BMJ Open

Relationship between vaginal and gut microbiome and pregnancy outcomes in eastern Ethiopia: a protocol for a longitudinal maternal-infant cohort study (the EthiOMICS study)

Fitsum Weldegebreal ,[1,2] Desalegn Admassu Ayana,[1] Tara Wilfong,[3] Merga Dheresa,[4] Tesfaye Assebe Yadeta ,[4] Akewok Sime Negesa,[1] Yohannes Mulugeta Demmu ,[5] Tewodros Tesfa,[1] Tegbaru Nibrat Alemu,[5] Tadesse Gure Eticha ,[6] Abraham Geremew,[5] Kedir Teji Roba,[4] Alemseged Abdissa,[7] Nega Assefa ,[4] Abel Abera Negash,[7,8] Piet Cools,[2] Abera Kenay Tura [4,9,10]

To cite: Weldegebreal F, Ayana DA, Wilfong T, *et al*. Relationship between vaginal and gut microbiome and pregnancy outcomes in eastern Ethiopia: a protocol for a longitudinal maternal-infant cohort study (the EthiOMICS study). *BMJ Open* 2025;**15**:e092461. doi:10.1136/bmjopen-2024-092461

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-092461).

Received 14 August 2024
Accepted 04 December 2024

For numbered affiliations see end of article.

Correspondence to
Fitsum Weldegebreal;
fwmlab2000@gmail.com

ABSTRACT

Introduction Although evidence exists on the impact of microbiota on pregnancy outcomes in many high-resource settings, there is a lack of research in many low-resource settings like Ethiopia. This study aims to fill this gap by studying the gut and vaginal microbiota changes throughout pregnancy and assess how these changes relate to pregnancy outcomes among a cohort of pregnant women in eastern Ethiopia.

Methods and analysis Vaginal and stool samples will be collected using DNA/RNA Shield Collection kits three times starting at 12–22 weeks, 28–36 weeks and at birth (within 7 days). Postnatally, newborns' skin swabs (at birth) and rectal swabs will be obtained until 2 years of age. Moreover, breast milk samples at birth and 6 months and environmental samples (water, indoor air and soil) will be collected at enrolment, birth, 6, 12 and 24 months post partum. DNA will be extracted using Roche kits. Metagenomic sequencing will be performed to identify metataxonomic profiling and assess variations in microbial profiles, and α and β diversity of the microbiota. Information on socioeconomic, behavioural, household and biological factors will be collected at enrolment. The collected data will be coded, entered into EpiData 3.1 and analysed using Stata 17.

Ethics and dissemination The Institutional Health Research Ethics Review Committee (Ref No. IHRERC/033/2022) of Haramaya University, Ethiopia has approved this study ethically. Written informed consent regarding the study and sample storage for biobanking will be obtained from all participants. Results will be published in international peer-reviewed journals, and summaries will be provided to the study funders. Clinical study data will be submitted to Data Compass (https://datacompass.lshtm.ac.uk/), and molecular profiles of the microbiome and whole-genome sequences will be submitted to the European Nucleotide Archive (https://www. ebi.ac.uk/ena). Requests for data should be directed to daberaf@gmail.com. The decision to share data will be made by the study steering committee under the College of Health and Medical Sciences, Haramaya University, Ethiopia.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Enrolment of women in early pregnancy until 2 years will enable us to document the microbiome changes during pregnancy and post-pregnancy period.
⇒ Inclusion of maternal-infant dyad is essential to comprehensively characterise maternal and infant microbiome and its relationship and associated pregnancy outcomes.
⇒ The study is being conducted within the Health and Demographic Surveillance Sites, making it representative of other similar settings.
⇒ Since recruitment is in early pregnancy and at health facilities, women who are not aware of their pregnancy status or did not seek antenatal care might be missed.

INTRODUCTION

Despite many interventions, maternal and child health remains the unfinished agenda in many low-resource settings. Although the maternal mortality ratio was reduced to 267 per 100 000 in 2020, an estimated 10 000 maternal deaths occur every year in Ethiopia.[1 2] While the two-third reduction in child mortality was achieved as part of the Minimum Development Goals (MDG), post-MDG, neonatal mortality remained unchanged, making the journey towards the achievement of the Sustainable Development Goals uncertain.[3 4] Since most studies exploring the causes of the high rates of maternal and neonatal mortality and morbidity are often directed

towards socioeconomic, nutritional and health service-related factors, the role of the microbiota in maternal and neonatal remains poorly understood, especially in lower middle income countries (LMIC).[5] Hence, the need to study microbiota is emerging, given the importance of these factors on maternal and child health.[6 7]

Microbiota refers to the collection of microorganisms in each environment, including (parts of) the human body.[8] The newborn birth process plays a vital role in colonising the infant's gut and is highly influenced by the maternal microbiota. Moreover, the birth environment contributes to the newborns' microbiota colonisation,[9 10] and the diversity of the microbiota varies with the age of the newborn.[11] The vaginal microbiota forms a homeostatic and mutualistic relationship with the human host and plays an important role in health and disease. The most dominant vaginal bacterial species, *Lactobacillus* spp.,[12 13] for example, protects reproductive-age women from non-indigenous pathogens.[12–14] The composition of the microbiota depends on its location and is influenced by several factors,[15 16] such as age, vaginal pH, hormonal secretions, antibiotic use, body mass index, hormones and sexual activity, which mainly affect the vaginal microbiota.[16] Even though vaginal microbiota during pregnancy may vary due to many factors, the members of the *Lactobacillus* genus and five of its species are significantly more abundant, and several genera, such as *Prevotella*, *Atopobium*, *Acinetobacter* and *Sneathia*, are substantially less abundant during healthy pregnancy.[17 18]

