Metatranscriptome analysis of the vaginal microbiota reveals potential mechanisms for recurrence and protection against metronidazole in bacterial vaginosis

Zhi-Luo Deng[*1+] (Zhiluo.Deng@helmholtz-hzi.de),

Cornelia Gottschick[1,4+] (cornelia.gottschick@uk-halle.de),

Sabin Bhuju[2] (sabinbhuju@gmail.com),

Clarissa Masur[3] (clarissa.masur@drwolffgroup.com),

Christoph Abels[3] (christoph.abels@drwolffgroup.com),

and Irene Wagner-Döbler[1] (Irene.Wagner-Doebler@helmholtz-hzi.de)

[1] Research Group Microbial Communication, Department of Medical Microbiology, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany

[2] Genome Analytics, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany

[3] Dr. August Wolff GmbH & Co. KG Arzneimittel, Sudbrackstrasse 56, 33611, Bielefeld, Germany

[4] Current address: Institute for Medical Epidemiology, Biometrics and Computer Science, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany

+ shared first authorship

*Corresponding author

Abstract

Background. Bacterial vaginosis (BV) is a prevalent multifactorial disease of women in their reproductive years characterized by a shift from the healthy *Lactobacillus sp.* dominated microbial community towards a highly diverse anaerobic community. BV can initially be cured by antibiotic therapy in most women, but the high recurrence frequency represents a serious challenge. Moreover, for unknown reasons, a small number of women does not respond to therapy. In a clinical study, out of 37 women diagnosed with BV, 31 were successfully treated with a single peroral dose of metronidazole, while 6 still had BV after treatment. Here, we performed a metatranscriptome analysis of the vaginal microbiota of the subgroup of those non-responding patients (N = 6), comparing them to patients (N = 8) who were successfully treated. Moreover, we followed the changes in the metatranscriptome composition over three months (5 time-points) in patients (N = 2) that developed recurrence in comparison to patients (N = 2) that were permanently cured.

Results. In health, *Lactobacillus iners* and *L. crispatus* contributed more than 90% of all bacterial transcripts, but they did not co-occur. Their functional profiles *in vivo* confirmed the pathogenic role of *L. iners.* In BV, *Gardnerella vaginalis* contributed on average 37% of all transcripts. In some patients, transcripts from *Gardnerella* species isolated from the bladder were abundant. In non-responding patients, *cas* and DNA methyltransferase genes of *G. vaginalis* were highly up-regulated, suggesting that they might play a role in defense against DNA damaging agents like metronidazole.

Conclusions. We hypothesize that colonization by *L. iners* and re-infection through the bladder contribute to recurrence of BV. Our data suggest that Cas genes of *G. vaginalis*, in addition to protecting against phages, might be involved in DNA repair thus mitigating the bactericidal effect of DNA damaging agents like metronidazole.

Trial registration: ClinicalTrials.gov NCT02687789. Retrospectively registered on 12 October 2015.

Keywords

Vaginal microbiome, bacterial vaginosis, antibiotic treatment, resistance, recurrence, metatranscriptome, CRISPR-Cas, DNA repair

Background

The healthy vaginal microbiome (VMB) is characterized by low pH and low diversity and can be categorized into community state types (CSTs) that are dominated by different *Lactobacillus sp.* such as *L. crispatus*, *L. iners*, *L. gasseri* and less frequently *L. jensenii* or a more diverse community [1]. Bacterial vaginosis (BV) is a frequent multifactorial disease of women in their reproductive years that is characterized by a shift of this *Lactobacillus sp.* dominated bacterial community to a community of various mostly anaerobic bacteria [2]. BV is associated with a higher risk of acquiring sexually transmitted infections such as HIV, miscarriage and preterm birth [3]. The most common bacteria found in BV, identified by 16S rRNA gene sequencing, are *Gardnerella*, *Atopobium*, *Prevotella*, *Bacteroides*, *Peptostreptococcus*, *Mobiluncus*, *Sneathia*, *Leptotrichia*, *Mycoplasma* and BV associated bacterium 1 (BVAB1) to BVAB3 of the order Clostridiales. Recently, three CSTs dominated by *Gardnerella vaginalis*, Lachnospiraceae and *Sneathia sanguinegens*, respectively, have been described [4] and our recent study identified *S. amnii* as the best biomarker for BV [5].

Metronidazole is a widely applied chemotherapeutic agent used to treat infectious diseases caused by anaerobic bacteria, and it is the first-line antibiotic for treating BV [6,7]. It is a prodrug which requires enzymatic reduction within the cell to transform it into an active form [8]. Metronidazole acts by damaging DNA, and resistance mechanisms have been studied in various pathogens [8]. Previous studies have shown that failure of BV treatment by metronidazole is relatively rare [5,9], and it is unclear if it is caused by resistance of the BV pathogens to metronidazole, and which mechanisms are acting *in vivo*. A recent study has demonstrated that failure of treatment of BV with metronidazole is not associated with higher loads of *G. vaginalis* and *A. vaginae* [10]. Therefore, the lack of response may result from the activity of the community instead of the abundance of the pathogens and their taxonomic

composition. The high rate of recurrence is another crucial problem for BV treatment. It has been reported that the one-year recurrence rate of BV ranges from 40% to 80% after treatment with metronidazole [11]. The recurrent rate is similar for other therapies such as clindamycin cream [12]. It has been hypothesized that patches of surviving bacterial biofilm may be one of the reasons for recurrence [11]. The predominant species of this multi-species biofilm is *G. vaginalis* [13]. *G. vaginalis* is frequently found in biofilms during BV but it is also a member of the healthy VMB [14]. *L. iners* is characteristic for a community state type in health but may be disadvantageous when the community shifts to a dysbiotic state [14-16]. Until now, the majority of studies regarding the VMB have focused on 16S rRNA gene sequencing, answering only questions on the taxonomic composition of bacterial communities but not on their functions [2]. A metatranscriptome analysis comparing vaginal swabs from two women with BV with two healthy subjects showed that *L. iners* upregulates transcription of the cholesterol-dependent cytolysin (CDC) and of genes belonging to the CRISPR system in BV [16]. No study has investigated the functional shifts of the VMB during antibiotic treatment of BV.

We had previously analyzed the VMB in the context of a clinical trial using 16S rRNA gene sequencing [5]. Of 37 patients diagnosed with BV and included in this study, 31 were cured by a single oral dose of metronidazole. Six patients did not respond, i.e. they were diagnosed with BV according to Nugent score at visit 2. Moreover, some patients developed recurrent BV during the three months of the study. Since we did not find differences in the taxonomic composition of the microbiota of non-responding and recurrent patients, we hypothesized that shifts in microbiota activity might be responsible for those different responses. We therefore chose the non-responding patients (N = 6) from the clinical study and compared their metatranscriptomes to those of 8 patients that were initially cured. For the analysis of

recurrence, we chose two recurrent patients from the lactic acid arm of the study (see below) and compared their metatranscriptomes to those of two permanently cured patients that were also provided with lactic acid pessaries, and followed their transcriptomes over the three months of the study (5 time-points). Our analysis is thus comprised of 14 patients, of which six were cured, six did not respond to treatment, and two developed recurrence. Including the time series samples, we analyzed 40 samples in total. We extracted the RNA and performed an in depth analysis of the metatranscriptomes of the VMB.

We could identify genes potentially leading to lack of response to metronidazole treatment and suggest a novel mechanism of *G. vaginalis* to mitigate the DNA damaging effect of metranidazole. Moreover, we demonstrate the importance of *G. vaginalis* for BV, which can be massively underestimated using 16S rRNA gene sequencing, shed light on the beneficial role of *L. crispatus* in comparison to *L. iners*, and hypothesize that recolonization of the vagina by *Gardnerella* strains residing in the bladder may contribute to recurrence. All of these findings are strongly supported by the metatranscriptome analyses reported here and should be validated in larger patient cohorts.

Material and Methods

Study design

Vaginal fluid samples of women analyzed here were a subset of the samples obtained during a randomized controlled clinical trial described previously [5]. The trial protocol was approved by the local ethics committee (Ärztekammer Nordrhein - Medical Association North Rhine) and written consent was obtained from all participants. The clinical trial was conducted in accordance with the Declaration of Helsinki on Ethical Principles for Medical Research

Involving Human Subjects. Principles and guidelines for good clinical practice were followed. The study was registered on ClinicalTrials.gov with the identifier NCT02687789. Briefly, women were included into the clinical trial if they were BV positive according to Amsel criteria and Nugent score and were biofilm positive on vaginal epithelial cells and positive for extracellular polysaccharides (EPS) in urine. For treatment of acute BV, they received 2 g of metronidazole orally and were afterwards treated with an intravaginal pessary for three weeks, twice a week. Samples were taken during acute BV (visit 1), after receiving metronidazole 7 to 28 days after visit 1 (visit 2), after pessary application one week after visit 2 (visit 3), after continued pessary application two weeks after visit 3 (visit 4) and during follow up three months after visit 4 (visit 5).

The aim of the clinical trial had been to compare the effectiveness of two different types of pessary. The results and the taxonomic composition of the vaginal microbial communities have been reported [5]. For the metatranscriptome analysis reported here, we chose a subset of 14 patients from the clinical trial. These 14 patients consisted of two groups named “with response to treatment” (N = 8) and “no response to treatment” (N = 6) (Fig. 1). Among the eight patients which responded to treatment, six were permanently cured and two experienced recurrence during the three months follow-up. For the analysis of recurrence, samples from all 5 visits were analyzed for 4 patients (two cured and two with recurrence) (20 samples in total). These four patients all received the commercially available lactic acid pessary after metronidazole therapy at visit 3 and 4. For the analysis of lack of response to metronidazole, samples from all 14 patients were analyzed at two time points, acute BV (visit 1) and 7 to 28 days after antibiotic treatment (visit 2) (28 samples in total). Nugent score > 6 was used to determine BV status, since it is considered the gold standard for BV diagnosis [17] (Table S1 sheet 1). The BV status at visit 5 was determined by Amsel criteria since there was no Nugent

score available at that time point. The sample ID was obtained by concatenating the patient ID and letters “a” to “e” indicating visit 1 to 5.

Sample collection and transport

Vaginal fluid was obtained by infusing 2 ml of saline solution into the vagina followed by rotation against the vaginal wall with a speculum and then collecting the vaginal fluid with a syringe. Approximately 700 µl were immediately transferred to a tube containing 2 ml RNAprotect (Qiagen, Germany). Tubes were immediately frozen at -20°C, transported at -20°C within a week and stored at -70°C.

