients, respectively;
149 benefit ratio: 1.16; 95% CI: 0.85 - 1.59), but did at week 24 (58% and 33% of LBP and placebo recipients,
150 respectively; benefit ratio: 1.76; 95% CI: 1.15 - 2.68; **Table 1**, **Figure 1B-C**).

151 We next investigated how these microbiota parameters related to the clinical diagnosis of BV by
152 comparing clinical and microbiota results for all post-randomization visits (weeks 4 and beyond, n = 729
153 visits). Among the 209 visits in which participants had ≥50% *L. crispatus* colonization, none had
154 concurrent clinical BV (**Table S3** and **Figure 1C**), and among 194 visits with <50% *L. crispatus* but ≥50%
155 total *Lactobacillus*, only 1% (n=2) had BV. However, of the 326 visits with <50% total *Lactobacillus*, 50%
156 (n=164) had clinical BV (**Table S3**). Thus, *Lactobacillus*-dominance (including *L. crispatus*-dominance)

157 was highly specific for the absence of BV, while non-*Lactobacillus*-dominance was highly sensitive but
158 less specific for BV.

159 In addition to sequencing-based microbiota composition analysis, total bacterial load in vaginal swab
160 samples was assessed via quantitative PCR (qPCR) at the post-MTZ and post-randomization visits.
161 Median bacterial load was significantly lower and more variable at the post-MTZ visit (median $10^{2.69}$
162 copies/swab; IQR $10^{0.08}$ - $10^{6.44}$) than at subsequent visits (median $10^{7.52}$ copies/swab; IQR $10^{7.09}$ - $10^{7.93}$
163 for all subsequent visits), consistent with antibiotic-mediated depletion of vaginal microbiota biomass
164 during BV treatment, followed by bacterial repopulation of the vaginal environment (**Figure 1D**). We
165 assessed whether post-MTZ microbiota composition (as determined by 16S rRNA gene sequencing) was
166 correlated with total bacterial load using a multi-table multivariate generalization of the squared Pearson
167 correlation, the RV coefficient (Josse and Holmes, 2016; Robert and Escoufier, 1976). This analysis
168 revealed no significant correlation between post-MTZ microbiota composition and bacterial load (RV
169 coefficient = 0.02, *p*-value > 0.1).

170 **Microbiota composition trajectories throughout the trial**

171 To characterize microbiota trajectories with greater resolution, we summarized microbiota composition
172 using topic mixtures (Sankaran and Holmes, 2019; Symul et al., 2023). Compared to clustering or
173 categorizing samples into community state types (CSTs) (Ravel et al., 2011), cervicotypes (CTs)
174 (Gosmann et al., 2017), or sub-CSTs (France et al., 2020), this method provides probabilistic (rather than
175 binary) weightings, allowing for more accurate descriptions of sample taxonomic composition and
176 improved characterization of longitudinal dynamics (Symul et al., 2023). In our cohort, assignment to
177 reference CST performed poorly (median Bray-Curtis dissimilarity with assigned CSTs > 0.33, **SI**),
178 especially for *Prevotella*-dominated samples (median BC dissimilarity > 0.5, **SI**), and with over 40% of
179 samples being equally similar to two or more CSTs (**SI**). To identify topics, which can be interpreted as
180 bacterial subcommunities (*i.e.,* consortia of bacterial taxa that co-occur with each other or with other

taxa), and estimate their proportions in each sample, we fitted a latent Dirichlet allocation (LDA) model (a Bayesian "topic model") (Blei et al., 2003) to the taxonomic counts. We adapted a previously described approach (Symul et al., 2023) in which topics were constrained to be fully composed of either *Lactobacillus* or non-*Lactobacillus* species. Four non-*Lactobacillus* topics were inferred from the data (**Figure S2A, B**) and we defined four *Lactobacillus*-dominated topics, including three exclusively composed of a single *Lactobacillus* species (*L. crispatus*, *L. iners*, and *L. jensenii*/*mulieris*) and one composed of a mixture of the remaining *Lactobacillus* species (**Figure S2C**). **Figure 2A** shows inferred composition of each non-*Lactobacillus* topic, while **Figure 2B** displays topic proportions for each participant throughout the trial.

Three non-*Lactobacillus* topics were highly prevalent and abundant in both arms at the pre-MTZ visit, including topics in which the predominant taxa were *Candidatus* Lachnocurva vaginae (BVAB1), *Gardnerella swidsinskii/leopoldii*, and *Prevotella amnii* (**Figure 2C** and **S2D**). Microbiota composition shifted substantially at the post-MTZ visit (**Figure 2D**), where besides high *L. iners* relative abundance, the non-*Lactobacillus* topic with the highest relative abundance was the topic in which *Gardnerella swindsinskii/leopoldii* was predominant (**Figure 2C** and **S2D**). Another major shift in microbiota composition occurred between the post-MTZ visit and the week 4 visit, which was larger in the LBP arm (median pairwise Bray-Curtis dissimilarity: 0.81; IQR: 0.50 - 0.97) than the placebo arm (median Bray-Curtis dissimilarity: 0.58; IQR 0.41 - 0.80), driven primarily by increased *L. crispatus* abundance among LBP recipients (**Figure 2C-D** and **S2D**). Microbiota composition was more stable after week 4, with median Bray-Curtis dissimilarity <0.55 at week 8 and all subsequent visits, which was comparable between arms (**Figure 2D**). LBP recipients who achieved *L. crispatus*-dominant colonization had more stable microbiota than their counterparts (**Figure S2E, SI**).

The observation that microbiota composition became more stable after week 4 was also supported by longitudinal patterns in microbiota categories. Among 115 participants with available data from both week

205 4 and week 12 visits, 71% (25 of 35; 95% CI: 53%-85%) of LBP recipients who achieved ≥50% *L.*
206 *crispatus* at week 12 also had *L. crispatus*-dominance at week 4 (**Figure 2B, Table S4**). Similarly, 71%
207 (27 of 38; 95% CI: 54%-84%) of LBP recipients who achieved *L. crispatus*-dominance at week 24 also
208 had *L. crispatus*-dominance at week 4 (**Figure 2B**, **Table S5**). However, early *L. crispatus*-dominant
209 colonization did not guarantee persistence, as only 49% (25 of 51) and 57% (27 of 47) of LBP recipients
210 with *L. crispatus*-dominance at week 4 retained it at weeks 12 and 24, respectively. By contrast, among
211 38 LBP recipients with non-*Lactobacillus*-dominant microbiota at week 4, 79% remained non-
212 *Lactobacillus* dominant at week 12 compared to just 10.5% each with *L. crispatus*-dominance or other
213 *Lactobacillus*-dominant colonization (**Table S4**). Thus, microbiota communities established early tended
214 to persist and early *Lactobacillus*-dominance – particularly *L. crispatus*-dominance – was strongly linked
215 to ongoing *Lactobacillus*-dominance.

216 **Metagenomic identification of the CTV-05 strain and assessment of *L. crispatus* strain dynamics**

217 The observation that LBP treatment promoted *L. crispatus* colonization prompted us to examine what
218 fraction of *L. crispatus* observed in LBP recipients represented the CTV-05 strain versus non-LBP native
219 strains. We used metagenomic sequencing to assess *L. crispatus* strain dynamics, with strain genotypes
220 and proportions inferred using StrainFacts (Smith et al., 2022). In a separate analysis, simulations and
221 experimental spike-in data showed that StrainFacts performs well at inferring LBP (including CTV-05)
222 strain genotypes and proportions in samples with ≥5% overall relative abundance of *L. crispatus* (Shih et
223 al., 2025). In the LACTIN-V trial samples, *L. crispatus* strain inference was successful in 313 samples,
224 identifying 24 genotypically distinct strains with at least 10% fractional strain abundance in at least one
225 sample. To determine which strain represented CTV-05, we generated a closed assembly of the CTV-05
226 genome, which nearly perfectly matched the StrainFacts genotype of a widely prevalent inferred strain
227 within the metagenomic dataset (Jaccard similarity >0.999). Among 57 samples with inferred *L. crispatus*

228 strain composition from placebo recipients (any visit) or from LBP recipients prior to LBP administration,
229 StrainFacts inferred CTV-05 presence in just 5 total samples from 5 unique participants (**Figure 3A**). This
230 result suggested our strain inference approach had a low false positive CTV-05 detection rate; we cannot
231 rule out that some "false positives" represented native strains with high genotypic similarity to CTV-05.
232 Collectively, these results support that CTV-05 detection was achieved with high accuracy in sample
233 metagenomes.

234 To further characterize fidelity of the strain-tracing approach, we analyzed inference results for native *L.*
235 *crispatus* strains, supplemented by targeted bacterial isolations and genome sequencing to
236 experimentally test bioinformatic predictions. Most participants in whom a native strain was inferred
237 exhibited colonization by only a single native strain at any point during the study, with a few participants
238 colonized by two or three native strains (**Figure 3B**). Since StrainFacts inferences are performed blinded
239 to whether samples are derived from the same participant (Shih et al., 2025; Smith et al., 2022),
240 consistent detection of the same 1-3 native strains in multiple samples from the same participant
241 additionally supported validity of the inferences. We also performed targeted bacterial isolations of *L.*
242 *crispatus* from a subset of samples inferred to contain CTV-05, one or more native strains, or a mixture
243 of CTV-05 and native strains. The resulting isolates were highly genotypically similar to inferred strains
244 from the same samples, including both CTV-05 and native strains (results for an example participant
245 shown in **Figure 3C-E**, with additional participants detailed in **Figure S3A**). These experimental results
246 further validated our metagenomic strain inference approach.

247 We examined *L. crispatus* strain dynamics by classifying samples with inferred strain data into three
248 categories: samples with >90% CTV-05 fractional strain abundance ("High-CTV-05"), samples with 10-
249 90% CTV-05 ("Mixed"), and samples with <10% CTV-05 ("Low-CTV-05"; **Figure 3A**). These thresholds
250 were defined based on simulations to determine parameters maximizing reliability of strain detection
251 (Shih et al., 2025). At each post-randomization timepoint, the majority of LBP recipients with analyzable

252 *L. crispatus* strains had high-CTV-05 strain abundance, but the proportion with high-CTV-05 or mixed
253 strains decreased over time, while the proportion with high native *L. crispatus* strain abundance (*i.e.*, low-
254 CTV-05) increased (**Figure 3A**). At week 4, 96.1% (n=74) of LBP recipients with ≥5% *L. crispatus* had
255 high-CTV-05 or mixed strains, while just 3.9% (n=3) had high native strains. However, by week 24 only
256 63.2% (n=31) had high-CTV-05 or mixed strains while those with high native strains increased to 36.8%
257 (n=18). To characterize this phenomenon in greater detail, we examined CTV-05 fractional strain
258 abundance for each pair of consecutive post-randomization visits (**Figure 3F-G**). Among 144 visits with
259 high-CTV-05, two main outcomes were observed at the next visit: 68% (n=98) maintained high-CTV-05
260 and 26% (n=38) transitioned to <5% total *L. crispatus* colonization (strain composition undetermined).
261 Among the 18 visits with high native strain abundance, 61% (n=11) retained high native strain abundance
262 at the next visit and none transitioned to CTV-05 dominance. Interestingly, among 23 visits with mixed
263 CTV-05/native strain composition, only one (4%) transitioned to high-CTV-05, while 39% (n=9) remained
264 mixed and 35% (n=8) transitioned to high native strain abundance, including some transitions to high
265 native strain abundance that occurred while participants were still receiving LACTIN-V (**Figure S3B**).
266 These results show CTV-05 was frequently the sole or dominant *L. crispatus* strain in LBP recipients, but
267 when native strains dominated or co-occurred with CTV-05, they often replaced CTV-05 at subsequent
268 visits but were almost never replaced by it.

269 **Intervention effects on vaginal inflammation**

270 We assessed the effects of MTZ and LBP treatment on vaginal mucosal immune state and inflammation
271 by measuring cytokine and chemokine concentrations in vaginal swab eluates using a previously
272 described custom Luminex assay (Gosmann et al., 2017; Symul et al., 2023). Concentrations of 18
273 cytokines and chemokines were log-transformed for variance stabilization and imputed when outside
274 their respective limits of quantification (**Figure S4A**). Concentrations of all chemokines and cytokines
275 were strongly correlated, with the first principal component (PC1) accounting for 50% of the total variation

276 (**Figure S4B**). This phenomenon suggested a size-effect (Jolicoeur and Mosimann, 1960) that we
277 attributed to the (unmeasured) variation in biomass collection between swabs, which was addressed
278 through PC1-subtraction (**SI**). Seven cytokines were excluded from analysis because a large proportion
279 of values were outside the quantification range (**Figure S4A**). Mucosal immune dynamics during
280 treatment were determined by comparing adjusted concentrations of each cytokine or chemokine from
281 each participant's baseline (pre-MTZ) visit with concentrations at subsequent visits.

282 We first examined how chemokines and cytokines with known relationships to BV and vaginal microbiota
283 composition (Anahtar et al., 2015; Gosmann et al., 2017; Masson et al., 2019, 2014) changed during
284 treatment. IL-1β concentrations decreased post-MTZ in both arms, while IP-10 showed the opposite
285 pattern, consistent with previously reported pattern with BV treatment (Masson et al., 2014) (**Figure 4A-**
286 **B**). These effects persisted in both arms through week 12, but IL-1β and IP-10 returned to pre-MTZ levels
287 by week 24 in placebo recipients, leading to a difference of -0.39 (95% CI: -0.66 – -0.14) between
288 adjusted concentrations in the LBP versus placebo arm for IL-1β and of 0.51 (95% CI: 0.21 - 0.79) for IP-
289 10. By contrast, IL-6 concentrations – which have been reported not to differ between women with and
290 without BV (Masson et al., 2014) – remained unchanged from pre-MTZ levels in both treatment arms
291 throughout the study (**Figure 4C**). Sensitivity analyses examining the data prior to adjustment for size-
292 effect via PC1-subtraction revealed findings that were qualitatively similar, but with reduced magnitude
293 (**SI**).

294 The reversion of treatment-induced effects for IL-1β and IP-10 concentrations in placebo recipients by
295 week 24 suggested an association with total *Lactobacillus* relative abundance, which differed between
296 treatment arms at this visit (**Table 1**). We therefore examined the correlation of microbiota composition
297 (summarized by topic proportions) with mucosal cytokine/chemokine concentrations. Their RV coefficient
298 was 0.18 with p-value <0.005 (permutation test), indicating significant correlation. Similar RV coefficient
299 values were obtained when repeating analysis independently for each visit (**Figure S4C**) except for a

300 substantially lower correlation at the post-MTZ visit (where microbiota absolute abundance varied widely,
301 **Figure 1D**), suggesting microbiota-immune correlation was not driven primarily by a participant effect.
302 To characterize the microbiota-immune relationship in greater detail, we employed DiSTATIS (Abdi et
303 al., 2009, 2005), a method that uses between-sample dissimilarity matrices to infer a consensus matrix
304 across two datasets and associated latent components. Projecting microbiota topic relative abundances
305 and transformed cytokine/chemokine concentrations onto this latent subspace using correlation circles
306 revealed that the first latent dimension discriminated *Lactobacillus*-dominated samples from non-
307 *Lactobacillus*-dominated samples (**Figure 4D-F**). Chemokines MIG (CXCL9), IP-10 (CXCL10), and ITAC
308 (CXCL11) had high positive correlation along the first latent dimension, indicating positive covariation
309 with *Lactobacillus*, while IL-1β, TNF-α, and IL-1α had high negative correlations, indicating positive
310 covariation with non-*Lactobacillus* species (**Figure 4E**). The second latent dimension discriminated *L.*
311 *crispatus* from *L. iners* but most cytokines and chemokines had little or no correlation with that
312 component, indicating minimal association of inflammatory markers with individual *Lactobacillus* species
313 (**Figure 4E**). Most cytokines and chemokines had high correlations with the 3rd and/or 4th latent
314 components but the microbiota topics did not, indicating a degree of immune variation that was
315 microbiota-independent (**Figure 4F**). In these dimensions, IFNγ and IL-17 correlated with each other but
316 were anti-correlated with IL-6, MIP-3α, MIP-1β, and MIP-1α. Notably, although IP-10 levels were
317 previously reported to differ in a small subset of LACTIN-V recipients based on whether participants were
318 highly colonized by CTV-05 versus non-CTV-05 *L. crispatus* strains (as measured by qPCR) (Armstrong
319 et al., 2022), we observed no clear differences in microbiota correlation with cytokine/chemokine levels
320 (including IP-10) in analysis that distinguished CTV-05 from other *L. crispatus* strains (**Figure S4D-G**).

321 **Heterogeneity in LBP treatment effect based on baseline microbiota composition**

322 We next performed exploratory *post-hoc* analyses to investigate whether differences in pre-MTZ
323 microbiota composition corresponded to differences in response to the LBP (*i.e.,* heterogeneity of
324 treatment effect). We stratified participants by their most abundant pre-MTZ taxon (aggregated at the

325 genus-level), identifying four groups for analysis: *Lactobacillus*-predominant, *Gardnerella*-predominant,
326 *Prevotella*-predominant, and *Ca.* Lachnocurva vaginae (BVAB1)-predominant. LBP treatment benefit
327 was evaluated for each group with respect to two distinct outcomes: ≥50% *L. crispatus* colonization (at
328 week 12 or 24) and BV recurrence (by week 12 or 24) (**Figure 5A**). Heterogeneity in benefit among pre-
329 MTZ microbiota groups (**Figure 5B**) was tested using an analysis of deviance comparing nested logistic
330 regression models allowing (or not) for heterogeneity and adjusting for multiple testing by controlling for
331 the false discovery rate (Methods, **SI**).

332 Analysis of the relationship between pre-MTZ microbiota groups and establishment of *L. crispatus*-
333 dominant colonization suggested some groups benefitted more from LBP treatment than others (adjusted
334 *p*-values = 0.11 at week 12 and 0.03 at week 24; **Figure 5C, D**). LACTIN-V recipients with pre-MTZ
335 predominance of *Lactobacillus*, *Gardnerella*, or *Prevotella* attained higher rates of *L. crispatus*-dominant
336 microbiota at both weeks 12 and 24 compared to their placebo counterparts, with modestly larger rate
337 differences at week 24 (**Figure 5D**). In contrast, participants with pre-MTZ *Ca.* Lachnocurva vaginae
338 (BVAB1) did not differ in *L. crispatus*-dominant colonization between LBP and placebo arms. Analysis of
339 BV recurrence also showed differences in benefit from LBP treatment based on pre-MTZ microbiota
340 (adjusted *p*-values <0.05 at week 12 and <0.01 at week 2; **Figure 5E-F**). Among LBP recipients with pre-
341 MTZ *Prevotella*-predominance, BV recurred in almost all placebo recipients by week 24, compared to
342 just over half of LBP recipients. By contrast, LBP recipients whose most prevalent taxa pre-MTZ was Ca.
343 *Lachnocurva vaginae* (BVAB1) were unique in exhibiting slightly higher absolute rates of rBV than those
344 who received placebo (**Figure 5E-F**). We used similar methods to perform a sensitivity analysis in which
345 participants were grouped based on pre-MTZ CSTs (France et al., 2020) or using a model-based
346 approach relying on topic relative abundances. Findings from this analysis were largely consistent with
347 results of our predominant-taxon analysis above (**Figure S5, SI**). Together, these exploratory analyses
348 showed that benefits of LBP treatment in achieving *L. crispatus*-dominance and preventing BV recurrence

349 differed depending on pre-MTZ microbiota composition, with participants who had pre-MTZ
350 predominance of *Ca.* Lachnocurva vaginae (BVAB1) exhibiting a lack of benefit compared to placebo,
351 whereas other participants experienced more favorable responses.

352 **Factors associated with achieving *L. crispatus* dominance in LBP recipients**

353 To identify factors associated with establishing *L. crispatus* dominance among LBP recipients, we
354 performed a series of multiblock partial least square discriminant analyses (MB-PLS-DA) on the
355 longitudinal LACTIN-V arm data (**Figure 6A-C**, **S6**, **S7A**). Based on the observation that early *L.*
356 *crispatus*-dominant colonization correlated with *L. crispatus* colonization at later visits (**Table S4**, **S5**), we
357 analyzed factors from prior visits that associated with the participants' microbiota category by week 4, as
358 well as separate analyses of factors associated with microbiota category at ensuing visits. We defined
359 three models corresponding to the different phases of the trial: the "initial phase" spanned from the post-
360 MTZ visit to week 4, the "continuation phase" spanned from week 4 to week 12, and the "follow-up phase"
361 spanned from week 12 to week 24 (**Figure S6A**). For each phase, factors of interest (*i.e.,* explanatory
362 variables) were grouped into 13 thematic blocks (**Figure 6A**, **Table S6**). The first block characterized
363 participant demographics, blocks 2-4 described participant baseline vaginal ecosystem (*i.e.*, pre-MTZ
364 microbiota composition and diversity, pH, cytokines concentrations), blocks 5-8 quantified the vaginal
365 ecosystem at the previous visit (*e.g.,* at the post-MTZ visit in the initial phase, when the response is week
366 4 colonization status), while blocks 9-13 characterized participant sexual behavior, douching/bleeding,
367 antibiotic use, and product adherence (**Table S6**). For each phase, MB-PLS-DA was used to evaluate
368 the ability of these variables to discriminate between the following microbiota categories at each visit
369 starting from week 4: ≥50% *L. crispatus*, ≥ 50% *Lactobacillus* but <50 % *L. crispatus*, or < 50%
370 *Lactobacillus* (see also **Figure 1B**). We also used nested models to assess the additional contribution of
371 specific blocks in discriminating colonization status (**SI**, **Figure S6A-C**).

372 In the model for the LBP arm initial phase, the most important block for discriminating week 4 colonization
373 categories was the previous (*i.e.,* post-MTZ) vaginal environment (**Figure 6A**, left panel). This block
374 included post-MTZ total bacterial load, α-diversity, and pH. Lower values of these three factors were
375 associated with higher chances of ≥50% *L. crispatus* at week 4, and were among the 5 most important
376 variables in the model (**Figure 6B-C** and **S7A**). The next most important blocks were demographics and
377 blocks characterizing participant pre-MTZ vaginal environment, including pre-MTZ vaginal pH, α-
378 diversity, cytokine concentrations, and microbiota composition (as described by topic proportions, **Figure**
379 **6A** and **S7A**). Specifically, high pre-MTZ α-diversity and pH were negatively associated with *L. crispatus*
380 colonization at week 4 (**Figure 6B**). Interestingly, pre-MTZ adjusted concentrations of the chemokine IL-
381 1β were positively associated with achieving *L. crispatus*-dominance at week 4 while pre-MTZ IP-10 and
382 MIG levels showed an opposite association (**Figure 6B**, **6D**). Self-declared race and education were the
383 most important demographic block variables, with white and more highly educated participants having
384 higher rates of *L. crispatus*-dominance (**Figure S7A**). However, race and education were highly
385 correlated in the cohort (**Figure S7B**) such that their individual contributions to explaining *L. crispatus*
386 colonization could not be clearly distinguished. While these two demographic variables were also
387 correlated with several pre-MTZ vaginal characteristics that also explained colonization such as pre-MTZ
388 α-diversity (**Figure S7C**), they still mildly contributed to model explanatory power (**Figure S6C**). The
389 remaining blocks or variables were not important for discriminating between initial phase colonization
390 categories (**Figure 6A-B**, **Figure S7A**), possibly because variables in these blocks contributed limited
391 information as overall adherence was high, use of antibiotics and douching was relatively rare, and some
392 birth control options were rare in this cohort (**SI**). In contrast, within these blocks, factors with higher
393 heterogeneity across participants such as bleeding and sexual activity showed greater variable
394 importance (**Figure S7A**).

395 In the continuation phase model for the LBP arm (from week 4 to week 12), the colonization category at
396 the previous visit was by far the most important block associated with colonization category at the
397 following visit (**Figure 6A**), consistent with higher stability of microbiota composition from week 4 to 12
398 (**Figure 2D**). While no other blocks significantly improved model performance (**Figure S6C**), individual
399 variables including sexual activity, bleeding, douching, and non-hormonal IUDs were all negatively
400 associated with high-level *L. crispatus* colonization (**Figure S7A**). In the model for the LBP follow-up
401 phase (from week 12 to week 24), the colonization category at the previous visit (*i.e.,* week 12) was again
402 strongly associated with the colonization category at week 24. However, in contrast to the previous phase,
403 sexual behavior and antibiotic use only mildly contributed in explaining week 24 colonization categories
404 (**Figure 6A**, right panel, **Figure S6C**, **S7A**).

405 Analogous analyses for placebo recipients did not show significant predictive value for either the initial
406 phase or the follow-up phase in cross-validation (**Figure S8A-C**, **S9**). The placebo continuation phase
407 model had modest predictive value, with microbiota category at the previous visit serving as the best
408 predictor of microbiota category at the next visit (**Figure S8A-C**, **S9**).