Moreover, the gut microbiota also undergoes changes associated with different factors.[19] The gut microbiota in pregnant women changes significantly to adapt to the changing metabolism and immunity, which is important in maintaining maternal health and foetal development.[20] During gestation, the female body undergoes several changes, including hormonal, immunological and metabolic changes to support foetal growth and development.[21] In a healthy pregnancy, the bacterial load and profound alterations in the composition of gut microbiota increase.[22 23] Most gut microbes in a healthy pregnancy belong to one of the two phyla: *Bacteroidetes* and *Firmicutes*.[16 24 25] While the composition of gut microbiota in the first trimester of pregnancy is like that of healthy non-pregnant women, it changes from the first to the third trimester of pregnancy, including an increased abundance of members of the *Actinobacteria* and *Proteobacteria phyla*, as well as a reduction in individual richness (alpha diversity).[23]

Investigating the interface between maternal/newborn complications and microbiota colonisation would minimise adverse pregnancy outcomes, child mortality, pregnancy-related complications and death.[23] Maternal vaginal microbiota and its relationship with the health of infants have been shown in several studies.[9 10] Alterations of the human microbiota are a known characteristic of various inflammatory disease states and have been linked to spontaneous preterm birth and other adverse pregnancy outcomes, such as preterm birth.[26–28] Recent studies have examined the role of vaginal microbiota with possible associations with preterm birth.[27 29] The healthy vaginal microbiota plays a role in preventing bacterial vaginosis, sexually transmitted infections, urinary tract infections and HIV. Protection is attributed to lactic acid-producing bacteria, mainly *Lactobacillus* spp.[27]

The gut microbiota of infants is influenced by multiple perinatal and early-life environmental factors, such as pregnancy, gestational age at birth, maternal health, mode of delivery, birth weight, and breast milk feeding duration,[30] exposure to antibiotics before, during and after delivery, and exposure to pets.[31] The maternal microbiota is an essential factor in maternal and child health outcomes, and a better understanding of the role of the microbiota and sources of pathogens will guide interventions to reduce adverse pregnancy outcomes.[23] However, the microbiota's effect during pregnancy and later life is being investigated, mainly in high-resource settings. As such, low-resource settings, like Ethiopia, have a limited contribution to the global knowledge of general microbiota-host-environment studies during pregnancy. Therefore, we are establishing a pregnancy and early-life cohort in eastern Ethiopia to fill this gap.

Aims (primary)

- Characterise the metataxonomic profile and the changes across trimesters and post partum.
- Assess the association of microbiota changes with pregnancy outcomes.
- Identify the common microbiota in childhood until 2 years and their link with child growth and development.
- Establish a microbiota biobank for future inquiries in eastern Ethiopia.
- Identify determinants of adverse perinatal outcomes through a metagenomic approach.

Aims (secondary)

Assess the prevalence of sexually transmitted infections, bacterial vaginosis (BV), vaginal candidiasis (VC) and their associated factors among pregnant women.

METHODS AND ANALYSIS

Study settings

This study will be integrated into the Hararghe Health and Demographic Surveillance System (HDSS).[32] The HDSS is an open cohort that maintains a continuous demographic and health data record for a known population within a defined geographic area. It regularly tracks vital events such as births, deaths, migration and marital status in urban, semiurban and rural districts. Individuals enter into the surveillance system through enumeration, birth or immigration and will leave the system through out-migration or death.[32] Midwives trained in basic obstetric ultrasound techniques perform ultrasound scans, estimate gestational age and invite those in the early trimester (12–22 weeks) at each facility. Moreover,

we will use village pregnancy detectors to identify pregnant women and refer them to facilities for ultrasound screening, thereby including them in the study.

Study design and population

This is a longitudinal multi-centre cohort study with embedded cross-sectional and case-control studies from August 2024 to January 2027. The primary cohort will consist of all women in early pregnancy (12–22 weeks) who will be followed throughout pregnancy up to 2 years post partum to study changes in the microbiota and the impact on pregnancy outcomes. Within this main cohort, a cross-sectional study, including repeated surveys, will be used to identify the microbiota, its changes and associated factors at a certain point in time. Additionally, we will use a case-control study to identify determinants of adverse pregnancy outcomes (stillbirth, preterm birth, neonatal death, congenital anomaly, low birth weight, macrosomia and small for gestational age).

The sample size is calculated based on the primary and secondary outcome variables. For the primary outcome variables, the calculation is done using the G* Power V.3.1.9.4 software with the assumption of effect size (Shannon diversity) of 0.25, 80% power and 95% CI, yielding a total sample size of 197. The primary outcome variable is the change in vaginal and gut microbiota from the first trimester of pregnancy to immediately post partum. To identify major determinants of adverse birth outcomes, the sample size is calculated using the OpenEpi (V.3) software by considering the prevalence of adverse birth outcomes, which was 27.47% from a previous study conducted in Tefera Hailu Memorial Hospital, Sekota Town, Northern Ethiopia[33], 80% power, 95% CI and a case-to-control ratio of 1:1. A total sample size will be 38 (19 for adverse birth outcomes cases (stillbirth, preterm birth, neonatal death, congenital anomaly, low birth weight, macrosomia and small for gestational age) and 19 for controls (term birth outcomes).

