Vaginal microbiome and metabolome highlight specific signatures of bacterial vaginosis

1 **Vaginal microbiome and metabolome highlight specific signatures of bacterial vaginosis**

2

3

4 Beatrice Vitali · Federica Cruciani · Gianfranco Picone · Carola Parolin · Gilbert Donders · Luca Laghi

5

6 Beatrice Vitali · Federica Cruciani · Carola Parolin

7 Department of Pharmacy and Biotechnology, University of Bologna, Via San Donato 19/2, 40127

8 Bologna, Italy

9

10 Gianfranco Picone · Luca Laghi

11 Department of Agro-Food Science and Technology, University of Bologna, P.za Goidanich 60, 47522

12 Cesena, Italy

13

14 Gilbert Donders

15 Department of Obstetrics and Gynecology, General Hospital Heilig Hart, Kliniekstraat 45, 3300

16 Tienen, and University Hospital Antwerp, Wilrijkstraat 10, 2650 Edegem, Belgium

17

18 Corresponding author: Beatrice Vitali, Department of Pharmacy and Biotechnology, University of

19 Bologna, Via San Donato 19/2, 40127, Bologna, Italy.

20 Tel: +39 051 2088750; Fax: +39 051 2099734

21 E-mail: b.vitali@unibo.it

22

23 Federica Cruciani present address: CRA-CER S.S. 673 Km 25+200, 71122 Foggia, Italy.

Abstract

Purpose. In the present study we sought to find novel bacterial and metabolic hallmarks for bacterial vaginosis (BV).

Methods. We studied the vaginal microbiome and metabolome of vaginal fluids from BV-affected patients ($n = 43$) and healthy controls ($n = 37$) by means of an integrated approach based on quantitative PCR (qPCR) and proton Nuclear Magnetic Resonance (^{1}H-NMR). The correlations between the clinical condition and vaginal bacterial communities were investigated by principal component analysis (PCA). To define the metabolomics signatures of BV, 100 discriminant analysis by projection on latent structure (PLS-DA) models were calculated.

Results. Bacterial signatures distinguishing the health condition and BV were identified by qPCR. *L. crispatus* strongly featured the healthy vagina while increased concentrations of *Prevotella*, *Atopobium* and *Mycoplasma hominis* specifically marked the infection. ^{1}H-NMR analysis has led to the identification and quantification of 17 previously unreported molecules. BV was associated with changes in the concentration of metabolites belonging to the families of amines, organic acids, short chain fatty acids, amino acids, nitrogenous bases and monosaccharides. In particular, maltose, kynurenine and NAD+ primarily characterized the healthy status while nicotinate, malonate and acetate were the best metabolic hallmarks of BV.

Conclusions. This study helps to better understand the role of the vaginal microbiota and metabolome in the development of BV infection. We propose a molecular approach for diagnosis of BV based on quantitative detection in vaginal fluids of *Atopobium*, *Prevotella* and *Mycoplasma hominis*, and nicotinate, malonate and acetate by combining qPCR and ^{1}H-NMR.

Keywords: vaginal microbiome; vaginal metabolome; bacterial vaginosis; lactobacilli; qPCR; ^{1}H-NMR

Introduction

The microbiota of the human vagina can significantly impact the health of women, their fetuses and newborn infants [1]. The vaginal microbiota of healthy reproductive age women is dominated by *Lactobacillus* species. Five distinct vaginal bacterial biotypes, characterized by the dominance of *L. crispatus*, *L. gasseri*, *L. iners*, *L. jensenii*, or an increased proportion of other strictly anaerobic bacteria, were described [2, 3]. Lactobacilli play key protective roles through different mechanisms, such as production of various antibacterial compounds, co-aggregation, competitive exclusion and immunomodulation [4-6].

Bacterial vaginosis (BV) is a common polymicrobial disorder of the vaginal microbiota characterized by loss of lactobacilli and increasing numbers of anaerobes and gram-negative rods [7, 8]. BV is associated with adverse outcomes, such as ascending reproductive tract infections, enhanced acquisition of HIV and other sexually transmitted diseases, spontaneous abortion and preterm birth [9-12]. Both a defined etiology and optimal treatment strategies for BV have remained elusive.

BV is typically diagnosed using either the Nugent scoring method [13] that examines bacterial composition via a Gram smear or the Amsel criteria [14] that considers factors such as presence of discharge, amine production, presence of clue cells and a vaginal pH greater than 4.5.

The microbiology of BV has been better characterized through microbiome studies based on 16S rRNA gene-directed PCR assays [15]. Although these approaches are able to provide a comprehensive understanding of the bacterial community membership, they are not able to determine the changes occurring in the vaginal environment at a metabolic level. Metabolomics analyzes complex systems, using high-throughput analytical methods, such as NMR spectroscopy that allows robust and sensitive identification of metabolites produced by microbes and host cells. This tool allows researchers to determine the effects caused by perturbations on the host's metabolic profile by analyzing the presence

71 and quantity of thousands of metabolites simultaneously. Metabolites that are significantly affected by

72 experimental variables can be identified by multivariate statistics and placed into the larger context of

73 how the host was affected overall [16-19].

74 In the perspective to discern potential novel determinants of BV, we investigated the relationship

75 between bacterial community composition and metabolic profiles of healthy and BV-associated vaginal

76 ecosystems by means of quantitative PCR (qPCR) of bacterial 16S rRNA genes and proton Nuclear

77 Magnetic Resonance (^{1}H-NMR)-based metabolomics.

78

79 **Materials and Methods**

80

81 Study participants and sample preparation

82

83 Subjects for this study were a cohort of 80 Belgian pre-menopausal, non-pregnant women, aged

84 between 18 and 50 years (mean age: 37). The enrolled women belonged to two groups: BV-affected

85 patients ($n = 43$) presenting positivity for at least three of four Amsel's criteria and a Nugent score > 3,

86 and age-matched healthy subjects ($n = 37$) who had no signs of vaginal tract infection and had never

87 had BV. All women tested negative to *Candida*, sexually transmitted infections and abnormal findings

88 on cervical Pap smears. They signed an informed consent in accordance with the approval and

89 guidelines of the Ethics committee of the Heilig Hart Hospital of Tienen, Belgium.

90 Standardized vaginal fluids were collected by flushing and re-aspirating 2 ml of saline through a 22-

91 gauge needle in the left, central, and right upper vaginal vaults [20] and stored at -80°C until use.

92 Vaginal samples were centrifuged at $9,500 \times g$ for 15 min. The supernatants were used for

93 metabolomic analysis, while the pellets were processed for DNA isolation by using a DNeasy blood

94 and tissue kit (Qiagen, Hilden, Germany) [21, 22]. DNA amount was quantified using NanoDrop ND-

95   1000 (NanoDrop® Technologies, Wilmington, DE).

96

97   Quantitative PCR (qPCR)

98

99   qPCR was performed on DNA samples extracted from the vaginal fluids using a LightCycler

100   instrument (Roche, Mannheim, Germany) and SYBR Green I as the reporter fluorophore. Genus- or

101   species- specific primer sets targeted to 16S rRNA gene or 16S-23S rRNA spacer region were used to

102   amplify bacteria belonging to *L. crispatus* [23], *L. iners* [24], *L. gasseri* [25], *L. jensenii* [24],

103   *Gardnerella vaginalis* [26], *Atopobium* [27], *Prevotella* [28], *Veillonella* [29], *Mycoplasma hominis*

104   [23] and *Mobiluncus* [30].