RNA extraction and mRNA enrichment

RNA was extracted from 1 ml vaginal fluid suspension using the MO BIO PowerMicrobiome™ RNA Isolation Kit (Qiagen, Germany) with pretreatment: Vaginal fluid was centrifuged at 13,000 rpm for 1 minute. The pellet was resuspended in MoBio lysis buffer and this suspension was added to the supplied bead tubes filled with 500 µl ice-cold phenol-choloroformisoamylalcohol solution (Carl Roth, Germany). The bead-suspension mix was shaken at 5 m/s for 1 minute in 3 intervals which were 2 minutes apart using the MO BIO PowerLyzer™ (Qiagen, Germany). After centrifugation for 1 minute at 13,000 rpm and 4°C the upper phase containing the RNA was further processed according to the manufacturer’s instructions including DNAse I treatment. RNA was eluted in 100µl nuclease free water and vaccum concentrated to 50 µl. The Ribo-Zero Gold rRNA Removal Kit (Epidemiology) by Illumina (USA) was then used for mRNA enrichment with ethanol precipitation according to the manufacturer’s instructions. Integrity of RNA was evaluated using a Bioanalyzer 2100 (Agilent, Germany).

Library preparation, sequencing and preprocessing of sequencing data

Paired-end mRNA Illumina sequencing libraries were constructed with the Script Seq Illumina Kit. Strand specific paired end sequencing was performed on the HiSeq 2500 Sequencer to yield 2 × 110 bp paired-end reads. Primers and sequencing adaptors were removed from raw sequencing data, followed by clipping the bases with quality score < 20 from the reads using Fastq-Mcf [18]. After clipping, the remaining reads shorter than 50 were removed. Thereafter, the ribosomal RNA reads were eliminated using SortMeRNA v2.0 [19] with the default parameters.

Taxonomy assignment using Kraken

Kraken [20], an accurate and ultra-fast taxonomy assignment tool for metagenomes was used to determine the taxonomic composition of the metatranscriptome data. Kraken uses the K-mer strategy and the lowest common ancestor (LCA) algorithm to affiliate a given read to a taxon. The standard Kraken database from the Kraken program with addition of the human genome was used to identify human reads and determine the predominant species in the communities for building the refined Kraken database. This standard database consists of prokaryote genomes (2786), virus genomes (4418) and with addition of human genome (ver. GRCh38) was downloaded from NCBI.

The refined Kraken database contained 19 species of bacteria (Table S1 sheet 3), including all species determined by the standard Kraken database which contributed more than 1% of classified reads, species identified by HMP database with abundance higher than 1% using BWA, *S. amnii*, *S. sanguinegens* detected by amplicon sequencing and two species of *Gardnerella*, *Gardnerella sp.* 26-12 and *Gardnerella sp.* 30-4 which were isolated from the bladder recently [21]. Subsequently, the cleaned putative microbial mRNAs reads were

mapped against this refined database. All results on the taxonomic composition in this study were achieved based on this database.

Short reads alignment by BWA

Since Kraken was only used for taxonomy classification, BWA was applied to determine the expression of functional genes. A customized reference gene database (Table S1 sheet 4) combining the genes from the urogenital tract subset of the HMP reference genome sequence data (HMRGD) containing genes of 147 genomes [22] and 9 additional genomes determined by taxonomy assignment (Table S1 sheet 4, in red and green) as well as the amplicon sequencing results was used as mapping reference for BWA to map all cleaned putative mRNA reads for functional profiling. The genes of *Gardnerella sp.* 26-12, *Gardnerella sp.* 30-4 were not included due to the poor quality of the genomes. The short reads alignment was performed using BWA with the BWA-MEM [23] algorithm. A mapping seed length of 31 which is much longer than the default seed length 19 was applied to achieve more reliable alignments. To avoid a biased mapping, the reads mapped with mapping quality score (MAPQ) lower than 10 were excluded. MAPQ contains the Phred-scaled posterior probability that the mapping position is wrong [24]. A MAPQ of 10 indicates a probability of 10% that the alignment position is wrong.

KEGG Orthologous (KO) gene annotation of customized reference database

The customized short reads alignment reference gene database which was used for BWA was annotated using KEGG prokaryote protein sequences to calculate the expression level of KO genes for differential KO gene expression analysis and KEGG pathway enrichment analysis. The KEGG prokaryote protein sequence database represents a non-redundant protein dataset of prokaryote (Bacteria and Archaea) on the species level and contains about 7 million non-

redundant peptide sequences grouped into 14,390 distinct KO genes. A KO gene contains several genes from different species with similar function. DIAMOND [25], a much faster alternative to BLASTX was applied to map the customized short reads alignment reference sequences against the KEGG prokaryote protein sequence database with its "more sensitive mode". To obtain reliable annotation, only alignments with sequence identity >=50 and E-value <= 1e-5 and query coverage >= 70% were taken into account. By annotating the reference genes to KO genes, we were able to determine the expression profile of KO genes, and investigate the functional shifts in BV based on differential expression analysis of KO genes and perform KEGG pathway enrichment analysis.

Differential expression (DE) analysis

All differential expression (DE) analyses were performed using the R package edgeR [26]. The Benjamin Hochberg (BH) method was used to correct the p value of DE analysis with the false discovery rate (FDR) for multiple comparisons. Genes with FDR smaller than 0.05 were considered as significantly differentially regulated, which were then used for KEGG pathway enrichment analysis. The sample groups defined for each comparison are listed in Table S1 sheet 1. The groups defined in column "group for G vaginalis" is for the DE analysis of KO genes between *G. vaginalis* in communities with no response to metronidazole and with response. The samples with relative abundance of *G. vaginalis* higher than 20% according to the taxonomy assignment were taken into account. In these two columns, "NR1" represents samples at visit 1 from patients with no response to metronidazole after treatment, "NR2" indicates the samples from the same patients at visit 2. And "WR1" denotes samples at visit 1 from patients with response to metronidazole after treatment, while "WR2" means samples from the recurrent time point. Column "group for LC LI" is for the DE analysis for KO genes

between *L. crispatus* in the communities dominated by *L. crispstus* ("LC") and *L. iners* in the communities dominated by *L. iners* ("LI").

KEGG pathway enrichment analysis

KEGG pathway enrichment analysis was performed based on differentially expressed KO genes (genes with same KO number were grouped together) using the kegga function of the R package limma [27]. This R package provides the kegga function that enables access to the KEGG database via REST API. The Fisher's exact test based on hypergeometric distribution was used to compute the significance of the enrichment and the. The BH method was used to correct the p value to FDR for multiple comparison.

Detection of putative metronidazole resistance related genes

To detect the expression of previously reported putative metronidazole resistant genes such as recA, recA-mediated autopeptidase (Rma), peroxiredoxin, nitroimidazole resistance protein (NIM), ferredoxin/ferredoxin-NADP reductase (FNR), nitroreductase, sialidase and ferredoxin, we examined the expression level of these genes for *G. vaginalis* in the vaginal community from patients without response to metronidazole treatment (n=6 patients) as well as with response (n=4 patients). As most of these genes do not have corresponding KO genes, we annotated the BWA reference database based on the sequences of these genes using BLASTN. These sequences were retrieved from ENA by key words of each "ferredoxin, ferric uptake regulator, NADPH flavin oxidoreductase, nitroreductase, peroxiredoxin, pyruvate ferredoxin oxidoreductase, recA, nitroimidazole resistance, sialidase" plus *G. vaginalis*. In total, 155 unique sequences of *G. vaginalis* were obtained for the annotation of BWA reference database. The identification of duplicate sequences was done by SeqKit [28].

Results

Study population and overview of sequencing results

We studied the vaginal microbiome of 14 patients during and after metronidazole treatment of BV using metatranscriptome sequencing (Fig. 1). Patients were part of a clinical trial described elsewhere [5]. For ten patients, samples were derived from two time points (before (visit 1) and after metronidazole treatment (visit 2). Six of those patients did not respond to treatment with the antibiotic, so were still BV positive according to Nugent score and four of them responded to the treatment and were cured. For the other four patients, 5 time points were analyzed, covering a period of 3 months after treatment. Two of those patients experienced recurrence, while the other two were cured without recurrence (details in Table S1 sheet 1). In total, we analyzed 40 vaginal fluid samples, 22 with BV status and 18 without. Metatranscriptome sequencing resulted in a total of 1,879,945,342 reads. Of these, 1,377,516,082 reads (73%) were left after quality filtering and removal of ribosomal RNA (Table 1). On average, 34 million reads were analyzed per sample.

Construction of the reference genome and gene databases for taxonomic and functional profiling

Based on the standard Kraken database, ~11% of the total putative mRNA reads could be assigned to human for BV and ~56% for health (Table S1 sheet 2). This suggests that the bacterial load is much lower in health than in BV since the human contamination read is much higher in health. Interestingly, transcripts from the most abundant 10 species contributed 90% of the mapped reads (Fig. S1), which is much less diverse than the periodontal metatranscriptome in which more than 100 species were required to cover 90% of mappable reads [29]. Using this standard database, 41% of total putative microbial (non-human) mRNA

281 reads could be assigned taxonomically (Table 2). To improve the fraction of taxonomically
282 assignable reads we then constructed a refined database consisting of all species detected by
283 the standard Kraken database with an abundance > 1% and species detected with more than
284 1% of the total putative microbial mRNA reads using the urogenital subset of the HMP
285 database by BWA. We also added genomes of *S. amnii* and *S. sanguinegens* which had
286 previously been shown to be highly abundant based on 16S rRNA gene sequencing [5] but
287 were not contained in either HMP or the standard Kraken reference database. To be able to
288 identify transcripts from bladder isolates, we added the genomes of *Gardnerella sp.* 26-12 and
289 *Gardnerella sp.* 30-4 which were isolated from the bladder recently [21]. To make sure that
290 the transcripts mapped to the *Gardnerella sp.* bladder isolates did not actually originate from
291 *G. vaginalis*, we included 8 genomes from distinct *G. vaginalis* strains in the refined Kraken
292 database. For the rest of the species, only the representative genome was used. In total, this
293 database contained 19 species and 26 genomes (Table S1 sheet 3). Using this refined database
294 for taxonomic assignment mapping of all putative microbial mRNA reads could be improved
295 to 77% on average (Table 2). These data show that the active microbiota in BV is much less
296 diverse than suggested by 16S rRNA amplicon sequencing.

297 For functional assignment, we used a customized reference gene database (Table S1 sheet 4)
298 containing the genes from the HMP urogenital tract subset database, and added the genes of
299 additional 9 genomes (see method section for details).To investigate the functional shifts of
300 the communities, we mapped the cleaned metatranscriptomic reads to this customized
301 reference gene database using BWA. In total, 31% of total putative mRNA reads for BV and
302 5% for health could be mapped to this microbial reference gene database with a MAPQ score
303 higher than 10. Most of the reads were of human origin. However, the sequencing provided
304 on average 8.9 million reads per sample, which allowed the functional profiling for highly

abundant species in the community. The mapping results per sample can be seen in Table S1 sheet 2.