For secondary outcome variables (ie, the prevalence of BV, VC and sexually transmitted infections), the sample size is estimated using a single population proportion formula for a proportion from the previous studies by using a 95% CI and 5% marginal error. For its associated factors, it is estimated by a double proportion formula using the EPI Info V.7 statistical software, with the assumptions of 95% CI, 80% power and a 1:1 ratio. The final sample size is 217, including a 10% loss to follow-up.

Inclusion and exclusion criteria

Inclusion criteria

- Women within the reproductive age (18–49 years).
- Between 12 and 22 weeks of gestational age (confirmed by obstetric ultrasound).
- Residing in the selected sites for at least 3 months during the pregnancy.
- Not planning to move out of the study area within the next 2 years.
- Singleton pregnancy (foetus).

Exclusion criteria

- Prolonged (more than 3 months) use of systemic antibiotics, antifungals, antivirals or antiparasitic therapy.
- Observed abnormalities during enrolment via ultrasound, such as anencephaly and spina bifida.

Data collection and measurements

Data will be collected on Redcap by trained female research midwives proficient in the local languages (Afan Oromo and Amharic) and knowledgeable about the local cultural contexts. At enrolment, data will be collected on the current pregnancy and obstetric history, food and nutrition, and household conditions through interviewer-administered questionnaires and observation. The questionnaire and checklist are adapted from the Ethiopian Demographic and Health Survey and the WHO indicators for the health of women during pregnancy, labour and delivery.[34 35] Initially prepared in English, local experts will translate the questionnaire into local languages (Afaan Oromo and Amharic) and back-translate it into English. The questionnaire will be pretested on 5% of the sample of pregnant women residing outside the study area. The research midwives will be trained on the standard operating procedure (SOP) and the overall aspects of the study. The questionnaire is found in online supplemental information 1 and solicits information on the following: livelihoods, wealth, animal ownership, management and disease, water, sanitation and hygiene, child health and nutrition and women's empowerment. We will further collect data on substance use (including khat), dietary diversity, household food insecurity (HFIAS), antibiotic use and diet. Furthermore, comprehensive obstetric and reproductive health conditions will be obtained at enrolment (12–22 weeks gestational age), 28–36 weeks and birth (within 7 days). Details of the data to be collected at each phase are described in figure 1.

Specimen collection, transportation and storage

An SOP on specimen collection, transportation, storage and analysis is prepared for leading the work. At enrolment, three vaginal swabs and one stool sample will be collected. At visit two (28–36 weeks), one vaginal swab and one stool sample will be collected. At birth, one vaginal and breast milk sample from the mother, as well as one rectal and skin swab from the newborn, will be collected.

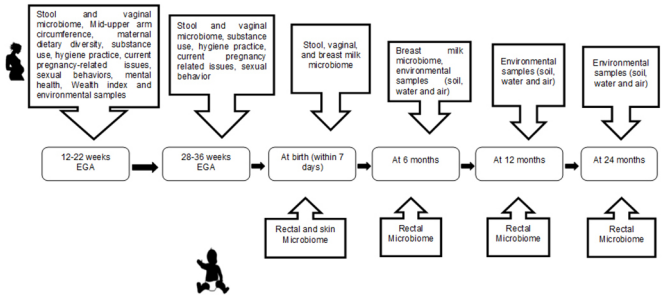

Figure 1 Diagrammatic presentation of the study flow chart and samples to be collected. EGA, estimated gestational age.

Rectal swabs will be collected from the newborn at 6, 12 and 24 months. Environmental samples, including water, indoor air and soil, will be collected during enrolment at birth, 6, 12 and 24 months post partum at the participants' home (figure 1).

Vaginal, rectal and skin swab samples will be collected using DNA/RNA Shield Collection (2 mL fill) swabs (Zymo Research, Irvine, CA, USA). The mother will lie down on the couch, and the sample will be collected by inserting the swab into the midpoint of the vagina.[36–38] The skin swab of the newborn infant will be collected from the retro-auricular crease (ear)[38] and the alar crease (nose).[39]

At enrolment, two additional vaginal swabs will be collected to investigate BV and VC using sterile Dacron swabs from the lateral and posterior vaginal fornices.[40] One of the vaginal swabs will immediately be dipped into Amies transport media and labelled with the patient's barcode number during collection before it is dispatched, then transported to the Hararghe Health Research (HHR) laboratory. The second vaginal swab will be used to perform a Gram stain for Nugent scoring and perform the whiff test.

The stool sample will be collected using the DNA/RNA Shield Fecal Collection Tube (Zymo Research). Place the plastic hat over the toilet seat to collect stool to prevent faeces from falling into the toilet water, ground/soil or urine to avoid sample contamination. If study participants are unable to produce the stool samples at health facilities, they will be provided with the collection kit and clear instructions for self-collection at home. All samples preserved in the DNA/RNA Shield reagent are stable and can be transported and stored at room temperature. Specimens received in the laboratory for nucleic acid detection will be stored at temperatures below −20°C (Zymo Research). Breast milk will be collected using a sterile container. The mothers will collect the sample by pump or manually following aseptic techniques. The collected sample will be stored immediately in a 2–8°C refrigerator and transported within 4 hours. Specimens received in the laboratory for nucleic acid detection will be stored at temperatures below −20°C.[41]

Approximately 200 g of soil samples will be collected from multiple points in the living compounds and indoors of the study participants. The samples will be the top or surface of 5–15 cm soil, where most of the microbial population is concentrated and their activity occurs. Using a sterilised spatula, the sample will be transferred to clean, dry and sterile polythene zip lock bags. To maintain accuracy and representativeness, wet patches, manure, urine areas, wood dumps burdening locations, fence rows, spoil banks, footpaths and the burn-row regions will be avoided.[42] A 250 mL drinking water sample will be collected using a sterile bottle from the point of use.[42] Moreover, indoor air samples will be collected to determine the concentration and diversity of fungus and bacteria-carrying particles in the living environments of pregnant mothers using a passive indoor air monitoring technique called a settling plate. The standard Petri dish containing MacConkey broth culture media will be exposed to the closed indoor air for 60 min. Multiple Petri dishes will be used based on the area of the house and 1 m^3 of air volume.[43]