105   Amplifications were carried out in a final volume of 20 μl containing each primer at 0.5 μM, 4 μl of

106   LightCycler-FastStart DNA Master SYBR green I (Roche) and 2 μl of template. The thermal cycling

107   conditions were optimized in the present study (Table S1, supplemental material). DNAs extracted

108   from *L. crispatus* DSM 20584, *L. gasseri* DSM 20243, *L. jensenii* DSM 20557, *L. iners* DSM 13335,

109   *Gardnerella vaginalis* ATCC 14018, *Prevotella bivia* ATCC 29303, *Veillonella parvula* ATCC 10790,

110   *Atopobium vaginae* ATCC BAA-55, *Mycoplasma hominis* DSM 19104, and *Mobiluncus curtisii* ATCC

111   43063 were used as standards for PCR quantification. The efficiency of qPCR was between 90% and

112   100%, and correlation coefficients for genomic DNA standards were > 0.99. The sensitivity of qPCR

113   assays ranged from 10^{-4} to 10^{-1} ng of target DNA. Melting curve analysis was carried out to confirm

114   that the PCR products from vaginal samples and standards had identical melting points.

115   DNA samples were amplified in triplicate for each primer set. Data were expressed as log ng of DNA

116   of the target genus or species per μg of total DNA extracted from the vaginal sample.

117

118   ^{1}H-NMR analysis

119

120   One ml of vaginal supernatant was added to 160 μl of a D_2O solution of 3-(trimethylsilyl)-propionic-

121   2,2,3,3-d4 acid sodium salt (TSP) 6.25 mM and the pH was adjusted to 7.00 with the addition of HCl or

122   NaOH (0.5 M). ^{1}H-NMR spectra were recorded at 298 K with an AVANCE spectrometer (Bruker,

123   Milan, Italy) operating at a frequency of 600.13 MHz, equipped with an autosampler with 60 holders

124   [17]. Each spectrum was acquired using 32K data points over a 7,211.54 Hz spectral width and adding

125   256 transients. A recycle delay of 5 s and a 90° pulse of 11.4 s were set up. Acquisition time (2.27 s)

126   and recycle delay were adjusted to be 5 times longer than the longitudinal relaxation time of the protons

127   under investigation, which was considered to be not longer than 1.4 s.

128   ^{1}H-NMR spectra baseline was adjusted by means of the simultaneous peak detection [31] and baseline

129   correction algorithm (SPDBC) implemented in the baseline R package [32]. Signals misalignments

130   were compensated by i-Coshift algorithm [33]. Differences in water concentration were compensated

131   by probabilistic quotient normalization procedure (PQN) [34]. The signals with an intensity of at least

132   five times that of noise were assigned by comparing their chemical shift and multiplicity with the

133   literature [17] and Chenomx software data bank (Chenomx Inc., Canada, ver 8.1). Their area was

134   employed for molecular quantification.

135

136   Statistical analysis

137

138   Statistical analyses were performed using R computational language [35]. Similarities among the

139   analyzed samples, as well as trends in the bacterial communities or metabolome profiles, were

140   investigated by means of principal components analysis (PCA) applied to the mean centered data.

141   Differences in bacterial amounts, metabolites concentrations and PCA data were analyzed using

142   Wilcoxon's signed rank test and Anova based on Tukey contrast [36]. A *P* value < 0.05 was considered

143 statistically significant. In addition, to better define the metabolomics signatures of BV, 100

144 discriminant analysis by projection on latent structure (PLS-DA) models in their sparse version were

145 calculated by means of mixOmics R package [37]. PLS-DA algorithm builds linear models with each

146 molecule concentration in order to maximize the possibility to predict BV presence. In its sparse

147 version, the most informative molecules are progressively added to the model until the best

148 compromise between prediction ability and model simplicity is reached. The robustness of the models

149 was expressed in terms of correct classification rate (CCR), that is the ratio between sum of true

150 positives and negatives predictions and the total number of samples.

151

152 **Results**

153

154 Distribution of bacterial communities

155

156 A qualitative analysis of the microbiota composition in healthy and BV-affected women is represented

157 in Fig. 1. The frequency of occurrence of *Lactobacillus* species and BV-related bacteria represented the

158 first element to clearly distinguish the two vaginal ecosystems (Fig. 1a). Healthy women primarily

159 harboured lactobacilli. Among these, *L. crispatus* was the most frequently detected bacterium

160 (70.27%), followed by *L. gasseri* (45.94%), *L. iners* (24.32%) and *L. jensenii* (18.92%). The detection

161 rate of *L. iners* was higher in BV-affected women (62.79%) than in healthy controls. The most

162 frequently BV-related bacteria detected in healthy women were *Prevotella* (67.57%), *M. hominis*

163 (62.16%) and *Atopobium* (54.05%). *G. vaginalis* was found in a small percentage (8.11%), while

164 *Veillonella* and *Mobiluncus* were not detected. BV patients possessed a diverse array of bacteria, with

165 the most frequent genera and species being *Atopobium* (100%), *G. vaginalis* (97.67%), *Prevotella*

166 (97.67%) and *M. hominis* (83.72%). Despite *Veillonella* (41.86%) and *Mobiluncus* (23.26%) have been

8

167   found at a lower frequency than the other BV-related bacteria, their presence seemed to indicate

168   unequivocally the existence of BV disorder. Differences in frequency of occurrence between healthy

169   and BV-affected women were significant for all bacterial groups ($P < 0.05$), except for *L. jensenii*.

170   The relative abundance of the main *Lactobacillus* species associated with human vaginal mucosa has

171   been explored as an additional criterion for distinguishing between health and BV (Fig. 1b). The

172   majority of women was characterized by the presence of a single *Lactobacillus* species. For women

173   who harboured more than one *Lactobacillus* species, the predominant species was identified based on

174   the bacterial amount calculated by qPCR. The vaginal microbiota of the totality of healthy women was

175   colonized by at least one *Lactobacillus* species. *L. crispatus* (59%) highly prevailed over other

176   lactobacilli, followed by *L. iners* (22%), *L. gasseri* (16%), and *L. jensenii* (3%). Unlike health

177   condition, the vaginal microbiota of a fair number of BV patients was not colonized by any of the four

178   main *Lactobacillus* species (N: 33%) and *L. iners* was the predominant species (63%). The microbiota

179   of a low percentage of BV-affected women was dominated by *L. gasseri* (5%), while *L. jensenii* and *L.*

180   *crispatus* were never predominant.

181

182   Quantification of *Lactobacillus* species and BV-related bacteria

183

184   The qPCR results are summarized in box blots representing concentrations of specific DNA of

185   *Lactobacillus* species and BV-related bacteria in vaginal samples of healthy controls and BV patients

186   (Fig. 2).

187   Regarding lactobacilli, only *L. crispatus* showed a significant difference between healthy and BV

188   women: the amount of this species was higher in healthy group (7.75 log ng/µg) than in BV group

189   (6.47 log ng/µg) ($P < 0.0001$). As already reported [38], *L. iners* had the highest median concentration

190   of any assayed species/genus in samples from both healthy (8.49 log ng/µg) and BV (8.58 log ng/µg)

191   women.

