STUDY PROTOCOL

Version 2 – 12 June 2026

The Isala programme: Genetic and functional characterisation of the female microbiome

Research study at the University of Antwerp, in a collaboration between the Faculty of Medicine and Health Sciences and the Faculty of Science, Department of Bioscience Engineering

Inclusion/exclusion criteria

Volunteers
Age: 18+
Sex: F/X

No further exclusion criteria will be defined in advance. However, other data (based on the questionnaire) will be taken into account in the analysis of the collected data.

Recruitment of participants

The entire study course, with recruitment at random times, spans a period of five years. The volunteers (n=10,000) will be recruited, among others, from the staff and students of UAntwerp and UZA via posters, *ad valvas* and e-mails. No students will be recruited who still have to take exams with the members involved in the study. Other channels will also be used to obtain a representative group covering all age categories, including (but not limited to) the EOS portal, 'Citizen Science-iedereen wetenschapper' (www.iedereenwetenschapper.be), relevant UAntwerp web pages, e-mails to university colleges associated with UAntwerp, other Flemish universities and university colleges, etc. Recruitment will also take place via the project-specific website (www.isala.be) and related social media channels (Isala Instagram and LinkedIn). The volunteers will receive information about the study and the consent form during (several) information evenings, via online communication or by post, where the *self-sampling* kit can also be distributed on site or sent by post.

Questionnaires

The volunteers will complete questionnaires assessing factors that influence the micro-organisms in and on the body. This is important to correctly interpret the results of the study. The study is designed so that personal data cannot be directly linked to questionnaire answers in order to ensure the privacy of the volunteers, more specifically by using pseudonymisation, whereby personal data are converted into encrypted data by means of a code. The key of the code linking to the personal data is password-protected and only available to the responsible physician, principal investigator and executing researchers of this study. All these data will be stored on the research group's server with the necessary data protection. The retention period is 20 years for the questionnaire data.

Sample collection

The samples will be collected by the volunteers themselves using a swab (vagina, skin or saliva) and a menstrual cup for vaginal fluid. The *self-sampling* kit contains detailed instructions, an information brochure and all materials the volunteers need for sample collection, storage and delivery to the lab of Prof. Lebeer. The retention period is 10 years for the collected biological samples.

SoHO legislation

All procedures involving samples of human origin will comply with the EU regulation on "Substances of Human Origin (SoHO), Regulation (EU) 2024/1938", which sets standards for quality, safety, traceability and transparency in the use of human biological material. The collection, processing, storage and possible secondary use of material derived from the microbiome will take place in accordance with this regulation and its national implementing provisions. These procedures are complemented by full compliance with the GDPR (Regulation (EU) 2016/679) for data protection and are approved by the relevant institutional ethics committees. Informed consent procedures explicitly cover the use, storage and possible future applications of samples, including the isolation of microbial strains and translational research. Where applicable, material transfer and data sharing will take place according to established EU and international guidelines to ensure responsible and ethical use.

Sample size calculation

Given the high inter-sample variability and insufficient knowledge of the expected differences in abundances between groups, it is not possible to perform a power calculation for this study. Studies in the gut have shown that different enterotypes exist for the human microbiota (1). Consequently, it is important to make the number of participants in this population sufficiently large to obtain a general picture of the microbiota in the different body fluids of the volunteers. In similar studies, the number of individuals studied ranged from 30 to 564 women (2,3). Based on this literature, and because our study is rather exploratory, it was decided that a total of at least 1,000 volunteers will be needed to study common high-abundance members of the microbiota, and 10,000 volunteers to also characterise common low-abundance (but potentially highly relevant) members of the microbiota at a given time point. A subset of these, at least 100 women, will be included in the subgroup to enable longitudinal sample collection at different time points spread over two consecutive menstrual cycles. Vaginal fluid will be collected using a menstrual cup.

Analysis of the samples

The studies and analyses will be carried out under the direction of the University of Antwerp (UAntwerp), within the Department of Bioscience Engineering in the Laboratory of Applied Microbiology and Biotechnology led by Prof. Sarah Lebeer, in collaboration with Prof. Veronique Verhoeven (Department of Family Medicine and Population Health, FAMPOP) as the responsible physician. In addition, where relevant, pseudonymised samples and personal data will be transferred to and processed by academic or industrial collaboration partners or service providers, both within and outside Belgium, to enable high-tech sample and data processing. Data security is guaranteed by Belgian legislation.

