STUDY PROTOCOL

Version 2 – February 6, 2025

The Maggie project: Exploring the origin and heredity of the vaginal microbiome

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

Inclusion and exclusion criteria of the central participants

- **Inclusion criteria:**
  - Adults aged 18 and over at the time of enrollment;
  - Sex: Female;
  - Pre-menopausal;
  - Self-reported good health (including the absence of general infection);
  - Absence of reproductive disorders (e.g. endometriosis, polycystic ovary syndrome), vaginal infections and symptoms diagnosed by a medical professional;
  - Absence of gastrointestinal disorders, infections and symptoms diagnosed by a medical professional;
  - Living in Belgium;
  - Sufficient knowledge of the Dutch language;
  - Consent form signed;
  - Participating alongside her mother and at least one additional co-participant in the study;
- **Exclusion criteria:**
  - Current pregnancy or planned pregnancy at the beginning and during of the study;
  - Breastfeeding at the start and during the study;
  - Current diagnosis of cancer and/or immunosuppressive therapy in the 6 months before the study;
  - Clinically significant abnormalities of the reproductive organs or any other medical condition at the discretion of the principal investigator;
  - Use of oral/vaginal antibiotics/antifungals in the 2 months before and during the study;
  - Use of oral/vaginal pre- and/or probiotics in the 2 weeks before the study and during the study;
  - Vaginal showering during the study;
  - Participation in an intervention study.

Furthermore, selections will be made based on the pre-selection survey to achieve a balanced study population and networks with greater diversity in relationships between the central participants and

co-participants. This process is further detailed in the section "Recruitment of participants and cohort assembly".

Inclusion and exclusion criteria of the co-participants

- **Inclusion criteria:**
  - Closely related to or interacting with the central participant of the network at least during the last six months or more before the study, e.g. her mother, aunts, female cousins, sisters, daughters, housemates, partners or close friends;
  - Aged 18 or over at the time of enrollment; **OR** 10 or over if the co-participant's mother is also participating in the study;
  - Sex: Female; **OR** male aged 18 and over who is the partner or one of the partners (in case of polygamous relationships) of the central participant;
  - Self-reported good health (including the absence of general infection);
  - Living in Belgium;
  - Sufficient knowledge of the Dutch language;
  - Consent form signed;
- **Exclusion criteria:**
  - Use of oral/vaginal antibiotics/antifungals in the 2 months before the study;
  - Parallel participation in an intervention study;
  - Vaginal showering during the study;
  - Clinically significant abnormalities of the reproductive organs or any other medical condition that, in the opinion of the principal investigator, warrants exclusion from the study.

Recruitment of participants and cohort assembly

To recruit the central participants (n=100), an open call will be launched through the website of the citizen-science project Isala (https://isala.be/). This call will be actively shared across various channels including (but not limited to): the Instagram, LinkedIn and X accounts of Isala (@Isala_UAntwerp); posters, *ad valvas* and emails to the staff and students from the University of Antwerp; the EOS portal; *Citizen Science-iedereen wetenschapper* (http://www.iedereenwetenschapper.be/); websites of the LAMB research group (https://www.uantwerpen.be/en/research-groups/endemic/) or other UAntwerp research groups; emails to UAntwerp associated colleges, other Flemish universities and colleges; and new outlets. The call will also be disseminated among volunteers who have previously expressed their interest in participating in Isala daughter projects, as well as among participants from previous clinical studies conducted under the supervision of Prof. Lebeer. A project-specific web page created for the Maggie project on the Isala website (https://isala.be/studie/maggie) will serve as the source of information for potential volunteers and the public (mainly, objectives and methodology).

Participants interested in participating in Maggie as central participants (N= 1 000) will be requested to complete a pre-selection survey encompassing questions about general health and

related to the inclusion and exclusion criteria established. Based on this questionnaire, potential central participants who meet all inclusion criteria will be identified. These volunteers will receive an email containing an alphanumeric code and an invitation template, which they can then share with their contacts who are interested in joining the study with them. Potential co-participants older than 18 years old will gain access to their own pre-selection survey after filling out the code provided by their central participant. Participants aged 10 to 17 can only be enrolled after their mothers have joined the study.

A network will be considered assembled when the central participant integrates it with her mother and between one and fourteen additional contacts (with a maximum of fifteen co-participants). We aim to recruit 100 central participants and approximately 500 participants in total. If more than 100 women meet the criteria established for the central participants, those with the highest number of co-participants with whom they have different types of relationships will be selected. After each network is assembled, all members will receive the confirmation survey to verify if any significant aspect relevant to the study has changed since they filled out the pre-selection survey.