Shifts in the taxonomic composition of the active community during 3 months following metronidazole treatment

Using the refined Kraken database, the taxonomic composition of transcripts in the 40 analyzed samples was determined. Fig. 1 shows that all BV communities (before metronidazole treatment, at recurrence, and in patients with no response to treatment) were dominated by *G. vaginalis*, *A. vaginae*, *S. amnii* and *G. sp.* 26-12. In patients that responded to treatment and were cured, the metatranscriptomes were dominated by *L. crispatus*, *L. iners* and *L. jensenii*.

Comparison of the taxonomic composition of vaginal fluid samples between metatranscriptome and 16S rRNA sequencing

To discover transcriptionally active bacteria which may contribute to the functional shifts in BV, we compared the taxonomic composition determined using 16S rRNA sequencing [5] and the taxonomic composition of the metatranscriptome determined by Kraken with the refined reference database. In health, we did not observe any considerable difference (Table S1 sheet 6), while in BV large differences between the two datasets were found. Figure 2 shows that most of the abundant species identified in the mRNA sequencing data set were also identified using 16S rRNA gene amplicon sequencing, although usually at different abundances. For example, *A. vaginae* comprised 15% of all reads based on 16S rRNA gene sequencing, but less than 10% in the metatranscriptome dataset. The most pronounced difference was observed for *G. vaginalis* which comprised on average 37% of relative abundance in the metatranscriptome and on average only 6% in the 16S rRNA sequencing

data. By contrast, the two bladder isolates of *Gardnerella sp.* and *Megasphaera sp.* type 1 were only identified in the metatranscriptome, probably due to the low resolution of the 16S rRNA gene. This finding suggests that *G. vaginalis* and other *Gardnerella* species are extremely transcriptionally active in the BV community. In addition, the commonly used 27F primer has been shown to underrepresent *G. vaginalis* [30].

L. iners and _L. crispatus_ were mutually exclusive in health and _G. vaginalis_ was the most active species in BV

The most abundant active species after metronidazole treatment were *Lactobacillus* species of which *L. iners* and *L. crispatus* were the most abundant ones (Fig. 2C). They seem to exclude each other, because high abundant *L. iners* was never accompanied by *L. crispatus* and high abundant *L. crispatus* was only twice accompanied by low abundant (lower than a quarter of the abundance of *L. crispatus*) *L. iners*. *L. gasseri* and *L. jensenii* had much lower abundances and were found together with both *L. crispatus* and *L. iners.* In one patient, the active healthy community was dominated by *L. jensenii*.

In BV, *G. vaginalis* was the most abundant active species. Only two communities of patients with BV were dominated by other species, namely *A. vaginae* and *S. amnii*. Additional BV associated species such as *S. sanguinegens* or different *Prevotella sp.* were frequently found in BV samples, and *Megasphaera sp.* type1 and M*obiluncus mulieris* were occasionally found. The only species of *Lactobacillus* found to be active in BV was *L. iners*.

Gardnerella species isolated from the bladder were active in BV communities

The bladder was thought to be sterile in healthy women, but it has now been shown that specific microorganisms are residing in the healthy bladder [21]. Therefore, the bladder has to be considered as a possible reservoir for pathogens. The high pH of urine may be a favorable

environment for several pathogens, including *Gardnerella sp.*. Here we observed that *Gardnerella sp.* isolated from the bladder (*Gardnerells sp*. 26-12 and *Gardnerella sp.* 30-4) contributed 15% of all classified reads in BV. Especially, *Gardnerella sp.* 26-12 comprised on average 12% of all classified reads in BV. These two species were abundant (between 16-18% of classified reads) in the communities from patients with BV recurrence (Fig. 1, samples 04_001_e and 06_004_a). From the six patients that did not respond to metronidazole therapy, three had a high abundance of transcripts from the bladder isolates in their vaginal fluid samples (samples 05_001_a/b, 15_006_a/b, 17_004_b). In several samples of BV patients, transcripts from bladder isolates were extremely abundant (between 25-49% of all transcripts) (samples 13_022_a, 07_001_a/b, 15_006_a/b, 17_004_a). Some samples ((04_001_a, 06_004_c (recurrence) and 06_001_a/b, 06_006_a/b (without response to treatment)) were dominated by *S. amnii* which has been discovered in the urine microbiome previously [31,32]. While 16S rRNA gene based studies rely on PCR amplification and are thus extremely sensitive and also detect dead pathogens, here we report populations which are not only relatively abundant, but also actively growing. This finding implies that *Gardnerella* species isolated from the bladder may contribute to recurrence of BV. We hypothesize that the *Gardnerella* as well as *Sneathia* species taking shelter in the bladder might be able to invade the vagina once the pH shifts to neutral or alkaline conditions and in such a way they are one factor causing BV recurrence.

Functional profiles in health and disease

In order to profile the function of the communities, all cleaned putative mRNA reads (Table S1 sheet 2, column C) were mapped using BWA onto a customized reference database annotated with KEGG ortholog (KO) genes. We compared the functional profiles of samples from patients in BV (14 patients, 22 samples) with samples from health (after successful

metronidazole treatment, (14 patients, 18 samples). The PCA based on the expression of KO genes clearly showed that there was a large functional difference between BV and health; the functional profiles formed clearly separated clusters (Fig. 3A). Moreover, the functional composition in health was more variable than in BV. This is a striking finding, because 16S rRNA gene sequencing showed a massive increase in diversity in BV [5]. The reason might be that the different species of *Lactobacillus* which dominate in different individuals in health have different functional profiles, while the different species that dominate in BV all have similar functional profiles.

Time course of functional profiles during recurrence

To identify functional difference between women with and without recurrence, the functional profiles of two women with recurrence (patients 04_001 and 06_004) were compared to those of two women without (patients 08_006 and 13_019) over a period of three months. Five time points were analyzed of which the first represented acute BV and the others were after metronidazole treatment. The principal components analysis of the functional profiles is shown in Fig. 3B. In acute BV, samples from all four patients clustered together (red circle). After the treatment, samples from patient 08_006 who was cured without recurrence moved into a very dense and distinct cluster (illustrated by the arrow 1 and enclosed by a green circle). Samples from patient 13_019 who was also cured, moved to a different cluster after treatment, shown by arrow 2 and encircled blue.

The functional shifts in patient 06_004, who experienced recurrence at visit 3, were especially noteworthy: After treatment, the community moved towards a functional profile distinct from all others (arrow 3). The recurrence of BV caused the community to shift back to the BV cluster (red circle, arrow 4). At visit 4, the community moved to the heathy cluster (blue circle) and the patient became healthy according to the Nugent score. We speculate that there

was a medical intervention after visit 3 which changed the microbiome but this was not recorded. Interestingly, the other case of recurrence (patient 04_001) had a different progression. From visit 2 to 4, patient 04_001 seemed to be cured, because these samples clustered together in the healthy cluster indicated by the blue circle. At visit 5, however, patient 04_001 had recurrent BV and the community shifted back again to the BV cluster.

Interestingly, the microbial community of patient 08_006 was the only one dominated by *L. crispatus* after metronidazole treatment, while all others, including those in the second healthy cluster, were dominated by *L. iners*. This confirms the previous observation that *L. iners* and *L. crispatus* are mutually exclusive and suggests that communities dominated by *L. crispatus* are protected from recurrence.

The inerolysin gene was highly expressed in *L. iners* and genes involved in the production of H_2O_2 were highly expressed in *L. crispatus*

As shown above, *L. iners* and *L. crispatus* dominated different patients after treatment, i.e. they were mutually exclusive, and *L. crispatus* seemed to have a protective role against recurrence. Therefore, we compared their genomes and functional profiles with each other. *L. iners* has the smallest genome encoding around 1500 proteins compared to other *Lactobacillus* species like *L. crispatus* which encodes more than twice as many proteins [33]. To gain more understanding of their different roles in the healthy vaginal microbial community, we performed a differential expression (DE) analysis based on their KO genes comparing the expression of KO genes of *L. crispatus* in *L. crispatus* dominated samples (n = 4) with the expression of KO genes of *L. iners* in *L. iners* dominated samples (n = 11). Fig. 3C shows that the two species share 569 KO genes, while 58 are unique to *L. iners* and 244 are unique for *L. crispatus*. The DE analysis identified 584 significantly differentially expressed KO genes, of which 355 were up-regulated in *L. crispatus* (Table S1 sheet 10).

Among the top 100 most differentially expressed KO genes in terms of FDR value, 89 were up-regulated in *L. crispatus*, indicating *L. crispatus* possesses far more diverse functions than *L. iners.* Remarkably, genes assumed to be involved in the production of H_2O_2 (pyruvate oxidase, NADH oxidase, glycolate oxidase) [34,35] were highly expressed in *L. crispatus* (Fig. 3C) suggesting a high level of hydrogen peroxide in vaginal microbiota dominated by *L. crispatus.* Hydrogen peroxide provides protection against pathogens for the symbiotic vaginal community [36]. An iron transport system was found to be highly expressed in *L. crispatus* and it is absent in the genome of *L. iners*. On the other hand, we identified a gene encoding inerolysin (INY) which was highly expressed (log2 counts per million mapped reads =10.18) in *L. iners*, but absent in the genome of *L. crispatus* (Fig. 3C, Table S1 sheet 10). INY, a cholesterol-dependent cytolysin (CDC) produced by *L. iners*, is a pore-forming toxin that is activated by reducing agents and inhibited by excess cholesterol [15]. CDCs were formerly annotated as "thiol-activated cytolysins" [37]. Interestingly, we also found the orthologous gene (K11031) of inerolysin expressed in *G. vaginalis* (details in Table S1 sheet 7). These findings show that *L. iners* exerts a similar activity as the BV pathogen *G. vaginalis* indicating it may initiate the dysbiotic shift of the vaginal flora thus contribute to the pathogenesis of BV as well as to recurrence.