Sample processing

Two of the vaginal samples collected at enrolment will be used for screening BV and VC. One of the swabs will be used to perform Gram staining. The vaginal swab will be smeared, dried, heat-fixed and Gram-stained using the SOP.[44] Next, it will be scored using the Nugent score system for laboratory diagnosis of BV.[45] The other swab will be used to test VC using other laboratory methods. In summary, the vaginal swab will be initially inoculated on Sabouraud Dextrose Agar (Oxoid, Hampshire, UK) and incubated at 37°C for 24–48 hours. Candida species identification will be performed by subculturing the identified fungus onto the selective and differential Chromogenic candida agar medium, which enables us to differentiate candida species based on their colour on the media[46]; also, the conventional biochemical test and sugar fermentation will be used for fungal identification. Antifungal susceptibility testing will be performed using a modified disc diffusion method for all Candida isolates as per the Clinical and Laboratory Standards Institute guidelines by adding 2% glucose and 0.5 µg/mL methylene blue dye into Mueller-Hinton agar. Finally, the zones of inhibition will be measured and interpreted according to CLSI guidelines (M44-A2).[47] The Gram staining and Candida species identification procedures will be conducted at the HHR laboratory (figure 2).

Environmental sample analysis

The drinking water collected from households will be analysed for its physicochemical and bacteriological quality. Physicochemical quality parameters such as pH, temperature, dissolved oxygen, nitrate, ammonia, hardness, fluoride and residual chlorine will be determined. Bacteriological quality indicators mainly fecal coliform and total coliform will be conducted using a membrane filtration technique and incubated at 37°C for total coliform and 44°C using proper media membrane lauryl sulfate. The results will be reported as colony-forming units (CFUs). To analyse the microbial composition of the soil sample, a serial dilution of 10 to 1 will be used.[42]

The index of microbial air contamination from the indoor air sample will be explored through a 48-hour incubation at 37°C and reported as counts of CFUs. The possible bacteria and fungi that cause indoor air quality problems will be identified through Gram stain and biochemical analysis.[43]

Nucleic acid extraction and sequencing

DNA extraction from the vaginal, stool, breast milk, rectal and skin samples will be conducted using the Roche extraction kits, following the standard manufacturer's protocol at the HHR laboratory. We will appropriately

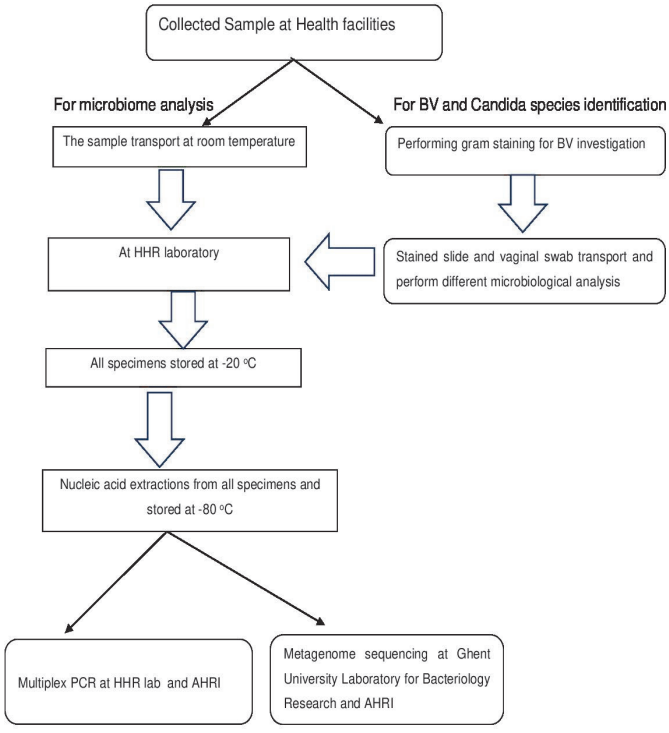

Figure 2 Laboratory analysis and sample storage flow chart. AHRI, Armauer Hansen Research Institute; BV, bacterial vaginosis; HHR, Hararghe Health Research.

document sample batches to control batch effects in our analysis. The DNA concentration will be measured using the Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). For those samples with a low load of DNA yield, we will amplify and sequence the 16S rRNA gene.

Library preparation will be performed using Nextera DNA Flex Library Preparation Kits. To allow pooled sample analysis, the samples will be indexed using IDT Illumina Nextera DNA Unique Dual Indexes. Pair-end sequencing will be done using the Illumina NextSeq2000 at the Ghent University Laboratory for Bacteriology Research and the Armauer Hansen Research Institute (AHRI). We will use rarefaction curves to determine the required sequencing depth to provide adequate coverage of microbial communities for taxonomic profiling. We will aim for a minimum yield of 2–5×10^6 read-pairs per sample. For samples of very low microbial load, that is, skin and breast milk, we will compute diversity indexes setting the minimum depth at 0.5 million reads. To identify potential environmental and laboratory contaminants, we will sequence negative controls and positive controls of mock communities along with the samples.