192   With respect to BV-related bacteria, the median concentrations of *Atopobium* (H: 0.13 log ng/µg; BV:

193   1.96 log ng/µg; *P* < 0.0001), *Prevotella* (H: -1.24 log ng/µg; BV: 1.20 log ng/µg; *P* < 0.0004) and *M.*

194   *hominis* (H: -1.22 log ng/µg; BV: 0.67 log ng/µg; *P* < 0.0413) were significantly higher in BV patients

195   compared to healthy controls. The amount of *G. vaginalis* did not vary significantly in the two groups

196   of women. *Veillonella* (-1.07 log ng/µg) and *Mobiluncus* (1 log ng/µg) were identified only in BV cases

197   at low concentrations.

198

199   Metabolic profiles

200

201   We sought a metabolic description of BV compared to health condition. 55 signals ascribable to single

202   molecules were identified (Table S2, supplemental material). 44 of these molecules were identified and

203   quantified, while other 6 were identified, but not reliably quantified, due to signals superimpositions or

204   artifacts introduced by signal alignment procedure. Concerning citrate, BV condition was not

205   associated to a variation in concentration, but to a shift in signals in the spectrum, a well-known

206   behavior caused by oscillations in divalent cations concentrations [39]. We identified and quantified 17

207   previously unreported small molecules, including the organic acids malonate and isovalerate, the

208   amino acids histidine, taurine and aspartate, the nitrogenous bases NAD+ and inosine.

209   Among the identified and quantified molecules, 32 significantly varied in concentration between

210   healthy and BV women (*P* < 0.05). These metabolites belong to the families of amines, organic acids,

211   short chain fatty acids (SCFAs), amino acids, nitrogenous bases and monosaccharides (Table 1). BV

212   was associated with a general increase of amines, organic acids and various SCFAs. BV was also

213   characterized by a general decrease of protein amino acids, with the exception of proline. Within the

214   other categories of metabolites no uniform trend of variation was observed.

215

216   Microbiome-metabolome correlation

217

218   The correlations between the clinical condition and vaginal bacterial communities have been studied by

219   means of a PCA built on the qPCR data related to *Lactobacillus* species and BV-related bacteria (Fig.

220   S1, supplemental material). The first two PCs accounted for the 32.3% of the whole variance of the

221   investigated samples.

222   Changes in bacterial populations were highlighted by plotting the medians of the two groups of women

223   (H and BV). H and BV groups were significantly different ($P < 0.05$), showing that most of the

224   variability was due to the onset of BV condition. BV samples were more widely distributed in the

225   multidimensional space compared to healthy samples, indicating a greater inter-individual variability of

226   microbiome and metabolome, as previously hypothesized [17, 40]. The vaginal microbiota of control

227   women was dominated by *L. crispatus*, *L. gasseri* and *L. jensenii* with the former most strongly

228   characterizing the state of health. *L. iners* was shifted toward the disease condition, suggesting its

229   correlation with the development of BV. The bacterial groups that primarily featured women with BV

230   were *Prevotella* and *Atopobium*, followed by *M. hominis*. Conversely, *Mobiluncus* and *Veillonella*,

231   which were identified in small quantities only in presence of BV, played a minor role as biomarkers of

232   the disease. The results of PCA show that concentration is a relevant factor in determining a distinctive

233   feature of BV in terms of diagnostic applications. *G. vaginalis* did not appear as one of the species most

234   closely associated with BV.

235   The implications of BV on vaginal metabolome were investigated by building 100 sPLS-DA models on

236   the metabolites that significantly varied in BV women compared to healthy controls (Fig. 3). At least in

237   90 times the variable reduction algorithm selected maltose, kynurenine, nicotinate, malonate, acetate

238   and NAD$^+$ as the most important metabolites in discriminating vaginal health from BV. In particular,

239   maltose, kynurenine and NAD+ were distinctive of healthy condition while nicotinate, malonate, and

240   acetate appeared to be specific hallmarks of BV. These molecules were also characterized by median

241   variable influence on projections (VIP) coefficients higher than 1, feature that was shared with

242   sarcosine and phenylalanine. PLS-DA model built on these 6 molecules only gave a percentage of 0.11

243   ± 0.42 (mean ± standard deviation) of misclassified samples pertaining to BV-affected women, and no

244   misclassified test samples pertaining to healthy women. This optimal performance was obtained in all

245   cases, including women dominated by *L. iners*. This means that, despite these women presented a BV-

246   like microbiome, the conditions regulating the concentration of maltose, kynurenine, nicotinate,

247   malonate, acetate and NAD$^+$ represent a common trait of BV onset and could be reliably used as

248   biomarkers and to better understand the biochemical bases of this disease.

249

250   Correlation between *Lactobacillus* species, BV-related bacteria and metabolites

251

252   In order to go into deeper detail about the health-promoting potential of *L. crispatus*, *L. iners*, *L.*

253   *gasseri* and *L. jensenii*, the correlations between the dominance of each *Lactobacillus* species and the

254   vaginal microbiome and metabolome were analyzed by PCA (Fig. S2, supplemental material). In the

255   biplot describing the correlations between the dominant *Lactobacillus* species and BV-related bacteria,

256   PC1 and PC2 accounted for the 47.4% of the whole variance of the investigated samples (Fig. S2a).

257   Medians of the four healthy groups were superimposed while the median of BVn was significantly

258   different from the medians of BVi and BVg ($P < 0.05$), showing that the variability of BV samples was

259   not only associated to the development of BV but also to the presence of lactobacilli. The more

260   different microbiota from health condition was found in BVn, followed by BVi. An intermediate

261   position in the two-dimensional space was occupied by BVg.

262   The six molecules best discriminating healthy condition from BV allowed to obtain a visual

263   correlations between the predominance of each *Lactobacillus* species and metabolome, as depicted in

264   Fig. S2b. PC1 and PC2 accounted for the 67.9% of the whole variance of the investigated samples. The

265   distribution in the PC plane of the samples from BV-affected women closely reproduced the trends

266   observed in Fig. S2a, with BVg samples appearing as most similar to the healthy samples, BVn

267   samples most different, and *L. iners* prevalence giving rise on average to an intermediate condition.

268

269   **Discussion**

270

271   In the present study we sought to identify correlations between the vaginal colonization of certain

272   bacterial populations and metabolic profiles, in order to find novel hallmarks of BV.

273   Our results support previous studies that BV is associated with dramatic compositional changes in the

274   vaginal microbiota, i.e. depletion of lactobacilli in conjunction with colonization of many diverse

275   bacteria, mainly strict anaerobes [7, 20, 39, 41, 42]. Healthy women primarily harbored lactobacilli: *L.*

276   *crispatus* was the most frequent species, detected at significantly higher concentration with respect to

277   BV patients. The frequency of occurrence of *L. iners* was higher in BV-affected women, confirming the

278   hypothesis that this species is common and abundant in vaginal communities characterized by high

279   concentrations of non-lactobacilli [22, 42]. Other studies of perturbed vaginal microflora suggest that *L.*

280   *iners* might be a transitional species, colonizing after disturbances to the vaginal environment [43].