Biofilm related pathways and vitamin B metabolism were up-regulated in BV

Using KEGG pathway enrichment analysis based on differentially expressed KO genes between BV and health, we found that in BV more pathways were enriched in the up-regulated KO genes than down-regulated genes (Table S1 sheet 9). As shown in Fig. 4, pathways related to the biosynthesis and metabolism of amino acids and vitamin B6 metabolism were highly enriched in up-regulated KO genes. Additionally, pathways of other B vitamins were enriched in up-regulated KO genes in BV as well, though the FDR did not

reach significance level, namely, thiamine metabolism (P=0.019, FDR=0.196), pantothenate and CoA biosynthesis (P=0.010, FDR=0.134), and riboflavin metabolism (P = 0.117, FDR = 0.526) (Table S1 sheet 9). Interestingly, lipopolysaccharide (LPS) biosynthesis, flagellar assembly and bacterial chemotaxis (P = 0.007, FDR = 0.111, Table S1 sheet 9) which are essential for living in biofilms were strongly up-regulated (Fig. 4).

In vivo **expression of putative metronidazole resistance associated genes in *G. vaginalis***

Metronidazole is a prodrug which requires anaerobic conditions for enzymatic reduction within the cell to transform it into an active form [8]. Several enzymes are involved in the reduction of metronidazole such as ferredoxin, ferredoxin/ferredoxin-NADP reductase (FNR), and nitroreductase [8,38-41]. The reduced metronidazole can introduce DNA strand breaks thus leading to the death of the pathogens. Accordingly, genes responsible for DNA repair like recA and recA-mediated autopeptidase (Rma) might contribute to resistance against metronidazole [42]. Moreover, a gene named nitroimidazole resistance gene (NIM) was shown to be related to the resistance against metronidazole in *Bacteroides fragilis* [43]. To clarify the possible contribution of these genes to metronidazole resistance of the vaginal microbiota, we examined the expression (Table S1 sheet 12) of these genes in *G. vaginalis* by comparing their expression in the communities with later response to metronidazole and without at visit 1 (acute BV). For this analysis, BV communities from 10 patients were analyzed in which the level of *G. vaginalis* transcripts was > 20%. Six of these patients did not respond to treatment and four were initially cured. Although *A. vaginae* and *S. amnii* are also key players in BV we did not analyze them here since there were two few samples dominated by them. As shown in Fig. S2, there was no clear expression pattern for most of these genes (detailed data in Table S1 sheet 12). However, ferredoxin/ferredoxin-NADP reductase (FNR) was clearly down-regulated in *G. vaginalis* in communities with no response

to the treatment (with relative abundance 0.014 ± 0.014% versus 0.059 ± 0.031%, Wilcoxon test P-value = 0.0381, relative abundance = read count of given genes of *G. vaginalis* / read count of *G. vaginalis*). FNR is involved in activating metronidazole [41] thus its weak expression indicates the lack of the active antibiotic, in accordance with the lack of response in those patients.

RecA and RecA-mediated autopeptidase (Rma) were up-regulated in some but not all communities that did not respond to metronidazole (with relative abundance 0.047 ± 0.015%, 0.034 ± 0.026% versus 0.036 ± 0.017%, 0.022 ± 0.028%, respectively). RecA and RecA-mediated autopeptidase are essential genes in SOS response [44] responsible for DNA repair in bacteria. They may play a crucial role in survival of bacteria after treatment with a DNA damaging agent like metronidazole.

CRISPR associated protein and adenine methyltransferase genes of *G. vaginalis* were strongly up-regulated in vaginal fluids of patients not responding to treatment

We then performed a global analysis of differential expression of KO genes of *G. vaginalis* in these same communities (visit 1, 10 BV samples with > 20% transcripts from *G. vaginalis* (6 patients that did not respond to treatment and 4 that were initially cured). We observed that there were 24 KO genes highly up-regulated with FDR <= 0.05 (log2 fold change up to 12.4) in communities without response. Strikingly, among the top ten most strongly up-regulated KO genes, seven were *cas* genes [45] (CRISPR-associated protein genes, *cas*1-3, *cas*A-D) (Fig. 5). Cas proteins have a dual function; they can perform homology-directed repair as well as degrade foreign DNA [46].

yhdJ and *licD* were the most strongly up-regulated genes with log2 fold change above 10, of which *yhdJ* encodes an adenine-specific DNA-methyltransferase and *licD* encodes a

lipopolysaccharide cholinephosphotransferase. The DNA adenine methyltransferase is an essential part of the restriction modification system that protects bacteria from digesting their own genome via their restriction enzymes [47].

There were 32 KO genes down-regulated but the fold change values were not as high as for the up-regulated genes. *galK* (galactokinase) was identified as the most strongly down-regulated gene with a log2 fold change of -5.6.

These findings suggest that CRISPR-Cas, a phage defense mechanism with a dual function, might mitigate the DNA damaging effect of metronidazole and thus may be one of the reasons for lack of response of the vaginal microbiota to treatment. In addition, lack of expression of an enzyme (FNR) that transforms the metronidazole prodrug into the active form might also contribute to lack of response to treatment.

Discussion

The aim of this study was to identify functional patterns in the vaginal fluid microbiota in health and BV. In particular, we compared metatranscriptomes of recurrent and non-recurrent patients and of patients who did and did not respond to metronidazole treatment. This is the first study to investigate the compositional as well as functional alterations of the vaginal microbiota from patients with BV during treatment with the antibiotic metronidazole *in vivo* using the metatranscriptomic approach. We found clear patterns which allowed us to formulate novel hypotheses. Since our analysis is based on a relatively small number of 14 patients, our findings should be followed up in larger patient groups with more cases of recurrence and more cases of lack of response to antibiotic treatment.

We observed that *G. vaginalis* was much more dominant based on the abundance of its transcripts than based on the 16S rRNA amplicon sequencing analysis. This suggests that *G.*

517 *vaginalis* is transcriptionally more active than other vaginal bacteria; moreover, the
518 commonly used 27F primer was previously shown to underrepresent *G. vaginalis* [30]. We
519 found that the functional profiles in health and BV were massively different. Interestingly, *L.*
520 *crispatus* and *L. iners* were the dominant species in health and communities were dominated
521 by either one or by the other. There were many factors observed by laboratory or genomic
522 studies [34,48,49] which show that more protection is provided by *L. crispatus* rather than by
523 *L. iners*. *L. crispatus* has the ability to produce hydrogen peroxide to prevent pathogen
524 colonization, whereas *L. iners* does not [34]. Furthermore, an iron transport system was found
525 in the genome of *L. crispatus* and not in *L. iners* suggesting that *L. crispatus* can capture iron
526 released by the host during menstruation away from potential pathogens [48]. On the other
527 hand, *L. iners* can produce inerolysin, a pore-forming cytolysin that enables nutrient
528 acquisition from host cells providing it with an advantage when conditions are scarce [49].
529 Strikingly, an ortholog of inerolysin was found highly expressed in *G. vaginalis* in BV. This
530 is the first *in vivo* study revealing that *L. iners* behaves pathogenic in health after antibiotic
531 treatment using metatranscriptome analysis. These findings suggest that *L. iners* can initiate
532 the dysbiotic shift of the vaginal flora and contribute to the pathogenesis of BV as well as to
533 recurrence. These observations may also explain why *L. crispatus* is more beneficial than *L.*
534 *iners* and why *L. iners* has the ability to thrive faster after antibiotic treatment.

535 Moreover, *Gardnerella* species isolated from the bladder were highly active in several
536 patients with BV, including two patients with recurrent BV. Thus re-colonization of the
537 vagina by pathogens residing in the bladder should be considered as an important mechanism
538 contributing to recurrence of BV.

539 Many potential reasons are discussed for recurrence and the occasional lack of response to
540 metronidazole treatment. For example, it has been reported that sexual activity can introduce

pathogens thus may also contribute to the recurrence of BV [11], dosage of metronidazole may have been too low for the patient, or highly virulent strains of *G. vaginalis* may be residing in this particular non-responding patient.

In our metatranscriptome analysis we found evidence for mechanisms that hinder the activation of the metronidazole prodrug, or mitigate the damage that metronidazole inflicts on DNA, and thus could be important reasons for the lack of response in some women.

We show for the first time that the FNR gene of *G. vaginalis* was down-regulated in those patients that did not respond to metronidazole. This gene is required for the activation of metronidazole in anaerobes [41].

Adenine methyltransferase was the most strongly up-regulated gene of *G. vaginalis* in communities from patients without response to metronidazole suggesting that it may play a role in protecting DNA from damage. Studies have provided compelling evidence for DNA methyltransferases 1 (DNMT1) acting as a regulator of genome integrity and as an early responder to DNA double strand breaks in human cells [50], but the explicit role of DNMTs in DNA strand break repair for bacteria remains to be elucidated.

Remarkably, *cas* genes of *G. vaginalis* were highly up-regulated in *G. vaginalis* from VMB communities of patients that did not respond to metronidazole treatment. The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated proteins) genes are present in about half of all Bacteria and most Archaea [51]; they represent a mechanism of adaptive immunity which protects the prokaryotic cell against foreign DNA and has been developed into a universal tool for genome editing [52]. The *cas* genes of *G. vaginalis* belong to the *E. coli* subtype and were found in about half of the clinical isolates [45]. Their up-regulation might reflect increased phage attacks in BV. Phages have been hypothesized to be crucial for the etiology of BV by causing the collapse of *Lactobacillus*

565 populations [53]; accordingly, *L. iners* upregulates its CRISPR-Cas system in BV [16]. More
566 than 400 annotated prophage sequences were found in 39 *Gardenerella* isolates from the
567 bladder [21]. They might be induced to enter the lytic cycle by the change in pH
568 accompanying the shift to BV.

569 The upregulation of CRISPR-Cas system genes in *G. vaginalis* from those patients that did
570 not respond to treatment by metronidazole suggests that the CRISPR-Cas system might have a
571 role in not only defending the host against phages, but also in mitigating the DNA damaging
572 effect of metronidazole. In addition to providing adaptive immunity, CRISPR-Cas systems
573 can have various additional functions [54] and it was shown that they can protect the cell
574 against DNA damaging agents [46]. The Cas1 enzyme of *E. coli* (YgbT) physically and
575 genetically interacts with the DNA repair system (RecBC, RuvB) and is recruited to DNA
576 double strand breaks; moreover, YgbT is necessary for resistance of *E. coli* to DNA damage
577 caused by the genotoxic antibiotic mitomycin C or UV light [46]. Our findings suggest that
578 the CRISPR-Cas system may protect the vaginal microbiota against the DNA damaging effect
579 of metronidazole. If experimentally confirmed, this finding might open a new path to fight
580 bacterial resistance against DNA damaging agents.