Metataxonomic profiling of the different microbiota

Obtained sequences will be trimmed, and sequencing adaptors will be removed using Trimmomatic.[48] Assembly will then be performed using SOAPdenovo2.[49] To filter and remove taxa that are found in a small number of samples and have small counts, we will use PERFect,[50] which, besides filtering, can be used to define and quantify loss due to filtering. Host sequences will be removed by mapping against the human genome with Bowtie2.[51] Reads will then be assigned to the most appropriate taxonomic level using MetaphlAn2.[52] Phylogenetic analysis will be performed using RAxML V.8.[53]

Integration of multi-omic microbiome data and disease prediction

To evaluate changes and interrelatedness of the breast milk, gut, skin and vaginal microbiome, we will use NetMoss.[54] The presence or absence of species across the range of microbiomes will be used to identify shared species. We will use SNP calling to detect shared species.[55] We will also apply phyLoSTM,[56] which is a deep learning framework that uses a combination of convolutional neural networks and long short-term memory networks to analyse the temporal dependency in our longitudinal microbiome sequencing data and environmental and clinical factors for the prediction of outcomes, including stillbirth, preterm birth, low birth weight and small for gestational age (SGA).

Pathogen detection

A multiplex PCR will be run for selected vaginal pathogens at the HHR and the AHRI.

Establishing the biobank

All nucleic acid extracts will be stored in a freezer at −80°C at the HHR laboratory for biobanking. All samples in the biobank will be archived according to strict standard practice protocols that will protect the participants' identities. The biobank will be managed according to the principles of transparency, ethics and national laws to support sustainability. Microbiome research is an extensive undertaking and requires most samples to be archived in the biobank.

Data processing and analysis

All data captured through the Redcap software will be exported and analysed using Stata 17. Descriptive data will be presented using frequency tables and figures. Categorical variables will be summarised using frequencies and proportions, while continuous variables will be described using means with SD and medians with the IQRs as appropriate. To compare the differences in proportions, we will use the χ^2 test, while Student's t-test will be used to compare the mean difference for normally distributed continuous data. A generalised linear mixed model that is flexible to address discrete variables with repeated observations will be used to assess the relationship between categorical outcomes and associated factors. We will use generalised estimating equations or mixed effects regression to analyse the regression relationship between normally distributed, repeatedly observed continuous outcome variables and covariates. Moreover, we will use a generalised linear model (binary logistic regression, Poisson regression or binomial) to assess the association between binary categorical variables measured once and covariates. Non-parametric variables will be considered for comparing data that are not normally distributed. The

strength of association will be determined by calculating the OR or prevalence/incidence ratio or rate and their 95%CIs based on the types of outcome variables under consideration.

For the statistical differences between the vaginal and gut microbiota throughout gestation and post partum, the Statistical Analysis of Metagenomics Profile software package will be used. A linear mixed model regression analysis will be used to find the statistical significance of microorganism abundance and community state types (CSTs) during pregnancy and post partum. A linear regression analysis will be performed for each microorganism to measure whether there is a significant change in bacterial species abundance and whether CSTs change significantly throughout pregnancy and post partum. For postpartum changes in microorganism abundance, we will create a binary time variable to distinguish between measurements taken during pregnancy and those collected post partum. To assess whether CSTs change significantly during pregnancy, we will conduct linear regression for each CST. Similarly, to examine CST changes during the puerperium, we will establish a binary time variable for measurements from both pregnancy and post partum, followed by linear regression for each CST. To identify the association between the composition of the vaginal and gut microbiota during pregnancy and post partum, a bivariate analysis will be done. The logistic regression model will also be used to test the significance of the interaction between other biobehavioural factors, sociodemographic factors and microbiota compositions. Multiple testing corrections will be applied using the Bonferroni Correction, which adjusts the significance level by dividing the desired alpha level (eg, 0.05) by the total number of tests performed.

Ethics and dissemination

Ethical approval was obtained from the Institutional Health Research Ethics Review Committee (Ref No. IHRERC/033/2022) and renewed (Ref No. IHRERC/103/2024) at Haramaya University, Ethiopia. The study subjects will be briefed on the purpose of the study, and their participation will be entirely voluntary, with the right to withdraw from the study at any point. Each participant will be assigned a participant identification number (code), and all information will be stored in a locked cabinet to maintain confidentiality. The mother/guardian will give consent on behalf of their children to participate in the study. Informed consent will be obtained from all study participants before data collection, as found in online supplemental information 2. Study participants will be asked to consent to storing samples and data as necessary for potential future usage. All samples in the biobank will be collected and archived according to a strict standard of practice, which will protect the participants' identity. The biobank will be managed according to the principles of transparency, ethics and national laws to support sustainability. Results will be published in international peer-reviewed journals, and summaries will be provided to the study funders. Clinical study data will be submitted to Data Compass (https://datacompass.lshtm.ac.uk/), and molecular profiles of the microbiome and whole-genome sequences will be submitted to the European Nucleotide Archive (https://www. ebi.ac.uk/ena). No new data have been created or analysed to date, as this is a study protocol, and therefore data sharing is not applicable. Once the data are collected and the biobank is established, all enquiries for data should be directed to Dr Abera Kenay (email: daberaf@gmail.com). Data-sharing decisions will be made by the project steering committee of the College of Health and Medical Sciences, Haramaya University, following the university's data-sharing guidelines.

Patient and Public Involvement (PPI)

There will not be patient or public involvement in the study.

DISCUSSION

In this study, we aim to establish a novel system to continuously register the microbiota data and its changes throughout pregnancy in eastern Ethiopia. Through enrolling women in the Hararghe Demographic Surveillance, we will collect primary socioeconomic, obstetric and faeto-maternal conditions, obstetric and medical history, and stool and vaginal samples throughout pregnancy and within 2 years post partum.