Study progress and sample collection

Maggie will include a cross-sectional study involving all participants from the networks, and a longitudinal study focusing only on the central participants.

Maggie Phase I: Cross-sectional study

Participants will receive a self-sampling kit containing all necessary materials and instructions for self-sampling, storage and delivery to Prof. Lebeer's lab. These kits will be picked up from the Laboratory of Applied Microbiology and Biotechnology at Campus Groenenborger (Groenenborgerlaan 171, 2020). If pickup is not possible, a study employee can optionally deliver the kits to the participant's home or workplace.

Each participant will self-collect two vaginal/penis swabs: one for microbial DNA characterization (and for the human DNA characterization, if the participant agrees), and one for microbial isolation and characterization of the metabolic environment. In addition, female participants will self-take an extra swab to measure the vaginal pH using a test strip (such as the Macherey-Nagel pH test strip or similar). For female participants under 18 years old, non-invasive first-void urine collectors will be included as an alternative to the vaginal swabs (Brown *et al.*, 2021; Datcu *et al.*, 2014). If these participants are comfortable, they can also provide both types of samples. To investigate strain sharing along the gut-vagina axis, central participants will be asked to provide a stool sample for bacterial and microbial DNA isolation.

Based on previous findings indicating the transmission of lactobacilli strains from mothers to reproductive-age daughters (Erreygers, Pinedo-Bardales *et al*., in preparation), female reproductive-age participants will be asked to self-take samples of approximately 14 days after the beginning of their menstruation (ovulation day). This timeframe corresponds with the expected high prevalence of lactobacilli, which is linked to elevated estrogen levels (Lebeer *et al.*, 2023). Samples for participants of the same network should be taken in a time frame of a maximum of

four weeks. The result of the vaginal pH measurement will be submitted via Qualtrics (GDPR-conform platform). Until transport, samples should be stored in the participants' fridge and later delivered to the lab at ambient temperature in the provided bags. This must be done within 24 hours of collection at Campus Groenenborger (Groenenborgerlaan 171, 2020). If delivery is not possible, a study employee can optionally collect them from the participant's home or workplace.

Participants will be also asked to complete a Qualtrics questionnaire based on the Isala survey and will cover factors known to impact the microbiome composition (e.g., diet, sexual activity, number of childbirths) (Lebeer *et al.*, 2023). Specific questions will be included to investigate potential transmission pathways within each network, such as lifestyle habits sharing between subjects. In addition, comparative genomics and advanced bioinformatics methods will be used to explore transmission at the strain level, both via different body sites in the individual and between participants from the same and distinct networks.

Maggie Phase II: Longitudinal study

These participants will self-collect weekly swabs at three time points each year over a period of five years. At each time point, they will take three swabs: one for microbial DNA characterization (and human DNA characterization, if consented), one for microbial isolation and analysis of the metabolic environment, and one for measuring pH. Additionally, participants will complete a survey and use an ovulation test to determine whether they are ovulating.

Estrogen stimulates glycogen production, providing a substrate for lactobacilli that may affect vaginal microbiome composition (Miller *et al.,* 2016). To mitigate this confounding factor, sampling will be uniformly conducted according to the menstrual phase of each participant. The first swab will be self-collected +/- 2 days after the end of the menstrual phase (or the stop week for participants using contraceptives) to avoid collecting residual blood. The following two samples will be taken approximately every 7 days. Subsequent two samples will be collected approximately every 7 days. One year later, the next three samples will be collected following the same criteria, and this process will continue until the five-year follow-up is complete. Samples will be stored in the participants' freezers until their yearly sampling is complete. Samples will be delivered to the lab at Campus Groenenborger (Groenenborgerlaan 171, 2020) at ambient temperature in the provided bags. If this is not possible, a study employee can optionally collect them from the participant's home or workplace. These longitudinal samples will be used to assess the persistence of the vaginal microbiome at the strain level and to identify potential factors that drive interactions at the intra- and inter-species levels in the vaginal microbiome, as well as for the characterization of the metabolic environment and its dynamics.

Sample size and power calculation

To determine the number of participants, we reviewed studies with similar objectives as those proposed in Maggie. To the best of our knowledge, only a few studies have been published evaluating the transmission and persistence of the vaginal microbiome at the strain-level. Therefore, we also reviewed studies where these aspects were analyzed at the species level. Since

our number of participants depends on the number of central participants, we decided first to calculate the sample size of this population, which will be the target of the longitudinal study. Only two previous longitudinal studies evaluated the persistence of the vaginal microbiome at the strain-level (M. France *et al.*, 2022) or using metagenomic reads (M. France, Fu, *et al.*, 2022). These studies included 2 and 39 healthy volunteers respectively. We decided to set-up of number of central participants at 100, considering a possible dropout rate of 10% (based on other longitudinal studies under the supervision of Prof. Lebeer), that will allow us to still collect longitudinal data for approximately 90 participants.