281   Compared to healthy controls, BV patients had in their vaginal microbiota higher prevalence of all the

282   considered BV-related bacteria and significantly higher abundances of *Atopobium*, *Prevotella* and *M.*

283   *hominis*. Because *Atopobium*, *Prevotella* and *M. hominis* were also frequently represented in healthy

284   vagina ($> 50\%$), high levels rather than the mere presence of these bacteria can be considered hallmarks

285   of BV disorder. *G. vaginalis*, which has long been considered distinctive of BV [44], in our study did

286   not seem to be particularly useful for diagnostic purposes. In fact, neither the presence nor an increased

287 concentration of this species were uniquely associated with the disease. The results of multivariate

288 statistical analysis reiterated the role of *L. crispatus* in strongly featuring the healthy vagina and the

289 functions of *Prevotella*, *Atopobium* and *M. hominis* as specific microbio-markers of BV disorder.

290 Conversely, *Mobiluncus* and *Veillonella*, even though highly specific for BV, seemed to play a minor

291 role in marking the disease, probably because the majority of BV-affected women did not harbor these

292 bacteria. It remains to be elucidated whether *Mobiluncus/Veillonella* positive BV expresses a subtype of

293 BV with specific features and pathogenicity. Notably, PCA highlighted the relevance of bacterial

294 concentration in determining a distinctive feature for proper diagnosis of BV.

295 The molecular data descriptive of the vaginal microbiome have been integrated with the ^{1}H-NMR data

296 descriptive of the metabolic activities of the host-bacteria meta-organisms, in order to obtain a

297 comprehensive picture of the evolution of the vaginal ecosystem associated with the infection. This led

298 to the identification and quantification of 17 previously unreported small molecules, including the

299 organic acids malonate and isovalerate, the amino acids histidine, taurine and aspartate, the nitrogenous

300 bases NAD^+ and inosine.

301 BV was associated with changes in several metabolites belonging to the families of amines, organic

302 acids, short chain fatty acids, amino acids, nitrogenous bases and monosaccharides. Most of these

303 molecules are known to be modulated by the balance between lactobacilli and BV-related bacteria, such

304 as biogenic amines, SCFAs and organic acids [17, 45-49]. In particular, we have observed in BV cases

305 a general increase of amines, including tyramine, trimethylamine and cadaverine, which are considered

306 the main responsible for the fishy odor of the vaginal discharge [8, 19]. Various SCFAs were

307 significantly higher in BV-positive women, which is not surprising given that BV is characterized by

308 the outgrowth of mostly anaerobic bacteria. It has been suggested that SCFAs may be involved in

309 recruitment and activation of the innate immune cells in the female genital tract [47]. This is in

310 agreement with the recent finding that BV is associated with an increase of proteins involved in the

311   innate immune response [50]. Organic acids were detected at higher concentrations in the vaginal fluids

312   of BV-positive women, as previously noted especially for acetate and succinate, typical metabolites

313   produced by *Prevotella* and *Mobiluncus* spp. [45]. In the present work, malonate also assumes a

314   particular emphasis being together acetate the organic acids that mostly differ BV from healthy state.

315   BV was also characterized by a decrease of certain protein amino acids, probably due to their

316   decarboxylation to biogenic amines. The high availability of maltose in healthy vagina can be regarded

317   as the principal cause of the overgrowth of lactobacilli. It could be attributed to an efficient synthesis of

318   glycogen and/or secretion of α-amylase by healthy vaginal epithelium. In fact, the human α-amylase is

319   primarily responsible for the digestion of vaginal glycogen, making available glycogen-breakdown

320   products, including maltose [51]. On the other hand, we found an increase of glucose concentration in

321   BV-affected patients. This finding is in agreement with previous studies reporting the association of

322   dietary indices, i.e. glycemic load, with BV acquisition and persistence [52]. The high level of glucose

323   could favor the glycolysis which involves the consumption of NAD^+.

324   In conclusion, our work may help to better understand the role of the vaginal microbiota and

325   metabolome in BV infection. The identification of bacterial and metabolic markers described here is a

326   prerequisite for the design of new diagnostic kits, which, being based on molecular parameters, are

327   likely to be endowed with greater sensitivity and reproducibility compared to the methods of diagnosis

328   currently used in clinical practice (Amsel and Nugent). Specifically, we propose a molecular tool for

329   diagnosis of BV based on quantitative detection in vaginal fluids of *Atopobium*, *Prevotella* and

330   *Mycoplasma hominis*, and malonate, acetate, and nicotinate by combining qPCR and ^{1}H-NMR. The

331   combination of these selected bacterial/ metabolic hallmarks could provide the best diagnostic test in

332   terms of specificity, allowing to correctly diagnose cases of infection reducing false positive rate.

333   Regarding sensitivity, such a combination maintains the peculiarities of qPCR and ^{1}H-NMR

334   techniques, that are able to detect bacterial DNA and metabolites at concentrations below the ng and

335   mmol/l, respectively.

336

339

340   **Conflicts of interest** The authors declare that they have no conflict of interest.

341

342   **Ethical approval** All procedures performed in studies involving human participants were in

343   accordance with the ethical standards of the institutional and/or national research committee and with

344   the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

345

References

346

347   1. Hillier SL, Krohn MA, Cassen E, Easterling TR, Rabe LK, Eschenbach DA (1995) The role of

348   bacterial vaginosis and vaginal bacteria in amniotic fluid infection in women in preterm labor

349   with intact fetal membranes. Clin Infect Dis. 20 Suppl 2:S276-278.

350   2. Drell T, Lillsaar T, Tummeleht L, Simm J, Aaspõllu A, Väin E, Saarma I, Salumets A, Donders

351   GG, Metsis M (2013) Characterization of the vaginal micro- and mycobiome in asymptomatic

352   reproductive-age Estonian women. PLoS One 8(1):e54379.

353   3. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R,

354   Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ (2011) Vaginal

355   microbiome of reproductive-age women. Proc Natl Acad Sci USA 108 Suppl 1:4680-4687.

356   4. Kaewsrichan J, Peeyananjarassri K, Kongprasertkit J (2006) Selection and identification of

357   anaerobic lactobacilli producing inhibitory compounds against vaginal pathogens. FEMS

358   Immunol Med Microbiol 48:75–83.

359   5. Klebanoff SJ, Hillier SL, Eschenbach DA, Waltersdorph AM (1991) Control of the microbial

360   flora of the vagina by H_2O_2-generating lactobacilli. J Infect Dis 164:94–100.

361   6. Reid G, Younes JA, Van der Mei HC, Gloor GB, Knight R, Busscher HJ (2011) Microbiota

362   restoration: natural and supplemented recovery of human microbial communities. Nat Rev

363   Microbiol 9:27-38.

364   7. Fredricks DN, Fiedler TL, Marrazzo JM (2005) Molecular identification of bacteria associated

365   with bacterial vaginosis. N Engl J Med 353:1899-1911.

366   8. Sobel JD (2000) Bacterial vaginosis. Annu Rev Med 51:349-356.

367   9. Atashili J, Poole C, Ndumbe PM, Adimora AA, Smith JS (2008) Bacterial vaginosis and HIV

368   acquisition: a meta-analysis of published studies. AIDS 22:1493-1501.