Many complications during pregnancy, such as gestational diabetes, preeclampsia, digestive disorders, susceptibility to certain infectious diseases and autoimmune diseases, are related to disruptions of gut microbiota.[57 58] The microbiota in the mothers' gastrointestinal tract is closely related to the infants' microbiota, contributing to chronic illnesses of the gastrointestinal system, such as Crohn's disease and irritable bowel syndrome.[59 60] Moreover, the microbiota in the mother's breast milk and vagina are closely related to the colonisation of microbiota in the early life of infants.[61] A sharp decline in the total number of beneficial microbes, that is, *Lactobacillus*, and an increase in the concentration of anaerobic bacteria resulted in BV.[62–65] The BV is associated with different adverse pregnancy outcomes, including spontaneous abortion, recurrent pregnancy losses, preterm labour, preterm delivery, preterm pre-labour rupture of membranes, low birth weight, amniotic fluid colonisation, postpartum endometritis, increased risk of pelvic inflammatory disease and facilitated transmission of sexually transmitted infections (STIs).[66 67]

Despite understanding the multifaceted health problems associated with the microbiota during pregnancy, Africa, in general, and Ethiopia, in particular, are lagging in contributing to this evidence. Given that the maternal microbiota is diverse and is highly influenced by the environment, diet, health, mode and place of delivery, and genetics, it is essential to study this in a low-resource setting, such as Ethiopia. Evidence from longitudinal clinical

studies and animal models revealed that disruption in the gut microbiota in early life could programme later-life disease. A microbiota biobank would facilitate answers to many questions relevant to human development in a country with a high burden of communicable diseases and an emergence of non-communicable diseases. The EthiOMICS biobank would provide abundant evidence in the future. Identifying the dynamics of microbes in pregnant women and their children and their association with maternal and child health will guide policy that improves maternal and child health in Ethiopia.

Findings from this project will be essential to unravelling contextual host-agent-environment interactions. With global research leaning towards big data and bioinformatics, the EthiOMICS biobank would place Haramaya at the forefront of future genomics and big data in Ethiopia. Like the prominent UK Biobank—which produces large datasets[68] and generates revenue—EthiOMICS would be a unique data source owned by Haramaya University. It could generate income for the university through its genome database and further expand the research work.

Author affiliations
[1]School of Medical Laboratory Sciences, Haramaya University College of Health and Medical Sciences, Harar, Ethiopia
[2]Laboratory Bacteriology Research, Department of Diagnostic Sciences, Ghent University Faculty of Medicine and Health Sciences, Gent, Belgium
[3]School of Public Health, Haramaya University College of Health and Medical Sciences, Harar, Ethiopia
[4]School of Nursing, Haramaya University College of Health and Medical Sciences, Harar, Ethiopia
[5]School of Environmental Health, Haramaya University College of Health and Medical Sciences, Harar, Ethiopia
[6]School of Medicine, Haramaya University College of Health and Medical Sciences, Harar, Ethiopia
[7]Armauer Hansen Research Institute, Addis Ababa, Ethiopia
[8]Department of Microbiology, Immunology and Parasitology, School of Medicine, Addis Ababa University, Addis Ababa, Ethiopia
[9]Obstetrics and Gynaecology, University Medical Centre Groningen, Groningen, The Netherlands
[10]Department of International Public Health, Liverpool School of Tropical Medicine, Liverpool, UK

Contributors The study concept and design were conceived by AKT, DAA, TW, AG, TAY, MD, FW, KTR and NA. FW, AKT, ASN, TT, MD, TAY, NA, YMD, DAA, AAN, TGE, AA, AG, TW, KTR, PC and TNA refined the study questionnaires and study design, and supervised the data collection and sample processing. Analyses will be conducted by FW, AKT, PC, AAN, TGE and ASN with the help of statisticians. FW prepared the first draft with close support from AKT. AKT is the guarantor. All authors critically revised the review and intellectual content of the final protocol.

Funding This study is funded by Haramaya University (grant number: HUGG_2021_02_04_26) and co-funded by Ghent University and Armauer Hansen Research Institute (AHRI). This work was supported by Universitaire Stichting (Belgium). The funding body has no role in the design of the study, data collection, data analysis, or interpretation.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Consent obtained directly from patient(s).

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Fitsum Weldegebreal http://orcid.org/0000-0002-7867-1483
Tesfaye Assebe Yadeta http://orcid.org/0000-0003-3015-8979
Yohannes Mulugeta Demmu http://orcid.org/0000-0002-3185-8240
Tadesse Gure Eticha http://orcid.org/0000-0003-1213-1385
Nega Assefa http://orcid.org/0000-0003-0341-2329
Abera Kenay Tura http://orcid.org/0000-0002-2735-7523