The transmission of the vaginal microbiome was analyzed in 13-40 pairs of mothers-daughters (Bassis *et al.*, 2023; M. France, Brown, *et al.*, 2022) and in 23 heterosexual couples (Mändar *et al.*, 2015) at the species level. At strain-level, this aspect was only analyzed in 22 pairs of mothers-daughters (M. France, Brown, *et al.*, 2022). If each network (n=100) includes at least 2 co-participants, our cohort will include by 200 pairs of central and co-participants to analyze transmission at strain-level between them, without considering additional pathways of transmission that can be analyzed within and between the networks. Given the diversity and large number of members in certain families, we decided to set up the maximum number of co-participants in each network to 15, with a total of 500 participants.

Power analyses are challenging for microbiome studies. This is mainly because the magnitude and variance of effects on the microbiome are unknown and highly dependent on the individual and different for each species. In the cross-sectional study, we will include approximately 500 participants and transmission will be analyzed within and between each network. To consider variation in each species and to analyze at strain-level, metagenome-assembled-genomes (MAGs) will be generated from the metagenome data. Representative reference genomes for each species will also be included in each analysis. This will allow us to perform a strain-level analysis with a high-resolution. In contrast to the Isala longitudinal phase, which included 200 participants, our longitudinal study will focus on only 100 core participants followed over five years. Despite the smaller sample size, we will generate extensive longitudinal metagenome and metabolome data which will be integrated. In addition, we will perform paired analyses on vaginal and stool samples that will allow us to consider intra-individual variation in gut/vagina axis to increase the power of the study.

Analysis of the samples

Lab Applied Microbiology and Biotechnology Prof. Sarah Lebeer, Faculty of Science, Department of Bioscience Engineering, University of Antwerp

Part 1: Microbiome analysis

Microbial DNA will be characterized from stool and vaginal/penile/urine samples (the latter depending on the self-sampling device used by the participant), collected during the cross-sectional and longitudinal studies. The samples collected for this purpose will be stored at -20°C until the total DNA of the samples is isolated using commercially available kits such as the HostZERO

Microbial DNA, Qiagen PowerSoil PRO DNA kit or similar. If needed, an additional enrichment step will be applied to yield a higher percentage of bacterial DNA. The isolated DNA will be stored at -20°C after which it can be used for further analyses. To quantify the bacterial DNA, qPCR including primers for human and bacterial DNA will be performed as described previously by Ahannach *et al.* (2021). If during lab processing it is observed that the samples for microbial DNA isolation do not have enough quality (in terms of the percentage of bacterial DNA) to perform the proposed analyses, the participant will be contacted to ask if they can donate an additional sample. Library preparation and metagenome sequencing will be outsourced to a subcontractor (e.g., Prebiomics, BaseClear). Raw metagenomic reads will be quality-filtered and human reads will be removed prior to subsequent analysis. Sequencing data will be stored on the research group's server and will be only accessible to staff officially involved in the project. Metagenomic reads without human reads will be made publicly available in public databases (e.g., ENA) during the publication process of the manuscripts resulting of this project. The composition of the microbiomes will be analyzed at the species level using in-house developed protocols. Relative abundances of each species will be examined, along with measures of alpha and beta diversity. MAGs will be reconstructed using in-house and publicly available pipelines. All this data will be stored on the research group's server with the necessary data protection.

Milestones: (1) Characterization of metagenomes from the different body sites/fluids of 500 participants; **(2)** Characterization of longitudinal metagenome data from a cohort of 100 healthy women.

Part 2: Cultivation of lactic acid bacteria and other beneficial microorganisms from stool and vaginal/urine samples

A swab will be collected from each stool (for central participants) or urine sample (for female participants younger than 18 years old). Swabs from the stool, vagina, penis and/or urine will be stored at -20°C in transport medium which contribute to maintain the viability of the bacteria until the swab is processed. Swabs will be plated on selective media for bacteria commonly identified in the vaginal niche, such as MRS (Man–Rogosa–Sharpe), LAMVAB medium, and blood agar. For this the swab will be dissolved in PBS (Phosphate Buffered Saline) and after thorough mixing a dilution series will be plated. Several colonies will be spotted and further identification at species level will be done by colony PCR with universal primers for the *16S rRNA* gene (e.g. 8F-1525R), followed by Sanger sequencing and alignment to a reference genome (e.g. using BLAST, EZBioCloud or an in-house pipeline). Whole genome sequencing of the microbial DNA will be outsourced to a subcontractor (e.g. BGI Group, BaseClear). Sequencing reads will be filtered and processed using in-house and publicly available pipelines. The isolated bacteria will then be stored at -80°C for possible later use. The details of these bacteria (species, growth medium, etc.) will be stored in an online protected inventory of the lab.