409 **Discussion**

410 Treatment with vaginal live biotherapeutic products (LBPs) offers significant promise to improve health,
411 but mechanisms, correlates, and strain dynamics of LBP colonization and efficacy remain incompletely
412 understood (Bradshaw et al., 2025). We performed comprehensive analyses of samples and data from
413 a randomized, double-blind, placebo-controlled phase 2b trial of LACTIN-V, a single-strain *L. crispatus*
414 LBP for prevention of BV recurrence (Cohen et al., 2020). We employed microbiome sequencing, strain
415 tracing, and immunologic characterization together with multi-block analysis of biological, clinical,
416 demographic, and behavioral parameters to investigate LBP effects and correlates of treatment success.
417 LACTIN-V treatment resulted in *L. crispatus*-dominant vaginal bacterial communities in 30% and 35% of
418 recipients at weeks 12 and 24, respectively, representing 3.4- and 4.5-fold higher rates than in placebo

419 recipients. *L. crispatus* colonization among LBP recipients was substantially due to the LACTIN-V strain
420 CTV-05, although dominance by native *L. crispatus* strains increased over time. Microbiota trajectory
421 analysis showed that microbial communities established early frequently persisted. Achieving
422 *Lactobacillus*-dominance – particularly *L. crispatus*-dominance – at early timepoints was strongly linked
423 to persistence of *Lactobacillus*-dominance, while early non-*Lactobacillus*-dominance also tended to be
424 maintained. In exploratory post-hoc analyses, we found that the LBP's benefits differed based on baseline
425 vaginal microbiota composition, and we identified specific microbial, immune, demographic, and
426 behavioral factors associated with high-level *L. crispatus* colonization among LBP recipients.

427 To enhance resolution and power, we relied, for most of our analyses, on the observed relative
428 abundances of species, genera, or *de novo* topics rather than on *de novo* clusters or assignment to
429 reference community state types (CSTs). This approach is based on prior work showing that diverse
430 vaginal microbiotas do not exhibit well-defined clusters (Lebeer et al., 2023; Symul et al., 2023) which we
431 also observed in this cohort. We also favored taxa-based reporting to facilitate interpretation and
432 comparison with past or future studies, noting that CST definitions remain in flux as datasets expand
433 beyond predominantly U.S. populations.

434 To elucidate mechanisms by which the LBP promoted *L. crispatus* colonization, we investigated *L.*
435 *crispatus* strain dynamics using both metagenomic analysis and cultivation-based strain characterization.
436 Metagenomic strain inferences showed the CTV-05 strain accounted for the majority of *L. crispatus*
437 colonization among LBP recipients, but native *L. crispatus* strains were also observed. Interestingly,
438 among LBP recipients with ascertainable *L. crispatus* strain composition, 96.1% had CTV-05 as their
439 dominant or co-dominant strain at week 4, but this proportion progressively declined, while the number
440 of participants with dominance by one or more native *L. crispatus* strains rose from 3.9% at week 4 to
441 36.7% at week 24. Recipients with CTV-05 strain dominance at a given visit tended to either maintain
442 CTV-05 dominance or exhibit substantial loss of *L. crispatus* at the ensuing visit. However, recipients with

443 co-dominance of CTV-05 and native strains frequently experienced replacement of CTV-05 by the native
444 strain at the next visit, but very rarely exhibited replacement of the native strain by CTV-05. Further
445 research is needed to determine whether native *L. crispatus* strains that outcompeted CTV-05 have
446 functional characteristics providing selective advantages and whether initial CTV-05 colonization
447 establishes a more permissive environment for eventual transition to native *L. crispatus* strains.

448 We also examined impacts of MTZ and LBP treatment on mucosal inflammation by measuring vaginal
449 cytokines and chemokines. Significant changes in several cytokines and chemokines were observed in
450 participants for whom treatment successfully shifted microbiota composition to *Lactobacillus*-dominance,
451 and these changes reverted in participants who shifted back to non-*Lactobacillus*-dominance. Most
452 notably, inflammatory cytokines including IL-1α, IL-1β, and TNFα increased as *Lactobacillus* abundance
453 decreased, whereas IP-10, MIG, and ITAC showed the opposite pattern, consistent with prior reports
454 (Anahtar et al., 2015; Gosmann et al., 2017; Masson et al., 2019, 2014). These effects appeared primarily
455 driven by genus-level *Lactobacillus* abundance rather than a particular species. A prior study examining
456 a subset of LACTIN-V trial participants in whom strain abundance was assessed by qPCR reported that
457 IP-10 levels were lower at week 24 in participants with CTV-05 colonization than those colonized by other
458 *L. crispatus* strains (Armstrong et al., 2022), but we did not observe a similar association in our analysis
459 of the full cohort. Treatment-induced cytokine changes persisted in the LBP arm at week 24 compared
460 to placebo, but this was driven primarily by the LBP's effect in promoting higher rates of *Lactobacillus*-
461 dominance rather than a strain-specific effect of CTV-05.

462 Our *post-hoc* analysis of treatment outcomes showed LACTIN-V's incomplete efficacy in achieving high-
463 level *L. crispatus* colonization and preventing BV recurrence was linked to differences in pre-MTZ
464 microbiota composition. Stratifying participants by their predominant bacterial genus at the pre-MTZ visit
465 revealed between-group differences in achieving *L. crispatus*-dominance at week 24 and in preventing
466 rBV by both weeks 12 and 24. Most prominently, these analyses showed a particular benefit from the

467 LBP in preventing BV recurrence among participants with pre-MTZ *Prevotella*-predominance, whereas
468 participants with pre-MTZ *Ca.* Lachnocurva vaginae (BVAB1)-predominance uniquely exhibited a trend
469 toward higher BV recurrence rates with LBP treatment and little or no benefit from the LBP in achieving
470 *L. crispatus*-dominance. The reasons for these patterns are unclear, but may reflect greater competitive
471 ability or MTZ resistance in *Ca.* Lachnocurva vaginae or its co-occurring species (Alauzet et al., 2010;
472 Aldridge et al., 2001). Since *Ca.* Lachnocurva vaginae remains uncultured (Fredricks et al., 2005; Holm
473 et al., 2020), experimentally testing its antimicrobial susceptibility and ability to compete with other
474 species is currently infeasible, but our results highlight its cultivation and phenotypic characterization as
475 key research priorities. However, we emphasize that these results showing pre-MTZ microbiota effects
476 on LBP efficacy are based on exploratory analyses of a relatively small study cohort, which can introduce
477 arm imbalances through *post-hoc* stratification. They should therefore be regarded as hypothesis-
478 generating observations that require replication and validation in future studies.

479 Our integrated multiblock analysis identified both expected and unanticipated biological, clinical,
480 demographic, and behavioral correlates of *L. crispatus*-dominant colonization in LBP recipients. Low total
481 bacterial load, low vaginal pH, and low microbiota α-diversity at the post-MTZ visit were associated with
482 *L. crispatus*-dominant colonization, findings that help elucidate microbiota dynamic underlying a *post hoc*
483 analysis of clinical parameters which showed that participants achieving BV cure at the post-MTZ visit
484 had lower rates of recurrence at later timepoints (Hemmerling et al., 2024). Our analysis also showed
485 low pre-MTZ vaginal concentrations of MIG and IP-10 and high concentrations of IL-1β were also
486 associated with *L. crispatus*-dominance at week 4 – an unexpected finding given their opposite
487 correlations with *Lactobacillus* abundance during concurrent visits. The explanation for this observation
488 is unclear, but may involve a more vigorous pre-MTZ mucosal immune response helping to clear BV-
489 associated bacteria, presence of inflammatory but MTZ- or LBP-responsive bacteria in some participants
490 at baseline, or other mechanisms. After week 4, the strongest predictor of a participants' microbiota
491 category was their microbiota category at the previous timepoint, indicating that microbiota composition

492 established early after LBP treatment best predicts subsequent outcomes. Demographic data including
493 self-declared race and education levels were also linked to microbiota composition at week 4 and beyond,
494 but since race and education were closely correlated and likely associated with unmeasured variables,
495 this precluded clear mechanistic interpretation. Despite evidence for sexual activity as a significant
496 influence on BV and vaginal microbiota composition (Vodstrcil et al., 2025), we did not identify a strong
497 signal for sexual behavior as an explanatory variable for microbiota composition in LBP recipients. This
498 may be because of little variation in self-reported sexual behavior throughout the trial such that effects of
499 protected or unprotected sex may have occurred early and been obscured by one of the many factors
500 associated with early *L. crispatus* colonization. It is also recognized that self-reported sexual activity in
501 clinical trials may not be a reliable measure of actual sexual activity (Jewanraj et al., 2020; Schroder et
502 al., 2003a, 2003b; Zenilman et al., 1995), and reporting biases or inaccuracies can decrease power to
503 detect effects.

504 In summary, our detailed analysis of the vaginal microbiota effects of a single-strain *L. crispatus* LBP
505 reveals patterns of microbiota composition and strain dynamics underlying the clinical effects of LACTIN-
506 V. We identify key microbiota and host factors associated with treatment success and *L. crispatus*
507 colonization. If validated, these findings could help identify patients most likely to benefit from LBP
508 treatment and guide the discovery of novel therapeutic targets to enhance LBP efficacy and improve
509 women's health outcomes globally.

510 **Limitations**

511 This study has several limitations. First, samples from a few participants in the primary clinical trial were
512 not available for microbiome and cytokine analysis (samples were unavailable from 10 of 152 originally
513 reported participants in the LBP arm and 5 of 76 participants in the placebo arm). In addition, technical
514 failures of 16S rRNA gene sequencing precluded microbiota compositional analysis in four additional
515 samples, so our analysis cohort was slightly smaller than the primary trial cohort. Second, the limited size

516 of this Phase 2b study reduced statistical power, especially for heterogeneity and between-arm *L.*
517 *crispatus* colonization comparisons. Third, the study's relatively infrequent sampling schedule did not
518 permit high temporal resolution to assess microbiota and strain dynamics. Finally, although detailed data
519 were available regarding sexual, contraceptive, hygiene, and clinical parameters, participants were
520 surveyed on only a narrow range of environmental and sociodemographic variables, precluding detailed
521 investigation of factors that might help explain associations with microbiota composition.

522 **Methods**

523 **Study design and sample collection**

524 Samples analyzed in this study were obtained as part of a previously reported phase 2b, randomized,
525 double-blind, placebo-controlled trial of the *L. crispatus* LBP LACTIN-V (Osel, Inc., Mountain View, CA)
526 for prevention of rBV (Cohen et al., 2020). Use of samples was approved by the Mass General Brigham
527 Institutional Review Board (IRB Protocol #:2020P002237) as well as the UCSF Institutional Review Board
528 (IRB Protocol#: 19-28337). LACTIN-V is a single-strain LBP formulated as a powder containing a
529 preservation matrix and $2x10^9$ colony-forming units (CFU) per dose of the *L. crispatus* strain CTV-05,
530 which was isolated from a human vaginal sample (Cohen et al., 2020). CTV-05 is administered using a
531 pre-filled vaginal applicator, and was compared in the trial to a placebo formulation consisting of the
532 preservation matrix without CTV-05. The trial was conducted at four centers within the USA and enrolled
533 premenopausal, non-pregnant women aged 18-45 years. Eligibility criteria were previously described
534 (Cohen et al., 2020). Briefly, potential participants attended a screening (pre-MTZ) visit at which they
535 were determined to be eligible for the study if testing revealed presence of BV as determined by both a
536 Nugent score of 4-10 on Gram stain of a vaginal smear (Nugent et al., 1991) and presence of at least
537 three of four Amsel criteria (characteristic vaginal discharge, >20% clue cells on microscopy of a vaginal
538 wet prep, vaginal fluid pH >4.5, and presence of a fishy odor upon addition of 10% potassium hydroxide
539 to a vaginal specimen) (Amsel et al., 1983), as well as negative testing for HIV, syphilis, gonorrhea,
540 chlamydia, trichomonas, and urinary tract infection. Women found to be eligible based on this evaluation
541 completed 5 days of intravaginal MTZ therapy within 30 days of their pre-MTZ visit (**Figure 1A, S1A)**.
542 They then returned to the trial clinic within 48 hours of completing antibiotics (post-MTZ visit) and were
543 randomized in a 2:1 ratio to receive either LBP or placebo after providing written informed consent. The
544 first dose of LBP or placebo was clinician-administered at the randomization (post-MTZ) visit, then doses
545 were vaginally self-delivered daily for the next four days, then twice weekly for ten additional weeks. In-
546 person study visits were scheduled 4, 8, 12, and 24 weeks after randomization, at which vaginal swabs

547 were collected and stored (details below) and clinical report forms (CRFs) were completed. In addition,
548 two phone visits were planned at week 16 and 20 during which a subset of the clinical report forms were
549 filled to capture information on adverse events, menstruation, concomitant medication use, and sexual
550 behavior protected or unprotected by condoms. Participants who desired additional in-person visits were
551 invited to present to the clinics. Swabs and CRFs were collected at these additional visits.

552 Clinician-collected vaginal swab samples were obtained via speculum exam at in-person study visits.
553 Two types of swabs were collected in parallel for analysis (Cohen et al., 2020). One set of swabs was
554 collected at all in-person visits including the pre-MTZ (screening) visit using the Starplex™ Scientific
555 Multitrans™ Collection and Transportation System (Starplex™ Scientific S1600), which comprises a
556 plastic-shaft Dacron™-tipped swab stored in a glass bead-containing transport medium consisting gelatin
557 (5.0 g/L), sucrose (68.46 g/L), glutamic acid (0.70 g/L), HEPES sodium salt (3.4 g/L), modified Hank's
558 balanced salts (9.8 g/L), sodium bicarbonate (0.35 g/L), bovine serum albumin (10 g/L) and the
559 antimicrobial agents vancomycin (0.1 g/L), amphotericin B (2.5 mg/L), and colistin (0.015 g/L) at a pH of
560 7.2-7.8. The other set of swabs was collected at the post-MTZ (randomization) visit and all subsequent
561 visits in the ESwab® Liquid Based Collection and Transport System (Copan ESwab 480C®). Samples
562 were stored at room temperature for 1-4 hours after collection, then frozen at -80°C.

563 Microbiota sequencing, bacterial isolation, and cytokine analysis was performed on samples stored using
564 the Starplex™ system. Samples were thawed on ice, vortexed at maximum speed for 5 seconds, then
565 the swabs were removed from the transport media and media was divided into aliquots and re-frozen at
566 -80°C. Subsequent processing was performed as described below. Measurement of bacterial load via
567 qPCR was performed using swabs collected via the Copan ESwab® system.

568 **Bacterial isolations and cultivation**

569 Bacterial isolation and cultivation was performed under anaerobic conditions at 37°C in an AS-580
570 anaerobic chamber (Anaerobe Systems) with an atmosphere of 5% carbon dioxide, 5% hydrogen, and
571 90% nitrogen (Airgas, Inc.). All culture media was pre-reduced prior to use by being placed in the
572 anaerobic chamber overnight. Bacteria were isolated and cultivated on solid media including
573 *Lactobacillus* MRS agar (Hardy Diagnostics, #G117), Columbia Blood Agar ("CBA", Hardy Diagnostics,
574 #A16), or CDC Anaerobe Laked Sheep Blood Agar with Kanamycin and Vancomycin ("LKV", BD BBL™
575 Prepared Plated Media, #221846). Known *L. crispatus* strains and not-yet-identified isolates obtained on
576 *Lactobacillus* MRS agar were expanded by culture in liquid media consisting of *Lactobacillus* MRS broth
577 (BD #288130) prepared according to manufacturer instructions. Isolates obtained on CBA agar or LKV
578 agar were expanded in liquid media consisting of either Wilkins-Chalgren Anaerobe Broth (Thermo
579 Scientific™ Oxoid™, #CM0643B; prepared according to manufacturer instructions) or of NYCIII broth
580 (American Type Culture Collection (ATCC) medium 1685), whichever produced better growth. NYCIII
581 broth was prepared using a slightly modified version of the standard ATCC protocol (Bloom et al., 2022).
582 Pre-media consisted of 4 g/L HEPES (Fisher Scientific, #BP310-500), 15 g/L proteose peptone no. 3 (BD
583 Biosciences, #BD 211693), and 5 g/L sodium chloride in 875 ml distilled water, which was pH-adjusted
584 to 7.3 and autoclaved on liquid protocol at 121°C for 15 minutes, then cooled and stored at 4°C. One day
585 before use, complete NYCIII broth was prepared from the autoclaved, cooled pre-media by adding
586 dextrose (from a stock of 3 g per 45 ml; Fisher Chemical™, #D16-500) at 7.5% v/v, yeast extract solution
587 (Gibco, #18180-059) at 2.5% v/v, and heat-inactivated horse serum (Gibco, #26050070) at 10% v/v, then
588 sterilized by passage through a 0.22 μm vacuum filter.

589 To establish a genome sequence for the CTV-05 strain, a cryopreserved pure culture of CTV-05 was
590 obtained from Osel, Inc. The strain was streaked for isolation on *Lactobacillus* MRS agar and cultured
591 for 48 hours, then a single colony was picked into *Lactobacillus* MRS broth and incubated for 20 hours.

592 The broth culture was harvested by centrifugation to obtain bacterial pellets for genomic DNA extraction
593 and sequencing.

594 Bacterial isolations were performed from twelve selected trial samples (see **Figure 3C-E**, **S3A**) using a
595 modification of previously described methods (Bloom et al., 2022). Since samples had been collected
596 into Starplex™ transport medium containing antibiotics (see above), the samples were thawed on ice,
597 immediately diluted into pre-reduced Dulbecco's Phosphate Buffered Saline ("PBS", Millipore Sigma,
598 #D8537) at 1:12 v/v, centrifuged for 10 minutes at 10,000 rcf, then supernatant was removed. Pellets
599 were re-suspended in PBS and re-centrifuged with removal of supernatant two more times, then
600 resuspended in PBS, diluted in serial 10-fold dilutions, and 100 μl aliquots from each dilution were plated
601 evenly on MRS, LKV, and CBA agar in parallel. Plates were incubated for 7 days and multiple examples
602 of each distinct colony morphology from each sample were picked and subcultured onto solid media of
603 the same type as the source media, with an emphasis on colonies from MRS agar with characteristic *L.*
604 *crispatus* morphology (e.g., **Figure 3D**). After sub-culture for 3-7 days, colonies of the sub-cultured
605 bacteria were picked into MRS broth (for colonies isolated on MRS agar) or into both NYCIII broth and
606 Wilkens-Chalgren broth in parallel and incubated for 1-4 days, depending on growth rate. The resulting
607 liquid cultures were then cryopreserved, with aliquots of each culture centrifuged and pellets saved for
608 genomic DNA extraction and sequencing (see below).

609 **Nucleic acid extraction for short-read sequencing**

610 Total nucleic acids (TNA) extraction from cervicovaginal swabs for microbiota profiling was performed via
611 a phenol-chloroform method, which includes a previously described bead beating process to disrupt
612 bacteria[70] and modified for processing in 96-well plate format (Phenol:Chloroform:IAA, 25:24:1, pH 6.6,
613 Invitrogen, #AM9730, which has since been discontinued; Sodium Dodecyl Sulfate 20% Solution, Fisher
614 Scientific, #BP1311-200; EDTA, Invitrogen, #AM9260G; 2-Propanol, Sigma, #I9516-500ML; 3M Sodium
615 Acetate, pH 5.5, Life Technologies, #AM9740). Aliquoted samples were thawed on ice, then TNA

616 extraction was performed using 200uL of well-mixed Star media. The extracted TNA sample was eluted
617 into 80uL of TE buffer (Promega, #V6321).

618 To extract bacterial genomic DNA (gDNA) for genome sequencing of cultured bacterial strains, each
619 isolated strain was streaked on the indicated solid media and a single, clonal colony was picked into broth
620 culture and incubated in static culture under anaerobic conditions for between 18 and 120 hours
621 (depending on strain growth kinetics). Cultures were centrifuged and gDNA was extracted from the pellets
622 using a plate-based protocol including a bead beating process and combining phenol-chloroform isolation
623 (Anahtar et al., 2016) with QIAamp 96 DNA QIAcube HT kit (Qiagen, # 51331) procedures.

624 **Bacterial 16S ribosomal RNA (rRNA) gene amplification and sequencing**

625 Bacterial microbiota taxonomic composition in cervicovaginal samples was determined by sequencing
626 the V4 region of the bacterial 16S rRNA gene. The V4 region was amplified via polymerase chain reaction
627 (PCR) using the primer set 515F/806R at 200 pM each (515F primer sequence
628 5'- AATGATACGGCGACCACCGAGACGTACGTACGGTGTGCCAGCMGCCGCGGTAA-3' and
629 barcoded 806R primer sequence 5'-
630 CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXXXAGTCAGTCAGCCGGACTACHVGGGTWTCT
631 AAT-3', in which the underlined sequences in each primer represent the regions of complementarity to
632 5' and 3' ends of the V4 region of the bacterial 16S rRNA gene, respectively, and the barcode positions
633 in the 806R primer are indicated by X; IDT), with the 806R primers barcoded for multiplexing (Bloom et
634 al., 2022; Caporaso et al., 2011). PCR was performed in 25 µl reactions containing 1X Q5 reaction buffer
635 (NEB, #B9027), 0.2 mM of dNTPs (NEB, #N0447), 0.2 µM of each primer, 0.5 unit of Q5 high-fidelity
636 DNA polymerase (NEB, #M0491), and 2 µl of the TNA sample. PCR was performed in triplicate for each
637 sample in the following program: 98°C for 30 seconds, followed by 30 cycles of 98°C for 10 seconds,
638 60°C for 30 seconds, and 72°C for 20 seconds, followed by a final extension at 72°C for 2 minutes. The
639 triplicate PCR reactions for each sample were combined and amplicon production and size were

640 confirmed on an agarose gel. Negative controls with PCR-quality water (Invitrogen, #10977) as a
641 template were amplified in parallel for each primer barcode mix and assessed in parallel by gel
642 electrophoresis to confirm absence of contamination and non-specific amplification. PCR products were
643 pooled, with the amount for each sample semi-quantitatively adjusted based on its gel band intensity,
644 then purified with a QIAquick PCR purification kit (Qiagen, #28104) and quality controlled with the Qubit™
645 4 Fluorometer (Invitrogen #Q33226), and TapeStation (Agilent Technologies, 4200 TapeStation).
646 Libraries were mixed with 10% PhiX and single-end sequenced on an Illumina MiSeq using a 300-cycle
647 v2 kit (Illumina, #MS-102-2002) employing the custom Earth Microbiome Project sequencing primers
648 (Read 1 sequencing primer sequence: 5'-ACGTACGTACGGTGTGCCAGCMGCCGCGGTAA-3'; read 2
649 sequencing primer sequence: 5'-ACGTACGTACCCGGACTACHVGGGTWTCTAAT-3'; index
650 sequencing primer sequence: 5'- ATTAGAWACCCBDGTAGTCCGGCTGACTGACT-3'; IDT) (Caporaso
651 et al., 2011). Negative controls for TNA extractions and PCRs were included in each sequencing library.
652 Study samples were sequenced in a total of six libraries. Samples with read counts < 10,000 in initial
653 libraries were re-amplified and re-pooled into subsequent libraries and data from the run producing the
654 highest number of reads for each sample (if the sample was sequenced multiple times) were selected for
655 subsequent analysis. Analyzable 16S rRNA gene data was generated from a total of 1152 out of 1156
656 available trial samples, with the remaining 4 samples failing due to technical challenges with extraction,
657 amplification, or sequencing.

658 **Bacterial short-read shotgun whole genome library preparation and sequencing**

659 Shotgun metagenomic and genomic libraries were prepared following a modified protocol of Baym et. al
660 (Baym et al., 2015), using the Nextera DNA Library Preparation Kit (Illumina, #20034211) and KAPA HiFi
661 Library Amplification Kit (Kapa Biosystems, #KK2602). In brief, DNA from each sample was standardized
662 to a concentration of 1ng/mL after quantification with SYBR Green (Invitrogen, #S7653), followed by
663 simultaneous fragmentation and sequencing adaptor incorporation by mixing 1ng of DNA (1 mL) with
664 1.25 ml TD buffer and 0.25 mL TDE1 provided in the Nextera kit and incubating for 9min at 55°C.

665 Tagmented DNA fragments were amplified in PCR using the KAPA high fidelity library amplification
666 reagents, with Illumina adaptor sequences and sample barcodes incorporated in primers. PCR products
667 were pooled, purified with magnetic beads (MagBio Genomics #AC-60050) and paired-end sequenced
668 on Illumina NovaSeq X with a 300-cycle kit (Psomagen, Inc.).

669 **Bacterial 16S rRNA gene sequence processing and annotation**

670 Demultiplexing of Illumina MiSeq bacterial 16S rRNA gene sequence data was performed using QIIME
671 1 version 1.9.188 (Caporaso et al., 2010). Mapping files created in QIIME 1 format were validated using
672 validate_mapping_file.py, then sequences were demultiplexed with split_libraries_fastq.py using
673 parameter store_demultiplexed_fastq and no quality filtering or trimming, and demultiplexed sequences
674 were organized into individual fastq files using split_sequence_file_on_sample_ids.py. Sequence reads
675 were trimmed and filtered using dada2 version 1.6.0 (Callahan et al., 2016), trimming at positions 10 (left)
676 and 230 (right) using the filterAndTrim function with truncQ = 11, MaxEE = 2, and MaxN = 0. Sequences
677 were then inferred, then initial taxonomy assigned using the dada2 assignTaxonomy function with the
678 RDP training database rdp_train_set_16.fa.gz (https://www.mothur.org/wiki/RDP_reference_files).
679 Amplicon sequence variant (ASV) taxonomic assignments were refined via extensive manual review. The
680 resulting annotated sequences were analyzed in R using phyloseq version 1.30.0 (McMurdie and
681 Holmes, 2013) and custom R scripts. Sequence processing and taxonomy assignment was performed
682 blinded to information about participants' and samples' clinical and demographic characteristics,
683 treatments, and trial outcomes.

