HELSINGIN YLIOPISTO

Quantitative analysis of fungi in the female reproductive tract in recurrent pregnancy loss

Fungal qPCR optimization
of gynecological samples

Genetics and Molecular Biosciences,
Molecular and Analytical Health Biosciences
Master's thesis

Author:
Maria Sofia Norppa

Supervisors:
Docent Anne Salonen
PhD Rebecka Ventin-Holmberg

11.11.2024
Helsinki

Faculty: Faculty of Biological and Environmental Sciences, University of Helsinki
Degree program: Genetics and Molecular Biosciences
Study track: Molecular and Analytical Health Biosciences
Author: Maria Sofia Norppa
Title: Quantitative analysis of fungi in the female reproductive tract in recurrent pregnancy loss – Fungal qPCR optimization of gynecological samples
Level: Master's thesis
Month and year: November 2024
Number of pages: 66
Keywords: mycobiota, fungi, yeast, Candida albicans, female reproductive tract, vaginal and endometrial microbiota, gynecological microbiota, mycobiome, recurrent pregnancy loss (RPL), miscarriage, dysbiosis, qPCR, qNGS, qPCR optimization, qPCR troubleshooting
Supervisors: Docent Anne Salonen, PhD Rebecka Ventin-Holmberg
Where deposited: University of Helsinki, E-thesis
Abstract:

The human microbiome research has focused mainly on bacteria, dismissing the mycobiota (the fungal microbiota), yet it is an important part of the human microbiota. The dysbiosis of a female reproductive tract microbiota predisposes women to a variety of gynecological health issues and diseases. Severe reproductive problems, such as an increased risk of miscarriage and recurrent pregnancy loss (RPL), have been connected to altered microbiota. Although the most investigated species of vaginal microbiota are bacteria, there is a growing interest to characterize the mycobiota also beyond *Candida albicans,* an abundant yeast species in gynecological samples.

This MSc thesis aimed to investigate the absolute amounts of fungi in vaginal and endometrial samples in women with a history of RPL and their controls, and specifically study if there is a correlation between fungal quantities of the two sample types from the same patient, and whether there are differences in the fungal load between RPL-patients and their controls. The vaginal mycobiota of the control group members was abundant, whereas the endometrium hardly harbored fungi. The RPL-patients had significantly more fungi in the endometrium compared to the controls, whereas the vaginal samples were more similar among both groups. The results suggested an association between altered fungal quantities of endometrium mycobiota and RPL.

The main method used in the thesis was quantitative PCR (qPCR), which required extensive optimization for gynecological samples. The qPCR data was generated by targeting the internal transcribed spacer 1 (ITS1) region in all fungi and coupled to next-generation sequencing (NGS) data on ITS1 amplicons from the same samples. This method enabled the quantitative NGS (qNGS) approach, which allows the estimation of the absolute amounts of microbial taxa identified by sequencing. Previously, qNGS has been used to study bacterial microbiota but not human mycobiota. Furthermore, a troubleshooting matrix for problems in qPCR is introduced in this thesis. This thesis took steps towards a more functional protocol for fungal qPCR and provided novel insights into the mycobiota as part of the human microbiota, potentially enhancing the mycobiome research as part of female reproductive health.

Tiedekunta: Bio- ja ympäristötieteellinen tiedekunta, Helsingin yliopisto
Koulutusohjelma: Genetiikka ja molekulaariset biotieteet
Opintosuunta: Molekulaariset ja analyyttiset terveysbiotieteet
Tekijä: Maria Sofia Norppa
Työn nimi: Naisen lisääntymiskanavan hiivojen kvantitatiivinen analyysi toistuvissa keskenmenoissa – Gynekologisten näytteiden qPCR optimointi hiivoille
Työn laji: Maisterintutkielma
Kuukausi ja vuosi: Marraskuu 2024
Sivumäärä: 66
Avainsanat: mykobiota, sienet, hiivat, Candida albicans, naisen lisääntymiskanava, vaginan ja kohdun limakalvon mikrobisto, gynekologinen mikrobisto, mykobiomi, toistuva keskenmeno, keskenmeno, dysbioosi, qPCR, qNGS, qPCR optimointi, qPCR vianmääritys
Ohjaajat: Dosentti Anne Salonen, FT Rebecka Ventin-Holmberg
Säilytyspaikka: Helsingin yliopisto, E-thesis
Tiivistelmä:

Ihmisen mikrobistotutkimus on keskittynyt pääasiassa bakteereihin, jättäen mykobiotan (sienet) vähemmälle huomiolle siitä huolimatta, että se on tärkeä osa ihmisen mikrobistoa. Naisen lisääntymiskanavan mikrobiston epätasapaino altistaa erilaisille gynekologisille terveysongelmille ja sairauksille. Merkittävät lisääntymisongelmat, kuten kohonnut toistuvan keskenmenon riski, on yhdistetty mikrobiston muutoksiin. Vaikka emättimen mikrobiston tutkiminen on keskittynyt bakteereihin, kiinnostus mykobiotan karakterisoimiseen on kasvanut ja laajentunut käsittämään muitakin sieniä kuin *Candida albicans* -lajin, joka on gynekologisten näytteiden runsain hiivalaji.

Tämän opinnäytetyön tavoitteena oli tutkia sienten absoluuttisia määriä emättimen ja kohdun limakalvon näytteistä. Tarkoituksena oli selvittää, onko sienipositiivisuuden ja saman potilaan kahden näytetyypin määrien välillä korrelaatiota ja onko sienien esiintyvyydessä ja määrissä eroja keskenmenopotilaiden ja heidän verrokkiensa välillä. Verrokkiryhmän emättimen mykobiota oli runsasta, kun taas kohdun limakalvolla ei juurikaan ollut sieniä tai hiivoja. Keskenmenopotilailla oli merkitsevästi enemmän sieniä kohdun limakalvolla kuin verrokeilla, kun taas emätinnäytteet olivat samankaltaisempia kaikilla potilailla. Tulokset viittaavat yhteyteen kohdun limakalvon mykobiotan sienimäärien ja toistuvien keskenmenojen välillä.

Opinnäytetyössä käytettiin keskeisenä menetelmänä kvantitatiivista PCR:ää (qPCR), mikä edellytti laajaa optimointia gynekologisten näytteiden osalta. Näytteistä monistettiin sienten *internal transcribed spacer 1* (ITS1) -aluetta, ja saatu qPCR data yhdistettiin samojen näytteiden sekvensointitietoihin (NGS) ITS1-amplikoneista. Menetelmä mahdollisti kvantitatiivisen NGS (qNGS) lähestymistavan, jossa arvioidaan sekvensoinnilla tunnistettujen mikrobilajien absoluuttisia määriä. Aiemmin qNGS menetelmää on käytetty bakteerimikrobiston tutkimiseen, mutta ei ihmisen mykobiotaan. Lisäksi opinnäytetyössä esitellään vianmääritysmatriisi qPCR-ongelmille. Tämä työ on askel kohti toimivampaa protokollaa sienien ja hiivojen qPCR -tutkimuksessa, ja siten edistää mykobiotan tutkimusta osana ihmisen mikrobistoa, sekä parantaa naisten lisääntymisterveyttä.

Table of contents

Abbreviations:

A-T	adenine-thymine
bp	base pair
C_q	quantification cycle
C_T	threshold cycle
CN	copy-number
CNV	copy-number variation
DMSO	dimethyl sulfoxide
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded deoxyribonucleic acid
-d(RFU)/dT	negative first derivative of fluorescence versus the temperature
E%	amplification efficiency
FAM	fluorescein amidites
G-C	guanine-cytosine
HMP	Human Microbiome Project
HT	high-throughput
ITS	internal transcribed spacer
MCA	melt curve analysis
MQ-water	molecular biology water
NIH	National Institutes of Health
NGS	next-generation sequencing
OPATHY	Omics of pathogenic yeasts
PCR	polymerase chain reaction
PD	primer-dimer
qNGS	quantitative next-generation sequencing
qPCR	quantitative polymerase chain reaction
R^2	coefficient of determination
rDNA	ribosomal deoxyribonucleic acid
RFU	fluorescence units
RFU/T	fluorescence of each RFU per temperature as a melt curve
RPL	recurrent pregnancy loss
rRNA	ribosomal ribonucleic acid
RT-PCR	real-time polymerase chain reaction
SQ	starting quantity
ssDNA	single-stranded deoxyribonucleic acid
T_m	melting temperature
TE-buffer	Tris-EDTA -buffer
VVC	vulvovaginal candidiasis
WGS	whole genome sequencing