369   10. Brotman RM, Klebanoff MA, Nansel TR, Yu KF, Andrews WW, Zhang J, Schwebke JR (2010)

370 Bacterial vaginosis assessed by gram stain and diminished colonization resistance to incident

371 gonococcal, chlamydial, and trichomonal genital infection. J Infect Dis 202:1907-1915.

372 11. Donders GG, Van Bulck B, Caudron J, Londers L, Vereecken A, Spitz B (2000) Relationship of

373 bacterial vaginosis and mycoplasmas to the risk of spontaneous abortion. Am J Obstet Gynecol

374 183:431-437.

375 12. Hillier SL, Martius J, Krohn M, Kiviat N, Holmes KK, Eschenbach DA (1988) A case-control

376 study of chorioamnionic infection and histologic chorioamnionitis in prematurity. N Engl J Med

377 319:972-978.

378 13. Nugent RP, Krohn MA, Hillier SL (1991) Reliability of diagnosing bacterial vaginosis is

379 improved by a standardized method of gram stain interpretation. J Clin Microbiol 29:297–301.

380 14. Amsel R, Totten PA, Spiegel CA, Chen KC, Eschenbach D, Holmes KK (1983) Nonspecific

381 vaginitis. Diagnostic criteria and microbial and epidemiologic associations. Am J Med 74:14-

382 22.

383 15. Lamont RF, Sobel JD, Akins RA, Hassan SS, Chaiworapongsa T, Kusanovic JP, Romero R

384 (2011) The vaginal microbiome: new information about genital tract flora using molecular

385 based techniques. BJOG 118:533-549.

386 16. Coen M, O'Sullivan M, Bubb WA, Kuchel PW, Sorrell T (2005) Proton nuclear magnetic

387 resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. Clin Infect

388 Dis 41:1582-1590.

389 17. Laghi L, Picone G, Cruciani F, Brigidi P, Calanni F, Donders G, Capozzi F, Vitali B (2014)

390 Rifaximin modulates the vaginal microbiome and metabolome in women affected by bacterial

391 vaginosis. Antimicrob Agents Chemother 58:3411–3420.

392 18. Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L,

393 Fernie AR (2003) Parallel analysis of transcript and metabolic profiles: a new approach in

394   systems biology. EMBO Rep 4:989-993.

395   19. Yeoman CJ, Thomas SM, Miller ME, Ulanov AV, Torralba M, Lucas S, Gillis M, Cregger M,

396   Gomez A, Ho M, Leigh SR, Stumpf R, Creedon DJ, Smith MA, Weisbaum JS, Nelson KE,

397   Wilson BA, White BA (2013) A multi-omic systems-based approach reveals metabolic markers

398   of bacterial vaginosis and insight into the disease. PLoS One 8(2):e56111.

399   20. Vitali B, Pugliese C, Biagi E, Candela M, Turroni S, Bellen G, Donders GG, Brigidi P (2007)

400   Dynamics of vaginal bacterial communities in women developing bacterial vaginosis,

401   candidiasis, or no infection, analyzed by PCR-denaturing gradient gel electrophoresis and real-

402   time PCR. Appl Environ Microbiol 73:5731-5741.

403   21. Vitali B, Biagi E, Brigidi P (2012) Protocol for the use of PCR-denaturing gradient gel

404   electrophoresis and quantitative PCR to determine vaginal microflora constitution and

405   pathogens in bacterial vaginosis. In: MacKenzie CR, Henrich B (eds.) Diagnosis of Sexually

406   Transmitted Diseases. Springer, New York, vol 903, pp 177-193.

407   22. Cruciani F, Brigidi P, Calanni F, Lauro V, Tacchi R, Donders G, Peters K, Guaschino S, Vitali B

408   (2012) Efficacy of rifaximin vaginal tablets in treatment of bacterial vaginosis: a molecular

409   characterization of the vaginal microbiota. Antimicrob Agents Chemother 56:4062-4070.

410   23. Zozaya-Hinchliffe M, Lillis R, Martin DH, Ferris MJ (2010) Quantitative PCR assessments of

411   bacterial species in women with and without bacterial vaginosis. J Clin Microbiol 48:1812-

412   1819.

413   24. De Backer E, Verhelst R, Verstraelen H, Alqumber MA, Burton JP, Tagg JR, Temmerman M,

414   Vaneechoutte M (2007) Quantitative determination by real-time PCR of four vaginal

415   *Lactobacillus* species, *Gardnerella vaginalis* and *Atopobium vaginae* indicates an inverse

416   relationship between *L. gasseri* and *L. iners*. BMC Microbiol 7:115.

417   25. Byun R, Nadkarni MA, Chhour KL, Martin FE, Jacques NA, Hunter N (2004) Quantitative

418   analysis of diverse *Lactobacillus* species present in advanced dental caries. J Clin Microbiol 42:

419   3128-3136.

420   26. Zariffard MR, Saifuddin M, Sha BE, Spear GT (2002) Detection of bacterial vaginosis-related

421   organisms by real-time PCR for Lactobacilli, *Gardnerella vaginalis* and *Mycoplasma hominis.*

422   FEMS Immunol Med Microbiol 34: 277-281.

423   27. Matsuki T, Watanabe K, Fujimoto J, Takada T, Tanaka R (2004) Use of 16S rRNA gene-targeted

424   group-specific primers for Real-Time PCR analysis of predominant bacteria in human feces.

425   Appl Environ Microbiol 70:7220-7228.

426   28. Matsuki T, Watanabe K, Fujimoto J, Miyamoto Y, Takada T, Matsumoto K, Oyaizu H, Tanaka R

427   (2002) Development of 16S rRNA-gene-targeted group-specific primers for the detection and

428   identification of predominant bacteria in human feces Appl Environ Microbiol 68:5445-5451.

429   29. Rinttilä T, Kassinen A, Malinen E, Krogius L, Palva A (2004) Development of an extensive set

430   of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal

431   samples by real-time PCR. J Appl Microbiol 97:1166-1177.

432   30. Tiveljung A, Forsum U, Monstein HJ (1996) Classification of the genus *Mobiluncus* based on

433   comparative partial 16S rRNA gene analysis. Int J Syst Bacteriol 46:332-336.

434   31. Coombes KR, Fritsche HA, Clarke C, Chen J-N, Baggerly KA, Morris JS, Xiao LC, Hung MC,

435   Kuerer HM (2003) Quality control and peak finding for proteomics data collected from nipple

436   spirate fluid by surface-enhanced laser desorption and ionization. Clin Chem 49:1615-1623.

437   32. Liland KH, Almøy T, Mevik BH (2010) Optimal choice of baseline correction for multivariate

438   calibration of spectra. Appl Spectrosc 64:1007-1016.

439   33. Savorani F, Tomasi G, Engelsen SB (2010) icoshift: A versatile tool for the rapid alignment of

440   1D NMR spectra. J Magn Reson 202:190-202.

441   34. Dieterle F, Ross A, Schlotterbeck G, Senn H. (2006) Probabilistic quotient normalization as

442   robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR

443   metabonomics. Anal Chem 78:4281-4290.

444   35. Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comput Graph

445   Stat 5:299-314.

446   36. Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models.