Milestones: Characterization of lactic acid bacteria and other beneficial microorganisms isolated from different body sites/fluids of 500 participants.

Part 3: Identification and quantification of metabolites from the vaginal swabs and urine samples

To investigate which metabolites are produced by the micro-organisms in the vagina and in what concentration, ("ultra-high performance") liquid chromatography and (high-resolution) mass spectrometry will be applied to the collected supernatants of solutions of the vaginal swabs for metabolomics from both phases of Maggie that are stored at -80°C. These will first be further prepared for this analysis by extraction with a cold organic solvent after addition of an internal standard mixture. Suitable software (for example the MassHunter Quantitative / Qualitative Analysis Software) will then be used for metabolite identification and quantification, followed by peak alignment, allowing metabolite annotation and statistical analyses. Subsequently, the data will be further analyzed and visualized using bioinformatic tools. In addition to these microbially produced metabolites, metabolites originating (possible) from food, environment, etc. will also be studied with corresponding identification and/or quantification. Furthermore, the acidity of the vagina will be evaluated using pH test strips (e.g. Macherey-Nagel, pH-Fix 3.6-6.1 or similar) to obtain an indication of vaginal lactic acid levels. In addition, hormones, neurotransmitters, and other metabolites will be measured from the urine samples using special kits (e.g., ZRT laboratory or similar).

Milestone: Identification and quantification of metabolites in the vaginal samples from both phases of Maggie.

Part 4: Evaluation of the origin and persistence of vaginal microbiome

To investigate the origin of the vaginal microbiome, various sources of transmission will be explored. As an initial analysis, microbiome profiles from vaginal and stool samples of central participants will be compared. This comparison will be complemented by whole-genome sequencing data from bacteria isolated from both sample types, allowing for the identification of shared strains between the gut and vaginal microbiomes within participants, using both in-house and publicly available pipelines.

In addition, microbiome profiles from vaginal, penile, and urine samples will be compared between central participants and their co-participants within each network. Statistical analyses will be conducted to determine the significance of any observed similarities. This analysis will then be extended to assess strain similarities between central and co-participants across the entire cohort, followed by broader comparisons across all participants. Persistence will be evaluated by applying the same methods to longitudinal samples collected from central participants. If needed, other strategies such as the use of strain specific primers and polymerase chain reaction will be used.

Functional and metabolic profiles will also be compared between samples with and without shared strains, using statistical methods to assess the significance of these similarities. Moreover, potential factors associated with the presence and persistence of shared strains in the vaginal microbiome will be examined by analyzing questionnaire metadata provided by participants. All data will be securely stored on the research group's server with appropriate data protection measures in place.

Milestones: (1) Evaluation of shared strains in the gut-vagina/penis axis at individual level and between individuals; **(2)** Evaluation of the persistence of vaginal microbiome at strain level in 100 healthy women followed over five years.

Part 5: Identification of genetic associations with the transmission of the vaginal microbiome

To better understand the genetic factors that may influence the establishment, transmission, and persistence of specific bacteria in the vaginal environment, a general genetic analysis will be performed. If the participant agrees, human reads will be extracted from the metagenomic data of both phases of Maggie and processed using both in-house and publicly available pipelines. During data analysis, a focused approach will be applied, concentrating on common genetic variations that may be associated with the colonization, transmission and persistence of particular strains in the vaginal microbiome. Known genes and mutations linked to serious diseases will not be examined.

Part 6: Characterization of the determinants of the vaginal microbiome and metabolome

Factors associated with changes in the vaginal microbiome, as well as the determinants of interactions within the microbial community, will be investigated by integrating metagenome data and questionnaire responses. The goal is to understand the factors influencing interactions between and within species in the vaginal microbiome, shedding light on the functions of each microbial member by analyzing their functional profiles. Additionally, by integrating longitudinal metabolome and metagenome data with metadata from surveys, multiple hypotheses will be tested. Specifically, the analysis will determine if the production of certain metabolites is influenced, for instance, by: (1) stable, host-related factors (e.g., age, smoking status); (2) dynamic, host-related factors (e.g., recent sexual intercourse, menstrual cycle phase); or (3) the microbiome community (e.g., community state type or eigentaxa abundances). All data generated in this step will be securely stored on the research group's server with appropriate data protection measures.