Part 1: Microbiome analysis

Microbial DNA and RNA will be isolated from the swabs and cups from different body sites and from fluids using a commercial extraction kit such as the Qiagen PowerFecal DNA kit and the Qiagen RNeasy kits (cf. Human Microbiome Project). The isolated DNA and RNA (converted to cDNA) will then be analysed by 16S *rRNA* gene sequencing, ITS sequencing, shotgun sequencing and (meta)transcriptomics. With 16S *rRNA* gene and ITS sequencing, the composition of microbial communities can be determined in terms of bacteria and fungi, respectively (4,5). Shotgun sequencing and (meta)transcriptomics make it possible to identify the members present in the microbial communities at a higher resolution and to study the functional profiles of these microbial communities. Combined, these techniques will provide valuable insights into better understanding the role of microbial communities in the context of health. The 16S *rRNA* gene consists of 9 different variable regions, which can be used for microbial characterisation and which are flanked by conserved regions. The latter are used as targets for 'polymerase chain reaction' (PCR) primers with a universal specificity for almost all bacteria. Use will be made of the Illumina sequencing platform at the Center of Medical Genetics or other sequencing platforms available at the applicable (international) collaboration partner or (international) service provider (e.g. Eurofins Genomics, Prebiomics). Sequencing data will be stored on the research group's server and are only accessible to staff officially involved in the project. In the run-up to publication of these data, the sequences may also be made publicly available in public databases (e.g. ENA).

Milestone: Identification of differences in the (functional) microbiome composition of the different body sites and fluids of adults

Part 2: Quantitative analysis of the concentration of specific micro-organisms

Using "quantitative real-time polymerase chain reaction" (qPCR) with genus-, species- or strain-specific primers, the concentration of certain micro-organisms will be estimated. In qPCR, the amount of a specific piece of DNA is measured in 'real time'. This allows the initial concentration of this specific DNA to be determined. The results thus give a picture of the concentration of different micro-organisms in a sample. The "Minimum Information for Publication of Quantitative Real-Time PCR Experiments" (MIQE) guidelines will be used for processing these data. All these data will be stored on the research group's server with the necessary data protection.

Milestone: Absolute abundance of micro-organisms in different body sites and fluids

Part 3: Culturing of lactic acid bacteria and other micro-organisms from swabs and cultivation of specific micro-organisms in vaginal fluid

Current probiotic strains, mostly lactic acid bacteria, are typically isolated from the gastrointestinal niche of our body. For probiotic applications in the vagina, however, it would be interesting to isolate (lactic acid) bacteria from the vagina and the skin. The expected advantage is that niche-specific strains are better adapted to the conditions in the vagina and certain skin sites, and better able to withstand the pathogenic micro-organisms also present there. Lactic acid bacteria are interesting because they have GRAS status ('Generally Recognized As Safe') and numerous health effects have been extensively described, but also because recent studies have shown that lower amounts of lactic acid bacteria are associated with more vaginal infections. Consequently, we will

plate the volunteers' vaginal swabs and cups on general growth media and selective media for lactic acid bacteria, e.g. MRS (de Man Rogosa and Sharpe), LAMVAB medium and blood agar. For this, the collected body fluid will be dissolved in PBS (Phosphate Buffered Saline) and, after thorough mixing, a dilution series will be plated. Several colonies will be picked and further identification at species level will be performed by colony PCR with universal primers for the *16S rRNA* gene (e.g. 8F-1525R) or ITS gene, followed by Sanger sequencing and alignment against a reference genome (e.g. BLAST and EZBioCloud). The aim is to analyse at least 500 isolates from different "healthy" individuals to obtain a diverse population. In addition, we also want to isolate some potentially pathogenic strains, such as GBS (Group B Streptococcus) strains, to better understand their interaction with lactic acid bacteria and their role in vaginal health and disease. This is of interest because GBS colonisation is associated with adverse pregnancy outcomes. In the same way as vaginal swabs, we will also culture and identify micro-organisms from other intimate areas, such as the skin around the mouth, groin and breast. This time, however, the swabs will also be plated on media typical for skin micro-organisms (e.g. blood agar or brain heart infusion agar) in addition to general growth media. Culturing micro-organisms from intimate areas helps us gain new insights into potential probiotics for these body sites. It also gives us the opportunity to investigate how micro-organisms vary between the different intimate areas.