447   Biom J 50:346-363.

448   37. Yao F1, Coquery J, Lê Cao KA (2012) Independent Principal Component Analysis for

449   biologically meaningful dimension reduction of large biological data sets. BMC Bioinformatics

450   13:24.

451   38. Shipitsyna E, Roos A, Datcu R, Hallén A, Fredlund H, Jensen JS, Engstrand L, Unemo M,

452   (2013) Composition of the vaginal microbiota in women of reproductive age-sensitive and

453   specific molecular diagnosis of bacterial vaginosis is possible? PLoS One 8(4):e60670.

454   39. Spraul M, Schütz B, Humpfer E, Mörtter M, Schäfer H, Koswig S, Rinke P (2009) Mixture

455   analysis by NMR as applied to fruit juice quality control. Magn Reson Chem 47 Suppl 1:S130-

456   S137.

457   40. Biagi E, Vitali B, Pugliese C, Candela M, Donders GG, Brigidi P (2009) Quantitative variations

458   in the vaginal bacterial population associated with asymptomatic infections: a real-time

459   polymerase chain reaction study. Eur J Clin Microbiol Infect Dis 28:281-285.

460   41. Turovskiy Y, Sutyak Noll K, Chikindas ML (2011) The aetiology of bacterial vaginosis. J Appl

461   Microbiol 110:1105-1128.

462   42. Ling Z, Kong J, Liu F, Zhu H, Chen X, Wang Y, Li L, Nelson KE, Xia Y, Xiang C (2010)

463   Molecular analysis of the diversity of vaginal microbiota associated with bacterial vaginosis.

464   BMC Genomics 11:488.

465   43. Macklaim JM, Gloor GB, Anukam KC, Cribby S, Reid G (2011) At the crossroads of vaginal

466   health and disease, the genome sequence of *Lactobacillus iners* AB-1. Proc Natl Acad Sci U S A

467   108 Suppl 1:4688-95.

468   44. Gardner HL, Dukes CD (1955) Haemophilus vaginalis vaginitis: a newly defined specific

469   infection previously classified non-specific vaginitis. Am J Obstet Gynecol 69:962–976.

470   45. Al-Mushrif S, Eley A, Jones BM (2000) Inhibition of chemotaxis by organic acids from

471   anaerobes may prevent a purulent response in bacterial vaginosis. J Med Microbiol 49:1023-

472   1030.

473   46. Chaudry AN, Travers PJ, Yuenger J, Colletta L, Evans P, Zenilman JM, Tummon A (2004)

474   Analysis of vaginal acetic acid in patients undergoing treatment for bacterial vaginosis. J Clin

475   Microbiol 42:5170-5175.

476   47. Mirmonsef P, Gilbert D, Zariffard MR, Hamaker BR, Kaur A, Landay AL, Spear GT (2011) The

477   effects of commensal bacteria on innate immune responses in the female genital tract. Am J

478   Reprod Immunol 65:190-195.

479   48. Sobel JD, Karpas Z, Lorber A (2012) Diagnosing vaginal infections through measurement of

480   biogenic amines by ion mobility spectrometry. Eur J Obstet Gynecol Reprod Biol 163:81-84.

481   49. Wolrath H, Forsum U, Larsson PG, Borén H (2001) Analysis of bacterial vaginosis-related

482   amines in vaginal fluid by gas chromatography and mass spectrometry. J Clin Microbiol

483   39:4026-4031.

484   50. Cruciani F, Wasinger V, Turroni S, Calanni F, Donders G, Brigidi P, Vitali B (2013) Proteome

485   profiles of vaginal fluids from women affected by bacterial vaginosis and healthy controls:

486   outcomes of rifaximin treatment. J Antimicrob Chemother 68:2648-2659.

487   51. Spear GT, French AL, Gilbert D, Zariffard MR, Mirmonsef P, Sullivan TH, Spear WW, Landay

488   A, Micci S, Lee BH, Hamaker BR (2014) Human α-amylase present in lower-genital-tract

489   mucosal fluid processes glycogen to support vaginal colonization by *Lactobacillus*. J Infect Dis

490   210:1019-1028.

491   52. Thoma ME, Klebanoff MA, Rovner AJ, Nansel TR, Neggers Y, Andrews WW, Schwebke JR

492   (2011) Bacterial vaginosis is associated with variation in dietary indices. J Nutr 141:1698-1704.

493

494   **Figure captions**

495

496   **Fig. 1** Molecular analysis of the vaginal microbiota composition in healthy (H) and BV-affected (BV)

497   women. (a) Frequency of occurrence of *L. crispatus*, *L. iners*, *L. gasseri*, *L. jensenii*, *Atopobium*, *G.*

498   *vaginalis*, *Prevotella*, *Veillonella*, *Mobiluncus* and *M. hominis*, calculated as the percentage of women

499   hosting each bacterial group. **, $P < 0.01$; *, $P < 0.05$. (b) Relative abundance of *L. crispatus, L. iners,*

500   *L. gasseri* and *L. jensenii*, calculated as percentage of women who hosted each species as dominant.

501   (the predominant species was identified based on the bacterial amount calculated by qPCR). N indicate

502   the percentage of women that is not colonized by any of the four analysed species

503

504   **Fig. 2** Quantification of *L. crispatus, L. iners, L. gasseri, L. jensenii, Atopobium, G. vaginalis,*

505   *Prevotella, Veillonella, Mobiluncus* and *M. hominis* in healthy (H) and BV-affected (BV) women.

506   qPCR data are expressed as log ng of DNA of the target genus or species per μg of total DNA extracted

507   from the vaginal sample. The box for each bacterial group represents the interquartile range (25th to 75th

508   percentile), and the line within this box is the median value. Bottom and top bars indicate the 10th and

509   90th percentiles, respectively. Outlier values are indicated (black circles). Only positive samples for

510   each bacterial group analysed were considered. **, $P < 0.01$; *, $P < 0.05$

511

512

513   **Fig. 3** Frequency of inclusion of the 32 molecules of Table 1 calculated by the sPLS-DA predictive

514   model. The median variable influence on projections (VIP) coefficients are indicated in parentheses

515

516

517

Supplemental material

520 **Fig. S1** Biplot of a PCA performed on the autoscaled qPCR data. Median values of the samples groups

521 corresponding to healthy and BV-affected women are indicated as H and BV. Empty circles and filled

522 squares indicate samples from healthy and BV women, respectively. Expl. Var, explained variance

523

524 **Fig. S2** Biplot of a PCA performed on the autoscaled qPCR data related to BV-associated bacteria (a)

525 and metabolites selected by sPLS-DA model. (b) Median values of the samples groups corresponding

526 to healthy women dominated by *L. crispatus*, *L. iners*, *L. gasseri*, *L. jensenii* and BV-affected women

527 dominated by *L. iners*, *L. gasseri* or none of the four considered species are indicated as Hc, Hi, Hg, Hj,

528 BVi, BVg and BVn, respectively. Empty circles and filled squares indicate samples from healthy and

529 BV women, respectively. Expl. Var, explained variance

Table 1 Concentration of metabolites which significantly varied ($P < 0.05$) in BV-affected patients