581 ## Conclusions

582 We hypothesize that colonization by *L. iners* and re-infection through the bladder contribute
583 to recurrence of BV. CRISPR-Cas, a phage defense mechanism with a dual function, might
584 mitigate the DNA damaging effect of metronidazole and thus may be one of the reasons for
585 lack of response of the vaginal microbiota to treatment. This first *in vivo* study of the
586 metatranscriptome landscape of vaginal fluid microbiota during metronidazole treatment
587 provides a wealth of new understanding towards the possible mechanisms of lack of response

to metronidazole and recurrence in BV. These findings may help to develop more efficient novel therapies to improve the treatment of BV and prevent recurrence. A completely new role for Cas proteins in microbial communities is hypothesized which warrants closer inspection.

Declarations

Ethics approval and consent to participate

The protocol of the clinical trial [5] was approved by the local ethics committee (Ärztekammer Nordrhein - Medical Association North Rhine) and written consent was obtained from all participants. The clinical trial was conducted in accordance with the Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects. Principles and guidelines for good clinical practice were followed.

Consent for publication

Not applicable

Availability of data and material

The clinical trial was registered on ClinicalTrials.gov with the identifier NCT02687789. The metatranscriptome sequencing data of this subgroup analysis have been deposited in European Nucleotide Archive and are accessible through study accession number PRJEB21446.

Competing interests

The authors have read the journal's policy. CM and CA have the following competing interests: they work for the company Dr. August Wolff GmbH & Co. KG Arzneimittel which

608 designed and conducted the clinical study. However, they had no influence on data analysis
609 and interpretation, or the decision to submit the work for publication.

610 **Funding**

611 This work was supported by ZIM project grant KF3134201MD3 of the Bundesministerium
612 für Wirtschaft und Energie (BMWi), Germany. The clinical study was funded by Dr. August
613 Wolff GmbH & Co. KG Arzneimittel within the ZIM project KF3134201MD3.

614 **Authors' contributions**

615 This study was designed by IWD and CG. CM and CA provided the clinical samples. RNA
616 extraction and mRNA enrichment was performed by CG. SB prepared cDNA libraries, and
617 performed Illumina sequencing. Z-LD performed all the data analyses. Data interpretation and
618 visualization were done by IWD, Z-LD and CG. Z-LD and CG wrote the manuscript draft,
619 and all authors reviewed the manuscript.

620 **Acknowledgements**

621 We would like to thank Dr. Juergen Tomasch for stimulating discussions and constructive
622 suggestions and iMED program in Helmholtz association for support.

623

624

625 Reference List

626

1. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL *et al.*: **Vaginal microbiome of reproductive-age women.** *Proc Natl Acad Sci U S A* 2011, **108 Suppl 1:** 4680-4687.

2. van de Wijgert JH, Borgdorff H, Verhelst R, Crucitti T, Francis S, Verstraelen H *et al.*: **The vaginal microbiota: what have we learned after a decade of molecular characterization?** *PLoS One* 2014, **9:** e105998.

3. Onderdonk AB, Delaney ML, Fichorova RN: **The Human Microbiome during Bacterial Vaginosis.** *Clin Microbiol Rev* 2016, **29:** 223-238.

4. Dols JA, Molenaar D, van der Helm JJ, Caspers MP, de KA-B, Schuren FH *et al.*: **Molecular assessment of bacterial vaginosis by Lactobacillus abundance and species diversity.** *BMC Infect Dis* 2016, **16:** 180.

5. Gottschick C, Deng Z-L, Vital M, Masur C, Abels C, Pieper D *et al.*: **Treatment of biofilms in bacterial vaginosis by an amphoteric tenside pessary - clinical study and microbiome analysis.** *Microbiome* 2017, **in press**.

6. Bradshaw CS, Pirotta M, De GD, Hocking JS, Morton AN, Garland SM *et al.*: **Efficacy of oral metronidazole with vaginal clindamycin or vaginal probiotic for bacterial vaginosis: randomised placebo-controlled double-blind trial.** *PLoS One* 2012, **7:** e34540.

7. Lofmark S, Edlund C, Nord CE: **Metronidazole is still the drug of choice for treatment of anaerobic infections.** *Clin Infect Dis* 2010, **50 Suppl 1:** S16-S23.

8. Mendz GL, Megraud F: **Is the molecular basis of metronidazole resistance in microaerophilic organisms understood?** *Trends Microbiol* 2002, **10:** 370-375.

9. Hilbert DW, Smith WL, Paulish-Miller TE, Chadwick SG, Toner G, Mordechai E *et al.*: **Utilization of molecular methods to identify prognostic markers for recurrent bacterial vaginosis.** *Diagn Microbiol Infect Dis* 2016, **86:** 231-242.

10. Ferreira CST, Donders GG, Parada CMGL, Tristao ADR, Fernandes T, da Silva MG *et al.*: **Treatment failure of bacterial vaginosis is not associated with higher loads of Atopobium vaginae and Gardnerella vaginalis.** *J Med Microbiol* 2017, **66:** 1217-1224.

11. Bradshaw CS, Morton AN, Hocking J, Garland SM, Morris MB, Moss LM *et al.*: **High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence.** *J Infect Dis* 2006, **193:** 1478-1486.

12. Sobel JD, Schmitt C, Meriwether C: **Long-term follow-up of patients with bacterial vaginosis treated with oral metronidazole and topical clindamycin.** *J Infect Dis* 1993, **167:** 783-784.

13. Machado D, Castro J, Palmeira-de-Oliveira A, Martinez-de-Oliveira J, Cerca N: **Bacterial Vaginosis Biofilms: Challenges to Current Therapies and Emerging Solutions.** *Front Microbiol* 2015, **6:** 1528.

14. Turovskiy Y, Sutyak NK, Chikindas ML: **The aetiology of bacterial vaginosis.** *J Appl Microbiol* 2011, **110:** 1105-1128.

15. Petrova MI, Reid G, Vaneechoutte M, Lebeer S: **Lactobacillus iners: Friend or Foe?** *Trends Microbiol* 2016.

16. Macklaim JM, Fernandes AD, Di Bella JM, Hammond JA, Reid G, Gloor GB: **Comparative meta-RNA-seq of the vaginal microbiota and differential expression by Lactobacillus iners in health and dysbiosis.** *Microbiome* 2013, **1:** 12.

17. Mohammadzadeh F, Dolatian M, Jorjani M, Alavi MH: **Diagnostic value of Amsel's clinical criteria for diagnosis of bacterial vaginosis.** *Glob J Health Sci* 2014, **7:** 8-14.

18. Aronesty E: **ea-utils: Command-line tools for processing biological sequencing data.** *Durham, NC: Expression Analysis* 2011.

19. Kopylova E, Noe L, Touzet H: **SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data.** *Bioinformatics* 2012, **28:** 3211-3217.

20. Wood DE, Salzberg SL: **Kraken: ultrafast metagenomic sequence classification using exact alignments.** *Genome Biol* 2014, **15:** R46.

21. Malki K, Shapiro JW, Price TK, Hilt EE, Thomas-White K, Sircar T *et al.*: **Genomes of Gardnerella Strains Reveal an Abundance of Prophages within the Bladder Microbiome.** *PLoS One* 2016, **11:** e0166757.

22. Methe BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C *et al.*: **A framework for human microbiome research.** *Nature* 2012, **486:** 215-221.

23. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.** *arXiv preprint arXiv:1303 3997* 2013.

24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.*: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25:** 2078-2079.

25. Buchfink B, Xie C, Huson DH: **Fast and sensitive protein alignment using DIAMOND.** *Nat Methods* 2015, **12:** 59-60.

26. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26:** 139-140.

27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W *et al.*: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43:** e47.

28. Shen W, Le S, Li Y, Hu F: **SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation.** *PLoS One* 2016, **11:** e0163962.

29. Deng ZL, Szafranski SP, Jarek M, Bhuju S, Wagner-Dobler I: **Dysbiosis in chronic periodontitis: Key microbial players and interactions with the human host.** *Sci Rep* 2017, **7:** 3703.

30. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ: **Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes.** *Appl Environ Microbiol* 2008, **74:** 2461-2470.

31. Whiteside SA, Razvi H, Dave S, Reid G, Burton JP: **The microbiome of the urinary tract--a role beyond infection.** *Nat Rev Urol* 2015, **12:** 81-90.

32. Gottschick C, Deng Z-L, Vital M, Masur C, Abels C, Pieper D *et al.*: **The urinary microbiota of men and women and its changes in women during bacterial vaginosis and antibiotic treatment.** *Microbiome* 2017, **5:** 99.

33. France MT, Mendes-Soares H, Forney LJ: **Genomic Comparisons of Lactobacillus crispatus and Lactobacillus iners Reveal Potential Ecological Drivers of Community Composition in the Vagina.** *Appl Environ Microbiol* 2016, **82:** 7063-7073.

34. Ojala T, Kankainen M, Castro J, Cerca N, Edelman S, Westerlund-Wikstrom B *et al.*: **Comparative genomics of Lactobacillus crispatus suggests novel mechanisms for the competitive exclusion of Gardnerella vaginalis.** *BMC Genomics* 2014, **15:** 1070.

35. Hertzberger R, Arents J, Dekker HL, Pridmore RD, Gysler C, Kleerebezem M *et al.*: **H(2)O(2) production in species of the Lactobacillus acidophilus group: a central role for a novel NADH-dependent flavin reductase.** *Appl Environ Microbiol* 2014, **80:** 2229-2239.

36. Cardone A, Zarcone R, Borrelli A, Di CA, Russo A, Tartaglia E: **Utilisation of hydrogen peroxide in the treatment of recurrent bacterial vaginosis.** *Minerva Ginecol* 2003, **55:** 483-492.

37. Macklaim JM, Gloor GB, Anukam KC, Cribby S, Reid G: **At the crossroads of vaginal health and disease, the genome sequence of Lactobacillus iners AB-1.** *Proc Natl Acad Sci U S A* 2011, **108 Suppl 1:** 4688-4695.

38. Tsugawa H, Suzuki H, Satoh K, Hirata K, Matsuzaki J, Saito Y *et al.*: **Two amino acids mutation of ferric uptake regulator determines Helicobacter pylori resistance to metronidazole.** *Antioxid Redox Signal* 2011, **14:** 15-23.

39. Wassmann C, Hellberg A, Tannich E, Bruchhaus I: **Metronidazole resistance in the protozoan parasite Entamoeba histolytica is associated with increased expression of iron-containing superoxide dismutase and peroxiredoxin and decreased expression of ferredoxin 1 and flavin reductase.** *J Biol Chem* 1999, **274:** 26051-26056.

40. Leitsch D, Burgess AG, Dunn LA, Krauer KG, Tan K, Duchene M *et al.*: **Pyruvate:ferredoxin oxidoreductase and thioredoxin reductase are involved in 5-nitroimidazole activation while flavin metabolism is linked to 5-nitroimidazole resistance in Giardia lamblia.** *J Antimicrob Chemother* 2011, **66:** 1756-1765.