684 **Total bacterial load via qPCR**

685 Quantification of total bacterial load via qPCR was performed and reported as part of the initial LACTIN-
686 V clinical trial[36]. In brief, DNA extracted from samples stored in the Copan ESwab® system were amplified
687 using bacterial 16S rRNA gene primers targeting total bacteria (16S ribosomal DNA,
688 AGAGTTTGATCCTGGCTCAG, GCTGCCTCCCGTAGGAGT, 312bp). Bacterial concentration was

689 calculated using a standard curve based on serial dilutions of the CTV-05 strain as previously described
690 (Cohen et al., 2020).

691 **Assessing balance between arms at the pre- and post-MTZ visits.**

692 Since pre-intervention (*i.e.,* pre-MTZ or post-MTZ) microbiota composition may be associated with
693 differential microbiota composition post-intervention, imbalances between arms could lead to biases in
694 our primary outcome benefit ratio estimates. To assess whether significant imbalances in microbiota
695 composition existed in this cohort, we performed a PERMANOVA analysis, as implemented in the vegan
696 R package, to test whether the intervention arm (LBP *vs* placebo) explained significant variability in
697 microbiota β-diversity, computed using the Bray-Curtis dissimilarity on ASV relative abundances.

698 **Benefit Ratios**

699 Benefit ratios, and associated confidence intervals and *p*-value were computed using Wald's method as
700 implemented in the epitools R package (Aragon, 2004).

701 **Identification of microbiota topics from 16S rRNA gene sequencing count data**

702 Microbiota topics were identified using a modification of a previously described approach (Symul et al.,
703 2023). One main difference is that here, *Lactobacillus* topics (i.e., topics composed exclusively of
704 *Lactobacillus* species) were defined independently from non-*Lactobacillus* topics (*i.e.,* topics composed
705 exclusively of non-*Lactobacillus* species), which were identified by fitting a Latent Dirichlet Allocation
706 model (Blei et al., 2003) to the non-*Lactobacillus* ASV counts aggregated at the species level. This was
707 done to facilitate the interpretation of topic composition and, specifically, to allow for "pure" topics for the
708 most prevalent *Lactobacillus* species in this cohort (*L. crispatus*, *L. iners*, and *L. jensenii*).

709 To determine *K*, the optimal number of non-*Lactobacillus* topics, we relied on a method called "topic
710 alignment" (Fukuyama et al., 2023) which examines the robustness of topics across resolutions

711 (increasing values of *K*) and provides diagnostics scores that facilitate the identification of spurious topics.
712 We selected *K* to minimize the number of spurious topics (flagged by low coherence scores) and such
713 that the number of paths (collection of similar topics across resolution) in the alignment presented a
714 plateau (Fukuyama et al., 2023) (**Figure S2A-B**). The estimated proportions (relative abundances) of
715 non-*Lactobacillus* topics in each sample ($\hat{p}_{ki}$ where k is the topic and i is the sample) were computed by
716 multiplying the proportions estimated by the model on non-*Lactobacillus* counts ($\hat{\pi}_{ki}$) by the total non-
717 *Lactobacillus* proportions in each samples ($\hat{\Pi}_i = \sum_{v \in V} p_{vi}$ where p_{vi} is the observed proportion of ASV
718 v in sample i and V is the set of non-*Lactobacillus* ASVs, such that $\hat{p}_{ki} = \hat{\pi}_{ki}\ \hat{\Pi}_i$).

719 *Lactobacillus* topics were defined as follows: *Lactobacillus* species which reached 50% of a microbiota
720 composition in at least 10 samples made up their own topic, while the remaining *Lactobacillus* species
721 were grouped into a single topic ("Other *L.*") as their total prevalence and abundance was overall small
722 (**Figure S2C**). The composition of this topic was estimated from the species average prevalences in this
723 cohort. The estimation of proportions of *Lactobacillus* topics in each sample was straightforward: they
724 were computed from the proportions of the corresponding species in each sample.

725 **Isolate genome assemblies from short-read sequencing**

726 Paired-end short-read genomic sequencing data from bacterial isolates were processed through a quality
727 control and assembly pipeline. Raw reads were first trimmed for adapter sequences using Cutadapt v4.6
728 with the Nextera adapter sequence (CTGTCTCTTAT). Quality filtering was then performed using Sickle-
729 trim v1.33 with a quality threshold of Q20 and minimum read length of 50 bp after trimming. Read quality
730 was assessed using FastQC. De novo genome assembly was performed using Unicycler v0.5.0 in
731 standard mode with default parameters. Unicycler produces high-quality assemblies from short reads
732 alone by optimizing SPAdes assembly, followed by graph simplification. The resulting assemblies were
733 annotated using Bakta v1.9.4 with the Bakta database v5, which provides comprehensive bacterial
734 genome annotation including coding sequences, rRNAs, tRNAs, and other genomic features. Assembly

735 completeness and quality were evaluated using BUSCO v5.5 with the Bacteroidales lineage dataset
736 (bacteroidales_odb10), providing a measure of genome completeness based on conserved single-copy
737 orthologs.

738 **Strain CTV-05 long-read DNA extraction, sequencing, and completed genome assembly**

739 For long-read (Oxford Nanopore Technologies) genome sequencing of the CTV-05 strain, a bacterial
740 pellet was prepared from a clonal colony cultured in *Lactobacillus* MRS broth media as described above.
741 gDNA was extracted using a non-bead-beating protocol to minimize shearing of DNA fragments, with
742 extractions performed and sequenced using five parallel aliquots of the same original broth culture to
743 maximize yield and consistency. Since *L. crispatus* is known to have a thick cell wall containing surface-
744 layer (S-layer) proteins, cell pellets were resuspended with 5M lithium chloride (Molecular Dimensions
745 #MD2-100-43) and then washed with PBS to begin to degrade the cell wall. Using the manufacturer's
746 protocol for the MasterPureTM Gram Positive DNA Purification Kit (Biosearch Technologies
747 #MGP04100), the bacterial cells were then lysed, proteins and RNA were digested, and genomic DNA
748 (gDNA) was extracted. Extracted gDNA concentration was diluted 1:5 and measured using the Qubit™
749 4 Fluorometer, then DNA libraries were prepared using Oxford Nanopore Sequencing's Rapid Barcoding
750 Kit (Oxford Nanopore Technologies #SQK-RBK004; note that this product has been discontinued as of
751 March 2024). In brief, 400 ng of gDNA from five replicate extractions was barcoded, pooled, purified, and
752 then loaded onto the MinION Flow Cell using R9.4.1 chemistry (Oxford Nanopore Technologies #FLO-
753 MIN106D; note that this product has been discontinued as of July 2024) in the MinION Sequencing
754 Device (Oxford Nanopore Technologies #MIN-101B). To produce the maximum number of reads possible
755 using the Rapid Barcoding Kit, the sequencing run was continued for a full 72 hours.

756 Demultiplexed genomic long-reads from MinKNOW v2.2 were concatenated, and reads were separately
757 filtered for quality and length using Filtlong v0.2.1 (https://github.com/rrwick/Filtlong). Filtered reads were
758 subsampled to make 12 different read sets using Trycycler v0.5.3 (Wick et al., 2021) and duplicate reads

759 from each read set were removed using BBMap v39.00 (Bushnell, 2014). The read sets were assembled
760 using Flye v2.9.1 (Kolmogorov et al., 2019), Minipolish v0.1.3 (Wick and Holt, 2021), Raven v1.8.1 (Vaser
761 and Šikić, 2021), and Canu v2.2 (Koren et al., 2017). Contigs from different assembly methods were
762 clustered based on similarity to generate consensus sequences and improve assembly quality using
763 Trycycler. Clusters were manually chosen and Trycycler was used to reconcile the contigs within each
764 cluster. Multiple sequence alignment and read partitioning were performed on the reconciled contigs
765 using Trycyler to understand sequence variation within each cluster and assign each read to the cluster
766 it best aligned with respectively. Trycycler was then used to generate a consensus contig sequence for
767 each cluster based on the multiple sequence alignment and read partitioning steps. Medaka v1.7.2
768 (https://github.com/nanoporetech/medaka), a tool for creating consensus sequences specifically from
769 nanopore sequence data, was used to polish the Trycycler consensus sequences. The consensus
770 sequences were then combined, after which Polypolish v0.5.0 (Wick and Holt, 2022) was used to polish
771 the long-read sequence with Illumina short-read data generated from an aliquot of the same clonal culture
772 of CTV-05 (extracted and sequenced as described above). The complete CTV-05 genome was annotated
773 using Bakta v1.7.0 (Schwengers et al., 2021) and published online as part of this manuscript under NCBI
774 BioProject reference number: PRJNA1303956.

775 **Construction of *L. crispatus* core SNP database for strain analysis**

776 A GT-Pro database of biallelic core sites was constructed using 236 *L. crispatus* genomes using k-mers
777 with thresholds for SNP prevalence of 0.95 and minor allele frequencies of 0.02 with MAAST (Shi et al.,
778 2023, 2022; Shih et al., 2025). Then k-mers for samples were counted to generate a metagenotype, or a
779 matrix of the number of reads at biallelic SNP sites for a single species (Smith et al., 2022). This matrix
780 was then used by StrainFacts to infer suitable allele and relative abundance matrices that represent
781 strains across all given samples (Shih et al., 2025).

782 ***L. crispatus* strain estimation in metagenomes**

783 Raw metagenomic paired-end reads from vaginal swab samples were quality-trimmed and adapter-
784 filtered using Cutadapt (v3.9) (Martin, 2011). Human reads were depleted by aligning reads to the T2T-
785 CHM13 reference genome using HISAT2 (v2.2.1) (Kim et al., 2019). Processed reads were profiled using
786 GT-Pro (v1.0.1) against the custom SNP database (Shi et al., 2022). Resulting metagenotype profiles
787 were analyzed with StrainFacts (v0.6.0) to infer strain genotypes and abundances for *L. crispatus*,
788 excluding samples with <25% coverage of the biallelic SNPs (Smith et al., 2022). The fit was then run
789 through a cleanup step to remove low abundance strains and strains with extremely similar genotypes
790 using sfacts cleanup_fit --abundance 0.01 --dissimilarity 0.01 --discretized. Analysis was performed on
791 samples with ≥5% overall relative abundance of *L. crispatus* as determined by 16S rRNA gene
792 sequencing based on evaluations of parameters supporting accurate StrainFacts LBP strain inference
793 (Shih et al., 2025). *L. crispatus* strain inference was successful in 313 of 343 samples in which species
794 relative abundance exceeded the analysis threshold, while strain analysis was unsuccessful in the
795 remaining 30 samples due to failed or inadequately deep shotgun metagenomic sequencing. A 10%
796 strain proportional abundance was used as a threshold for confident strain identification based on prior
797 simulations and experimental validation (Shih et al., 2025).

798 **Strain analysis of isolate genomes and comparison to metagenomically inferred strain genotypes**

799 Isolate genomes were taxonomically identified using GTDB-Tk v2 (Chaumeil et al., 2022). Gene-calling
800 was performed for *L. crispatus* isolate genomes, including the completed CTV-05 genome (see above)
801 using Prodigal (Hyatt et al., 2010). Core ribosomal proteins were identified using HMMER (Eddy, 2023),
802 aligned using mafft (Katoh et al., 2019; Kuraku et al., 2013), then concatenated and a phylogenetic tree
803 was constructed using fasttree (Price et al., 2010, 2009), which was used to compute branch lengths to
804 identify isolates representing distinct strains and determine which isolates represented CTV-05 or

805 endogenous strains. GT-Pro (Shi et al., 2022) was then used to calculate genotypes from representative
806 isolate reads for comparison to metagenomically inferred strain genotypes, calculated based on Jaccard
807 similarity of the StrainFacts genotypes (Shih et al., 2025).

808 **Measurement of vaginal mucosal cytokines and chemokines.**

809 Cytokines and chemokines were measured using a previously described custom 20-plex High Sensitivity
810 Luminex Assay Kit from EMD Millipore that measures interferon gamma-induced protein (IP-10),
811 interleukin 8 (IL-8), interleukin 6 (IL-6), monokine induced by interferon gamma (MIG), interferon-
812 inducible T cell alpha chemoattractant (ITAC), interleukin 1 alpha (IL-1α), interleukin 1 beta (IL-1β),
813 macrophage inflammatory protein-1 alpha (MIP-1α), macrophage inflammatory protein-1 beta (MIP-1β),
814 macrophage inflammatory protein-3 alpha (MIP-3α), tumour necrosis factor alpha (TNFα), interleukin 21
815 (IL-21), interleukin 17 (IL-17), interferon gamma (IFNγ), interleukin 23 (IL-23), interleukin 12 (IL-12 p70),
816 interleukin 13 (IL-13), interleukin 10 (IL-10), interleukin 4 (IL-4), and interleukin 5 (IL-5) (Gosmann et al.,
817 2017). Reagents were used as supplied by the manufacturer, except the Mixed Beads solution, detection
818 antibody mixture, and Streptavidin-phycoerythrin (Strep-PE) were diluted 3-fold (1:2 volume:volume ratio)
819 for use. The Mixed Beads were diluted in Bead Diluent, while the detection antibodies and Strep-PE were
820 diluted individually in Assay Buffer. All reagents were from a single manufacturer lot to minimize batch
821 effects.

822 Samples were first pre-processed to remove interfering mucus and debris. In brief, 200 μL aliquots of
823 each vaginal swab supernatant were thawed on ice, vortexed for 10 seconds to resuspend, then
824 centrifuged at 1000 rcf for 15 minutes at 4°C to pellet mucous and cells. The maximum amount of
825 supernatant recoverable from each sample without disturbing the resulting pellet was then transferred to
826 a well of a 0.22 m PVDF filter plate (EMD Millipore #MSGVS2210), filtered by centrifugation at 2,451 rcf
827 for 1 hour at 4°C, then transferred to freezer-safe tubes and stored at -80°C until assayed. Samples were
828 assayed in 96-well plates as per manufacturer protocol with the addition of periodic sonication steps to

829 minimize bead clumping. Each plate assay included one blank background control, 7 standard serial
830 dilutions assayed in duplicate, two manufacturer-supplied Quality Control (QC) samples assayed in
831 duplicate, 72 experimental samples, and 4 biological control samples used as internal quality controls on
832 all plates. Batches of standards and QC samples were prepared from the kit on the day of the assay
833 according to manufacturer protocol. For each sample or standard, 25 μL each of Assay Buffer, Mixed
834 Beads, and sample were aliquoted into each plate well. After the addition of Mixed Beads, all incubation
835 steps were performed while the plate was protected from light. Plates were sonicated for 30 seconds at
836 room temperature in a bath sonicator, then incubated for 16-18 hours on a horizontal plate shaker at 600
837 rpm in 4°C. An additional 30 second sonication at room temperature was then performed. A handheld
838 plate magnet was applied to the base of the plate to retain the magnetic beads and excess sample and
839 reagents were removed via three consecutive washes using manufacturer-supplied 1X Wash Buffer.
840 Detection antibodies were added as per manufacturer protocol and plates were incubated for 1 hour at
841 room temperature on a horizontal plate shaker at 600 rpm. Next, Strep-PE reagent was added, followed
842 by a 30 minute incubation at room temperature on a horizontal plate shaker at 600 rpm. Plates were
843 washed 3 times as above using the magnet. After the final wash, 150 μL of sheath fluid was added to all
844 wells and a final 30 second sonication at room temperature was performed. Analyte concentrations were
845 measured using a FLEXMAP 3D instrument with Luminex xPONENT software.

846 **Cytokine and chemokine concentration transformations**

847 Cytokine and chemokine concentrations were log-transformed (log) to stabilize the mean-variance
848 relationship. Compounds with concentration values below the lower limit of quantification (LLOQ) were
849 imputed at half the LLOQ; those with concentration values above the upper limit of quantification (ULOQ)
850 were imputed at the ULOQ (**Figure S4A**).

851 Measurement of soluble molecules in swab samples can be complicated by a “size effect” phenomenon
852 whereby a component of between-swab variation in concentrations is due to technical differences in the

853 amount of material collected on each swab. In such situations, the first principal component (PC1), which
854 reflects the "size" of observations (Jolicoeur and Mosimann, 1960), can primarily be driven by the amount
855 of material on the swab rather than biological variation. Consistent with this size effect phenomenon, we
856 observed that per-sample cytokine/chemokine concentrations were highly collinear, with PC1 of the
857 standardized log-transformed cytokine concentrations accounting for over 50% of the total variance
858 (**Figure S4B**) but showed no substantial correlation with clinical and biological factors such as proportions
859 of total *Lactobacillus* or of *Lactobacillus crispatus*, or participants' contraceptives (**SI**). Such size effects
860 can be addressed by subtracting PC1 from the data. We therefore performed PC1 subtraction by re-
861 assigning scores corresponding to the 1st PC to a value of 0, then transforming the data back to its
862 original variable space using the transposed rotation matrix (**SI**). Sensitivity analyses were performed to
863 assess the impact of this transformation on findings (**SI**). Finally, due to the large uncertainty regarding
864 their distributions, cytokines with values below the LLOQ in ≥60% of the samples or with values above
865 the ULOQ in ≥30% of the samples were excluded from the analysis (**Figure S4A**).

866 **Associations between microbiota composition and cytokine profiles**

867 To quantify the association between microbiota composition and cytokine profiles, we computed the RV
868 coefficient between the microbiota composition as expressed as topic proportions and cytokine
869 transformed $\log_{10}$-concentrations. Associated *p*-value was computed with a permutation test (Bougeard
870 and Dray, 2018; Dray and Dufour, 2007). To confirm that correlation between these tables was not driven
871 by potential subject (longitudinal) effects, we also computed correlation between the two tables at each
872 visit independently. If correlations were found significant, further analyses were carried out to characterize
873 the relationship between tables. Specifically, DISTATIS (Abdi et al., 2012, 2005), as implemented in the
874 distatisR R package, was used. This method presents the advantage of estimating compromise scores
875 (and partial residual scores) directly from dissimilarity matrices such that ecological measures of

876 (dis)similarity, such as the Bray-Curtis dissimilarity, can be used for representing microbiota composition.
877 Associations with the original variables can then be assessed using correlations and displayed in
878 correlation circles as typically done with PCA, PCoA, or NMDS results. The minimum number of
879 components used for display was chosen based on the presence of an elbow in the DiSTATIS scree plot.

880 **Estimating effect heterogeneity**

881 Binary outcomes (Y: ≥50% colonization by *L. crispatus* at week 12 or 24 or absence of rBV by week 12
882 or 24) were predicted using logistic regression (logit link function) with input variables representing the
883 intervention arm (*A*: LBP or placebo), the baseline (pre-MTZ) microbiota (*V*: stratified by most prevalent
884 genus, CST, or topic proportions), and their interactions (*A:V*). In stratification analyses, strata with less
885 than two participants per arm were excluded from the analysis and visualizations. Since *Lactobacillus*
886 species other than *L. iners* had low pre-MTZ relative abundance, prior to determining fits relying on topic
887 proportions, *Lactobacillus*-dominated topics were agglomerated such that the proportion of total
888 *Lactobacillus* was considered, which improved fit convergence. The statistical significance of effect
889 heterogeneity was tested by comparing the null model (which only includes the intervention arm as
890 predictor) with the full model where pre-MTZ microbiota and interaction effects were included (analysis
891 of deviance F test). *P*-values were adjusted to account for multiple hypotheses testing using the
892 Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). For the stratified analyses, CI were
893 computed using Wilson scores due to the small sample size in each stratum. When pre-MTZ microbiota
894 composition was included as a quantitative multivariate variable, counterfactual probabilities of success
895 and associated 95% confidence intervals were predicted from the logistic regression fitted model by
896 setting the intervention arm to LBP or placebo. The predicted participant-level odd ratios were computed
897 as the ratio between predicted odds of *L. crispatus* colonization at week 12 and 24 or rBV by week 12 or
898 24 had participants been receiving LBP or the placebo.

899 **Identifying factors associated with successful *L. crispatus* colonization in intervention arm**

900 To identify demographic, clinical, behavioral, or microbiologic factors associated with successful *L.*
901 *crispatus* colonization (≥ 50%) in the intervention arm, we relied on a multiblock partial least square
902 discriminant analysis (MB-PLS-DA) (Brandolini-Bunlon et al., 2019) where the 3-category response
903 variable indicated whether (1) relative abundance of *L. crispatus* was ≥50%, (2) relative abundance of
904 total *Lactobacillus* (any species) was ≥50% but relative abundance of *L. crispatus* was <50%, or (3) the
905 relative abundance of total *Lactobacillus* was <50%. MB-PLS-DA relies on the same principles as PLS-
906 DA but allows for explanatory variables to be grouped into several thematic blocks, which enables the
907 computation of block importance indices and covariances with the response block (Bougeard et al., 2011;
908 Brandolini-Bunlon et al., 2019). The description of each block and associated variables is provided in
909 Table S6.

910 Several explanatory variables were correlated. For example, the abundances of some cytokines
911 correlated with microbiota composition (**Figure 4**), and the baseline α-diversity was lower in White
912 participants (**Figure S7C**, p-value < 0.05). We addressed these existing correlations between explanatory
913 blocks and/or variables in two different ways depending on our assumptions on the underlying correlation
914 source or cause. We either relied on nested models to evaluate the additive predictive power of specific
915 blocks (*e.g.,* demographics) that had variables correlated with microbiological blocks or used one variable
916 or one block to predict the values of another one and included the residuals instead of the observed
917 values in the model. This was indicated by the symbol (r) in the variable or block names (see **SI** for
918 details). We used this approach for the cytokine blocks whose residuals were computed using their PLS-
919 predicted abundances based on microbiota composition (**SI**). Similarly, we computed residual microbiota
920 composition, α-diversity, and pH at the previous visit such that topic proportions, α-diversity, or pH were
921 relative to those expected based on the participants' colonization status at the same visit.

Categorical variables such as race or birth control were included using one-hot encoding, and variables were standardized. To ensure convergence of the fits in cross-validation or using the bootstrap, we added Gaussian noise with very small variance ($\sigma_i^2 = 10^{-3} S_i^2$ where σ_i^2 is the variance of the Gaussian noise for variable i, and S_i^2 the empirical variance of the ith variable) to one-hot encoded variables that had few participants in some categories.

Given that our categorical response had three categories, we selected two latent components for the MB-PLS-DA models to avoid overfitting. Further, we performed cross-validation analyses (40 random 75-25% split into calibration and validation sets) and found that two latent components maximized the mean average F1 score on the validation sets for all models except for the initial phase of the placebo in which no variables were found to predict our response, leading to poor and unstable performances in cross-validation (**Figure S6B** and **S8A**). This was consistent with the presence of an elbow in the screeplots and robust to other choices of metrics such as accuracy or RMSE (**SI**). Relative cumulative block importance indices were computed by dividing cumulative block importance, the BIPC as in (Brandolini-Bunlon et al., 2019), by the total inertia of each block, which corresponds to the expected BIPC if all variables had a similar importance (**SI**).

Acknowledgments

We thank members of the Vaginal Microbiome Research Consortium for helpful discussions. Funding was obtained by D.S.K from the Gates Foundation (INV-033690). S.M.B. was partially supported by NIH grant 1K08AI171166.

Author Contributions

D.S.K., S.P.H., S.M.B., and L.S. conceptualized the research with input from J.E., C.M.M., A.H., and C.R.H.; D.S.K. and S.P.H. supervised the project; S.M.B., L.S., J.E., J.X., S.H., J.S., and S.P.H. developed and validated methods and models; S.M.B., J.X., S.H., A.S., C.M.M., T.P.P., A.K., and F.A.H.

945 generated samples and performed experiments; L.S., S.M.B., J.E., J.S., A.H., A.K., F.A.H., and S.P.H.
946 performed data analysis; L.S., S.M.B., and D.S.K. led writing of the manuscript with contributions from
947 S.P.H., J.E., J.X., S.H., C.M.M., A.H., T.P.P., A.K., F.A.H., and C.R.H.; All authors provided critical
948 feedback on methods, results, analysis, and writing.

949 **Declaration of Interests**

950 C.M.M. has a financial interest in Ancilia Biosciences, a company developing a new class of Live
951 Biotherapeutics and other bacterial products. C.M.M.'s interests were reviewed and are managed by
952 MGH and Mass General Brigham in accordance with their conflict-of-interest policies. C.M.M. serves on
953 the scientific advisory board for Concerto Bio, has served as a consultant for Scynexis and Ancilia
954 Biosciences, and has received royalties from Up to Date. C.R.C. has served as a scientific advisor for
955 Osel, Inc, and Evvy and has stock options from both. The UCSF Conflict of Interest Committee approved
956 a plan to minimize his potential conflict of interest. T.P.P. is an employee of Osel Inc, and inventor on US
957 patents US20210330721A1, US20200164007A1, and US20110066137A1 related to formulation and
958 application of LACTIN-V.

959 **Data and code availability**

960 Sequencing data and genome assemblies generated as part of this study are posted under NCBI
961 BioProject reference number: PRJNA1303956. Analysis scripts and SI are available on our publicly
962 available GitHub repository (https://github.com/lasy/IIb-LACTIN-V-omics-reanalysis/).