Vaginal fluid collected with the cup will also be used for small-scale cultivation of specific micro-organisms. Depending on the amount of vaginal fluid, samples from different participants may or may not be pooled and mixed with microbial media (e.g. de Man, Rogosa and Sharpe, MRS). After sterilisation, the vaginal fluid will be inoculated with cultures of specific micro-organisms (such as the vaginal isolates *Limosilactibacillus reuteri* AMBV339 and *Lacticaseibacillus rhamnosus* GR-1). The growth of the cultures will be monitored over time, and changes in gene expression of the micro-organisms will be monitored using qPCR and RNA sequencing techniques (5). For microbial RNA extraction, commercially available kits such as the RNeasy mini kit (Qiagen) will be used. Vaginal fluid without added micro-organisms will be included as a control. The samples will then be processed in-house or sent out for ribosomal RNA depletion, library preparation and sequencing. By using vaginal fluid instead of *simulated vaginal fluid*, which is often used in the lab to mimic human vaginal fluid, more insight can be gained into the different adaptation factors and gene expression profiles of micro-organisms in the vaginal niche.

The isolated micro-organisms will be stored at -80°C for possible later use and further research. Whole-genome sequencing of the microbial DNA will also be performed, in-house or outsourced to an (international) collaboration partner or (international) service provider (e.g. BGI Group, BaseClear). The resulting DNA sequences are filtered and processed using internal and publicly available pipelines. In the run-up to publication of these data, the sequences may also be made publicly available in public databases (e.g. ENA). The sequences and associated data on these micro-organisms (species, growth medium, etc.) are stored on the server and in a secure online inventory file of the research group, respectively, with the necessary data protection.

Milestone: Culturing of at least 500 lactic acid bacteria and other micro-organisms from swabs and cups. Growing specific micro-organisms in (sterilised) vaginal fluid to gain insights into adaptation factors and gene expression profiles in the vaginal niche

Part 4: Identification and quantification of metabolites from body fluids collected with swabs or cups

In addition to the identification and quantification of specific micro-organisms from swabs and cups, it is also interesting to look at which metabolites are present (e.g. metabolites produced by the micro-organisms present, by the host or by exogenous factors) and in what concentration. This can be determined using metabolomics. Specifically, gas and/or liquid chromatography and (tandem) mass spectrometry will be applied, in-house or outsourced to an (international) collaboration partner or (international) service provider, to body fluids (whether or not supernatant), which will be further prepared by extraction with, for example, a cold organic solvent after addition of an internal standard mixture. Suitable software (e.g. MassHunter Quantitative/Qualitative Analysis Software) will then be used for identification and quantification of metabolites, followed by peak alignment, enabling metabolite annotation and statistical analyses. These data will be stored on the research group's server with the necessary data protection.

In addition, vaginal acidity will be evaluated using pH test strips (e.g. Macherey-Nagel, pH-Fix 3.6-6.1) to obtain an indication of vaginal lactic acid levels.

Milestone: Identification and quantification of metabolites from swabs or cups

Part 5: Identification and quantification of immune factors in body fluids collected with swabs and cups