REFERENCES

1 Assefa Y, Damme WV, Williams OD, *et al*. Successes and challenges of the millennium development goals in Ethiopia: lessons for the sustainable development goals. *BMJ Glob Health* 2017;2:e000318.
2 WHO. Trends in maternal mortality 2000 to 2017: estimates by who. *UNICEF*; 2019.
3 Ruducha J, Mann C, Singh NS, *et al*. How Ethiopia achieved Millennium Development Goal 4 through multisectoral interventions: a Countdown to 2015 case study. *Lancet Glob Health* 2017;5:e1142–51.
4 Hug L, Sharrow D, Zhong K, *et al*. Unicef ea. levels & trends in child mortality: report 2018, estimates developed by the united nations children's fund. United nations inter-agency group for child mortality estimation; 2018.
5 Cunningham SA, Shaikh NI, Nhacolo A, *et al*. Health and Demographic Surveillance Systems Within the Child Health and Mortality Prevention Surveillance Network. *Clin Infect Dis* 2019;69:S274–9.
6 Gluckman PD, Hanson MA, Cooper C, *et al*. Effect of in utero and early-life conditions on adult health and disease. *N Engl J Med* 2008;359:61–73.
7 Muglia LJ, Benhalima K, Tong S, *et al*. Maternal factors during pregnancy influencing maternal, fetal, and childhood outcomes. *BMC Med* 2022;20:418.
8 Sekirov I, Russell SL, Antunes LCM, *et al*. Gut microbiota in health and disease. *Physiol Rev* 2010;90:859–904.
9 Dominguez-Bello MG, Costello EK, Contreras M, *et al*. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA* 2010;107:11971–5.
10 Dunn AB, Jordan S, Baker BJ, *et al*. The Maternal Infant Microbiome: Considerations for Labor and Birth. *MCN Am J Matern Child Nurs* 2017;42:318–25.
11 Oyedemi OT, Shaw S, Martin JC, *et al*. Changes in the gut microbiota of Nigerian infants within the first year of life. *PLoS ONE* 2022;17:e0265123.
12 Hickey RJ, Zhou X, Pierson JD, *et al*. Understanding vaginal microbiome complexity from an ecological perspective. *Transl Res* 2012;160:267–82.
13 Yamamoto T, Zhou X, Williams CJ, *et al*. Bacterial populations in the vaginas of healthy adolescent women. *J Pediatr Adolesc Gynecol* 2009;22:11–8.
14 McLEAN NW, Rosenstein IJ. Characterisation and selection of a Lactobacillus species to re-colonise the vagina of women with recurrent bacterial vaginosis. *J Med Microbiol* 2000;49:543–52.
15 Edwards SM, Cunningham SA, Dunlop AL, *et al*. The Maternal Gut Microbiome During Pregnancy. *MCN Am J Matern Child Nurs* 2017;42:310–7.
16 Taddei CR, Cortez RV, Mattar R, *et al*. Microbiome in normal and pathological pregnancies: A literature overview. *Am J Reprod Immunol* 2018;80:e12993.
17 Zhang X, Zhai Q, Wang J, *et al*. Variation of the Vaginal Microbiome During and After Pregnancy in Chinese Women. *Genomics Proteomics Bioinformatics* 2022;20:322–33.

