

Bacterial Vaginosis & microbiome: Predicting clinical outcomes and uncovering new relations with KMINE text mining

Whitepaper written by: Nils Hijlkema

For more info, contact: nils.hijlkema@tenwise.nl or +31 611718915.

Introduction

According to clinicaltrials.gov currently over 2,000 clinical trials are being run in which the detection of changes in the microbiome is a primary or secondary objective. These studies aim to connect bacterial species to the presence or absence of a disease. Whereas indeed clinical trials need to be done to discover new connections between micro-organisms and diseases, a large part of these connections may already be obtained from the available literature.

This whitepaper describes the use of the [KMAP platform](#) – TenWise’s proprietary knowledge network that contains over 200 million biological relationships - and its KMINE products, to predict the clinical relevance to bacterial vaginosis (BV) for a large number of micro-organisms. The vaginal microbiome in relation to BV is an upcoming research topic underscored by a few hundred papers in the last 2 years.

As a benchmark we use the paper of Ceccarani *et al.* (2019) ‘**Diversity of vaginal microbiome and metabolome during genital infections**’, that describes a trial in which microbiome profiles of women with BV, were compared with a healthy control group.

Clinical outcomes

The clinical outcomes, derived from the paper of Ceccarani *et al.* are shown in the table below in which the microbiome shifts from healthy (HC%) to BV patients (BV%) are described for those genera with a relative abundance of more than 1%.

Table 1: Clinical outcomes with shifts in relevant abundance of most important genera of the vaginal microbiome in healthy and diseases states.

| | Group | Relative abundance | | | |
|---|-------------------------|--------------------|--------|---------|--------|
| | | HC (%) | CT (%) | VVC (%) | BV (%) |
| ↓ | <i>Lactobacillus</i> | 79.16 | 67.45 | 56.69 | 18.80 |
| ↑ | <i>Gardnerella</i> | 2.72 | 3.65 | 7.68 | 11.44 |
| ↑ | <i>Prevotella</i> | 1.16 | 1.60 | 3.76 | 9.15 |
| ↑ | <i>Megasphaera</i> | 0.06 | 2.97 | 1.04 | 8.64 |
| ↑ | <i>Roseburia</i> | 1.09 | 4.42 | 3.51 | 3.51 |
| ↑ | <i>Sneathia</i> | 0.18 | 0.41 | 0.53 | 7.76 |
| = | <i>Shuttleworthia</i> | 0.03 | 0.54 | 0.31 | 7.48 |
| ↑ | <i>Faecalibacterium</i> | 1.49 | 2.31 | 2.14 | 2.09 |
| ↑ | <i>Atopobium</i> | 0.17 | 1.00 | 1.94 | 4.92 |

The table shows that in the BV cases, several genera rise in levels at the expense of the 80% relative abundance of *Lactobacillus* that is consequently lowered to about 20%. Based on these observed shifts, the paper concludes: “On the other hand, microbiota of BV-affected women was characterized by a major shift in the relative abundance of many groups, such as *Lactobacillus* (decreased), *Gardnerella*, *Prevotella*, *Megasphaera*, *Roseburia*, *Sneathia*, and *Atopobium* (all increased). Increase of *Mobiluncus* was also observed, even if at a low relative abundance (0.49% and 0.03% in BV and HC, respectively, $p=0.002$), confirming evidences from Parolin et al.”

Text mining outcomes

To compare the clinical outcome to the results obtained by text mining, we retrieved all organisms associated with Bacterial Vaginosis from the KMAP database. In the table below, the main clinically identified genera are matched to the species retrieved from the KMAP database.

Table 2: Table depicting the most important genera measured in BV samples in the clinical trial by Ceccarani *et al.* (2019) and the text mining results using KMINE by TenWise (2020). The hyperlinks in the last column are clickable and refer to *tenwiseapps* where Bacterial Vaginosis is linked to the specific species.