963 **References**

964

965 Abdi, H., Dunlop, J.P., Williams, L.J., 2009. How to compute reliability estimates and display confidence
966 and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional
967 scaling (DISTATIS). Neuroimage 45, 89–95. https://doi.org/10.1016/j.neuroimage.2008.11.008

968 Abdi, H., O'Toole, A.J., Valentin, D., Edelman, B., 2005. DISTATIS: The Analysis of Multiple Distance
969 Matrices, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern
970 Recognition (CVPR'05) - Workshops. Presented at the 2005 IEEE Computer Society Conference
971 on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, IEEE, San Diego, CA,
972 USA, pp. 42–42. https://doi.org/10.1109/CVPR.2005.445

973 Abdi, H., Williams, L.J., Valentin, D., Bennani-Dosse, M., 2012. STATIS and DISTATIS: optimum
974 multitable principal component analysis and three way metric multidimensional scaling. WIREs
975 Computational Stats 4, 124–167. https://doi.org/10.1002/wics.198

976 Alauzet, C., Mory, F., Teyssier, C., Hallage, H., Carlier, J.P., Grollier, G., Lozniewski, A., 2010.
977 Metronidazole resistance in Prevotella spp. and description of a new nim gene in Prevotella
978 baroniae. Antimicrob Agents Chemother 54, 60–64. https://doi.org/10.1128/AAC.01003-09

979 Aldridge, K.E., Ashcraft, D., Cambre, K., Pierson, C.L., Jenkins, S.G., Rosenblatt, J.E., 2001. Multicenter
980 survey of the changing in vitro antimicrobial susceptibilities of clinical isolates of Bacteroides
981 fragilis group, Prevotella, Fusobacterium, Porphyromonas, and Peptostreptococcus species.
982 Antimicrob Agents Chemother 45, 1238–1243. https://doi.org/10.1128/AAC.45.4.1238-
983 1243.2001

984 Allsworth, J.E., Peipert, J.F., 2007. Prevalence of bacterial vaginosis: 2001-2004 National Health and
985 Nutrition Examination Survey data. Obstet Gynecol 109, 114–120.
986 https://doi.org/10.1097/01.AOG.0000247627.84791.91

987 Amsel, R., Totten, P.A., Spiegel, C.A., Chen, K.C.S., Eschenbach, D., Holmes, K.K., 1983. Nonspecific
988 vaginitis: Diagnostic criteria and microbial and epidemiologic associations. The American Journal
989 of Medicine 74, 14–22. https://doi.org/10.1016/0002-9343(83)91112-9

990 Anahtar, M.N., Bowman, B.A., Kwon, D.S., 2016. Efficient Nucleic Acid Extraction and 16S rRNA Gene
991 Sequencing for Bacterial Community Characterization. J Vis Exp 53939.
992 https://doi.org/10.3791/53939

993 Anahtar, M.N., Byrne, E.H., Doherty, K.E., Bowman, B.A., Yamamoto, H.S., Soumillon, M., Padavattan,
994 N., Ismail, N., Moodley, A., Sabatini, M.E., Ghebremichael, M.S., Nusbaum, C., Huttenhower, C.,
995 Virgin, H.W., Ndung'u, T., Dong, K.L., Walker, B.D., Fichorova, R.N., Kwon, D.S., 2015.
996 Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female
997 genital tract. Immunity 42, 965–976. https://doi.org/10.1016/j.immuni.2015.04.019

998 Aragon, T.J., 2004. epitools: Epidemiology Tools. https://doi.org/10.32614/CRAN.package.epitools

999 Armstrong, E., Hemmerling, A., Miller, S., Burke, K.E., Newmann, S.J., Morris, S.R., Reno, H., Huibner,
1000 S., Kulikova, M., Nagelkerke, N., Coburn, B., Cohen, C.R., Kaul, R., 2022. Sustained effect of

1001 LACTIN-V (Lactobacillus crispatus CTV-05) on genital immunology following standard bacterial
1002 vaginosis treatment: results from a randomised, placebo-controlled trial. Lancet Microbe 3, e435–
1003 e442. https://doi.org/10.1016/S2666-5247(22)00043-X

1004 Atashili, J., Poole, C., Ndumbe, P.M., Adimora, A.A., Smith, J.S., 2008. Bacterial vaginosis and HIV
1005 acquisition: a meta-analysis of published studies. AIDS 22, 1493–1501.
1006 https://doi.org/10.1097/QAD.0b013e3283021a37

1007 Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M., Kishony, R., 2015. Inexpensive
1008 multiplexed library preparation for megabase-sized genomes. PLoS One 10, e0128036.
1009 https://doi.org/10.1371/journal.pone.0128036

1010 Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful
1011 Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological)
1012 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

1013 Bilardi, J.E., Walker, S., Temple-Smith, M., McNair, R., Mooney-Somers, J., Bellhouse, C., Fairley, C.K.,
1014 Chen, M.Y., Bradshaw, C., 2013. The burden of bacterial vaginosis: women's experience of the
1015 physical, emotional, sexual and social impact of living with recurrent bacterial vaginosis. PLoS
1016 One 8, e74378. https://doi.org/10.1371/journal.pone.0074378

1017 Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet Allocation. Journal of Machine Learning
1018 Research 30.

1019 Bloom, S.M., Mafunda, N.A., Woolston, B.M., Hayward, M.R., Frempong, J.F., Abai, A.B., Xu, J., Mitchell,
1020 A.J., Westergaard, X., Hussain, F.A., Xulu, N., Dong, M., Dong, K.L., Gumbi, T., Ceasar, F.X.,
1021 Rice, J.K., Choksi, N., Ismail, N., Ndung'u, T., Ghebremichael, M.S., Relman, D.A., Balskus, E.P.,
1022 Mitchell, C.M., Kwon, D.S., 2022. Cysteine dependence of Lactobacillus iners is a potential
1023 therapeutic target for vaginal microbiota modulation. Nat Microbiol 7, 434–450.
1024 https://doi.org/10.1038/s41564-022-01070-7

1025 Bougeard, S., Dray, S., 2018. Supervised Multiblock Analysis in *R* with the **ade4** Package. J. Stat. Soft.
1026 86. https://doi.org/10.18637/jss.v086.i01

1027 Bougeard, S., Qannari, E.M., Rose, N., 2011. Multiblock redundancy analysis: interpretation tools and
1028 application in epidemiology. J. Chemometrics 25, 467–475. https://doi.org/10.1002/cem.1392

1029 Bradshaw, C.S., Brotman, R.M., 2015. Making inroads into improving treatment of bacterial vaginosis -
1030 striving for long-term cure. BMC Infect Dis 15, 292. https://doi.org/10.1186/s12879-015-1027-4

1031 Bradshaw, C.S., Morton, A.N., Hocking, J., Garland, S.M., Morris, M.B., Moss, L.M., Horvath, L.B.,
1032 Kuzevska, I., Fairley, C.K., 2006. High recurrence rates of bacterial vaginosis over the course of
1033 12 months after oral metronidazole therapy and factors associated with recurrence. J Infect Dis
1034 193, 1478–1486. https://doi.org/10.1086/503780

1035 Bradshaw, C.S., Plummer, E.L., Muzny, C.A., Mitchell, C.M., Fredricks, D.N., Herbst-Kralovetz, M.M.,
1036 Vodstrcil, L.A., 2025. Bacterial vaginosis. Nat Rev Dis Primers 11, 43.
1037 https://doi.org/10.1038/s41572-025-00626-1

1038 Bradshaw, C.S., Sobel, J.D., 2016. Current Treatment of Bacterial Vaginosis—Limitations and Need for
1039 Innovation. J Infect Dis. 214, S14–S20. https://doi.org/10.1093/infdis/jiw159

1040 Brandolini-Bunlon, M., Pétéra, M., Gaudreau, P., Comte, B., Bougeard, S., Pujos-Guillot, E., 2019. Multi-
1041 block PLS discriminant analysis for the joint analysis of metabolomic and epidemiological data.
1042 Metabolomics 15, 134. https://doi.org/10.1007/s11306-019-1598-y

1043 Brusselaers, N., Shrestha, S., van de Wijgert, J., Verstraelen, H., 2019. Vaginal dysbiosis and the risk of
1044 human papillomavirus and cervical cancer: systematic review and meta-analysis. Am J Obstet
1045 Gynecol 221, 9-18.e8. https://doi.org/10.1016/j.ajog.2018.12.011

1046 Brusselmans, J., De Sutter, A., Devleesschauwer, B., Verstraelen, H., Cools, P., 2023. Scoping review
1047 of the association between bacterial vaginosis and emotional, sexual and social health. BMC
1048 Womens Health 23, 168. https://doi.org/10.1186/s12905-023-02260-z

1049 Bushnell, B., 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner, in: Report Number: LBNL-7065E.
1050 Research Org.: Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

1051 Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2:
1052 High-resolution sample inference from Illumina amplicon data. Nat Methods 13, 581–583.
1053 https://doi.org/10.1038/nmeth.3869

1054 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N.,
1055 Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley,
1056 R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R.,
1057 Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010.
1058 QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7, 335–336.
1059 https://doi.org/10.1038/nmeth.f.303

1060 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer,
1061 N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences
1062 per sample. Proc Natl Acad Sci U S A 108 Suppl 1, 4516–4522.
1063 https://doi.org/10.1073/pnas.1000080107

1064 Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2022. GTDB-Tk v2: memory friendly
1065 classification with the genome taxonomy database. Bioinformatics 38, 5315–5316.
1066 https://doi.org/10.1093/bioinformatics/btac672

1067 Chetty, C., Mafunda, N.A., Happel, A.-U., Khan, A., Demidkina, B.C., Yende-Zuma, N., Saidi, Y., Polliah,
1068 A.M., Lewis, L., Osman, F., Radebe, P., Passmore, J.-A.S., Kwon, D.S., Ravel, J., Ngcapu, S.,
1069 Liebenberg, L., Symul, L., Holmes, S., Mitchell, C.M., Potloane, D., 2025. Randomized Trial of
1070 Multi-Strain Lactobacillus crispatus Vaginal Live Biotherapeutic Products after Antibiotic Therapy
1071 for Bacterial Vaginosis: Study Protocol for VIBRANT (Vaginal Live Biotherapeutic RANdomized
1072 Trial). https://doi.org/10.2139/ssrn.5261044

1073 Cohen, C.R., Wierzbicki, M.R., French, A.L., Morris, S., Newmann, S., Reno, H., Green, L., Miller, S.,
1074 Powell, J., Parks, T., Hemmerling, A., 2020. Randomized Trial of Lactin-V to Prevent Recurrence
1075 of Bacterial Vaginosis. N Engl J Med 382, 1906–1915. https://doi.org/10.1056/NEJMoa1915254

1076 Colbert, L.E., El Alam, M.B., Wang, R., Karpinets, T., Lo, D., Lynn, E.J., Harris, T.A., Elnaggar, J.H.,
1077 Yoshida-Court, K., Tomasic, K., Bronk, J.K., Sammouri, J., Yanamandra, A.V., Olvera, A.V.,
1078 Carlin, L.G., Sims, T., Delgado Medrano, A.Y., Napravnik, T.C., O'Hara, M., Lin, D., Abana, C.O.,
1079 Li, H.X., Eifel, P.J., Jhingran, A., Joyner, M., Lin, L., Ramondetta, L.M., Futreal, A.M., Schmeler,
1080 K.M., Mathew, G., Dorta-Estremera, S., Zhang, J., Wu, X., Ajami, N.J., Wong, M., Taniguchi, C.,

1081 Petrosino, J.F., Sastry, K.J., Okhuysen, P.C., Martinez, S.A., Tan, L., Mahmud, I., Lorenzi, P.L.,
1082 Wargo, J.A., Klopp, A.H., 2023. Tumor-resident Lactobacillus iners confer chemoradiation
1083 resistance through lactate-induced metabolic rewiring. Cancer Cell 41, 1945-1962.e11.
1084 https://doi.org/10.1016/j.ccell.2023.09.012

1085 DiGiulio, D.B., Callahan, B.J., McMurdie, P.J., Costello, E.K., Lyell, D.J., Robaczewska, A., Sun, C.L.,
1086 Goltsman, D.S.A., Wong, R.J., Shaw, G., Stevenson, D.K., Holmes, S.P., Relman, D.A., 2015.
1087 Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci
1088 USA 112, 11060–11065. https://doi.org/10.1073/pnas.1502875112

1089 Dray, S., Dufour, A.-B., 2007. The **ade4** Package: Implementing the Duality Diagram for Ecologists. J.
1090 Stat. Soft. 22. https://doi.org/10.18637/jss.v022.i04

1091 Eddy, S.R., 2023. HMMER.

1092 Ferris, M.J., Norori, J., Zozaya-Hinchliffe, M., Martin, D.H., 2007. Cultivation-independent analysis of
1093 changes in bacterial vaginosis flora following metronidazole treatment. J Clin Microbiol 45, 1016–
1094 1018. https://doi.org/10.1128/JCM.02085-06

1095 France, M.T., Ma, B., Gajer, P., Brown, S., Humphrys, M.S., Holm, J.B., Waetjen, L.E., Brotman, R.M.,
1096 Ravel, J., 2020. VALENCIA: a nearest centroid classification method for vaginal microbial
1097 communities based on composition. Microbiome 8, 166. https://doi.org/10.1186/s40168-020-
1098 00934-6

1099 Fredricks, D.N., Fiedler, T.L., Marrazzo, J.M., 2005. Molecular Identification of Bacteria Associated with
1100 Bacterial Vaginosis. N Engl J Med 353, 1899–1911. https://doi.org/10.1056/nejmoa043802

1101 Fukuyama, J., Sankaran, K., Symul, L., 2023. Multiscale analysis of count data through topic alignment.
1102 Biostatistics 24, 1045–1065. https://doi.org/10.1093/biostatistics/kxac018

1103 Gosmann, C., Anahtar, M.N., Handley, S.A., Farcasanu, M., Abu-Ali, G., Bowman, B.A., Padavattan, N.,
1104 Desai, C., Droit, L., Moodley, A., Dong, M., Chen, Y., Ismail, N., Ndung'u, T., Ghebremichael,
1105 M.S., Wesemann, D.R., Mitchell, C., Dong, K.L., Huttenhower, C., Walker, B.D., Virgin, H.W.,
1106 Kwon, D.S., 2017. Lactobacillus-Deficient Cervicovaginal Bacterial Communities Are Associated
1107 with Increased HIV Acquisition in Young South African Women. Immunity 46, 29–37.
1108 https://doi.org/10.1016/j.immuni.2016.12.013

1109 Hemmerling, A., Harrison, W., Schroeder, A., Park, J., Korn, A., Shiboski, S., Cohen, C.R., 2009. Phase
1110 1 dose-ranging safety trial of Lactobacillus crispatus CTV-05 for the prevention of bacterial
1111 vaginosis. Sex Transm Dis 36, 564–569. https://doi.org/10.1097/OLQ.0b013e3181a74924

1112 Hemmerling, A., Wierzbicki, M.R., Armstrong, E., Cohen, C.R., 2024. Response to Antibiotic Treatment
1113 of Bacterial Vaginosis Predicts the Effectiveness of LACTIN-V (Lactobacillus crispatus CTV-05)
1114 in the Prevention of Recurrent Disease. Sex Transm Dis 51, 437–440.
1115 https://doi.org/10.1097/OLQ.0000000000001962

1116 Hillier, S.L., Martius, J., Krohn, M., Kiviat, N., Holmes, K.K., Eschenbach, D.A., 1988. A case-control
1117 study of chorioamnionic infection and histologic chorioamnionitis in prematurity. N Engl J Med
1118 319, 972–978. https://doi.org/10.1056/NEJM198810133191503

1119 Holm, J.B., France, M.T., Ma, B., McComb, E., Robinson, C.K., Mehta, A., Tallon, L.J., Brotman, R.M.,
1120 Ravel, J., 2020. Comparative Metagenome-Assembled Genome Analysis of “Candidatus
1121 Lachnocurva vaginae”, Formerly Known as Bacterial Vaginosis-Associated Bacterium-1
1122 (BVAB1). Front Cell Infect Microbiol 10, 117. https://doi.org/10.3389/fcimb.2020.00117

1123 Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal:
1124 prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11,
1125 119. https://doi.org/10.1186/1471-2105-11-119

1126 Jewanraj, J., Ngcapu, S., Osman, F., Mtshali, A., Singh, R., Mansoor, L.E., Abdool Karim, S.S., Abdool
1127 Karim, Q., Passmore, J.-A.S., Liebenberg, L.J.P., 2020. The Impact of Semen Exposure on the
1128 Immune and Microbial Environments of the Female Genital Tract. Front. Reprod. Health 2,
1129 566559. https://doi.org/10.3389/frph.2020.566559

1130 Joag, V., Obila, O., Gajer, P., Scott, M.C., Dizzell, S., Humphrys, M., Shahabi, K., Huibner, S., Shannon,
1131 B., Tharao, W., Mureithi, M., Oyugi, J., Kimani, J., Kaushic, C., Ravel, J., Anzala, O., Kaul, R.,
1132 2019. Impact of Standard Bacterial Vaginosis Treatment on the Genital Microbiota, Immune
1133 Milieu, and Ex Vivo Human Immunodeficiency Virus Susceptibility. Clinical Infectious Diseases
1134 68, 1675–1683. https://doi.org/10.1093/cid/ciy762

1135 Jolicoeur, P., Mosimann, J.E., 1960. Size and Shape Variation in The Painted Turtle. A Principal
1136 Component Analysis. Growth.

1137 Josse, J., Holmes, S., 2016. Measuring multivariate association and beyond. Statist. Surv. 10.
1138 https://doi.org/10.1214/16-SS116

1139 Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment,
1140 interactive sequence choice and visualization. Briefings in Bioinformatics 20, 1160–1166.
1141 https://doi.org/10.1093/bib/bbx108

1142 Kenyon, C., Colebunders, R., Crucitti, T., 2013. The global epidemiology of bacterial vaginosis: a
1143 systematic review. Am J Obstet Gynecol 209, 505–523.
1144 https://doi.org/10.1016/j.ajog.2013.05.006

1145 Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and
1146 genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907–915.
1147 https://doi.org/10.1038/s41587-019-0201-4

1148 Kindinger, L.M., Bennett, P.R., Lee, Y.S., Marchesi, J.R., Smith, A., Cacciatore, S., Holmes, E.,
1149 Nicholson, J.K., Teoh, T.G., MacIntyre, D.A., 2017. The interaction between vaginal microbiota,
1150 cervical length, and vaginal progesterone treatment for preterm birth risk. Microbiome 5, 6.
1151 https://doi.org/10.1186/s40168-016-0223-9

1152 Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat
1153 graphs. Nat Biotechnol 37, 540–546. https://doi.org/10.1038/s41587-019-0072-8

1154 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and
1155 accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. Genome Res.
1156 27, 722–736. https://doi.org/10.1101/gr.215087.116

1157 Kuraku, S., Zmasek, C.M., Nishimura, O., Katoh, K., 2013. aLeaves facilitates on-demand exploration of
1158 metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity.
1159 Nucleic Acids Research 41, W22–W28. https://doi.org/10.1093/nar/gkt389

1160 Lebeer, S., Ahannach, S., Gehrmann, T., Wittouck, S., Eilers, T., Oerlemans, E., Condori, S., Dillen, J.,
1161 Spacova, I., Vander Donck, L., Masquillier, C., Allonsius, C.N., Bron, P.A., Van Beeck, W., De
1162 Backer, C., Donders, G., Verhoeven, V., 2023. A citizen-science-enabled catalogue of the vaginal
1163 microbiome and associated factors. Nat Microbiol 8, 2183–2195. https://doi.org/10.1038/s41564-
1164 023-01500-0

1165 Leitich, H., Kiss, H., 2007. Asymptomatic bacterial vaginosis and intermediate flora as risk factors for
1166 adverse pregnancy outcome. Best Pract Res Clin Obstet Gynaecol 21, 375–390.
1167 https://doi.org/10.1016/j.bpobgyn.2006.12.005

1168 Lev-Sagie, A., Goldman-Wohl, D., Cohen, Y., Dori-Bachash, M., Leshem, A., Mor, U., Strahilevitz, J.,
1169 Moses, A.E., Shapiro, H., Yagel, S., Elinav, E., 2019. Vaginal microbiome transplantation in
1170 women with intractable bacterial vaginosis. Nat Med 25, 1500–1504.
1171 https://doi.org/10.1038/s41591-019-0600-6

1172 Marconi, C., Duarte, M.T.C., Silva, D.C., Silva, M.G., 2015. Prevalence of and risk factors for bacterial
1173 vaginosis among women of reproductive age attending cervical screening in southeastern Brazil.
1174 Int J Gynaecol Obstet 131, 137–141. https://doi.org/10.1016/j.ijgo.2015.05.016

1175 Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet
1176 j. 17, 10. https://doi.org/10.14806/ej.17.1.200

1177 Masson, L., Barnabas, S., Deese, J., Lennard, K., Dabee, S., Gamieldien, H., Jaumdally, S.Z.,
1178 Williamson, A.-L., Little, F., Van Damme, L., Ahmed, K., Crucitti, T., Abdellati, S., Bekker, L.-G.,
1179 Gray, G., Dietrich, J., Jaspan, H., Passmore, J.-A.S., 2019. Inflammatory cytokine biomarkers of
1180 asymptomatic sexually transmitted infections and vaginal dysbiosis: a multicentre validation
1181 study. Sex Transm Infect 95, 5–12. https://doi.org/10.1136/sextrans-2017-053506

1182 Masson, L., Mlisana, K., Little, F., Werner, L., Mkhize, N.N., Ronacher, K., Gamieldien, H., Williamson,
1183 C., Mckinnon, L.R., Walzl, G., Abdool Karim, Q., Abdool Karim, S.S., Passmore, J.-A.S., 2014.
1184 Defining genital tract cytokine signatures of sexually transmitted infections and bacterial vaginosis
1185 in women at high risk of HIV infection: a cross-sectional study. Sex Transm Infect 90, 580–587.
1186 https://doi.org/10.1136/sextrans-2014-051601

1187 McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and
1188 graphics of microbiome census data. PLoS One 8, e61217.
1189 https://doi.org/10.1371/journal.pone.0061217

1190 Mitchell, C., Manhart, L.E., Thomas, K., Fiedler, T., Fredricks, D.N., Marrazzo, J., 2012. Behavioral
1191 predictors of colonization with Lactobacillus crispatus or Lactobacillus jensenii after treatment for
1192 bacterial vaginosis: a cohort study. Infect Dis Obstet Gynecol 2012, 706540.
1193 https://doi.org/10.1155/2012/706540

1194 Munoz, A., Hayward, M.R., Bloom, S.M., Rocafort, M., Ngcapu, S., Mafunda, N.A., Xu, J., Xulu, N., Dong,
1195 M., Dong, K.L., Ismail, N., Ndung'u, T., Ghebremichael, M.S., Kwon, D.S., 2021. Modeling the
1196 temporal dynamics of cervicovaginal microbiota identifies targets that may promote reproductive
1197 health. Microbiome 9, 163. https://doi.org/10.1186/s40168-021-01096-9

1198 Nilsen, T., Swedek, I., Lagenaur, L.A., Parks, T.P., 2020. Novel Selective Inhibition of Lactobacillus iners
1199 by Lactobacillus-Derived Bacteriocins. Appl Environ Microbiol 86, e01594-20.
1200 https://doi.org/10.1128/AEM.01594-20

1201 Norenhag, J., Du, J., Olovsson, M., Verstraelen, H., Engstrand, L., Brusselaers, N., 2020. The vaginal
1202 microbiota, human papillomavirus and cervical dysplasia: a systematic review and network meta-
1203 analysis. BJOG 127, 171–180. https://doi.org/10.1111/1471-0528.15854

1204 Nugent, R.P., Krohn, M.A., Hillier, S.L., 1991. Reliability of diagnosing bacterial vaginosis is improved by
1205 a standardized method of gram stain interpretation. J Clin Microbiol 29, 297–301.
1206 https://doi.org/10.1128/jcm.29.2.297-301.1991

1207 Peebles, K., Velloza, J., Balkus, J.E., McClelland, R.S., Barnabas, R.V., 2019. High Global Burden and
1208 Costs of Bacterial Vaginosis: A Systematic Review and Meta-Analysis. Sexual Trans Dis 46, 304–
1209 311. https://doi.org/10.1097/OLQ.0000000000000972

1210 Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for
1211 Large Alignments. PLoS ONE 5, e9490. https://doi.org/10.1371/journal.pone.0009490

1212 Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: Computing Large Minimum Evolution Trees with
1213 Profiles instead of a Distance Matrix. Molecular Biology and Evolution 26, 1641–1650.
1214 https://doi.org/10.1093/molbev/msp077

1215 Ravel, J., Brotman, R.M., Gajer, P., Ma, B., Nandy, M., Fadrosh, D.W., Sakamoto, J., Koenig, S.S., Fu,
1216 L., Zhou, X., Hickey, R.J., Schwebke, J.R., Forney, L.J., 2013. Daily temporal dynamics of vaginal
1217 microbiota before, during and after episodes of bacterial vaginosis. Microbiome 1, 29.
1218 https://doi.org/10.1186/2049-2618-1-29

1219 Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S.K., McCulle, S.L., Karlebach, S., Gorle, R.,
1220 Russell, J., Tacket, C.O., Brotman, R.M., Davis, C.C., Ault, K., Peralta, L., Forney, L.J., 2011.
1221 Vaginal microbiome of reproductive-age women. Proceedings of the National Academy of
1222 Sciences 108, 4680–4687. https://doi.org/10.1073/pnas.1002611107

1223 Ravel, J., Simmons, S., Jaswa, E.G., Gottfried, S., Greene, M., Kellogg-Spadt, S., Gevers, D., Harper,
1224 D.M., 2025. Impact of a multi-strain L. crispatus-based vaginal synbiotic on the vaginal
1225 microbiome: a randomized placebo-controlled trial. NPJ Biofilms Microbiomes 11, 158.
1226 https://doi.org/10.1038/s41522-025-00788-6

1227 Robert, P., Escoufier, Y., 1976. A Unifying Tool for Linear Multivariate Statistical Methods: The RV-
1228 Coefficient. Journal of the Royal Statistical Society. Series C (Applied Statistics) 25, 257–265.
1229 https://doi.org/10.2307/2347233

1230 Sankaran, K., Holmes, S.P., 2019. Latent variable modeling for the microbiome. Biostatistics 20, 599–
1231 614. https://doi.org/10.1093/biostatistics/kxy018

1232 Schroder, K.E.E., Carey, M.P., Vanable, P.A., 2003a. Methodological challenges in research on sexual
1233 risk behavior: I. Item content, scaling, and data analytical options. Annals of Behavioral Medicine
1234 26, 76–103. https://doi.org/10.1207/S15324796ABM2602_02

Schroder, K.E.E., Carey, M.P., Vanable, P.A., 2003b. Methodological challenges in research on sexual risk behavior: II. Accuracy of self-reports. Annals of Behavioral Medicine 26, 104–123. https://doi.org/10.1207/S15324796ABM2602_03

Schwebke, J.R., Lensing, S.Y., Lee, J., Muzny, C.A., Pontius, A., Woznicki, N., Aguin, T., Sobel, J.D., 2021. Treatment of Male Sexual Partners of Women With Bacterial Vaginosis: A Randomized, Double-Blind, Placebo-Controlled Trial. Clin Infect Dis 73, e672–e679. https://doi.org/10.1093/cid/ciaa1903

Schwengers, O., Jelonek, L., Dieckmann, M.A., Beyvers, S., Blom, J., Goesmann, A., 2021. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification: Find out more about Bakta, the motivation, challenges and applications, here. Microbial Genomics 7. https://doi.org/10.1099/mgen.0.000685

Shi, Z.J., Dimitrov, B., Zhao, C., Nayfach, S., Pollard, K.S., 2022. Fast and accurate metagenotyping of the human gut microbiome with GT-Pro. Nat Biotechnol 40, 507–516. https://doi.org/10.1038/s41587-021-01102-3

Shi, Z.J., Nayfach, S., Pollard, K.S., 2023. Maast: genotyping thousands of microbial strains efficiently. Genome Biol 24, 186. https://doi.org/10.1186/s13059-023-03030-8

Shih, J., Bloom, S.M., Xu, J., Mitchell, C.M., Elsherbini, J., Kwon, D.S., 2025. StrainFacts accurately quantifies both endogenous and live biotherapeutic product strain abundances in simulated and clinical vaginal microbiota samples. bioRxiv. https://doi.org/10.1101/2025.08.15.670563

Smith, B.J., Li, X., Shi, Z.J., Abate, A., Pollard, K.S., 2022. Scalable Microbial Strain Inference in Metagenomic Data Using StrainFacts. Frontiers in Bioinformatics 2.