1 Introduction

1.1 Microbiota and mycobiota in health and disease

Although out of sight and observation, we are constantly affected by a myriad of particles, both inside and outside of us, potentially threatening, but also protecting, even being an inseparable part of ourselves. We are surrounded by the microbiota, defined as all the microorganisms of a certain environment, consisting of bacteria, viruses, archaea, protozoa, parasites, and fungi, whereas the microbiome refers to all the genetic material of the microbiota (Zhang *et al.,* 2021; Zoll *et.al.,* 2016; Gil *et al.*, 2006). The mycobiota, i.e., yeast and other fungi, and respectively the mycobiome, consist of the fungal branch of the microbiota (Virtanen *et al.,* 2023; Wiesmann *et al.*, 2022). Our cells are being outnumbered by the microbes: the abundance of mycobiome components is estimated to be 10^{12}–10^{13}, bacteriome 10^{13}–10^{14}, and virome even more (Rowan-Nash *et al.,* 2019). The microbiota is an entity and could be described as an organ-like structure, which has been proposed to be a part of the human genetic landscape (Gill *et al*, 2006). Indeed, the Human Microbiome Project (HMP) started a "second human genome project" in 2009, aiming to bring insight into our bacteriome, virome, and mycobiome in health and disease, concentrating on the mouth, gut, vagina, and skin, being the four major sites of microbial colonization (National Institutes of Health [NIH] HMP 2009).

The microbiota has a significant influence on human health in many aspects, as it both protects us from harmful colonization of pathogens, but also promotes our immune systems (Rowan-Nash *et al.,* 2019). An imbalance of microbiota, a dysbiosis, can predispose individuals to a variety of health issues and diseases (NIH HMP 2009; Han *et al.,* 2021; Rowan-Nash *et al.,* 2019; Zoll *et.al.,* 2016). According to the Consortium of Omics of Pathogenic Yeasts (OPATHY) fungi related issues have increased in recent years, creating a serious threat, especially to risk groups with reduced immunity (OPATHY & Gabaldon, 2019). The fungi may be commensal asymptomatic colonizers, restricted by the other microbiota (Achkar & Fries, 2010; Rowan-Nash *et al.,* 2019), but from the estimated 1.5 million existing fungal species, around 300 are known to cause diseases in humans, yet those responsible for the majority of serious diseases belong to few genera, namely *Candida*, *Aspergillus*, *Pneumocystis,* and *Cryptococcus* (Stop Neglecting Fungi, 2017; Tiew *et al.,* 2020). Indeed, *Candida* is the most common non-bacterial cause of invasive infections of the bloodstream, intra-abdominal, oral, and vulvovaginal environment

(Achkar & Fries, 2010; da Matta *et al.,* 2017; OPATHY & Gabaldon, 2019). The editorial in Nature Microbiology (Stop Neglecting Fungi, 2017) suggests placing fungi in the spotlight, as they are hardly discussed in public, despite fungal-related diseases are affecting over 300 million people worldwide, causing more fatalities than malaria, with 1.6 million fungi-related deaths annually. Despite the fungi are a ubiquitous and important part of the human microbiota, the focus of human microbiome research has mainly been on bacteria, antibiotic resistance, and viral outbreaks (NIH HMP 2006; Rowan-Nash *et al.,* 2019; Stop Neglecting Fungi, 2017; Virtanen *et al.,* 2023; Wiesmann *et al.*, 2022; Zoll *et.al.,* 2016).

1.2 Microbiota in the female reproductive tract

Several widely recognized health risks are associated with infections or imbalances in the reproductive tract microbiota (Haahr *et al.,* 2019; Han *et al.*, 2021; Peuranpää *et al.*, 2022). Severe reproductive problems, including an increased risk of miscarriage and recurrent pregnancy loss (RPL), have been connected to infections caused by microbial dysbiosis, such as the higher bacterial richness of endometrium (Liu *et al.,* 2022; Peuranpää *et al.*, 2022), chronic endometritis (McQueen *et al.,* 2021) and bacterial vaginosis (Haahr *et al.,* 2019).