41. Ishii M, Ueda Y, Yoon KS, Igarashi Y, Kodama T: **Purification and characterization of ferredoxin from Hydrogenobacter thermophilus strain TK-6.** *Biosci Biotechnol Biochem* 1996, **60:** 1513-1515.

42. Steffens LS, Nicholson S, Paul LV, Nord CE, Patrick S, Abratt VR: **Bacteroides fragilis RecA protein overexpression causes resistance to metronidazole.** *Res Microbiol* 2010, **161:** 346-354.

43. Lofmark S, Fang H, Hedberg M, Edlund C: **Inducible metronidazole resistance and nim genes in clinical Bacteroides fragilis group isolates.** *Antimicrob Agents Chemother* 2005, **49:** 1253-1256.

44. Michel B: **After 30 years of study, the bacterial SOS response still surprises us.** *PLoS Biol* 2005, **3:** e255.

45. Pleckaityte M, Zilnyte M, Zvirbliene A: **Insights into the CRISPR/Cas system of Gardnerella vaginalis.** *BMC Microbiol* 2012, **12:** 301.

46. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B *et al.*: **A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair.** *Mol Microbiol* 2011, **79:** 484-502.

47. Pukkila PJ, Peterson J, Herman G, Modrich P, Meselson M: **Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli.** *Genetics* 1983, **104:** 571-582.

48. France MT, Mendes-Soares H, Forney LJ: **Genomic Comparisons of Lactobacillus crispatus and Lactobacillus iners Reveal Potential Ecological Drivers of Community Composition in the Vagina.** *Appl Environ Microbiol* 2016, **82:** 7063-7073.

49. Rampersaud R, Planet PJ, Randis TM, Kulkarni R, Aguilar JL, Lehrer RI *et al.*: **Inerolysin, a cholesterol-dependent cytolysin produced by Lactobacillus iners.** *J Bacteriol* 2011, **193:** 1034-1041.

50. Ha K, Lee GE, Palii SS, Brown KD, Takeda Y, Liu K *et al.*: **Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery.** *Hum Mol Genet* 2011, **20:** 126-140.

51. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ *et al.*: **An updated evolutionary classification of CRISPR-Cas systems.** *Nat Rev Microbiol* 2015, **13:** 722-736.

52. Doudna JA, Charpentier E: **Genome editing. The new frontier of genome engineering with CRISPR-Cas9.** *Science* 2014, **346:** 1258096.

53. Blackwell AL: **Vaginal bacterial phaginosis?** *Sex Transm Infect* 1999, **75:** 352-353.

54. Westra ER, Buckling A, Fineran PC: **CRISPR-Cas systems: beyond adaptive immunity.** *Nat Rev Microbiol* 2014, **12:** 317-326.

Figures, tables and additional files

Fig. 1: Taxonomic composition of vaginal fluid metatranscriptomes in BV and after treatment with metronidazole. Species with relative abundance smaller than 2% in all samples were grouped into "Others". The red dot on top of the samples indicates BV. The digits indicate the patient ID, while the letters a-e denote visit 1-5. After the first sampling at visit 1 the patients were treated with metronidazole. Total putative bacterial mRNA reads were mapped to the refined reference database using Kraken (see Methods for details). BV status was determined by Nugent score.

Fig. 2: Average taxonomic composition of vaginal fluid samples in BV determined by 16S rRNA amplicon sequencing (A) and metatranscriptome sequencing (B) and taxonomic composition of communities on the species level determined by metatranscriptome sequencing in health and BV (C). In A, amplicon sequencing was performed as described in our previous study [5] using primers V1-V2. For B and C, taxonomy was assigned based on all cleaned reads after removal of human reads using Kraken and the refined reference database. Species with average relative abundance over 2% are shown in A and B. Relative average abundance was calculated based on all mapped reads. Mean and standard error are shown. In C, red and pink dots on top of the heatmap indicate paired samples of patients which developed recurrence.

Fig. 3: Distinct functional profiles in health and BV (A), temporal dynamics of the metatranscriptome during treatment (B), and differential KO gene expression of *Lactobacillus crispatus* and *L. iners* (C). A) Principal components analysis of KO gene expression shows highly similar community activity in BV which is clearly distinct but much more variable in health. **B)** Time course based on PCA of functional profiles from four women during treatment and recurrence over a period of three months (five sampling time

805 points). Two women with recurrence (pink and blue color range) and two women without
806 recurrence (green and orange color range) are shown. In the figure legend, BV indicates time
807 point with BV, R indicates the recurrence and H represents health. The green and blue circles
808 highlight healthy clusters, respectively, while the red circle highlights samples from BV. The
809 arrows denote the temporal shifts of the communities during the treatment. **C)** *L. crispatus*
810 and *L. iners* exhibit distinct function activities. The Venn diagram indicates the unique KO
811 genes of *L. crispatus* and *L. iners* as well as their shared KO genes. The innermost ring
812 denotes the expression of KO genes by log2CPM, the outer ring illustrates the fold change of
813 the expression of KO genes between *L. crispatus* dominated communities and *L. iners*
814 dominated communities by log2FC. The small red triangle marks the inerolysin gene, while
815 blue triangles mark the genes involved in the production of hydrogen peroxide.

816 **Fig. 4: The functional shifts on KEGG pathways in BV communities.** The size of the dot
817 indicates the number of up-regulated genes in each pathway, and the x-axis denotes the ratio
818 of number of up-regulated genes in a given pathway divided by number of expressed genes in
819 this pathway.

820 **Fig. 5: Seven *cas* genes and a methyltransferase gene of *G. vaginalis* were highly up-**
821 **regulated in communities which did not respond to the treatment**. KO genes with FDR
822 <= 0.05 are colored in red or turquoise (significantly differentially regulated) while FDR >
823 0.05 are in grey.

824

825 **Table 1. Overview of sequencing results**

Overview of sequencing results	BV (22 samples)	Health (18 samples)
raw reads	44,503,332 ± 8,934,125	50,048,447 ± 10,989,072
cleaned reads after quality control and rRNA removal	32,506,992 ± 5,955,546	36,797,904 ± 6,934,913
putative microbial mRNA reads	27,524,573 ± 6,035,462	8,686,750 ± 5,867,867

826

827 **Table 2. Mapping statistics using Kraken and BWA**

Mapping statistics	% mapped putative microbial mRNA reads
standard Kraken database	40.55 ± 23.93
refined Kraken database	77.22 ± 14.44
customized reference gene database using BWA	~ 78.02 ± 16.31

828

829 **Fig. S1: The dominance of the vaginal microbiome in BV and health based on**
830 **metatranscriptome with the standard Kraken database.** Red denotes the dominance curve
831 in BV while blue represents the dominance in health. Y-axis indicates the cumulative relative
832 abundance.

833 **Fig. S2: Differential expression of putative metronidazole resistance associated genes of**
834 ***G. vaginalis* in vaginal fluid microbiota that did not respond to treatment.** The expression
835 value was calculated based on relative abundance of reads mapped onto *G. vaginalis* using
836 BWA. "NR1" (No Response 1) indicates the samples from BV that did not respond to
837 metronidazole treatment later; "NR2" (No Response 2) indicates the samples after treatment
838 with no response to the treatment. "WR1" (With Response 1) represents the samples with
839 acute BV which afterwards responded to metronidazole.

840

Supplementary Table S1: All supplementary data

Sheet 1. Sample description

Sheet 2. Read summary

Sheet 3. Genomes in the refined Kraken reference genome database

Sheet 4. Customized reference gene database for BWA

Sheet 5. Species composition determined by Kraken based on refined database

Sheet 6. The comparison of the community composition identified by 16S and metatranscriptome (average relative abundance >= 2%)

Sheet 7. The gene expression based on customized reference gene database with KO annotation

Sheet 8. Differential expression of KO between BV and health

Sheet 9. KEGG pathway enriched in up-regulated KO genes in BV

Sheet 10. The differential expression of KO genes between *L. crispatus* and *L. iners*

Sheet 11. KEGG pathway enriched in up-regulated KO genes in *L. crispatus*

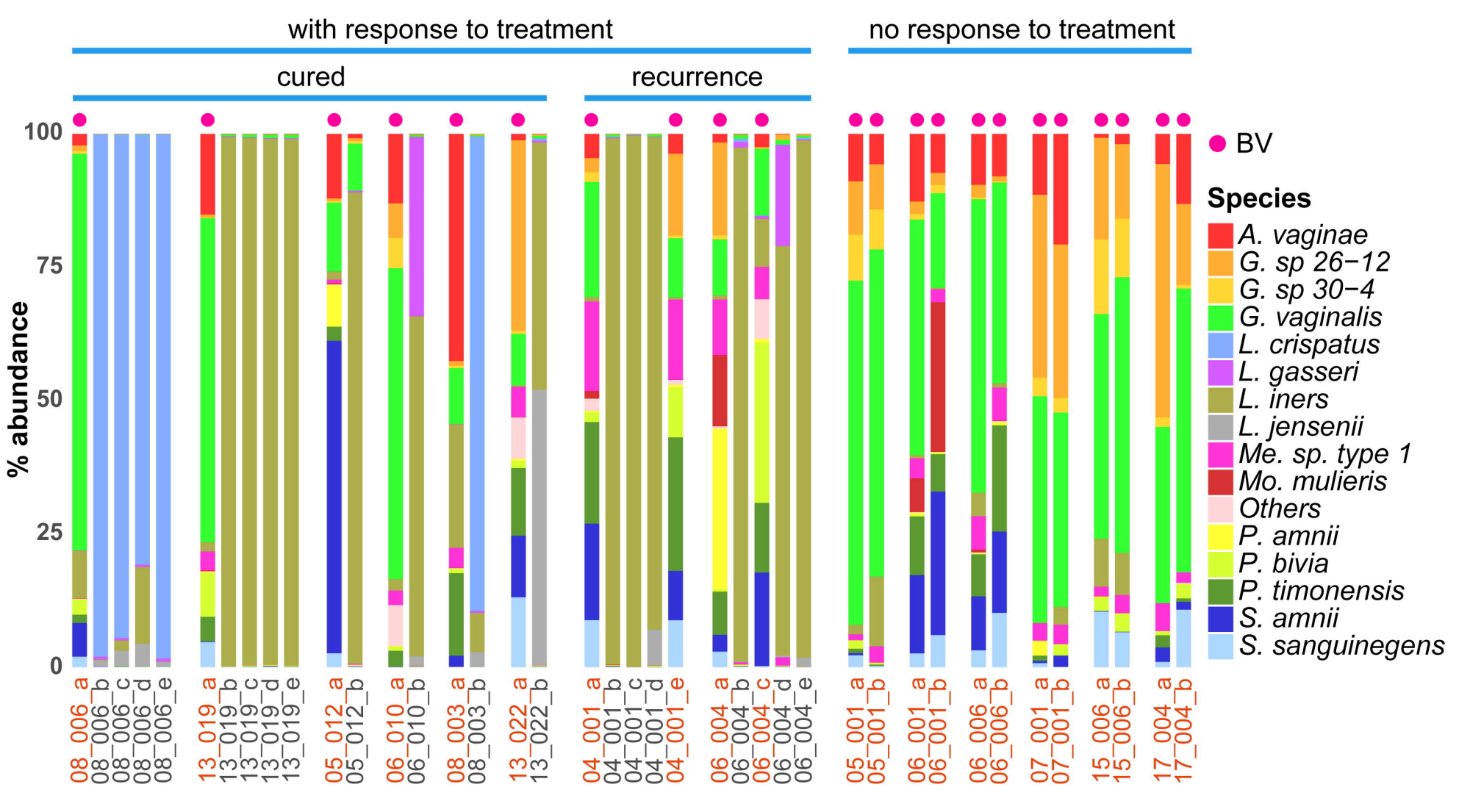
Sheet 12. Expression of metronidazole associated genes in *G. vaginalis*