Srinivasan, S., Liu, C., Mitchell, C.M., Fiedler, T.L., Thomas, K.K., Agnew, K.J., Marrazzo, J.M., Fredricks, D.N., 2010. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. PLoS One 5, e10197. https://doi.org/10.1371/journal.pone.0010197

Symul, L., Jeganathan, P., Costello, E.K., France, M., Bloom, S.M., Kwon, D.S., Ravel, J., Relman, D.A., Holmes, S., 2023. Sub-communities of the vaginal microbiota in pregnant and non-pregnant women. Proceedings of the Royal Society B: Biological Sciences 290, 20231461. https://doi.org/10.1098/rspb.2023.1461

Tamarelle, J., Shardell, M.D., Ravel, J., Brotman, R.M., 2022. Factors Associated With Incidence and Spontaneous Clearance of Molecular-Bacterial Vaginosis: Results From a Longitudinal Frequent-Sampling Observational Study. Sex Transm Dis 49, 649–656. https://doi.org/10.1097/OLQ.0000000000001662

van Houdt, R., Ma, B., Bruisten, S.M., Speksnijder, A.G.C.L., Ravel, J., de Vries, H.J.C., 2018. Lactobacillus iners-dominated vaginal microbiota is associated with increased susceptibility to Chlamydia trachomatis infection in Dutch women: a case-control study. Sex Transm Infect 94, 117–123. https://doi.org/10.1136/sextrans-2017-053133

Vaser, R., Šikić, M., 2021. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci 1, 332–336. https://doi.org/10.1038/s43588-021-00073-4

1273 Verwijs, M.C., Agaba, S.K., Darby, A.C., Van De Wijgert, J.H.H.M., 2020. Impact of oral metronidazole
1274 treatment on the vaginal microbiota and correlates of treatment failure. American Journal of
1275 Obstetrics and Gynecology 222, 157.e1-157.e13. https://doi.org/10.1016/j.ajog.2019.08.008

1276 Vodstrcil, L.A., Plummer, E.L., Fairley, C.K., Hocking, J.S., Law, M.G., Petoumenos, K., Bateson, D.,
1277 Murray, G.L., Donovan, B., Chow, E.P.F., Chen, M.Y., Kaldor, J., Bradshaw, C.S., 2025. Male-
1278 Partner Treatment to Prevent Recurrence of Bacterial Vaginosis. N Engl J Med 392, 947–957.
1279 https://doi.org/10.1056/NEJMoa2405404

1280 Wick, R.R., Holt, K.E., 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies.
1281 PLoS Comput Biol 18, e1009802. https://doi.org/10.1371/journal.pcbi.1009802

1282 Wick, R.R., Holt, K.E., 2021. Benchmarking of long-read assemblers for prokaryote whole genome
1283 sequencing. F1000Res 8, 2138. https://doi.org/10.12688/f1000research.21782.4

1284 Wick, R.R., Judd, L.M., Cerdeira, L.T., Hawkey, J., Méric, G., Vezina, B., Wyres, K.L., Holt, K.E., 2021.
1285 Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biol 22, 266.
1286 https://doi.org/10.1186/s13059-021-02483-z

1287 Workowski, K.A., Bachmann, L.H., Chan, P.A., Johnston, C.M., Muzny, C.A., Park, I., Reno, H.,
1288 Zenilman, J.M., Bolan, G.A., 2021. Sexually Transmitted Infections Treatment Guidelines, 2021.
1289 MMWR Recomm Rep 70, 1–187. https://doi.org/10.15585/mmwr.rr7004a1

1290 Yockey, L.J., Hussain, F.A., Bergerat, A., Reissis, A., Worrall, D., Xu, J., Gomez, I., Bloom, S.M.,
1291 Mafunda, N.A., Kelly, J., Kwon, D.S., Mitchell, C.M., 2022. Screening and characterization of
1292 vaginal fluid donations for vaginal microbiota transplantation. Sci Rep 12, 17948.
1293 https://doi.org/10.1038/s41598-022-22873-y

1294 Zenilman, J.M., Weisman, C.S., Rompalo, A.M., Ellish, N., Upchurch, D.M., Hook, E.W.I., Celentano, D.,
1295 1995. Condom Use to Prevent Incident STDs: The Validity of Self-Reported Condom Use.
1296 Sexually Transmitted Diseases 22, 15.

1297 Zhu, M., Frank, M.W., Radka, C.D., Jeanfavre, S., Xu, J., Tse, M.W., Pacheco, J.A., Kim, J.S., Pierce,
1298 K., Deik, A., Hussain, F.A., Elsherbini, J., Hussain, S., Xulu, N., Khan, N., Pillay, V., Mitchell, C.M.,
1299 Dong, K.L., Ndung'u, T., Clish, C.B., Rock, C.O., Blainey, P.C., Bloom, S.M., Kwon, D.S., 2024.
1300 Vaginal Lactobacillus fatty acid response mechanisms reveal a metabolite-targeted strategy for
1301 bacterial vaginosis treatment. Cell 187, 5413-5430.e29. https://doi.org/10.1016/j.cell.2024.07.029
1302

1303

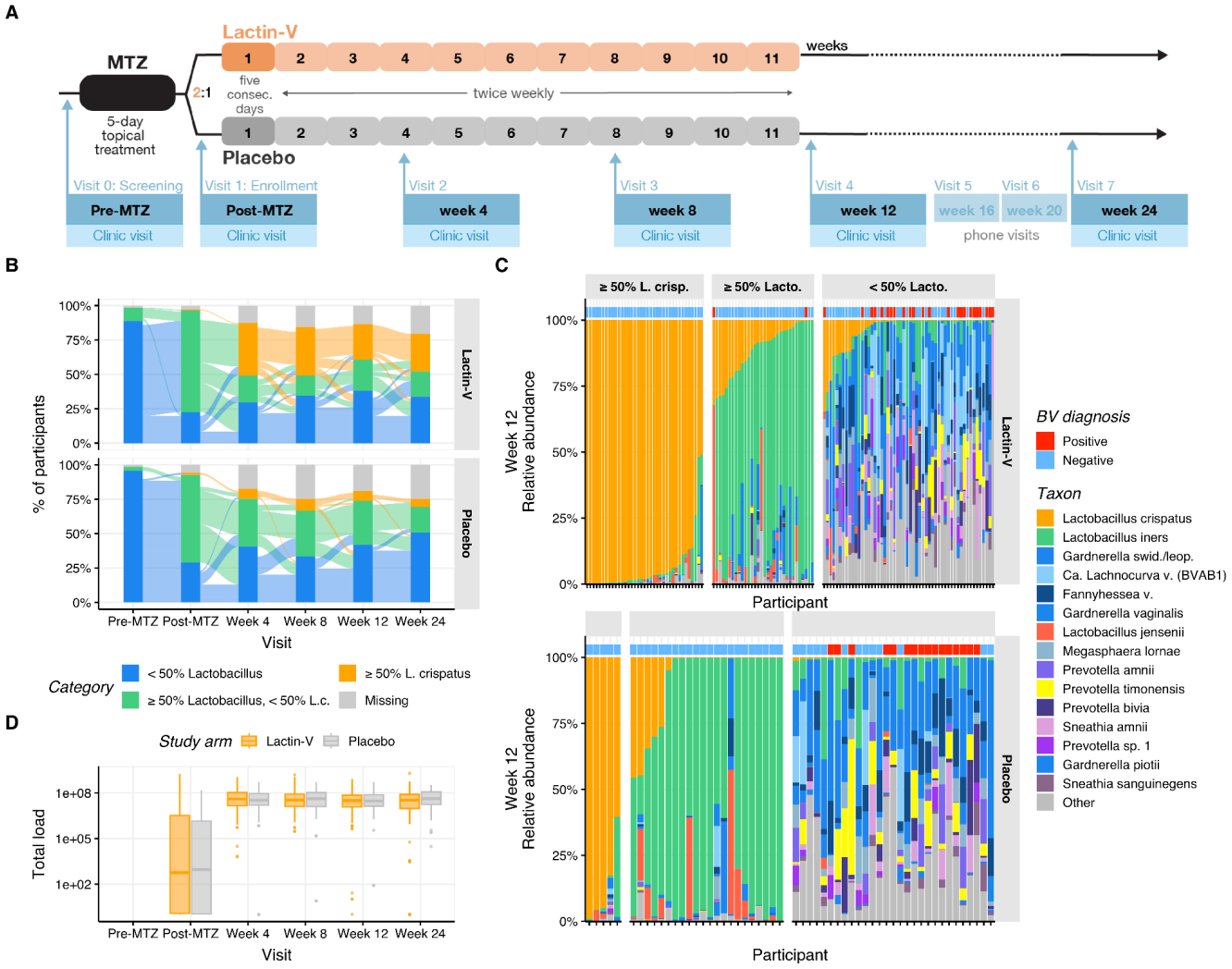
1304 **Main Tables and Figures**

1305 **Table 1: Benefit ratios for LBP microbiota effects**

Taxon	Week	Microbiota Endpoint	LACTIN-V	Placebo	Benefit Ratio (95% CI)	P Value
L. crispatus	Week 12	≥ 50% *L.c.*	37 (30%)	5 (9%)	3.37 (1.40 - 8.11)	< 0.005
		< 50% *L.c.*	86 (70%)	51 (91%)		
	Week 24	≥ 50% *L.c.*	39 (35%)	4 (8%)	4.49 (1.69 - 11.90)	
		< 50% *L.c.*	74 (65%)	48 (92%)		
Lactobacillus	Week 12	≥ 50% *Lacto*	69 (56%)	27 (48%)	1.16 (0.85 - 1.59)	
		< 50% *Lacto*	54 (44%)	29 (52%)		
	Week 24	≥ 50% *Lacto*	65 (58%)	17 (33%)	1.76 (1.15 - 2.68)	
		< 50% *Lacto*	48 (42%)	35 (67%)		

1306 Benefit Ratios and associated 95% confidence intervals were computed using Wald's method.

1307 **Figure 1: LACTIN-V trial design and microbiota endpoints**



1308

1309 A. Schematic of LACTIN-V trial study design and sampling visit timing.

1310 B. Participants were grouped into three microbiota categories: *L. crispatus*-dominant colonization (≥

1311 50% *L. crispatus* relative abundance), *Lactobacillus* (non-*crispatus*)-dominant colonization (≥

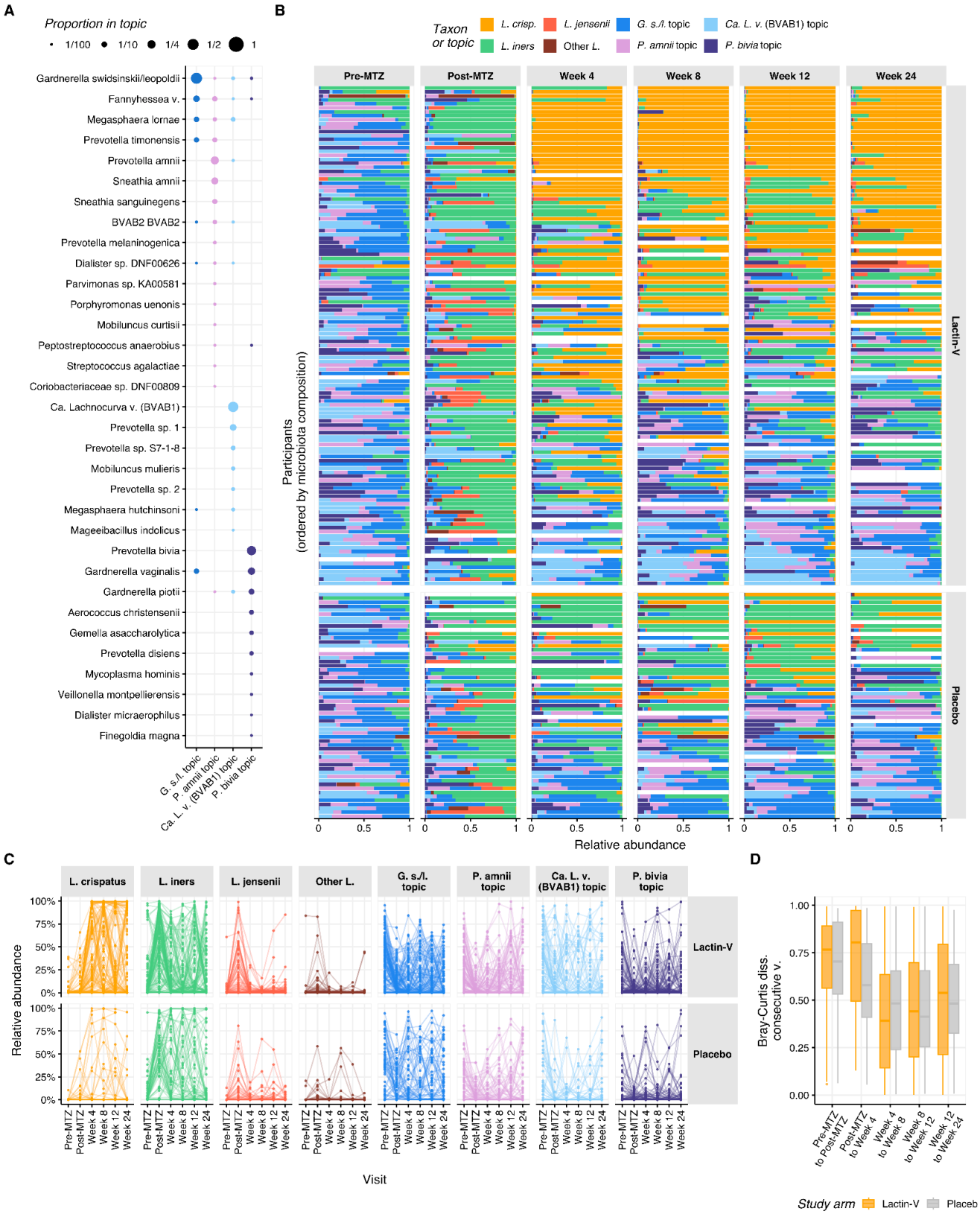
1312 50% *Lactobacillus,* < 50% *L. crispatus*), or non-*Lactobacillus*-dominant (< 50% *Lactobacillus*)

1313 based on results of bacterial 16S rRNA gene sequencing. The Sankey diagram displays how

1314 participants transitioned between these three categories throughout the trial.

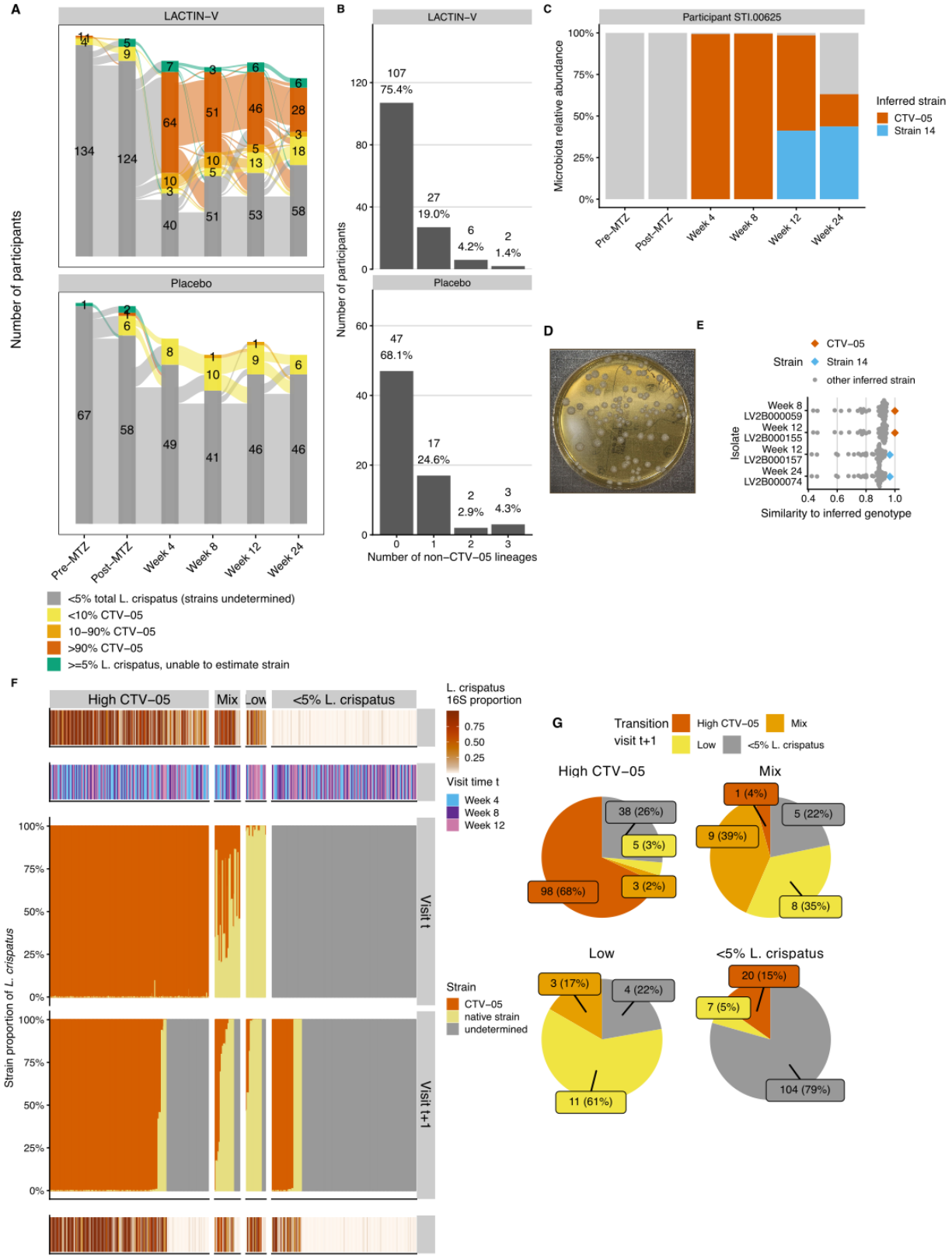
1315 C. Microbiota composition and BV diagnosis at week 12 for each participant in the LBP and placebo
1316 treatment arms. Relative abundances of the 15 most abundant taxa are shown, with remaining
1317 taxa summed as “Other”. Participants are grouped by treatment arm into the categories defined
1318 as in **Figure 1B,** with BV diagnosis indicated along the top margin.
1319 D. Total bacterial load (measured by qPCR) at each scheduled visit in each arm. Boxplots (here and
1320 in subsequent figures) represent the median (middle horizontal line), the 25th and 75th percentiles
1321 (lower and upper boundaries of boxes, respectively, “IQR”), measurements that fall within 1.5
1322 times the IQR (whiskers), and any individual measurement outside that range (dots).

1323 **Figure 2: Microbiota trajectories throughout the trial**


A
Proportion in topic
1/100
1/10
1/4
1/2
1
Gardnerella swidsinskii/leopoldii
Fannyhessea v.
Megasphaera lornae
Prevotella timonensis
Prevotella amnii
Sneathia amnii
Sneathia sanguinegens
BVAB2 BVAB2
Prevotella melaninogenica
Dialister sp. DNF00626
Parvimonas sp. KA00581
Porphyromonas uenonis
Mobiluncus curtisii
Peptostreptococcus anaerobius
Streptococcus agalactiae
Coriobacteriaceae sp. DNF00809
Ca. Lachnocurva v. (BVAB1)
Prevotella sp. 1
Prevotella sp. S7-1-8
Mobiluncus mulieris
Prevotella sp. 2
Megasphaera hutchinsoni
Mageeibacillus indolicus
Prevotella bivia
Gardnerella vaginalis
Gardnerella piotii
Aerococcus christensenii
Gemella asaccharolytica
Prevotella disiens
Mycoplasma hominis
Veillonella montpellierensis
Dialister micraerophilus
Finegoldia magna
G. s./l. topic
P. amnii topic
Ca. L. v. (BVAB1) topic
P. bivia topic
B
Taxon or topic
L. crisp.
L. iners
L. jensenii
Other L.
G. s./l. topic
P. amnii topic
Ca. L. v. (BVAB1) topic
P. bivia topic
Pre-MTZ
Post-MTZ
Week 4
Week 8
Week 12
Week 24
Lactin-V
Placebo
Participants
(ordered by microbiota composition)
0
0.5
1
Relative abundance
C
L. crispatus
L. iners
L. jensenii
Other L.
G. s./l. topic
P. amnii topic
Ca. L. v. (BVAB1) topic
P. bivia topic
Lactin-V
Placebo
Relative abundance
100%
75%
50%
25%
0%
Pre-MTZ
Post-MTZ
Week 4
Week 8
Week 12
Week 24
Visit
D
Bray-Curtis diss. consecutive v.
1.00
0.75
0.50
0.25
0.00
Pre-MTZ to Post-MTZ
Post-MTZ to Week 4
Week 4 to Week 8
Week 8 to Week 12
Week 12 to Week 24
Study arm
Lactin-V
Placeb

1325 A. Proportion of each taxon (y-axis) in each non-*Lactobacillus* topic (x-axis) as estimated by Latent
1326 Dirichlet Allocation (LDA) from 16S rRNA gene sequencing data. Proportions sum to one for each
1327 topic. Taxa were included if they made up at least 1% of any topic and non-*Lactobacillus* topics
1328 are named according to their most predominant taxon. "*G. s/l*": *Gardnerella swidsinskii-leopoldii*.
1329 "*P. amnii*": *Prevotella amnii*. "*Ca.* L. v. (BVAB1)": *Candidatus* Lachnocurva vaginae (BVAB1). "*P.*
1330 *bivia*": *Prevotella bivia*.
1331 B. Microbiota composition expressed as topic relative abundance (x-axis) for each participant at
1332 each scheduled visit. Results are depicted for participants with microbiota compositional data for
1333 at least three of the six visits. Results are grouped by visit (horizontal panels) and treatment arm
1334 (vertical panels), and the y-axis is ordered by microbiota composition trajectories (see **SI**).
1335 C. Relative abundance of each topic (horizontal panels) in each arm (vertical panels) at each visit.
1336 Samples from the same participant are connected by a line.
1337 D. Bray-Curtis dissimilarity between each participant's microbiota composition at the two indicated
1338 consecutive visits, calculated from 16S rRNA gene amplicon sequence variant (ASV) relative
1339 abundances. For each pair of consecutive visits, values are shown only for participants with
1340 available microbiota data from both visits.