| Microbiota analysis | | | KMINE text mining results | | | |
|---------------------|-----------------------|------------------------------|---|-------------|---------|----------------------|
| Up / Down | Genus | Relative abundance in BV (%) | Species | # relations | eScore | link |
| ↓ | Lactobacillus | 18.80 | <i>Lactobacillus crispatus</i> | 98 | 139160 | link |
| | | | <i>Lactobacillus iners</i> | 64 | 172608 | link |
| | | | <i>Lactobacillus jensenii</i> | 36 | 55332 | link |
| ↑ | Gardnerella | 11.44 | <i>Gardnerella vaginalis</i> | 557 | 1400855 | link |
| ↑ | Prevotella | 9.15 | <i>Prevotella bivia</i> | 60 | 102240 | link |
| ↑ | Megasphaera | 8.64 | <i>Megasphaera elsdenii</i> | 2 | | - |
| ↑ | Roseburia | 3.51 | not reported | 0 | | - |
| ↑ | Sneathia | 7.76 | <i>Sneathia amnii</i> <i>sneathia</i> <i>sanguinegens</i> | 15 | | - |
| = | Shuttleworthia | 7.48 | not reported | 0 | | - |
| ↑ | Faecalibacterium | 2.09 | not reported | 0 | | - |
| | not reported | | <i>Chlamydia trachomatis</i> | 384 | 73344 | link |
| ↑ | below threshold of 1% | | <i>Mycoplasma hominis</i> | 155 | 90520 | link |
| ↑ | Atopobium | 4.92 | <i>Atopobium vaginae</i> | 130 | 652600 | link |
| ↑ | below threshold of 1% | | <i>Mobiluncus curtisii</i> | 27 | 90396 | link |
| | | | <i>Mobiluncus mulieris</i> | 16 | 50016 | link |

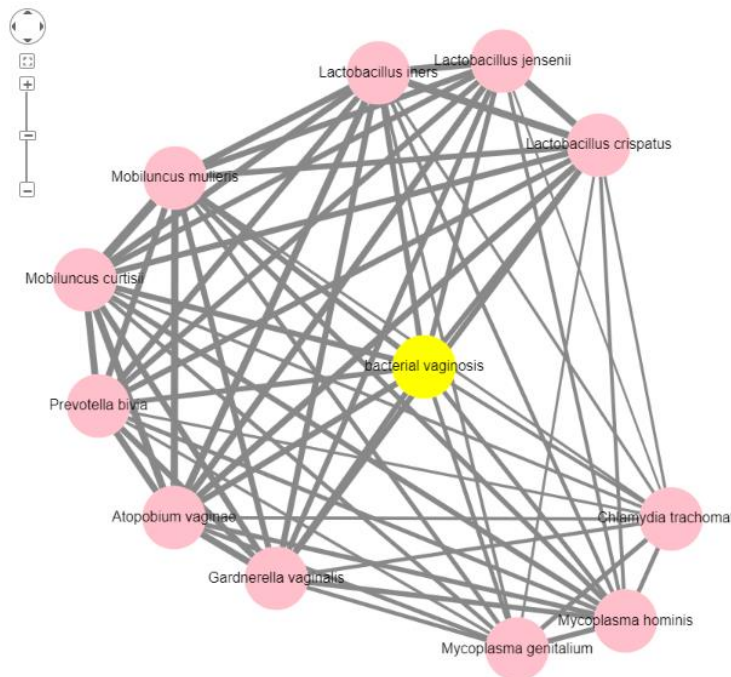


Figure 1: Literature network of the most important species related to Bacterial Vaginosis; The node color indicates the type of biological concepts; the width of the lines indicates the importance of the relations.

The merged table and network show three pockets of novel insights:

1. Overlapping conclusions: most important genera associated to BV

Our KMINE text mining outcomes confirm Ceccarani's conclusion that the genera *Gardnerella*, *Prevotella*, *Atopobium* and - at a low relative abundance - *Mobiluncus* are the main genera that have increased prevalence at the cost of the *Lactobacillus iners* and *crispatus*. This is visualized in figure 1.

2. Novel insights from the clinical study: new genera found that impact BV

The increase of *Roseburia*, *Megaspaera*, *Shuttleworthia* and *Sneathia* on BV are new insights coming from this clinical trial as there is no such descriptive relation found in the available literature.

3. Novel insights from the text mining case: From genera to species

KMINE suggest the most likely species for each genus. KMINE analysis identifies apart from the overly abundant *Lactobacillus iners* and *crispatus*, the species that have important relations with BV: *Gardnerella vaginalis*, *Atopobium vaginae*, *Prevotella bivia* and *Mobiluncus curtisii*. This is fully in line with the SPINGO based species-level analysis performed as part of the clinical study. Of these, *Gardnerella vaginalis* is one of the important pathobionts in the occurrence of BV as it is by far most described in relation to BV in literature.

Both *Chlamydia trachomatis* and *Mycoplasma hominis* and *genitalium* are associated with BV by looking at the relative high number of relations in KMINE. Ceccarani, however, states that 'among other bacteria, *Mycoplasma* was detected at very low relative abundance (0.07%)...' while 'Chlamydia, on the other hand, was detected only in CT subjects, as expected, although at a very low

relative abundance (0.02%) (data not shown).’ Surprisingly, the impact of *Chlamydia trachomatis* in BV in this clinical trial is described as low in comparison to what is expected from available literature derived from KMINE.