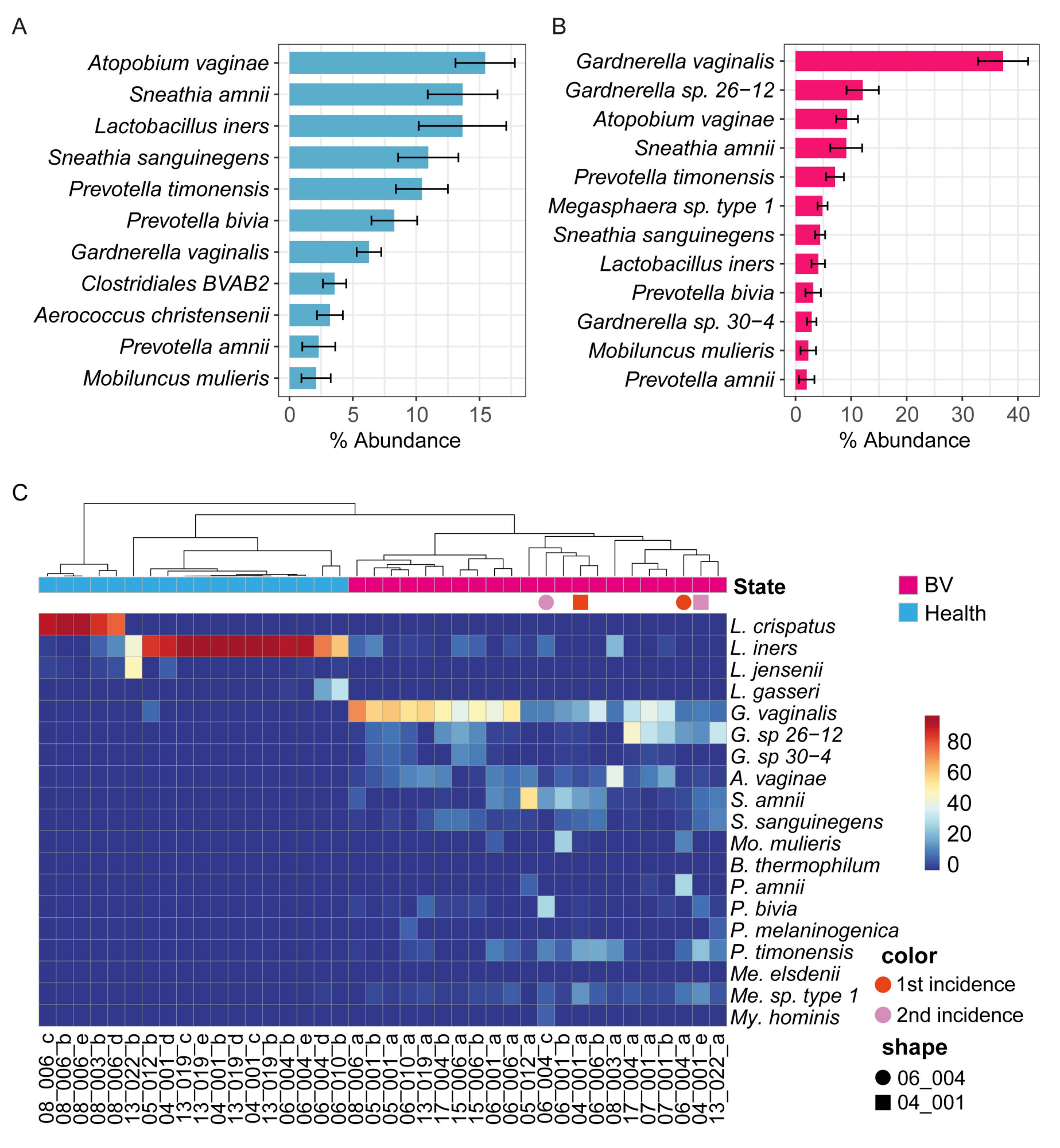
Sheet 13. Expression of metronidazole associated genes in *A. vaginae*

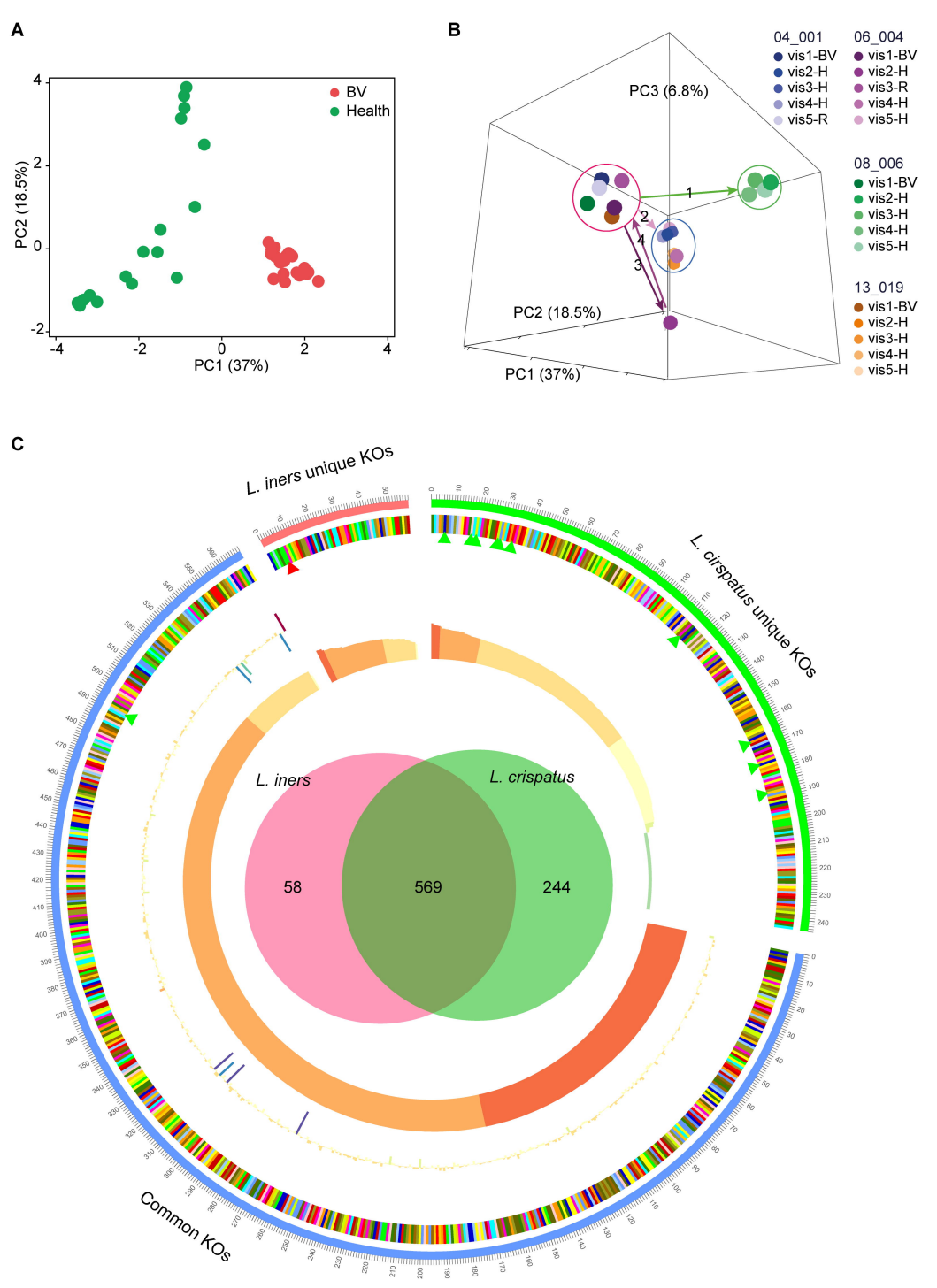
Sheet 14. The differential expression of KO genes between community without response to metronidazole treatment and with response