1341 **Figure 3: *L. crispatus* strain dynamics among LBP recipients**


A
LACTIN–V
Placebo
Number of participants
Pre–MTZ
Post–MTZ
Week 4
Week 8
Week 12
Week 24
<5% total L. crispatus (strains undetermined)
<10% CTV–05
10–90% CTV–05
>90% CTV–05
>=5% L. crispatus, unable to estimate strain
B
LACTIN–V
Placebo
Number of participants
107
75.4%
27
19.0%
6
4.2%
2
1.4%
47
68.1%
17
24.6%
2
2.9%
3
4.3%
Number of non–CTV–05 lineages
C
Participant STI.00625
Microbiota relative abundance
Inferred strain
CTV–05
Strain 14
D
E
Strain
CTV–05
Strain 14
other inferred strain
Isolate
Week 8
LV2B000059
Week 12
LV2B000155
Week 12
LV2B000157
Week 24
LV2B000074
Similarity to inferred genotype
F
High CTV–05
Mix
Low
<5% L. crispatus
L. crispatus 16S proportion
Visit time t
Week 4
Week 8
Week 12
Strain proportion of L. crispatus
Visit t
Visit t+1
Strain
CTV–05
native strain
undetermined
G
Transition visit t+1
High CTV–05
Mix
Low
<5% L. crispatus
High CTV–05
38 (26%)
5 (3%)
3 (2%)
98 (68%)
Mix
1 (4%)
5 (22%)
9 (39%)
8 (35%)
Low
3 (17%)
4 (22%)
11 (61%)
<5% L. crispatus
20 (15%)
7 (5%)
104 (79%)

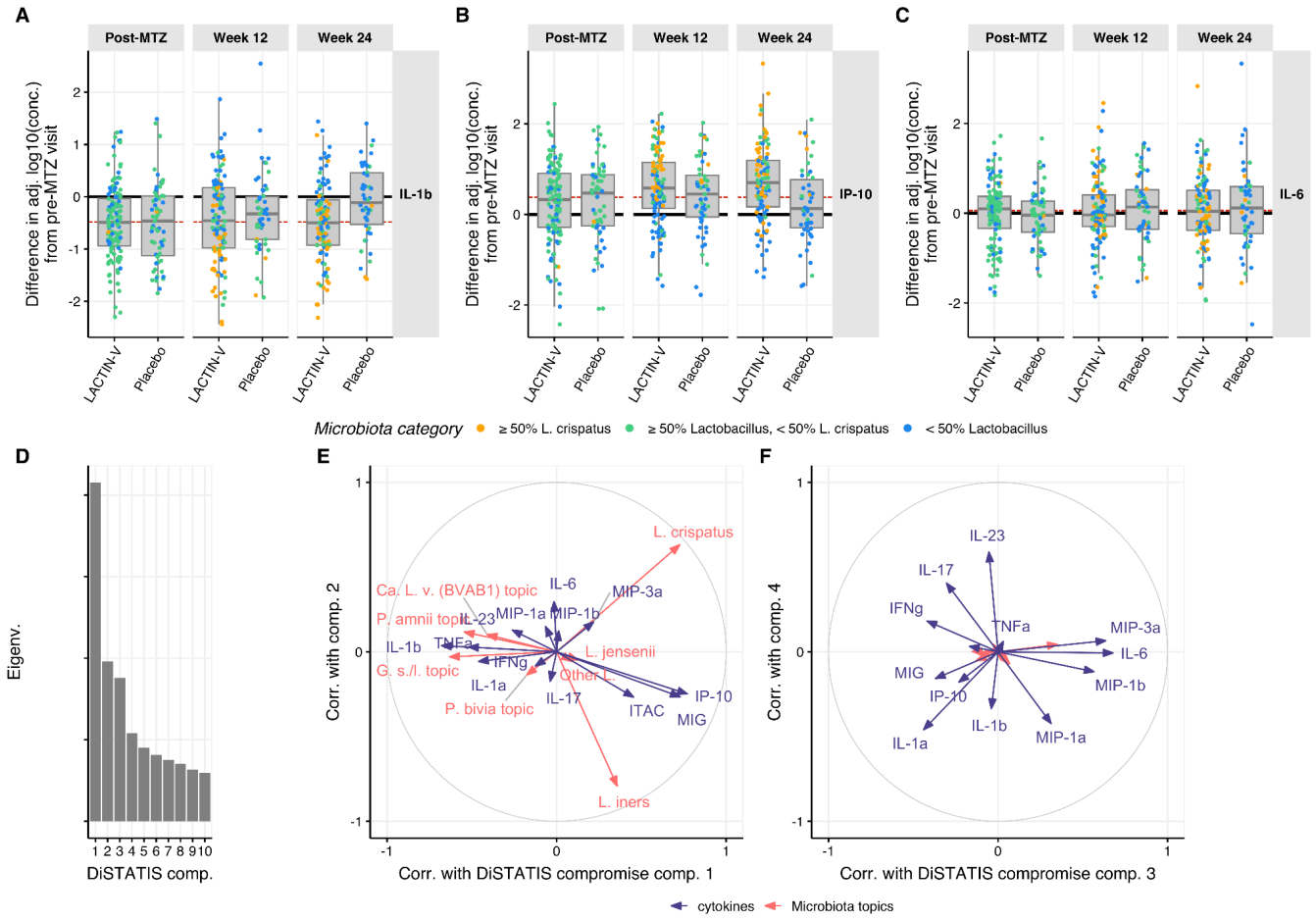
1342

1343 A. Sankey diagram showing dynamics of CTV-05 versus non-CTV-05 *L. crispatus* strain colonization
1344 among LBP recipients. Strains were inferred from shotgun metagenomic whole genome
1345 sequencing (mWGS) using StrainFacts for all samples with ≥5% total *L. crispatus* relative
1346 abundance as determined from 16S rRNA gene sequencing. Samples were categorized based
1347 on proportional *L. crispatus* strain abundance as >90% CTV-05 ("High CTV-05"), 10-90% CTV-
1348 05 ("Mixed"), <10% CTV-05 ("Low CTV-05"). Strain inference was unsuccessful for a small
1349 fraction of samples (depicted in green) due to technical failures of metagenomic sequencing.
1350 B. Histograms showing the number (percentage) of study participants in each treatment arm with
1351 the indicated number of inferred native (non-CTV-05) *L. crispatus* strains detected at least once
1352 during the trial.
1353 C. Metagenomically inferred *L. crispatus* strain composition throughout the trial in an example
1354 participant from whom targeted isolations were performed at weeks 8, 12, and 24. The plot depicts
1355 overall microbiota relative abundance, with non-*L. crispatus* species summed (gray) and inferred
1356 *L. crispatus* strains shown individually in red (CTV-05 strain) and blue (a native strain: "Strain 14").
1357 D. Example *Lactobacillus* MRS agar plate from bacterial isolations showing characteristic *L.*
1358 *crispatus* colony morphology.
1359 E. Genomes of *L. crispatus* isolated from the week 8, 12, and 24 visits of the participant in **Figure**
1360 **3C** identified two strains nearly identical to the metagenomically inferred strains in these samples
1361 (assessed by Jaccard similarity to StrainFacts-inferred genotypes) and dissimilar from all other
1362 inferred strains in the cohort, none of which were detected in this participant's samples. The
1363 isolated strains included the CTV-05 strain isolated from week 8 and 12 samples (red), and native
1364 Strain 14 isolated from week 12 and 24 samples (blue; see also **Figure S3A**).
1365 F. Transition plot for LBP recipients showing total *L. crispatus* relative abundance (determined by
1366 16S rRNA gene sequencing, top and bottom rows) and strain category proportions (third and

1367 fourth rows) at each pair of consecutively scheduled post-randomization visits with available strain
1368 inference data or at which strain inference was not performed due to <5% total *L. crispatus* relative
1369 abundance. Strain categories in the third and fourth rows show proportions of observed *L.*
1370 *crispatus* representing CTV-05 versus the sum of the native strain(s) at each visit; "undetermined"
1371 refers to samples in which strain inference was not performed. Pairs including a visit in which
1372 strain inference was unsuccessful due to technical failures of metagenomic sequencing are not
1373 shown. For each pair of visits, the top three panels show *L. crispatus* relative abundance, visit
1374 week, and strain category at the first visit in the pair ("Visit t") and the bottom two panels show
1375 strain category and relative abundance at the subsequent visit ("Visit t+1"). Visit pairs are grouped
1376 along the horizontal axis by strain category at Visit t and ordered within groups by CTV-05 strain
1377 proportion at Visit t+1.

1378 G. Pie charts for each Visit t strain category from **Figure 3F**, summarizing subsequent *L. crispatus*
1379 strain category frequencies at Visit t+1.

1380 **Figure 4: Treatment effects on mucosal inflammation and relationship to microbiota composition.**



1381

1382 A. Differences in log10-transformed, adjusted IL-1β concentrations between participants' pre-MTZ
1383 visit and their ensuing post-MTZ, week 12, and week 24 visits (horizontal panels) in both treatment
1384 arms. Each dot represents a single visit per participant.

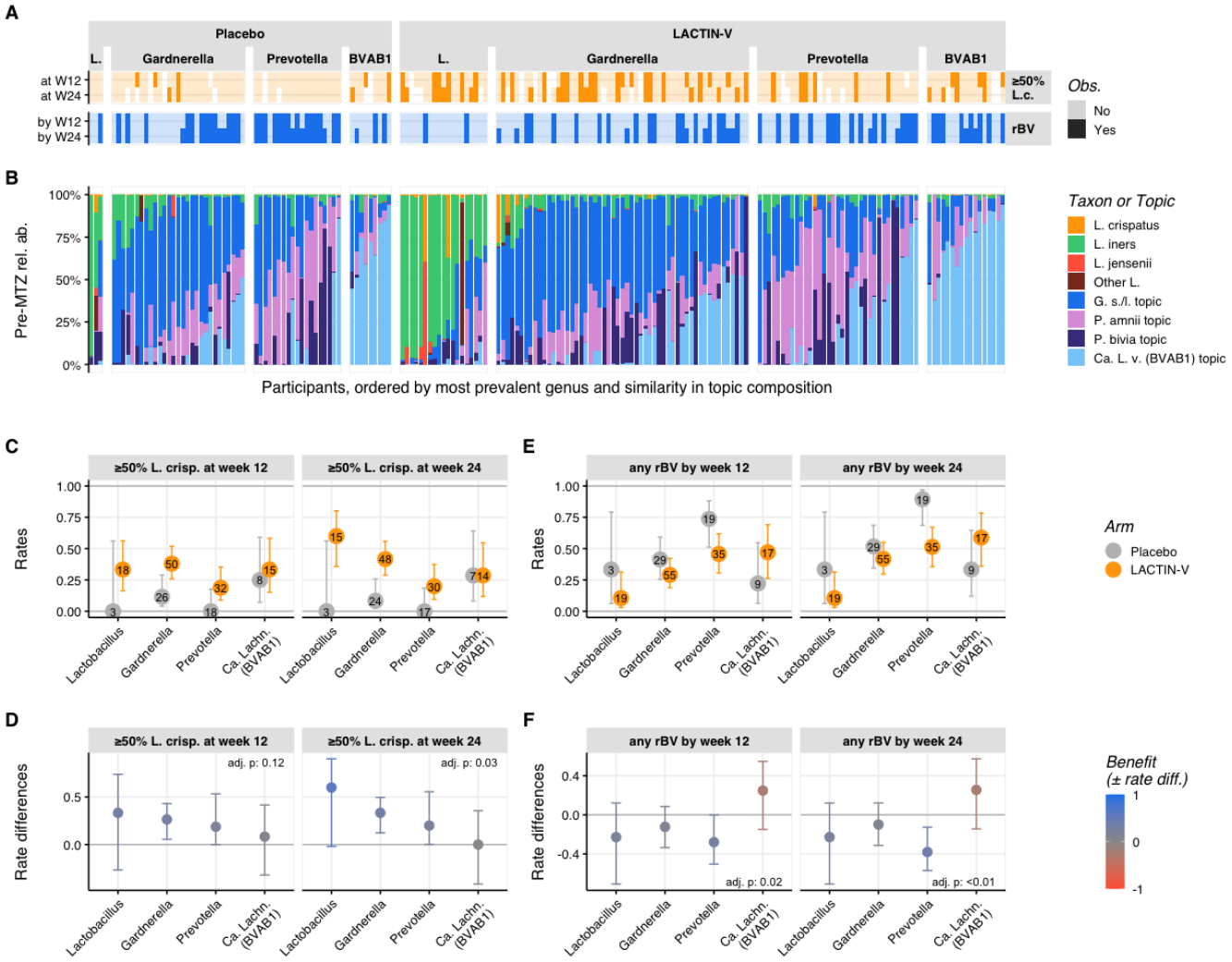
1385 B. Same as **Figure 4A**. for IP-10 adjusted concentrations.

1386 C. Same as **Figure 4A**. and **4B** for IL-6 adjusted concentrations. Similar displays are available for
1387 all cytokine adjusted concentrations and non-adjusted concentrations in the **SI**.

1388 D. DiSTATIS screeplot: eigenvalues (a.u.) of DiSTATIS compromise for the first 10 latent
1389 components.

1390 E. DiSTATIS correlation circle showing the correlations between the 1st (x-axis) and 2nd (y-axis)
1391 DiSTATIS compromise latent dimension and the microbiota topic proportions (red) or cytokine
1392 transformed concentrations (blue).
1393 F. Same as **Figure 4E** for DiSTATIS 3rd and 4th latent components.

1394 **Figure 5: Heterogeneity in intervention effects with respect to pre-MTZ microbiota**



1395

1396 A. Participant-level observed outcomes for attainment of *L. crispatus*-dominance at weeks 12 and

1397 24 (dark orange: ≥50% *L. crispatus*; light orange <50% *L. crispatus*) and BV recurrence by weeks

1398 12 and 24 (dark blue: rBV; light blue: no rBV) for each participant with week 12 and/or week 24

1399 data. White indicated data not available. Participants are grouped according to treatment arm

1400 and, within treatment arm, by the most predominant bacterial genus at their pre-MTZ visit ("L."

1401 indicates *Lactobacillus*; “BVAB1” indicates *Ca.* Lachnocurva vaginae). Predominant-genus
1402 groupings lacking at least two participants in each treatment arm were excluded.

1403 B. Pre-MTZ microbiota composition for each participant in **Figure 5A** (x-axis ordered and vertically
1404 aligned as in **5A**), expressed as topic proportions (see **Figure 2A** for topic definitions).

1405 C. Rates and associated 95% CI of *L. crispatus*-dominance (≥50% *L. crispatus*) at week 12 or 24
1406 (horizontal panels) by pre-MTZ microbiota group in each arm. Participants are grouped based on
1407 their most predominant genus as defined in **Figure 5A-B**. CI are computed using Wilson's scores
1408 due to low sample sizes in some groups.

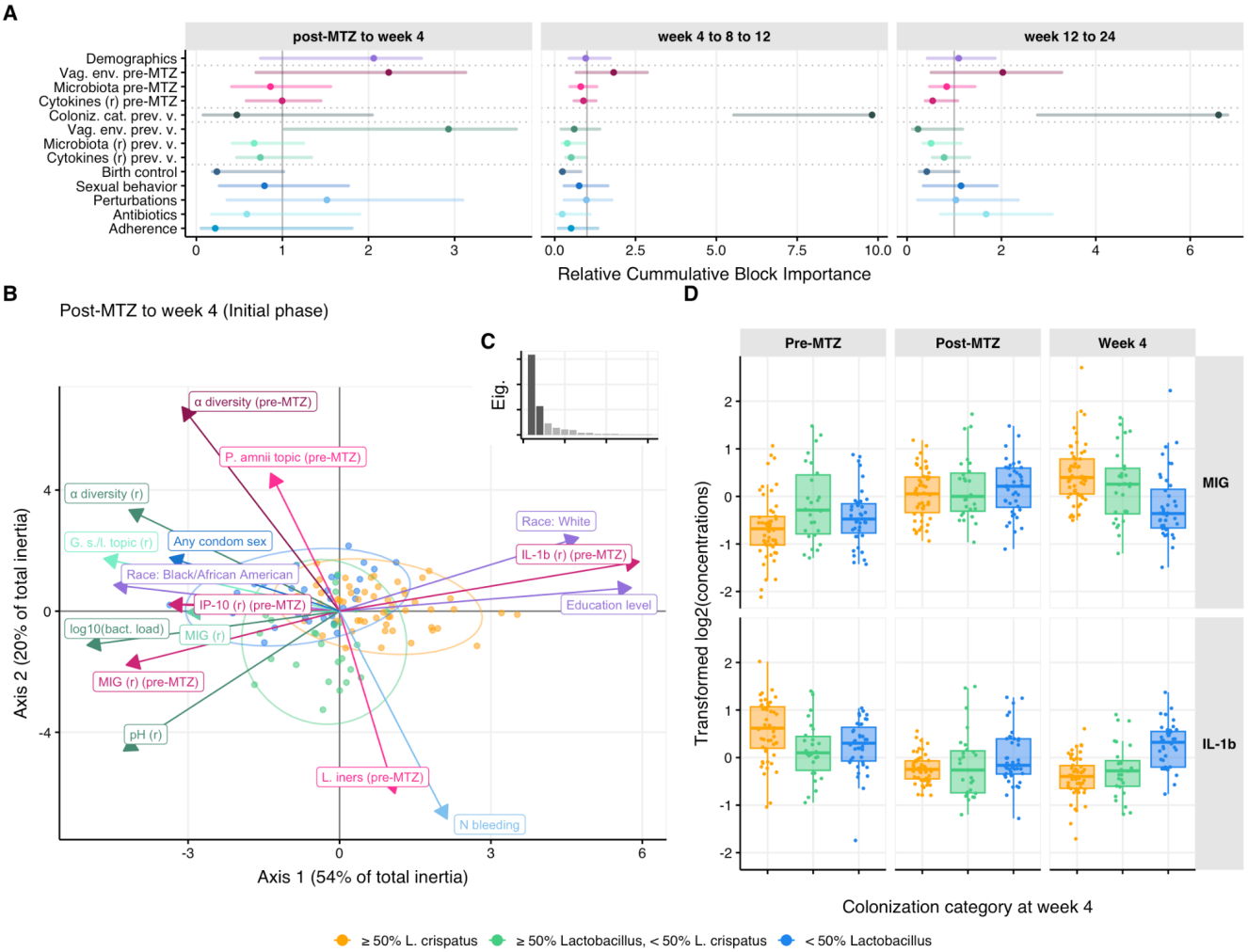
1409 D. Rate differences (rate in LBP arm - rate in placebo arm) and associated 95% CI (Wilson’s score
1410 method) between arms for results in **Figure 5C**. Color scale shows degree of benefit (defined as
1411 the difference in rates between the two arms) in achieving *L. crispatus*-dominance, with blue
1412 indicating benefit with LBP treatment and red indicating benefit with placebo. *P*-values adjusted
1413 for multiple testing (across panel **5D** and **5F**) using the Benjamini-Hochberg correction of the
1414 analysis of deviance test for heterogeneity in treatment effects when stratifying participants
1415 according to their most predominant genus at the pre-MTZ visit are shown in upper right corner.

1416 E. Same as (**Figure 5C**), but for BV recurrence by week 12 or 24.

1417 F. Same as (**Figure 5D**), but for BV recurrence by week 12 or 24. Color scale shows degree of
1418 benefit in reducing rBV, with blue indicating benefit with LBP treatment and red indicating benefit
1419 with placebo.

1420 **Figure 6: Multiblock analysis of factors associated with microbiota categories in LACTIN-V**

1421 **recipients.**



1422

1423 A. Multiblock partial least squared discriminant analysis (MB-PLS-DA, **Figure S6A**) of factors

1424 predictive of microbiota categories (≥50% *L. crispatus*; ≥50% *Lactobacillus* and <50% *L.*

1425 *crispatus*; or <50% *Lactobacillus*) at the indicated visits. Relative cumulative importance indices

1426 (x-axis, Methods) are shown for each thematic block of explanatory variables (y-axis, color) for

1427 the initial phase model (post-MTZ to Week 4; left panel), continuation phase model (Week 4 to 8

1428 to 12; middle panel), and follow-up phase model (Week 12 to 24; right panel). Dots indicate point

1429 estimates and translucent horizontal bars represent bootstrapped 95% CI. Vertical lines at x = 1
1430 show the expected value under the (null) hypothesis that all variables have the same importance.
1431 Block definitions along with the description of the variables included in each block are provided in
1432 **Table S6**. Horizontal dashed lines separate blocks included in the nested models (**Figure S6A**,
1433 **SI**, Methods).

1434 B. Bi-plot for the MB-PLS-DA initial phase model (from post-MTZ to week 4). Dots represent scores
1435 (each dot is a participant) while arrows represent loadings (each arrow is an explanatory variable,
1436 see **Figure S7A**) for the most important variables (Methods) in the space of the 2 first latent
1437 variables. Dots (participants) are colored by their corresponding response categories (*i.e.,* their
1438 colonization status at week 4). Perfect separation of the categories would indicate perfect
1439 prediction of colonization status at the week 4 visit. Arrows (explanatory variables) are colored by
1440 the block to which they belong to - **Figure 6A** and **Table S6** serve as color legend.

1441 C. Scree-plot of the first 15 eigenvalues of the model from **Figure 6B**. The first two latent variables
1442 (selected by cross-validation) are shown in dark gray.

1443 D. Transformed $\log_2$-concentrations (y-axis) of two cytokines (MIG and IL-1b) whose pre-MTZ levels
1444 (left panel) were identified as associated with participants' colonization status at week 4 (x-axis,
1445 colors) by the model (**Figure 6A** and **S7A**). Concentrations at the post-MTZ and week 4 visits are
1446 shown in the middle and right panels. Each dot is a participant, colored by the colonization status
1447 at week 4 for each panel.

1448 **Supplementary Tables and Figures**

1449 **Table S1**

	At least one visit	Pre-MTZ	Post-MTZ	Week 4	Week 8	Week 12	Week 24	All planned in-person visits
Total	211	208	203	181	172	179	165	139
LACTIN-V	142	140	138	124	120	123	113	99
Placebo	69	68	65	57	52	56	52	40

1450 Number of participants in each study arm (rows) with available 16S rRNA gene sequencing data at any,

1451 each, or all planned in-person visits (unscheduled clinic visits not shown).

1452 **Table S2**

	% (n) [both arm]	% (n) [LACTIN-V]	% (n) [Placebo]
≥ 50% *L. iners*	52 % (106)	52 % (72)	52 % (34)
≥ 50% *L. jensenii*	8 % (16)	9 % (13)	5 % (3)
≥ 50% *L. crispatus*	1 % (2)	1 % (1)	2 % (1)
≥ 50% *Lactobacillus* (no individual species ≥ 50%)	13 % (27)	14 % (20)	11 % (7)
< 50% *Lactobacillus*	26 % (52)	23 % (32)	31 % (20)

1453 At the post-MTZ visit, number and proportion of samples dominated by one of the top *Lactobacillus*
1454 species (top 3 rows), a mixture of *Lactobacillus* (4th row), or by non-*Lactobacillus* species in both arms
1455 (2nd column), the LACTIN-V arm (3rd column), or the placebo arm (4th column).

1456 **Table S3**

	Clinical diagnosis	
Microbiota outcome (16S rRNA sequencing-based)	no rBV n (%)	rBV n (%)
< 50% *Lactobacillus*	162 (50 %)	164 (50 %)
≥ 50% *Lactobacillus, <50% L. crispatus*	192 (99 %)	2 (1 %)
≥ 50% *L. crispatus*	209 (100 %)	0 (0 %)

1461

1462 Number of visits (n) at which participants did not (left column) or did (right column) meet diagnostic criteria
1463 for recurrent BV (rBV). Participants are shown according to microbiota category. Percentages sum to
1464 100% for each category.