(BV) compared to healthy control women (H)

| Molecule	Chemical shift (ppm)	H	BV	Variations in BV
Amines				
Tyramine	7.228	$1.52 \times 10^{-02} \pm 1.29 \times 10^{-02}$	$9.21 \times 10^{-02} \pm 1.02 \times 10^{-01}$	↑
Ethanolamine	3.151	$6.04 \times 10^{-02} \pm 1.85 \times 10^{-02}$	$1.18 \times 10^{-01} \pm 5.29 \times 10^{-02}$	↑
TMA	2.894	$6.50 \times 10^{-04} \pm 7.13 \times 10^{-04}$	$2.13 \times 10^{-02} \pm 1.99 \times 10^{-02}$	↑
Methylamine	2.61	$9.08 \times 10^{-04} \pm 8.91 \times 10^{-04}$	$1.97 \times 10^{-02} \pm 2.12 \times 10^{-02}$	↑
Cadaverine	1.722	$1.64 \times 10^{-01} \pm 6.66 \times 10^{-02}$	$3.80 \times 10^{-01} \pm 3.15 \times 10^{-01}$	↑
Organic acids				
Formate	8.459	$3.35 \times 10^{-02} \pm 7.96 \times 10^{-02}$	$4.50 \times 10^{-01} \pm 6.88 \times 10^{-01}$	↑
Malonate	3.134	$1.74 \times 10^{-02} \pm 7.29 \times 10^{-03}$	$4.98 \times 10^{-02} \pm 2.56 \times 10^{-02}$	↑
Succinate	2.409	$1.91 \times 10^{-01} \pm 1.63 \times 10^{-01}$	$9.70 \times 10^{-01} \pm 1.12 \times 10^{+00}$	↑
Pyruvate	2.380	$1.22 \times 10^{-01} \pm 1.48 \times 10^{-01}$	$3.38 \times 10^{-01} \pm 3.91 \times 10^{-01}$	↑
Acetate	1.916	$7.24 \times 10^{-01} \pm 5.33 \times 10^{-01}$	$4.73 \times 10^{+00} \pm 3.12 \times 10^{+00}$	↑
SCFAs				
Propionate	2.200	$1.56 \times 10^{-02} \pm 9.42 \times 10^{-03}$	$4.21 \times 10^{-01} \pm 7.61 \times 10^{-01}$	↑
Butyrate	2.150	$5.03 \times 10^{-02} \pm 1.95 \times 10^{-02}$	$1.62 \times 10^{-01} \pm 2.68 \times 10^{-01}$	↑
2-Hydroxyisovalerate	0.837	$2.01 \times 10^{-02} \pm 1.02 \times 10^{-02}$	$3.87 \times 10^{-02} \pm 2.56 \times 10^{-02}$	↑
Amino acids				
Tryptophan	7.749	$1.25 \times 10^{-02} \pm 5.95 \times 10^{-03}$	$9.20 \times 10^{-03} \pm 6.66 \times 10^{-03}$	↓
Phenylalanine	7.391	$9.46 \times 10^{-02} \pm 3.15 \times 10^{-02}$	$6.80 \times 10^{-02} \pm 3.43 \times 10^{-02}$	↓
Tyrosine	7.191	$6.77 \times 10^{-02} \pm 2.53 \times 10^{-02}$	$4.54 \times 10^{-02} \pm 3.51 \times 10^{-02}$	↓
Glutamate	2.354	$3.07 \times 10^{-01} \pm 1.28 \times 10^{-01}$	$1.77 \times 10^{-01} \pm 8.86 \times 10^{-02}$	↓

Proline	2.017	$2.95\times10^{-02} \pm 1.19\times10^{-02}$	$4.70\times10^{-02} \pm 2.88\times10^{-02}$	↑
Isoleucine	1.012	$8.27\times10^{-02} \pm 3.54\times10^{-02}$	$6.91\times10^{-02} \pm 8.39\times10^{-02}$	↓
Leucine	0.962	$4.65\times10^{-01} \pm 1.95\times10^{-01}$	$3.66\times10^{-01} \pm 3.43\times10^{-01}$	↓
Nitrogenous bases				
NAD+	9.341	$2.22\times10^{-03} \pm 1.58\times10^{-03}$	$1.83\times10^{-04} \pm 1.28\text{x}^{-03}$	↓
Nicotinate	8.943	$1.63\times10^{-03} \pm 1.30\times10^{-03}$	$4.77\times10^{-03} \pm 1.92\times10^{-03}$	↑
Uracil	7.543	$1.32\times10^{-02} \pm 6.01\times10^{-03}$	$3.12\times10^{-02} \pm 3.06\times10^{-02}$	↑
Inosine	6.098	$3.61\times10^{-03} \pm 7.46\times10^{-03}$	$8.12\times10^{-04} \pm 1.48\times10^{-03}$	↓
Sugars				
Glucose	3.537	$8.27\times10^{-02} \pm 6.48\times10^{-02}$	$1.67\times10^{-01} \pm 1.28\times10^{-01}$	↑
Maltose	3.298	$1.17\times10^{+00} \pm 6.51\times10^{-01}$	$6.27\times10^{-01} \pm 5.10\times10^{-01}$	↓
Others				
Urocanate	6.410	$6.15\times10^{-04} \pm 1.45\times10^{-03}$	$3.84\times10^{-03} \pm 4.07\times10^{-03}$	↑
Kynurenine	6.812	$5.72\times10^{-02} \pm 2.33\times10^{-02}$	$1.94\times10^{-02} \pm 1.84\times10^{-02}$	↓
sn-Glycero-3-phosphocholine	3.229	$2.45\times10^{-02} \pm 1.11\times10^{-02}$	$1.35\times10^{-02} \pm 1.72\times10^{-02}$	↓
Sarcosine	2.757	$2.99\times10^{-02} \pm 1.78\times10^{-02}$	$1.17\times10^{-02} \pm 6.89\times10^{-03}$	↓
2-Aminoadipate	2.241	$8.66\times10^{-02} \pm 6.09\times10^{-02}$	$2.58\times10^{-01} \pm 1.55\times10^{-01}$	↑
3-Methyl-2-oxovalerate	1.109	$3.34\times10^{-03} \pm 3.39\times10^{-03}$	$1.06\times10^{-02} \pm 1.17\times10^{-02}$	↑