The vaginal microbiota, despite being dynamic and influenced by a variety of elements, such as age, body mass index (BMI), ethnicity, pregnancy, and childbirth, is relatively well-known and widely described in previous studies, according to which the vaginal microbiome has a low diversity, *Lactobacillus* species being the most abundant (Boskey *et al.,* 1999; Godoy-Vitorino *et al.,* 2018; Ma *et al.,* 2020; Peuranpää *et al.,* 2022; Virtanen *et al.,* 2023; Virtanen *et al.*, 2019). The endometrium, the inner lining of the uterus, has a drastically lower number of microbes and has long been considered even a sterile environment in healthy women (Mitchell *et al.,* 2015). According to the meta-analysis by Haahr *et al.* (2019) the microbiome of endometrium differs from the vagina, both taxonomically and in abundance, although the vaginal microbiota is a source of ascending colonization.

The vaginal mycobiome is less recognized than the bacteriome (Bradford & Ravel, 2017; Drell *et al.,* 2013; Lehtoranta *et al.,* 2021), and the endometrial mycobiome is hardly studied (Peuranpää *et al.*, 2022). The discussion about the female reproductive tract mycobiome has mostly covered vulvovaginal candidiasis (VVC), a yeast infection caused by *Candida* (Rowan-Nash *et al.,* 2019), yet the vaginal mycobiota has been associated also

with other diseases, such as bacterial dysbiosis (Lehtoranta *et al.*, 2021) and abnormalities in the cells of the cervix, i.e., cervical dysplasia (Godoy-Vitorino *et al.,* 2018). The vaginal mycobiome has a lower taxonomic prevalence and abundance than the bacteriome: in a metagenomic study of more than 1,500 women, Ma *et al.* (2020) encountered that the vaginal mycobiome relative abundance was 0–0.2%, with five fungal species, *Candida* being the most prevalent. According to a pioneering study of the vaginal mycobiota by Drell *et al.* (2013), and many others, *Candida albicans* is the most frequent colonizer of vaginal fungal flora (Achkar & Fries, 2010; Bradford & Ravel, 2017; Guo *et al.,* 2012; Lehtoranta et al., 2021; Ma *et al.* 2020; Rowan-Nash *et al.,* 2019; Virtanen *et al.,* 2023). It has been suggested to be present as an asymptomatic colonizer in more than half of women, also within healthy individuals (Drell *et al.,* 2013; Virtanen *et al.,* 2023). Additional commonly prevalent vaginal mycobiota among healthy women are other *Candida* species, such as *C. glabrata, C. parapsilosis*, and *C. tropicalis* (Drell *et al.,* 2013; Ma *et al.,* 2020), and species from other fungal genera, such as *Malassezia, Saccharomyces, Dothideomycetes, Cladosporium, Eurotium, Epicoccum,* and *Alternaria* (Drell *et al.,* 2013; Guo *et al.,* 2012; Virtanen *et al.,* 2023).

According to a review by Achkar & Fries (2010), typically one fungal species was identified per individual in vulvovaginal diseases, and less than 5% of the patients were observed with more than one species. They encountered *C. albicans* the most common species in VVC and urinary infections, yet there was a notable increase in the portion of non-*albicans* species, such as *C. glabrata*, *C. tropicalis,* and *C. parapsilosis,* especially among recurrent VVC, immunodeficient patients, uncontrolled diabetes, and postmenopausal women. The emergence of non-albicans *Candida* species is concerning, as antifungal resistance is more common among them (Achkar & Fries, 2010). Furthermore, *Candida*, Basidiomycota, *Malassezia,* and *Sporidiobolaceae* have been associated with cervical dysplasia, such as papillomavirus infections (Godoy-Vitorino *et al.,* 2018).