1465 **Table S4**

Category at week 4	Category at **week 12**			
	≥ 50% *L. crispatus*	≥ 50% *Lactobacillus*, < 50% *L. crispatus*	< 50% *Lactobacillus*	Total
≥ 50% *L. crispatus*	25	14	12	51
≥ 50% *Lactobacillus*, < 50% *L. crispatus*	6	12	8	26
< 50% *Lactobacillus*	4	4	30	38
Total	35	30	50	115

1466 Contingency table showing the correspondence between Week 4 and Week 12 microbiota categories for

1467 LACTIN-V recipients with available data from both visits.

1468 **Table S5**

Category at week 4	Category at **week 24**			
	≥ 50% *L. crispatus*	≥ 50% *Lactobacillus*, < 50% *L. crispatus*	< 50% *Lactobacillus*	Total
≥ 50% *L. crispatus*	27	9	11	47
≥ 50% *Lactobacillus*, < 50% *L. crispatus*	6	6	11	23
< 50% *Lactobacillus*	5	10	22	37
Total	38	25	44	107

1469 Contingency table showing the correspondence between Week 4 and Week 24 microbiota categories for

1470 LACTIN-V recipients with available data from both visits.

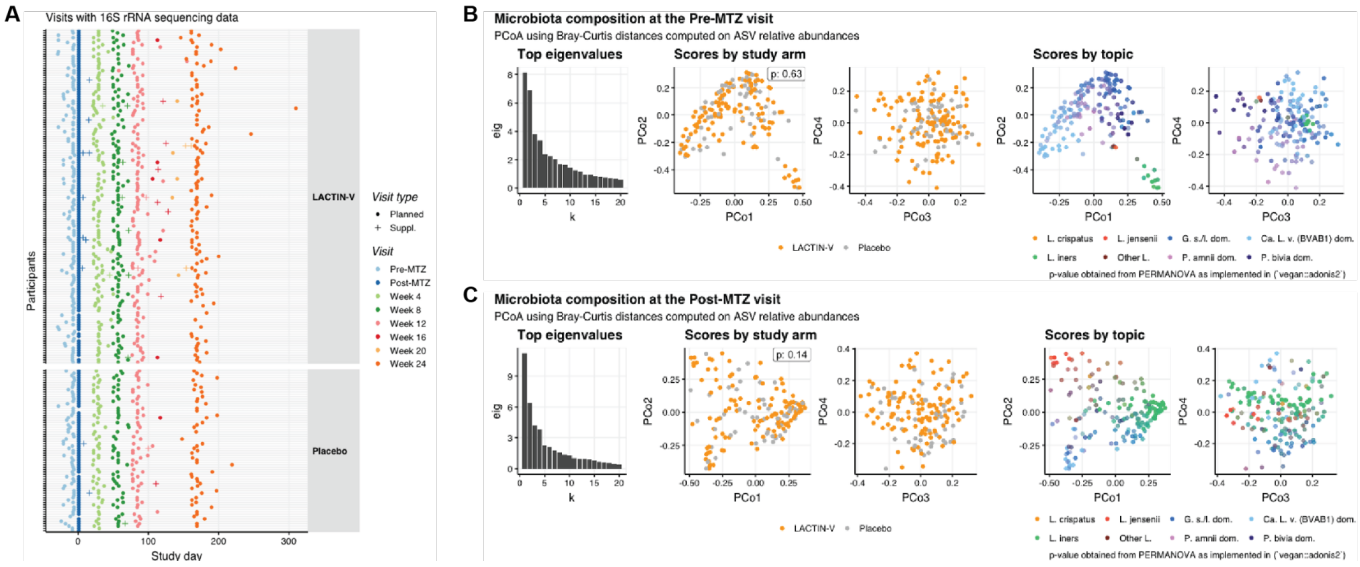
1471 **Table S6**

Block	Variable	Description
Demographics	Age	Age of the participant at enrollment.
	N past BV	Number of past BV episodes (coded as a number on the following scale: (1) None, (2) 1-2, (3) Unknown, (4) 3-4, (5) 5 or more).
	Education level	Education level of the participant (coded as a number on the following scale: (1) Did not complete high school, (2) Completed high school, (3) Completed junior college, (4) Completed college (undergraduate degree), (5) Completed graduate degree).
	Race	Self-declared race of the participant (Categories with few participants were merged into "Other").
Vag. env. pre-MTZ	α diversity	Shannon diversity index of baseline (pre-MTZ) microbiota, computed on taxa proportions quantified from 16S rRNA gene sequencing data.
	pH	Vaginal pH at the pre-MTZ visit, measured at study sites.
Microbiota pre-MTZ		Pre-MTZ microbiota composition described as topic proportions.
Cytokine (r) pre-MTZ		Pre-MTZ cytokine residual levels. These values are computed as the difference between the transformed cytokine/chemokines levels and their predicted value based on microbiota composition expressed as topic proportions.
Coloniz. cat. prev. v.		Colonization category at the previous visit. One of these mutually exclusive categories: ≥ 50% *L. crispatus*, ≥ 50% *Lactobacillus* and < 50% *L. crispatus*, or < 50% *Lactobacillus*.
Vag. env. prev. v.	α diversity (r)	Difference between the Shannon diversity index at the previous visit (w.r.t the visit of the response variable), computed on taxa proportions quantified from 16S rRNA gene sequencing data, and its expected value based on colonization category at the same visit.
	pH (r)	Difference between vaginal pH at the previous visit, measured at study sites, and its expected value based on colonization category at the same visit.
	log10(bact. load)	Log_{10} total bacterial load at the previous visit, quantified by qPCR.
Microbiota (r) prev. v.		Difference between the observed microbiota composition at the previous visit described as topic proportions and the average microbiota composition within each colonization category.
Cytokines (r) prev. v.		Cytokine residual levels at the previous visit. These values are computed as the difference between the transformed cytokine/chemokines levels and their predicted value based on microbiota composition expressed as topic proportions.
Birth control	Non-hormonal	A binary variable indicating whether participants reported exclusively using one of the following non-hormonal birth control methods including abstinence, use of condoms, or fertility awareness methods.
	IUD (Hormonal)	Same as above for hormonal intra-uterine device.
	IUD (Non-hormonal)	Same as above for non-hormonal intra-uterine device (copper IUD).

	Combined	A binary variable indicating whether participants reported using a combination of hormonal contraceptives or switched contraceptive method between the last visit and this one.
	P only	A binary variable indicating whether participants reported exclusively using progestin-only hormonal contraceptive, including the pill, patches, and implants.
	unknown	A binary variable indicating whether a participant's birth control is unknown (missing data). Each participant belongs to only one of each birth control category.
Sexual behavior	N new partners	Number of new sexual partners since the last visit.
	Any condomless sex	A binary flag indicating whether the participant reported any sexual intercourse without condoms since the last visit.
	Any condom sex	A binary flag indicating whether the participant reported any sexual intercourse with condoms since the last visit.
Perturbations	N douching	Number of days participant reported any douching episode since the last visit.
	N bleeding	Number of days participant reported vaginal bleeding since the last visit.
Antibiotics	N vag. abx	Number of days participant reported using vaginal antibiotics since the previous visit.
	N oral abx	Number of days participant reported taking oral antibiotics since the previous visit.
Adherence	N missed doses	Difference between the number of LACTIN-V doses that should have been taken between the previous visit and the current visit if the study protocol had been strictly followed and the actual number of doses that were taken.
	Days since last dose	Number of days since the most recent LACTIN-V dose was taken.

1472

Figure S1

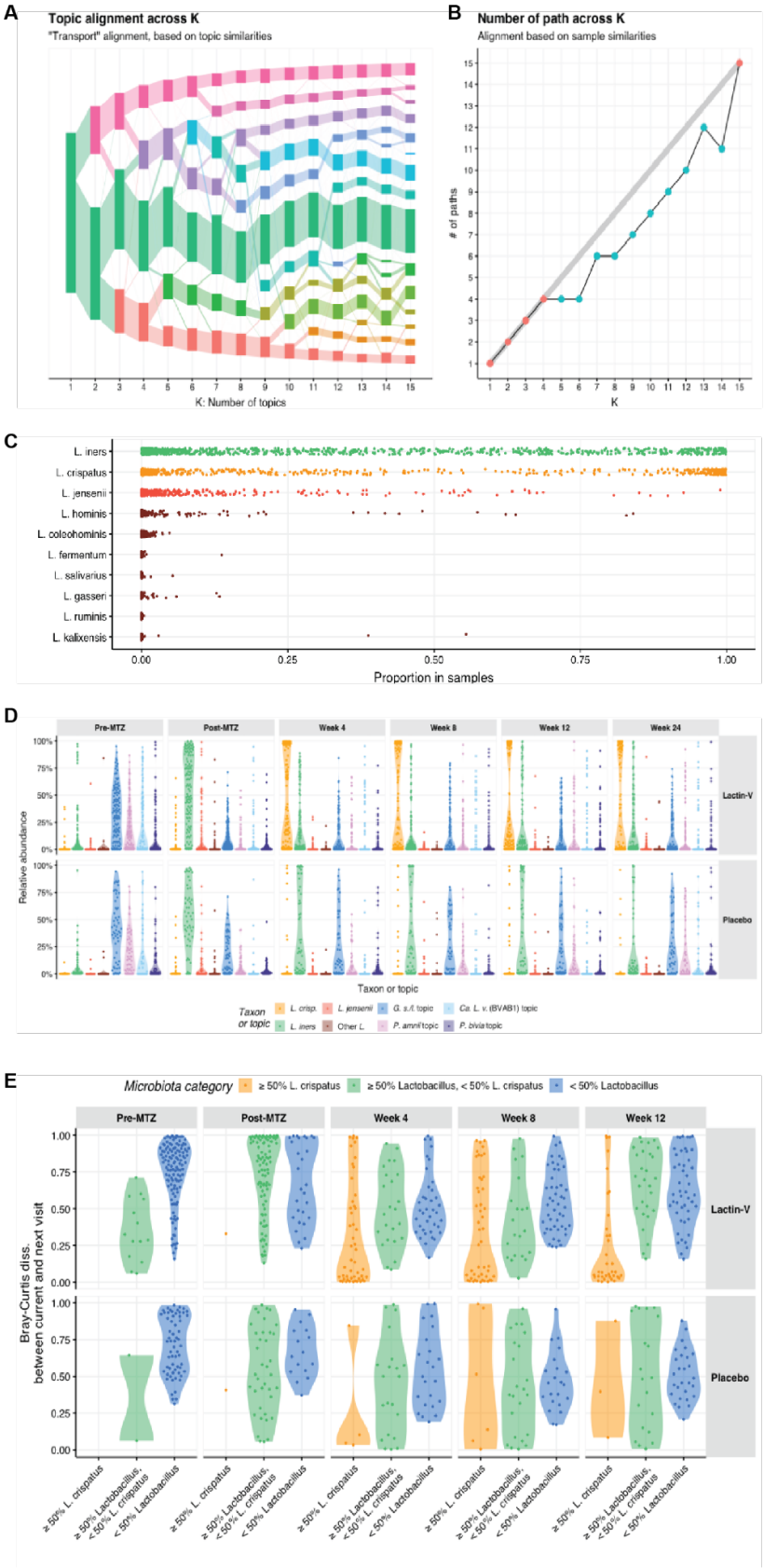


A. Time-line of collected swabs from which 16S rRNA gene sequencing data could be generated.

B. Balance between treatment arms in terms of microbiota composition at the pre-MTZ visit. The left panel shows the top 20 eigenvalues of the PCoA (Principal Coordinate Analysis) performed on the Bray-Curtis dissimilarity matrix, computed from the relative abundances of ASVs. The next two panels show the PCoA scores (projection of the samples, each dot is a sample) colored by the intervention arm for the first 4 principal coordinates. The *p*-value in the upper right corner corresponds to the PERMANOVA test on the study arm. The right panels show the same projections as the middle panels, but samples are here colored by topic relative abundance (**Figure 2A-B**).

C. Balance between treatment arms in terms of microbiota composition at the post-MTZ visit, labeled and arranged as for **Figure S1B**.

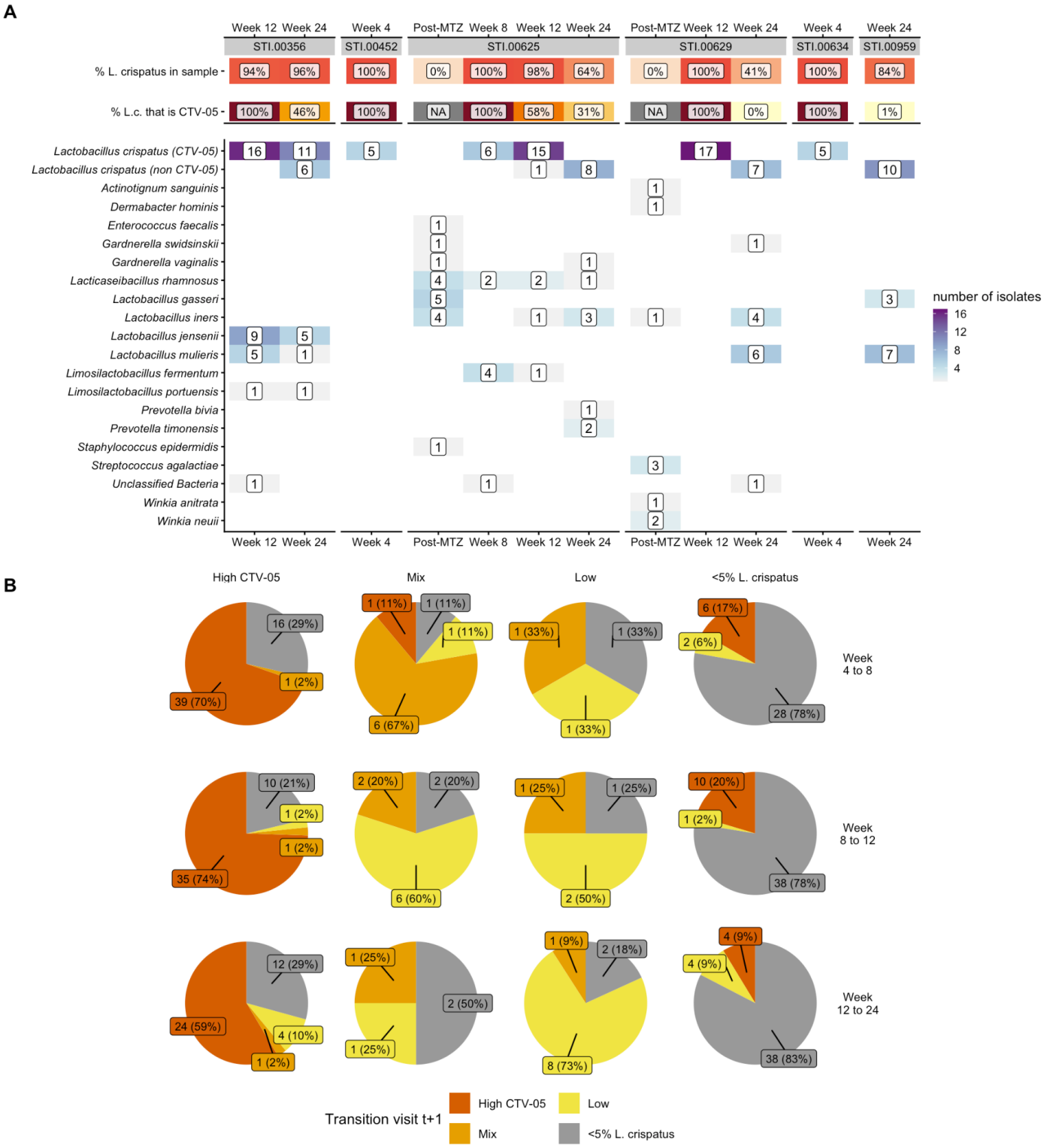
1486 **Figure S2**


A
Topic alignment across K
"Transport" alignment, based on topic similarities
K: Number of topics
B
Number of path across K
Alignment based on sample similarities
# of paths
K
C
L. iners
L. crispatus
L. jensenii
L. hominis
L. coleohominis
L. fermentum
L. salivarius
L. gasseri
L. ruminis
L. kalixensis
Proportion in samples
D
Pre-MTZ
Post-MTZ
Week 4
Week 8
Week 12
Week 24
Lactin-V
Placebo
Relative abundance
Taxon or topic
Taxon or topic
L. crisp.
L. jensenii
G. a./f. topic
Ca. L. v. (BVAB1) topic
L. iners
Other L.
P. amnii topic
P. bivia topic
E
Microbiota category
≥ 50% L. crispatus
≥ 50% Lactobacillus, < 50% L. crispatus
< 50% Lactobacillus
Pre-MTZ
Post-MTZ
Week 4
Week 8
Week 12
Lactin-V
Placebo
Bray-Curtis diss. between current and next visit

1487

1488 A. “Transport” alignment between topics identified by LDA models fitted on the non-*Lactobacillus*
1489 taxa counts for varying total number of topics (K, x-axis). Each rectangle represents a topic; the
1490 height of that rectangle represents the summed prevalence of that topic in the population, and the
1491 size of the connection between topics is proportional to the degree of similarity between the
1492 composition of two given topics.
1493 B. The number of paths (y-axis) identified at each resolution (K, the number of topics considered by
1494 the LDA models, x-axis) from the “product” alignment of topics; for details see (Fukuyama et al.,
1495 2023). A path identifies a series of topics with high alignment scores throughout resolutions. A
1496 plateau in the number of paths (here at K = 4) suggests that the 4-topic model is the model that
1497 identifies most robust topics (i.e., topics that remain similar across resolution despite the
1498 introduction of new, potentially spurious, topics).
1499 C. Proportion (x-axis) of each *Lactobacillus* species (y-axis) detected across all samples. Species
1500 with a relative abundance > 50% in at least 10 samples made up their own topic.
1501 D. Relative abundance of taxon or topic (x-axis, colors) in each arm at each visit.
1502 E. Microbiota stability by arm, visit, and microbiota category. Stability is quantified using the Bray-
1503 Curtis dissimilarity, described as relative abundances at the ASV level at the participant’s current
1504 visit (vertical panels) and subsequent scheduled visit. At each visit, participants are grouped into
1505 one of the three categories (colors, x-axis) defined earlier (*L. crispatus*-dominant, *Lactobacillus*-
1506 dominant but low-level *L. crispatus*, and low-level *Lactobacillus*).

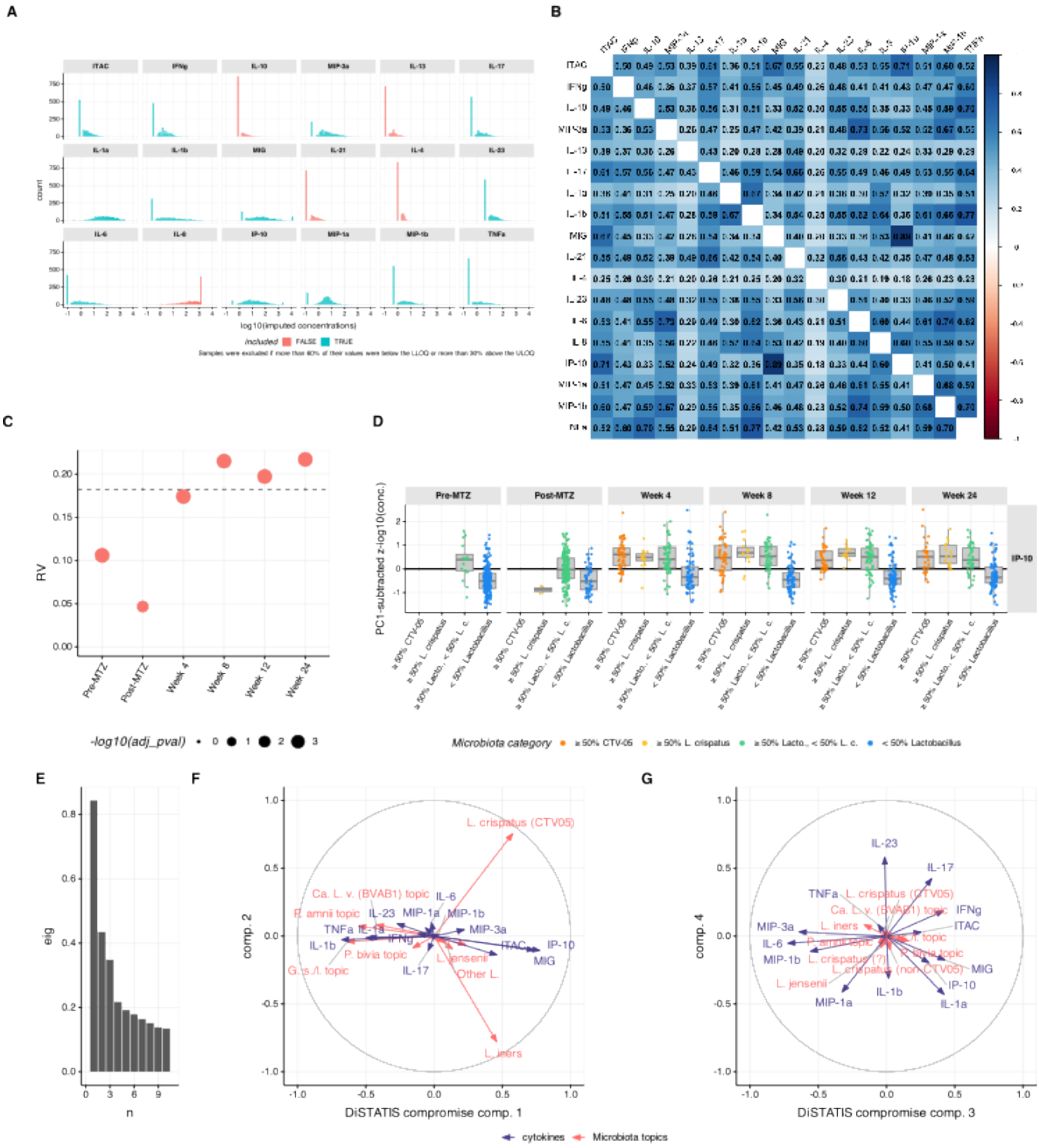
1507 **Figure S3**



1508

1509 A. Results of targeted isolations from LACTIN-V trial samples, showing source participant, visit
1510 number, sample *L. crispatus* relative abundance as estimated by 16S rRNA gene sequencing,
1511 CTV-05 proportional strain abundance, and the numbers and taxonomic assignments of
1512 genome-sequenced isolates from each sample. *L. crispatus* isolates are separated by whether
1513 genome sequencing showed them to be CTV-05 or non-CTV-05 (native) strains. Multiple
1514 isolates of genotypically identical strains (including of *L. crispatus* isolates) were obtained from
1515 some samples.
1516 B. Pie charts for each Visit t strain category from **Figure 3F** summarizing *L. crispatus* strain category
1517 frequencies at Visit t+1, shown individually for each Visit t.

Figure S4

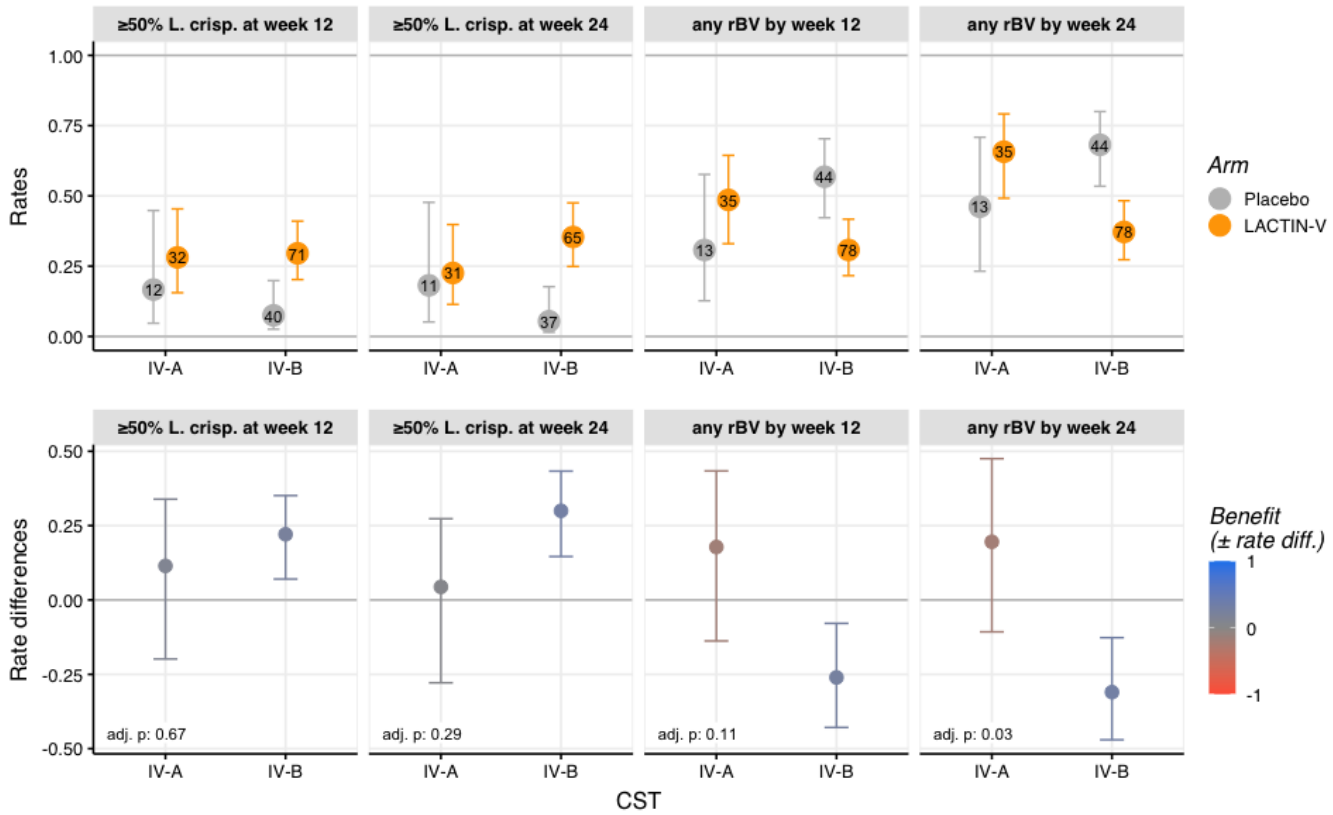

A
B
C
D
E
F
G
log10(imputed concentrations)
included FALSE TRUE
RV
Pre-MTZ
Post-MTZ
Week 4
Week 8
Week 12
Week 24
-log10(adj_pval)
PC1-subtracted z-log10(conc.)
IP-10
Microbiota category
≥ 50% CTV-05
≥ 50% L. crispatus
≥ 50% Lacto., < 50% L. c.
< 50% Lactobacillus
eig
n
DiSTATIS compromise comp. 1
comp. 2
L. crispatus (CTV05)
Ca. L. v. (BVAB1) topic
P. amnii topic
P. bivia topic
G. v./l. topic
L. jensenii
L. iners
Other L.
IL-6
IL-23
MIP-1a
MIP-1b
MIP-3a
TNFa
IL-1a
IFNg
IL-1b
IL-17
ITAC
MIG
DiSTATIS compromise comp. 3
comp. 4
cytokines
Microbiota topics

1520 A. Distribution of unadjusted cytokine and chemokine concentrations within the dataset, with values
1521 below the lower limit of quantification (LLOQ) imputed at half the LLOQ and values above the
1522 upper limit of quantification (ULOQ) imputed at the ULOQ. Analytes were excluded from further
1523 analysis if >60% of values were below the LLOQ or >30% were above the ULOQ.
1524 B. Pair-wise Pearson correlation coefficients between log10-transformed concentrations of each
1525 cytokine. All correlations were significant ($p < 0.05$).
1526 C. RV coefficients (y-axis) between microbiota composition described in terms of topic proportions
1527 and cytokine/chemokine transformed concentrations at each visit (x-axis). Permutation test *p*-
1528 values were adjusted for multiple testing to control for the false discovery rate. The size of the dot
1529 is inversely proportional to the $\log_{10}$ of the associated adjusted p-value. The horizontal dashed
1530 line provides the value of the RV coefficient when computed on all visits combined.
1531 D. IP-10 adjusted $\log_{10}$(concentrations) (y-axis) by microbiota composition categories (x-axis) at
1532 each visit (horizontal panels). Microbiota categories are mutually exclusive such that the first
1533 category (≥ 50% CTV-05) encompasses samples in which the relative abundance of CTV-05
1534 (within the overall microbiota) is larger or equal to 50%; the second category encompasses
1535 samples with ≥50% total *L. crispatus* but <50% CTV-05 relative abundance within the microbiota;
1536 the third category includes samples in which total *L. crispatus* relative abundance is <50%, but
1537 the total *Lactobacillus* relative abundance is ≥50%, and the remaining category includes samples
1538 with <50% *Lactobacillus*. Participants from both arms are included.
1539 E. DISTATIS screeplot from repeated DiSTATIS analysis, similar to analysis in **Figure 4D** but
1540 differentiating between CTV-05 and summed other *L. crispatus* strains. Eigenvalues (a.u.) of
1541 DISTATIS compromise for the first 10 latent components are shown.