18 Saraf VS, Sheikh SA, Ahmad A, *et al*. Vaginal microbiome: normalcy vs dysbiosis. *Arch Microbiol* 2021;203:3793–802.
19 Hasan N, Yang H. Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* 2019;7:e7502.
20 Jiang G, Zhou Z, Li X, *et al*. The Gut Microbiome During Pregnancy. *Maternal Fetal Med* 2023;5:36–43.
21 Kumar P, Magon N. Hormones in pregnancy. *Niger Med J* 2012;53:179–83.
22 Collado MC, Isolauri E, Laitinen K, *et al*. Distinct composition of gut microbiota during pregnancy in overweight and normal-weight women. *Am J Clin Nutr* 2008;88:894–9.
23 Koren O, Goodrich JK, Cullender TC, *et al*. Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* 2012;150:470–80.
24 Amir M, Brown JA, Rager SL, *et al*. Maternal Microbiome and Infections in Pregnancy. *Microorganisms* 1996;8:1996.
25 Tang M, Frank DN, Tshefu A, *et al*. Different Gut Microbial Profiles in Sub-Saharan African and South Asian Women of Childbearing Age Are Primarily Associated With Dietary Intakes. *Front Microbiol* 2019;10:1848.
26 Dahl C, Stanislawski M, Iszatt N, *et al*. Gut microbiome of mothers delivering prematurely shows reduced diversity and lower relative abundance of Bifidobacterium and Streptococcus. *PLoS One* 2017;12:e0184336.
27 Dunlop AL, Mulle JG, Ferranti EP, *et al*. Maternal Microbiome and Pregnancy Outcomes That Impact Infant Health: A Review. *Adv Neonatal Care* 2015;15:377–85.
28 Geldenhuys J, Redelinghuys MJ, Lombaard HA, *et al*. Diversity of the gut, vaginal and oral microbiome among pregnant women in South Africa with and without pre-eclampsia. *Front Glob Womens Health* 2022;3:810673.
29 Jandhyala SM, Talukdar R, Subramanyam C, *et al*. Role of the normal gut microbiota. *WJG* 2015;21:8787.
30 Milani C, Duranti S, Bottacini F, *et al*. The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiol Mol Biol Rev* 2017;81:e00036-17.
31 Kim H, Sitarik AR, Woodcroft K, *et al*. Birth Mode, Breastfeeding, Pet Exposure, and Antibiotic Use: Associations With the Gut Microbiome and Sensitization in Children. *Curr Allergy Asthma Rep* 2019;19:22:1–9:.
32 Assefa N, Oljira L, Baraki N, *et al*. HDSS Profile: The Kersa Health and Demographic Surveillance System. *Int J Epidemiol* 2016;45:94–101.
33 Seyoum E, Bekele A, Tsegaye AT, *et al*. Magnitude and Determinants of Adverse Perinatal Outcomes in Tefera Hailu Memorial Hospital, Sekota Town, Northern Ethiopia. *Glob Pediatr Health* 2021;8:2333794X211015524.
34 Countdown to 2015 Maternal Health Group, UNICEF. Monitoring maternal, newborn and child health: understanding key progress indicators. 2011.
35 CSACE I. Ethiopia Demographic and Health Survey 2016. Ethiopia, and Rockville, Maryland, USA: Addis Ababa, 2016.
36 Aagaard K, Riehle K, Ma J, *et al*. A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. *PLoS One* 2012;7:e36466.
37 HUTH-NHST. Instructions for Self-Collection of Vaginal Swabs: Hull University Teaching Hospitals NHS Trust, 2009. Available: https://www.hey.nhs.uk/pathology/departmentofinfection/virology/vaginal-swabs/
38 Number HP, Sites PL. Human microbiome project–core microbiome sampling protocol a. 2010.
39 Skin swab instructions: Children's Hospital of Philadelphia, 2021. Available: https://www.research.chop.edu/sites/default/files/2021-05/Skin%20Swab%20Sampling%20Instructions.pdf
40 Gaydos CA, Crotchfelt KA, Shah N, *et al*. Evaluation of dry and wet transported intravaginal swabs in detection of Chlamydia trachomatis and Neisseria gonorrhoeae infections in female soldiers by PCR. *J Clin Microbiol* 2002;40:758–61.
41 CDC. Storage and preparation of breast milk, 2024. Available: https://www.cdc.gov/breastfeeding/pdf/preparation-of-breast-milk_h.pdf
42 Burh PM. Microbial analysis of soil and water samples from Koel river in Rourkela. 2011.
43 Pasquarella C, Pitzurra O, Savino A. The index of microbial air contamination. *J Hosp Infect* 2000;46:241–56.
44 Odogwu NM, Onebunne CAC, Olayemi OO, *et al*. Protocol for a Longitudinal Analysis of the Vaginal Microbiome from a Pregnant Cohort of African Women in Nigeria. *Int J MCH AIDS* 2020;9:173–81.
45 Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J Clin Microbiol* 1991;29:297–301.
46 Gms I. In: *TM 1197-CHROMOGENIC CANDIDA AGAR [CHROMOGENIC CANDIDA DIFFRENTIAL AGAR]*.
47 CLSI. Method for antifungal disk diffusion susceptibility testing of yeasts; approved guideline second edition. CLSI document M44-A2. Wayne, PA.Clinical and Laboratory Standards Institute; 2009.
48 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20.
49 Luo R, Liu B, Xie Y, *et al*. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 2012;1:18.
50 Smirnova E, Huzurbazar S, Jafari F. PERFect: PERmutation Filtering test for microbiome data. *Biostatistics* 2019;20:615–31.
51 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
52 Truong DT, Franzosa EA, Tickle TL, *et al*. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 2015;12:902–3.
53 Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–3.
54 Xiao L, Zhang F, Zhao F. Large-scale microbiome data integration enables robust biomarker identification. *Nat Comput Sci* 2022;2:307–16.
55 Truong DT, Tett A, Pasolli E, *et al*. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res* 2017;27:626–38.
56 Sharma D, Xu W. phyLoSTM: a novel deep learning model on disease prediction from longitudinal microbiome data. *Bioinformatics* 2021;37:3707–14.
57 Hair M, Sharpe J. The Center for Ecogenetics and Environmental Health. 1/2014. NIEHS Grant #P30ES007033, 2014.
58 Gorczyca K, Obuchowska A, Kimber-Trojnar Ż, *et al*. Changes in the Gut Microbiome and Pathologies in Pregnancy. *Int J Environ Res Public Health* 2022;19:9961.
59 Andersson OP. Is the microbiome the cause of irritable bowel syndrome and inflammatory bowel disease? Lessons to consider from odontology. *Int J Colorectal Dis* 2023;38:117.
60 Yilmaz B, Juillerat P, Øyås O, *et al*. Microbial network disturbances in relapsing refractory Crohn's disease. *Nat Med* 2019;25:323–36.
61 Yao Y, Cai X, Chen C, *et al*. The Role of Microbiomes in Pregnant Women and Offspring: Research Progress of Recent Years. *Front Pharmacol* 2020;11:643.
62 Abdelaziz ZA, Ibrahim ME, Bilal NE, *et al*. Vaginal infections among pregnant women at Omdurman Maternity Hospital in Khartoum, Sudan. *J Infect Dev Ctries* 2014;8:490–7.
63 Castro J, Machado D, Cerca N. Unveiling the role of Gardnerella vaginalis in polymicrobial Bacterial Vaginosis biofilms: the impact of other vaginal pathogens living as neighbors. *ISME J* 2019;13:1306–17.
64 Drew RJ, Murphy T, Broderick D, *et al*. An interpretation algorithm for molecular diagnosis of bacterial vaginosis in a maternity hospital using machine learning: proof-of-concept study. *Diagn Microbiol Infect Dis* 2020;96:114950.
65 Muzny CA, Blanchard E, Taylor CM, *et al*. Identification of Key Bacteria Involved in the Induction of Incident Bacterial Vaginosis: A Prospective Study. *J Infect Dis* 2018;218:966–78.
66 Ugwumadu A. Role of antibiotic therapy for bacterial vaginosis and intermediate flora in pregnancy. *Best Practice & Research Clinical Obstetrics & Gynaecology* 2007;21:391–402.
67 Isik G, Demirezen Ş, Dönmez H, *et al*. Bacterial vaginosis in association with spontaneous abortion and recurrent pregnancy losses. *J Cytol* 2016;33:135.
68 Bycroft C, Freeman C, Petkova D, *et al*. The UK Biobank resource with deep phenotyping and genomic data. *Nature New Biol* 2018;562:203–9.