Immune factors in body fluids will be identified and quantified. Human pro- and anti-inflammatory markers (such as TNF-α and IL-10, respectively) and other biomarkers involved in inflammation (such as MMP, INF, integrins, apoptosis biomarkers, etc.) will be identified and quantified in the volunteers' samples using commercial immunoassays (e.g. ELISA, MSD, etc.) or using a (multiplex) analysis platform at an in-house or external (international) collaboration partner or (international) service provider (e.g. Luminex or Olink). For this purpose, the vaginal swab or cup will be rinsed in PBS (Phosphate Buffered Saline) with protease inhibitor. After thorough mixing, total protein is determined using a bicinchoninic acid (BCA) protein assay, and the sample is further processed depending on the analytical technique used. The human immune response can also be quantified in terms of biomarker expression levels by qPCR or RNA sequencing. For this purpose, the total RNA content will be extracted from the collected samples using commercially available kits. Specific primers (purchased, designed in-house or donated by partners) are used to quantify the expression levels of target genes. Human immune cells (such as macrophages, eosinophils, neutrophils and lymphocytes) in the collected samples will be differentially counted under the microscope after Romanowsky staining on slides. Microbial cell-surface proteins involved in the interaction with the human immune system (such as MR1) will be identified by flow cytometry. The collected data will be stored on the research group's server with the necessary data protection.

Milestone: Identification of immune factors in different body sites and fluids of the volunteers

Part 6: Statistical analysis of questionnaire data and sample results

The volunteers will complete specific questionnaires on their lifestyle, use of medication, etc. **Questionnaire data** will be processed statistically and correlated with the results of the sample analyses. All these data will be stored on the research group's server with the necessary data protection. More specifically, by combining data on micro-organisms, metabolites and immune factors with questionnaire answers, factors associated with differences in the microbiome and metabolome of different body sites will be investigated.

Analysis of microbiome data: Differential abundance of individual taxa will be tested. Since differential abundance testing tools often yield divergent results, we will adopt an approach in which multiple differential abundance testing tools are used, and a taxon/condition is only considered significant if at least three tools agree (5). Three tools are able to perform paired analyses by including random effects per volunteer. Secondly, for the microbiome, we will investigate modules of co-occurring taxa known to be associated with lifestyle factors (e.g. the *Lactobacillus crispatus* module). Such modules have been replicated in several large studies of the vaginal microbiome. This multi-layered approach is useful for detecting effects that may not be visible at the individual level. Furthermore, various methods (e.g. SparCC) will be applied to determine correlations between the abundance of taxa in the different samples. Microbiome data will be statistically analysed at different levels. For example, associations between microbial community composition (beta diversity) and questionnaire data will be investigated with an Adonis test, as implemented in the '*vegan* package' in R. Data processing and visualisation will mainly take place in Python and R, using the '*tidyverse* package' and the in-house developed '*tidyamplicons*/*tidytacos* packages'. The analyses will be adapted and extended according to the research questions and results.

Analysis of metabolite data: Metabolic differences in volunteers' samples will be assessed using an in-house developed 'multiverse' analysis in which different 'pre-processing' strategies are implemented (e.g. different data transformations) to verify the robustness and reproducibility of results. The analyses will be adapted and extended depending on the research questions and results.

Analysis of immune data: Concentrations of different immune factors will be normalised to the total protein content in the corresponding samples. 'Normalized Protein Expression' (NPX) values of different immune factors, obtained from a multiplex analysis such as via the Olink® platform, will be analysed per marker and further interpreted statistically when comparing different samples. The analyses will be adapted and extended depending on the research questions and results.

Milestone: Insights into the possible correlation of health status, lifestyle, medication use, diet or other requested data with microbiome, metabolome and immune profiles

References

1. Arumugam M, Bokr P, Raes J, Pelletier E, Paslier D Le, Batto J michel, et al. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174–80.

2. Pino A, Bartolo E, Caggia C, Cianci A, Randazzo CL. Detection of vaginal lactobacilli as probiotic candidates. Sci Rep. 2019;9(1):1–10.

3. Borgdorff H, van der Veer C, van Houdt R, Alberts CJ, de Vries HJ, Bruisten SM, et al. The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. Fredricks DN, editor. PLoS One. 2017 Jul 11;12(7):e0181135.

4. Condori S, Ahannach S, Vander Donck L, Oerlemans E, Dillen J, Dricot C, et al. Recent Insights into the Vaginal Microbiota. Microb Health Dis. 2022;4(3).

5. Lebeer S, Ahannach S, Gehrmann T, Wittouck S, Eilers T, Oerlemans E, et al. A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. Nat Microbiol. 2023 Oct 26;8(11):2183–95.