F. DISTATIS correlation circle for the analysis in **Figure S4E**, showing the 1st and 2nd latent components (similar to **Figure 4E**).

G. DISTATIS correlation circle for the analysis in **Figure S4E**, showing the 3rd and 4th latent components (similar to **Figure 4F**).

1546 **Figure S5**

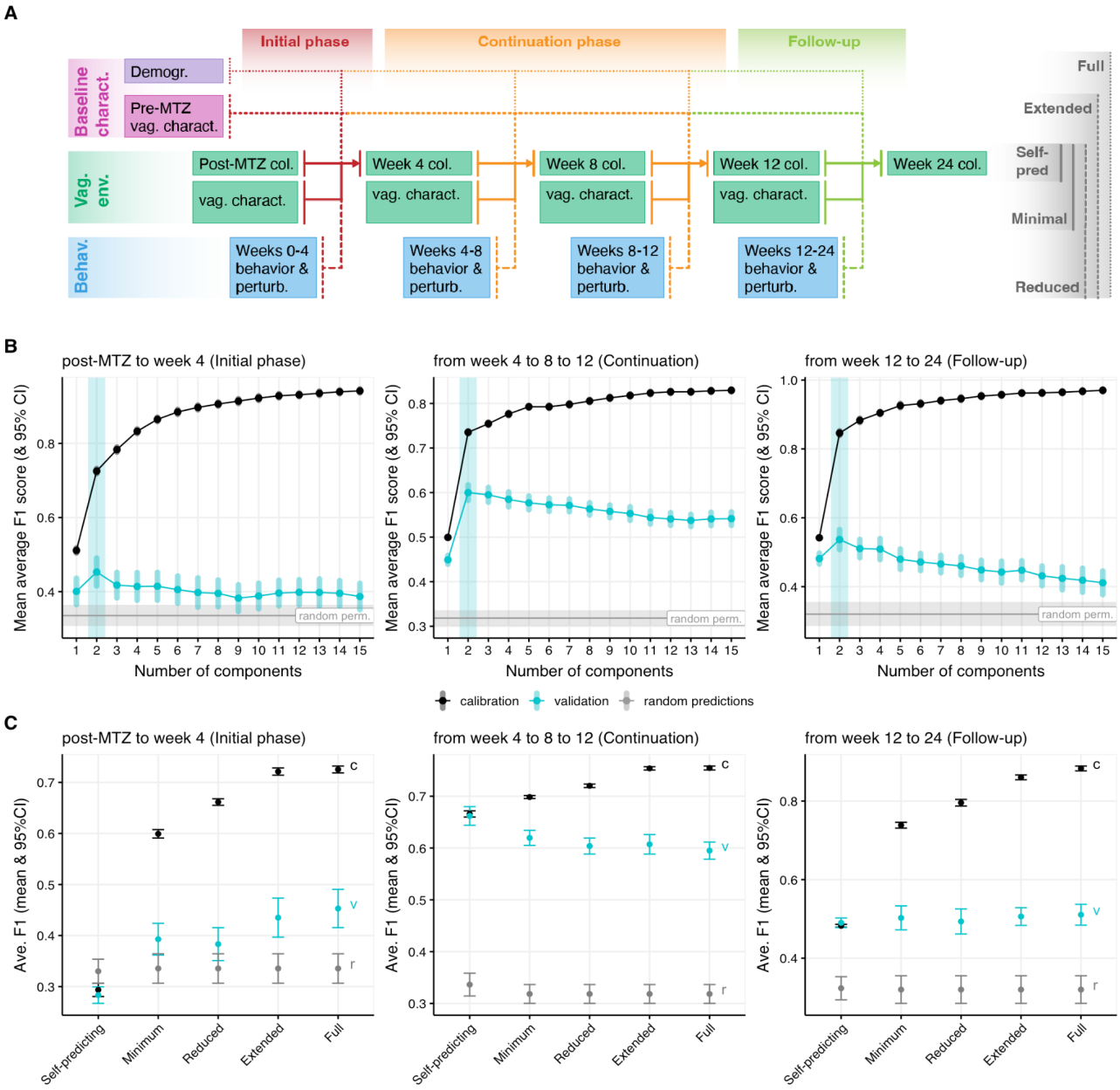


1547

1548 Rates and associated 95% CI of high-level *L. crispatus* colonization (≥50% of relative abundance) (y-
1549 axis) at week 12 or 24 (left two panels, top row) or rBV by Week 12 or 24 (right two panels, top row) by
1550 pre-MTZ CST in each arm (colors), and the corresponding rate differences and 95% CI between arms
1551 (bottom row), analyzed as in **Figure 5C-F**. Analysis is restricted to CSTs with at least two participants in
1552 each arm (*i.e.,* CSTs IV-A and IV-B). The color scale in the bottom row shows the degree of benefit
1553 (defined as the difference in rates between the two arms) in achieving *L. crispatus*-dominance (left
1554 panels) or reducing rBV (right panels), with blue indicating benefit with LBP treatment and red indicating
1555 benefit with placebo. *P*-values adjusted for multiple testing using the Benjamini-Hochberg correction of
1556 the analysis of deviance test for heterogeneity in treatment effects when stratifying participants by CST

1557 at the pre-MTZ visit. The centroid definition of CST IV-A includes *Ca*. Lachnocurva vaginae (BVAB1) as
1558 a predominant taxon while CST IV-B has higher abundance of *Gardnerella* and both share moderate
1559 *Prevotella* abundance (France et al., 2020).

1560 **Figure S6**



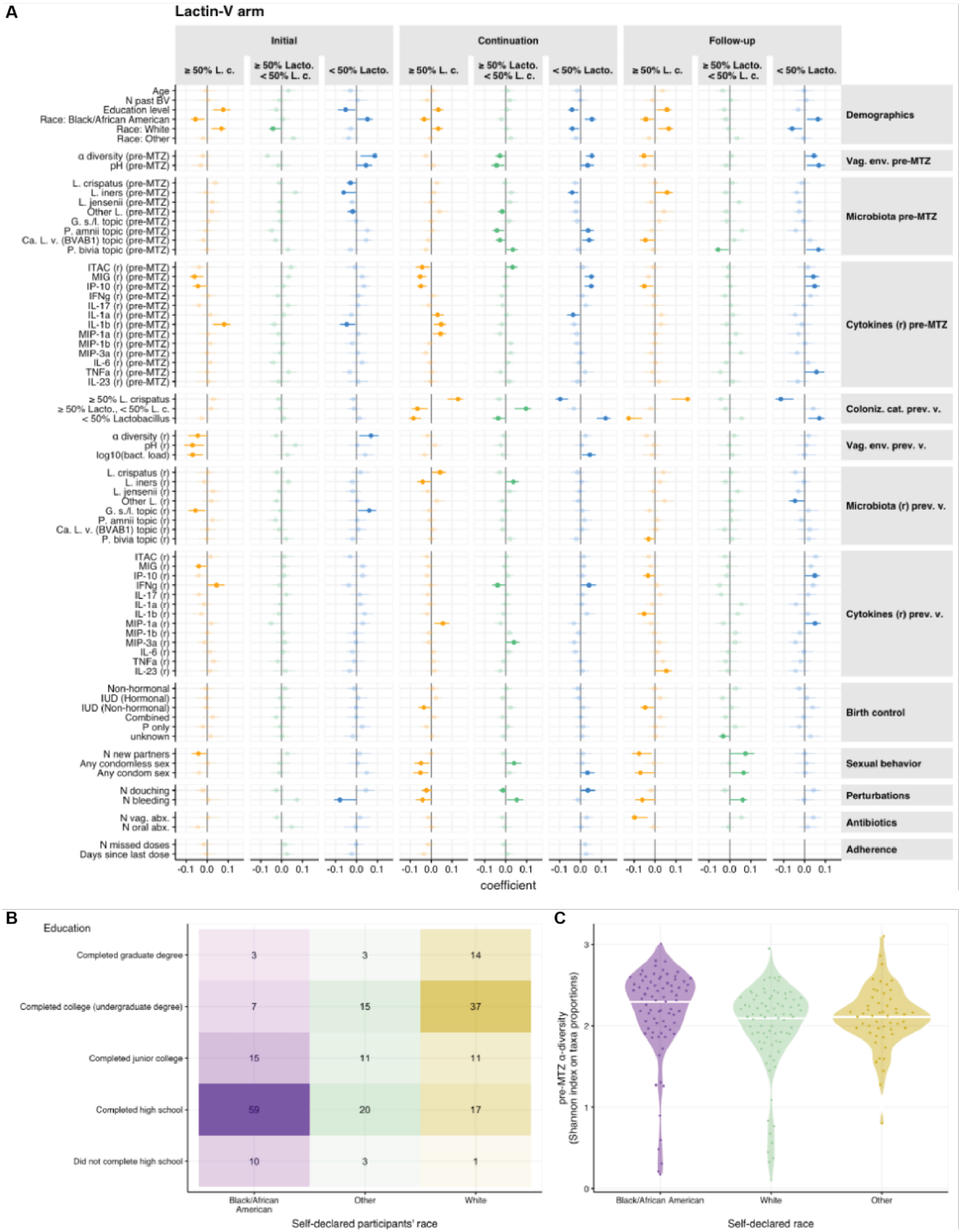
1561

1562 A. Diagram depicting MB-PLS-DA nested models. To assess whether blocks contributed to
1563 explaining colonization, we compared a series of nested models (far-right labels). The full model
1564 includes all blocks and variables, while the “self-predicting” model solely includes the microbiota
1565 categories at the previous visit.

B. Number of latent component selection by cross-validation for the initial phase (left), continuation phase (middle), and follow-up phase models (S6A). Mean (dots) and 95% CI (vertical segment) of the average F1 score (y-axis) obtained when including k latent components (x-axis) on 20 random calibration (black) or validation (turquoise) set. For reference, values obtained for random permutations of the response labels are shown by the horizontal gray line and band. The number of selected latent components is shown by the light turquoise vertical band (here k = 2 for all phases).

C. Contribution of nested models for the initial, continuation, and follow-up phase models. Mean and 95%CI of the average F1 score (y-axis) on the calibration (black, “c”), validation (turquoise, “v”), or for random predictions (gray, “r”) for each nested model (x-axis) including the number of latent components shown on the corresponding upper-panel.

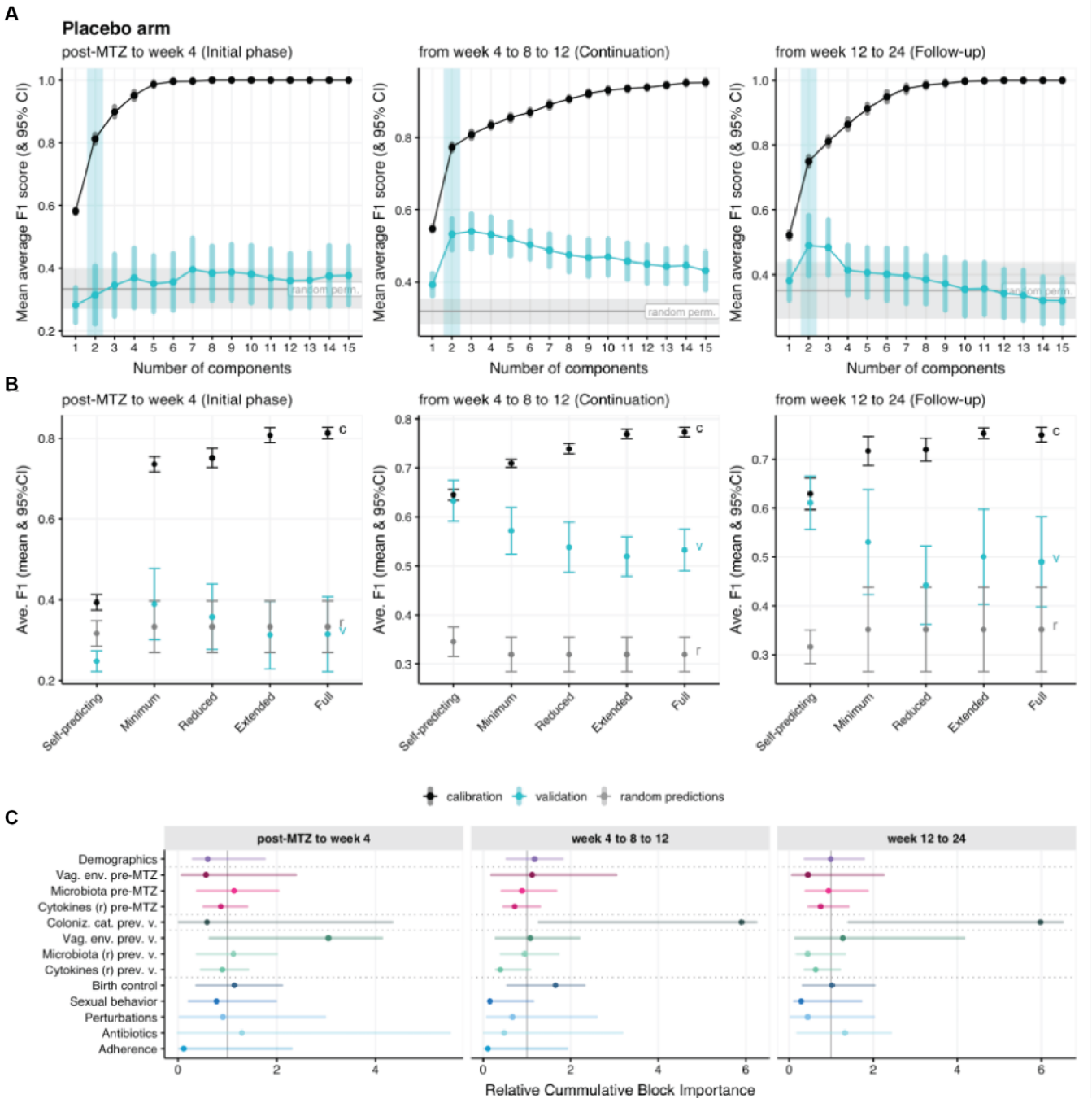
1577 **Figure S7: Variables associated with microbiota categories in LACTIN-V recipients**


A
Lactin-V arm
Initial
Continuation
Follow-up
≥ 50% L. c.
≥ 50% Lacto. < 50% L. c.
< 50% Lacto.
Age
N past BV
Education level
Race: Black/African American
Race: White
Race: Other
Demographics
α diversity (pre-MTZ)
pH (pre-MTZ)
Vag. env. pre-MTZ
L. crispatus (pre-MTZ)
L. iners (pre-MTZ)
L. jensenii (pre-MTZ)
Other L. (pre-MTZ)
G. s./l. topic (pre-MTZ)
P. amnii topic (pre-MTZ)
Ca. L. v. (BVAB1) topic (pre-MTZ)
P. bivia topic (pre-MTZ)
Microbiota pre-MTZ
ITAC (r) (pre-MTZ)
MIG (r) (pre-MTZ)
IP-10 (r) (pre-MTZ)
IFNg (r) (pre-MTZ)
IL-17 (r) (pre-MTZ)
IL-1a (r) (pre-MTZ)
IL-1b (r) (pre-MTZ)
MIP-1a (r) (pre-MTZ)
MIP-1b (r) (pre-MTZ)
MIP-3a (r) (pre-MTZ)
IL-6 (r) (pre-MTZ)
TNFa (r) (pre-MTZ)
IL-23 (r) (pre-MTZ)
Cytokines (r) pre-MTZ
≥ 50% L. crispatus
≥ 50% Lacto., < 50% L. c.
< 50% Lactobacillus
Coloniz. cat. prev. v.
α diversity (r)
pH (r)
log10(bact. load)
Vag. env. prev. v.
L. crispatus (r)
L. iners (r)
L. jensenii (r)
Other L. (r)
G. s./l. topic (r)
P. amnii topic (r)
Ca. L. v. (BVAB1) topic (r)
P. bivia topic (r)
Microbiota (r) prev. v.
ITAC (r)
MIG (r)
IP-10 (r)
IFNg (r)
IL-17 (r)
IL-1a (r)
IL-1b (r)
MIP-1a (r)
MIP-1b (r)
MIP-3a (r)
IL-6 (r)
TNFa (r)
IL-23 (r)
Cytokines (r) prev. v.
Non-hormonal
IUD (Hormonal)
IUD (Non-hormonal)
Combined
P only
unknown
Birth control
N new partners
Any condomless sex
Any condom sex
Sexual behavior
N douching
N bleeding
Perturbations
N vag. abx.
N oral abx.
Antibiotics
N missed doses
Days since last dose
Adherence
-0.1 0.0 0.1
coefficient
B
Education
Completed graduate degree
Completed college (undergraduate degree)
Completed junior college
Completed high school
Did not complete high school
3 3 14
7 15 37
15 11 11
59 20 17
10 3 1
Black/African American
Other
White
Self-declared participants' race
C
pre-MTZ α-diversity (Shannon index on taxa proportions)
0
1
2
3
Black/African American
White
Other
Self-declared race

1578

1579 A. Coefficient of association (dots, x-axis) and associated bootstrapped 95% CI (horizontal
1580 segments) between explanatory variables (y-axis) and microbiota categories (colors, horizontal
1581 panels, see **Figure 6A**) for each phase of the trial (horizontal panels) for MB-PLS-DA models
1582 fitted on LBP recipient data (**Figure S6A**). Variables whose 95% CI include 0 are shown in
1583 transparent/lighter shades. Variables are shown grouped according to their assigned thematic
1584 blocks (**Figure 6A**, **Table S6**).
1585 B. Self-reported education levels are significantly associated with self-reported race (χ^2 *p*-value <
1586 0.01)
1587 C. Pre-MTZ α-diversity is higher in Black or African American participants than in White participants
1588 (*p*-value of *t*-test on association coefficient < 0.05).

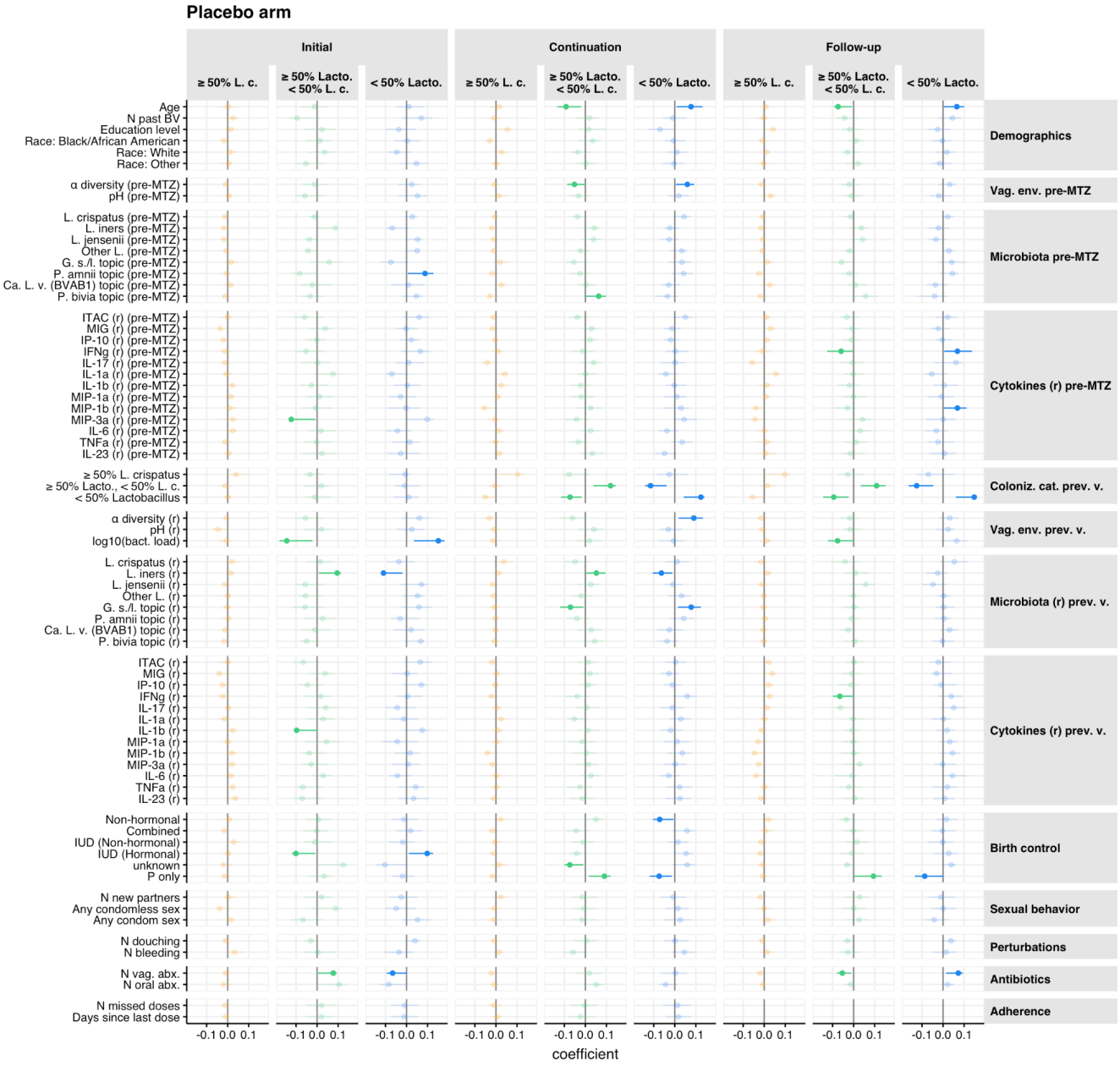
Figure S8: Multiblock analysis of factors associated with microbiota categories in placebo recipients



A. Number of latent component selection by cross-validation for the initial (left), continuation (middle), and follow-up phase (right) model fitted on the placebo arm data. Reads as **Figure S6B**.

1594 B. Contribution of nested models fitted on the placebo arm data for the initial (left), continuation
1595 (missing) and follow-up phase (right). Reads as **Figure S6C.**
1596 C. Relative cumulative importance indices in the placebo arm (x-axis, Methods) for each block (y-
1597 axis, color) for the initial phase model (post-MTZ to Week 4; left), continuation phase model (Week
1598 4 to 8 to 12; middle), and follow-up phase model (Week 12 to 24; right), analogous to portrayal of
1599 LBP arm in **Figure 6A**.

Figure S9: Variables associated with microbiota categories in placebo recipients



Coefficient of association (dots, x-axis) and associated bootstrapped 95% CI (horizontal segments) between explanatory variables (y-axis) and microbiota categories (colors, horizontal panels) for each phase of the trial (horizontal panels) when MB-PLS-DA models are fitted on the placebo arm participants' data. Variables whose 95% CI include 0 (suggesting no statistically significant associations) are shown

in transparent/lighter shades. Variables are shown grouped according to their assigned thematic blocks (**Figure 6A**, **S8C Table S6**).