Milestone: Characterization of the determinants of the vaginal metabolome of 100 healthy women followed over five years, integrating longitudinal metagenome and metabolome data.

References

Ahannach, S., Delanghe, L., Spacova, I., Wittouck, S., Van Beeck, W., De Boeck, I., & Lebeer, S. (2021). Microbial enrichment and storage for metagenomics of vaginal, skin, and saliva samples. *iScience*, *24*(11), Article 11. https://doi.org/10.1016/j.isci.2021.103306

Bassis, C. M., Bullock, K. A., Sack, D. E., Saund, K., Pirani, A., Snitkin, E. S., Alaniz, V. I., Quint, E. H., Bell, J. D., & Young, V. B. (2023). Vaginal microbiota of adolescents and their mothers: A preliminary study of vertical transmission and persistence. *Frontiers in Microbiomes*, *2*. https://www.frontiersin.org/articles/10.3389/frmbi.2023.1129394

Brown, S. E., Robinson, C. K., Shardell, M. D., Holm, J. B., Ravel, J., Ghanem, K. G., & Brotman, R. M. (2021). Assessing the Concordance Between Urogenital and Vaginal Microbiota: Can Urine Specimens Be Used as a Proxy for Vaginal Samples? *Frontiers in Cellular and Infection Microbiology*, *11*, 671413. https://doi.org/10.3389/fcimb.2021.671413

Datcu, R., Gesink, D., Mulvad, G., Montgomery-Andersen, R., Rink, E., Koch, A., Ahrens, P., & Jensen, J. S. (2014). Bacterial Vaginosis Diagnosed by Analysis of First-Void-Urine Specimens. *Journal of Clinical Microbiology*, *52*(1), 218–225. https://doi.org/10.1128/JCM.02347-13

France, M., Ma, B., & Ravel, J. (2022). Persistence and In Vivo Evolution of Vaginal Bacterial Strains over a Multiyear Time Period. *mSystems*, *7*(6), e0089322. https://doi.org/10.1128/msystems.00893-22

France, M. T., Brown, S. E., Rompalo, A. M., Brotman, R. M., & Ravel, J. (2022). Identification of shared bacterial strains in the vaginal microbiota of related and unrelated reproductive-age mothers and daughters using genome-resolved metagenomics. *PLOS ONE*, *17*(10), Article 10. https://doi.org/10.1371/journal.pone.0275908

France, M. T., Fu, L., Rutt, L., Yang, H., Humphrys, M. S., Narina, S., Gajer, P. M., Ma, B., Forney, L. J., & Ravel, J. (2022). Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data. *Genome Biology*, *23*(1), Article 1. https://doi.org/10.1186/s13059-022-02635-9

Lebeer, S., Ahannach, S., Gehrmann, T., Wittouck, S., Eilers, T., Oerlemans, E., Condori, S., Dillen, J., Spacova, I., Vander Donck, L., Masquillier, C., Allonsius, C. N., Bron, P. A., Van Beeck, W., De Backer, C., Donders, G., & Verhoeven, V. (2023). A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. *Nature Microbiology*, *8*(11), Article 11. https://doi.org/10.1038/s41564-023-01500-0

Ma, B., France, M. T., Crabtree, J., Holm, J. B., Humphrys, M. S., Brotman, R. M., & Ravel, J. (2020). A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-14677-3

Mändar, R., Punab, M., Borovkova, N., Lapp, E., Kiiker, R., Korrovits, P., Metspalu, A., Krjutškov, K., Nõlvak, H., Preem, J.-K., Oopkaup, K., Salumets, A., & Truu, J. (2015). Complementary seminovaginal microbiome in couples. *Research in Microbiology*, *166*(5), 440–447. https://doi.org/10.1016/j.resmic.2015.03.009

Miller, E. A., Beasley, D. E., Dunn, R. R., & Archie, E. A. (2016). Lactobacilli Dominance and Vaginal pH: Why Is the Human Vaginal Microbiome Unique? *Frontiers in Microbiology*, *7*, 1936. https://doi.org/10.3389/fmicb.2016.01936

Zhang, Y., Thompson, K. N., Huttenhower, C., & Franzosa, E. A. (2021). Statistical approaches for differential expression analysis in metatranscriptomics. *Bioinformatics*, *37*(Supplement_1), i34–i41. https://doi.org/10.1093/bioinformatics/btab327