532 Values are expressed as means of mmol/l ± standard deviation.

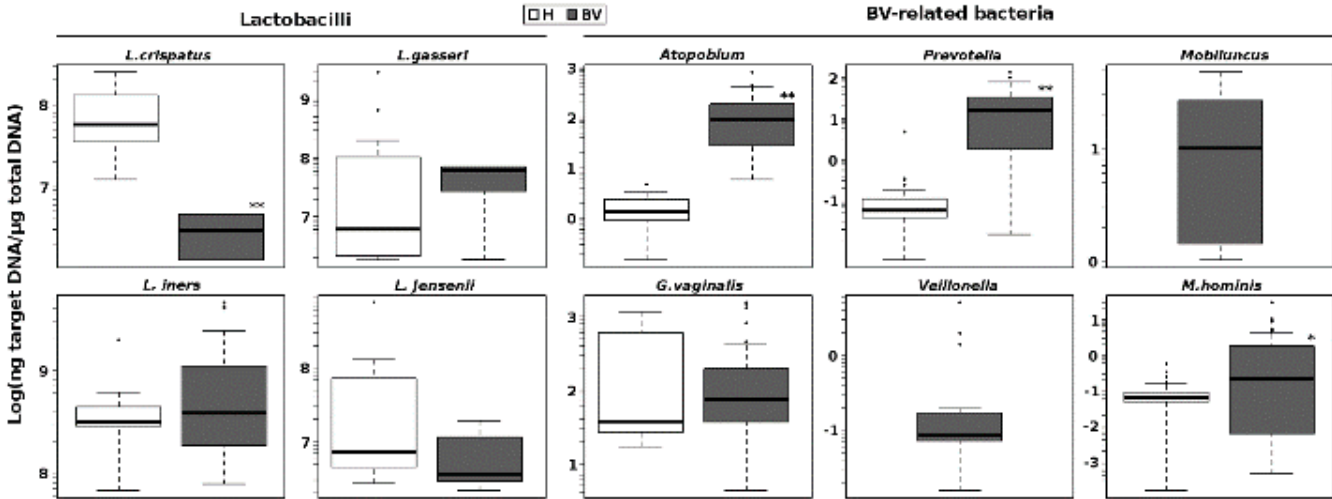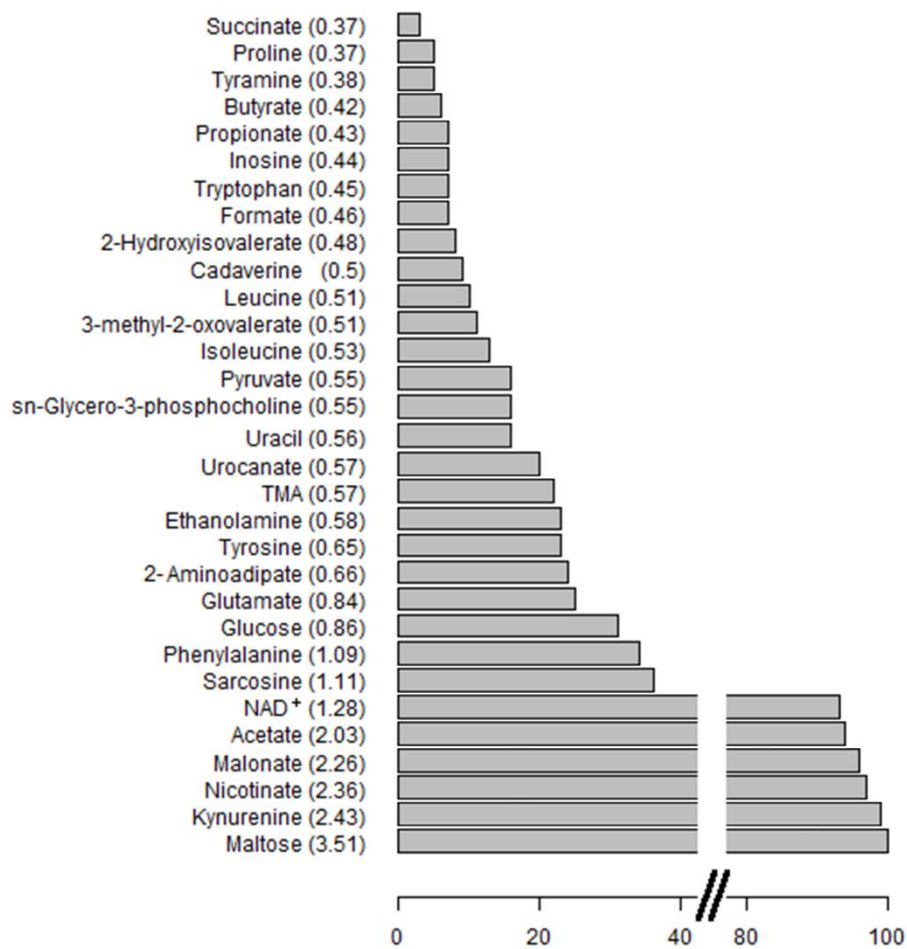533 TMA, Trimethylamine; SCFAs, short chain fatty acids.

534
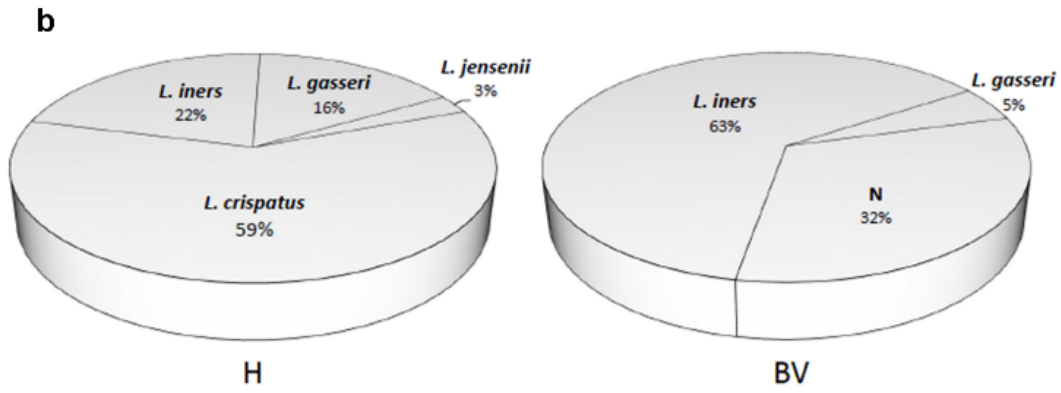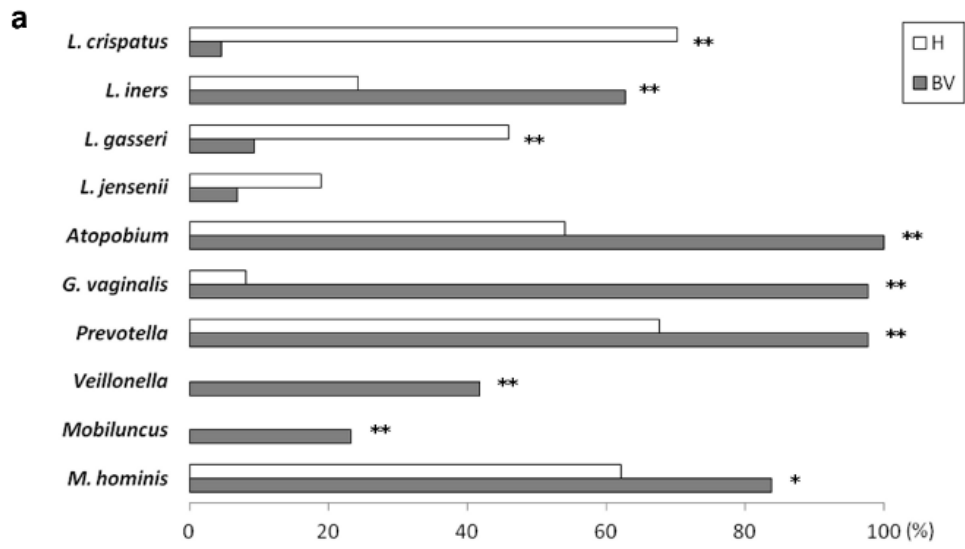
540   Figure 2

541

542

543   Figure 3

544

545