KMINE sentence analysis offers another level of insight

Finally, KMINE is able to look at both abstract level and sentence level in which it uncovers not only relations, but also the possible mode of actions of the found species in Bacterial Vaginosis. Examples shown below, reveal for example:

- *Gardnerella vaginalis* produces pore forming CDS (e.g. vagolysin) that affect membrane integrity.
- *Gardnerella vaginalis* forms polymicrobial biofilms which is a hallmark for BV biofilms. These biofilms also contain *Atopobium vaginae*, both of which are considered early colonizers in BV.
- *Atopobium vaginae* and *Prevotella bivia* are described in relation to increased levels of proinflammatory IL1 and IL12 cytokines in women’s suffering from BV.

Table 3: Suggestions for Mode of Action via KMINE fact analysis. The number of sentences describing relations are shown together with some typical examples that describe the relations in more detail.

| Genus from paper | Species from KMAP | # KMAP facts | Examples |
|--------------------|-------------------|--------------|--|
| <i>Gardnerella</i> | <i>vaginalis</i> | 77 | <ul style="list-style-type: none"> - <i>Gardnerella vaginalis</i>, the bacterial species most frequently isolated from women with bacterial vaginosis (BV), produces a cholesterol-dependent cytolysin (CDC), vaginolysin (VLY). - Biofilm-producing <i>Gardnerella vaginalis</i> are likely to play an important role in initiating the structured polymicrobial biofilm that is a hallmark of bacterial vaginosis. - FISH enabled us to show the presence of a polymicrobial biofilm in bacterial vaginosis, in which <i>Atopobium vaginae</i> is part of a <i>Gardnerella vaginalis</i>-dominated biofilm. |
| <i>Atopobium</i> | <i>vaginae</i> | 53 | <ul style="list-style-type: none"> - <i>Atopobium vaginae</i>, a fastidious, anaerobic, Gram-positive cocci-shaped bacterium that generates large quantities of lactic acid, is associated with bacterial vaginosis (BV). - New molecular techniques have increased our understanding of the numerous bacteria associated with bacterial vaginosis, and a biofilm containing mostly <i>Gardnerella</i> and <i>Atopobium vaginae</i>, which can persist after treatment has been described. |
| <i>Prevotella</i> | <i>bivia</i> | 16 | <ul style="list-style-type: none"> - <i>Prevotella bivia</i>, prevalent bacteria in the disturbed microbiota of bacterial vaginosis, induced strong upregulation of seven selected PIC-associated genes. - Women (N = 40) with incident bacterial vaginosis (Nugent 7-10) had significantly lower concentrations of lactobacilli and higher concentrations of <i>Gardnerella vaginalis</i>, <i>Atopobium vaginae</i>, and <i>Prevotella bivia</i>, at the incident visit and when concentrations of proinflammatory cytokines (IL-1β, IL-12p70) were increased and IP-10 and elafin were decreased. |

Conclusions

- KMINE **predicts** with high accuracy the species that are clinically associated with Bacterial Vaginosis whereas current sequencing often only reveals the genus level.
- KMINE **highlights new insights** coming from a study or experiment. In this whitepaper the roles of *Roseburia* and *Shuttleworthia* in the onset of BV are such a clear new insight as this relation has not been described in other, available literature to date.
- KMINE's sentence level analysis shows the possible **mode of actions** of identified species in Bacterial Vaginosis. We revealed indications of mode of actions of main players found in this clinical study whereas for others such indications clearly need to be researched and described in more detail.
- KMINE **saves time** by leveraging available literature by generating tables and networks with relations ranked on hits or importance, reducing the time that is spent on searching from an estimated 8-12 hours to 1-2 hours.

About TenWise

TenWise B.V. is set-up by Wynand Alkema PhD and Nils Hijlkema MSc who have a combined experience of over 35 years in food and life sciences research. They have been active in drug discovery, microbiome and probiotic programs in both industrial and academic settings. Dr. Alkema is also Professor of Data Science at Hanze University of Applied Sciences in Groningen.

TenWise offers text mining solutions by providing access to its KMAP via its KMINE Literature Explorer, the KMINE REST-API solution and a fast service for generating KMINE Literature Reports.

The company is based in Leiden, the Netherlands and works for clients around the world.

www.tenwise.nl