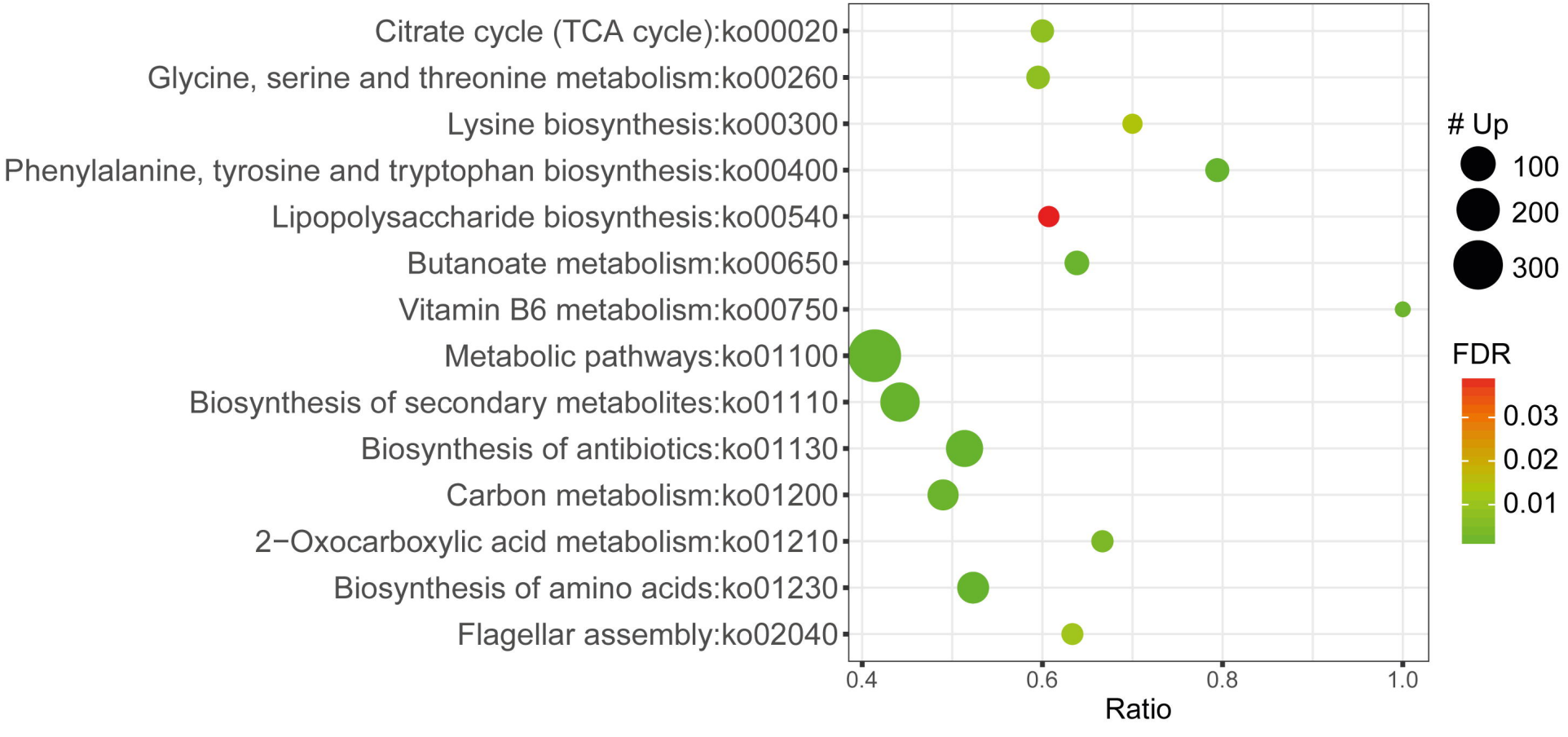
Sheet 15. The differential expression of KO genes of *G. vaginalis* from community without response to metronidazole treatment and with response

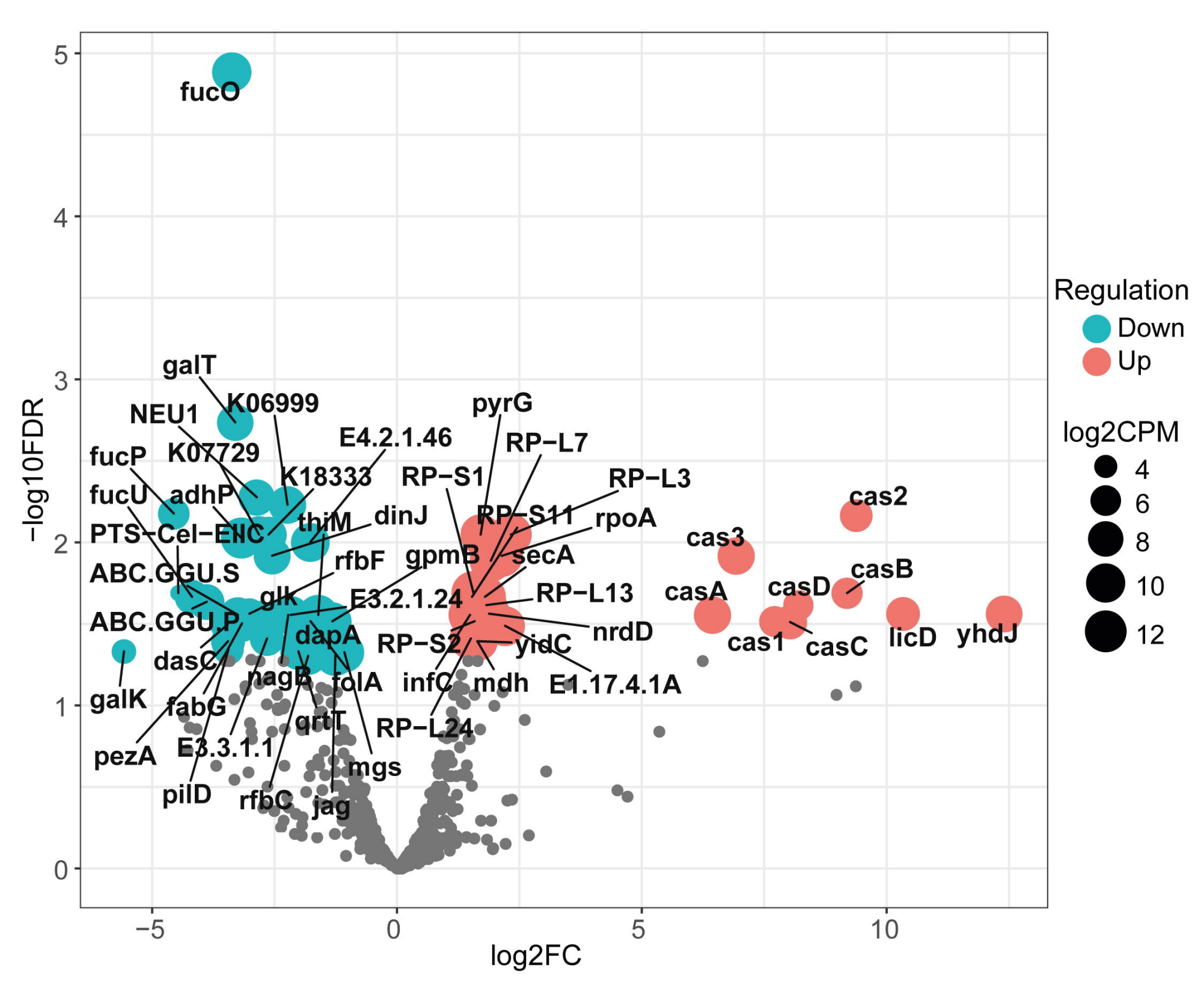
Sheet 16. The differential expression of KO genes of *G. vaginalis* from community after treatment without response to metronidazole and before the treatment with response


with response to treatment
no response to treatment
cured
recurrence
100
75
50
25
0
% abundance
BV
Species
A. vaginae
G. sp 26−12
G. sp 30−4
G. vaginalis
L. crispatus
L. gasseri
L. iners
L. jensenii
Me. sp. type 1
Mo. mulieris
Others
P. amnii
P. bivia
P. timonensis
S. amnii
S. sanguinegens
08_006_a
08_006_b
08_006_c
08_006_d
08_006_e
13_019_a
13_019_b
13_019_c
13_019_d
13_019_e
05_012_a
05_012_b
06_010_a
06_010_b
08_003_a
08_003_b
13_022_a
13_022_b
04_001_a
04_001_b
04_001_c
04_001_d
04_001_e
06_004_a
06_004_b
06_004_c
06_004_d
06_004_e
05_001_a
05_001_b
06_001_a
06_001_b
06_006_a
06_006_b
07_001_a
07_001_b
15_006_a
15_006_b
17_004_a
17_004_b


A
Atopobium vaginae
Sneathia amnii
Lactobacillus iners
Sneathia sanguinegens
Prevotella timonensis
Prevotella bivia
Gardnerella vaginalis
Clostridiales BVAB2
Aerococcus christensenii
Prevotella amnii
Mobiluncus mulieris
0
5
10
15
% Abundance
B
Gardnerella vaginalis
Gardnerella sp. 26−12
Atopobium vaginae
Sneathia amnii
Prevotella timonensis
Megasphaera sp. type 1
Sneathia sanguinegens
Lactobacillus iners
Prevotella bivia
Gardnerella sp. 30−4
Mobiluncus mulieris
Prevotella amnii
0
10
20
30
40
% Abundance
C
State
BV
Health
L. crispatus
L. iners
L. jensenii
L. gasseri
G. vaginalis
G. sp 26−12
G. sp 30−4
A. vaginae
S. amnii
S. sanguinegens
Mo. mulieris
B. thermophilum
P. amnii
P. bivia
P. melaninogenica
P. timonensis
Me. elsdenii
Me. sp. type 1
My. hominis
80
60
40
20
0
color
1st incidence
2nd incidence
shape
06_004
04_001


A
BV
Health
PC2 (18.5%)
PC1 (37%)
B
04_001
vis1-BV
vis2-H
vis3-H
vis4-H
vis5-R
06_004
vis1-BV
vis2-H
vis3-R
vis4-H
vis5-H
08_006
vis1-BV
vis2-H
vis3-H
vis4-H
vis5-H
13_019
vis1-BV
vis2-H
vis3-H
vis4-H
vis5-H
PC3 (6.8%)
PC2 (18.5%)
PC1 (37%)
C
L. iners unique KOs
L. cirspatus unique KOs
L. iners
L. crispatus
58
569
244
Common KOs


Citrate cycle (TCA cycle):ko00020
Glycine, serine and threonine metabolism:ko00260
Lysine biosynthesis:ko00300
Phenylalanine, tyrosine and tryptophan biosynthesis:ko00400
Lipopolysaccharide biosynthesis:ko00540
Butanoate metabolism:ko00650
Vitamin B6 metabolism:ko00750
Metabolic pathways:ko01100
Biosynthesis of secondary metabolites:ko01110
Biosynthesis of antibiotics:ko01130
Carbon metabolism:ko01200
2−Oxocarboxylic acid metabolism:ko01210
Biosynthesis of amino acids:ko01230
Flagellar assembly:ko02040
0.4
0.6
0.8
1.0
Ratio
Up
100
200
300
FDR
0.03
0.02
0.01


fucO
galT
NEU1
K06999
pyrG
E4.2.1.46
RP−L7
fucP
K07729
K18333
RP−S1
RP−L3
fucU
adhP
thiM
dinJ
RP−S11
rpoA
cas2
PTS−Cel−EIIC
cas3
rfbF
gpmB
secA
ABC.GGU.S
casB
glk
E3.2.1.24
RP−L13
casA
casD
ABC.GGU.P
dapA
RP−S2
nrdD
yidC
cas1
casC
licD
yhdJ
dasC
nagB
folA
infC
mdh
E1.17.4.1A
galK
fabG
grtT
RP−L24
pezA
E3.3.1.1
mgs
pilD
rfbC
jag
−log10FDR
log2FC
Regulation
Down
Up
log2CPM
4
6
8
10
12