Factors such as the menstrual cycle, antibiotic and probiotic use, history of bacterial vaginosis, and most strongly, the number of lifetime sex partners, affect the vaginal mycobiome (Virtanen *et.al.*, 2023). In addition, genetics and immune-based factors, as well as behavioral and nutritional factors have been suggested to modulate the vaginal mycobiome (Achkar & Fries, 2010). Furthermore, the mycobiome is affected by the bacteriome, and the interactions have been under intensive investigation, as it has been shown in various studies (Achkar & Fries, 2010; Rowan-Nash *et al.,* 2019) that *Candida* is

antagonized in several ways by bacteria of the *Lactobacillus* genus, among vaginal microbiota in particular, where lactobacilli lower the pH by producing lactic acid, thus providing a chemical barrier against pathogens (Boskey *et al.,* 1999). Virtanen *et al.,* (2023) have studied the correlations between the vaginal microbiota and mycobiota but did not find strong correlations. The disturbance, such as antibiotic treatment, of the balance between the fungal colonization and the host, might lead to fungal overgrowth, leading to the development of a disease, such as VVC (Achkar & Fries, 2010).

An understanding of the functional and compositional dynamics of the vaginal mycobiome might provide insights into the treatment of infections or imbalances in the reproductive tract microbiota and improve reproductive health and obstetric outcomes (Bradford & Ravel, 2017). Nevertheless, the mycobiome is a challenging area to explore, and the research of fungi in gynecological samples, aside from *C. albicans*, has not been wide.

1.3 Studying the mycobiota

In the field of mycobiota research, conventional culture-based or microscopic methods have traditionally been important. However, in addition to their inaccuracy, as they only capture a fraction of the complexity of the species, there are challenges in isolation and growth in culture media, they are time-consuming and relatively insensitive, requiring specific expertise, hence leading to ambiguous results and contingency in diagnostics (OPATHY & Gabaldon, 2019; Guo *et al.,* 2012; Khot & Fredricks, 2009; Tiew *et al.,* 2020; Zoll *et.al.,* 2016). Hence, non-culture deoxyribonucleic acid (DNA) and molecular-based methods have evolved, including the detection of general fungal markers, such as β-glucan, and most importantly, the polymerase chain reaction (PCR) based methods (OPATHY & Gabaldon, 2019; Khot & Fredricks, 2009; Tiew *et al.,* 2020; Zhu *et al.,* 2020). Culture-independent techniques for microbiota characterization enabled large-scale microbiota research, which expanded after the development of high-throughput (HT) sequencing techniques (NIH HMP 2009; Zoll *et.al.,* 2016), including next-generation sequencing (NGS) methods, such as 454 pyrosequencing by Roche, Ion Torrent PGM by Thermo Fisher Scientific, PacBio RSII and Sequel by Pacific Biosciences, MinION, GrifION and PrometION by Oxford Nanopore Technologies, and Illumina MiSeq, HiSeq, and NovaSeq by Illumina (Nilsson *et al.,* 2019). The substantial amount of data achieved using these methods is analyzed with advanced bioinformatic tools, enabling further conclusions, such as taxonomic characterizations and microbiota profiling.

The NGS methods, used to study both bacteria and fungi, include metagenomics, i.e., whole genome shotgun sequencing (WGS), and targeted gene amplicon sequencing (Rowan-Nash *et al.,* 2019; Virtanen *et al.,* 2023; Zoll *et.al.,* 2016). The detection of fungi is challenging as the mycobiome varies extensively both taxonomically and in abundance in different parts of the body due to pH and other environmental factors, in addition to which the relative abundances of fungal species might be low, less than 0.1% of the overall microbial community (Zoll *et.al.,* 2016). Compared to WGS, the targeted amplicon sequencing provides enhancement, as it encounters the low abundant communities more efficiently (Zoll *et.al.,* 2016; Virtanen *et al.,* 2023). The microbiome research has concentrated on PCR-based HT gene amplicon sequencing of bacterial 16S ribosomal RNA (rRNA) gene and the non-coding fungal internal transcribed spacer (ITS) region of ribosomal DNA (rDNA) gene complex, both of which are hypervariable genetic regions, flanked by highly conserved procaryotic genetic regions. These universal marker genes enable the design and use of universal primers, such as 16S rRNA gene and ITS primers, capturing the desired genetic region in an amplification process, enabling the microbiome profiling of samples containing also non-microbial DNA, such as host DNA (Contijoch *et al.,* 2019). Therefore, genetic marker regions have been extensively used for molecular and phylogenetic evaluation and characterization of microbial communities and their diversity on a genetic level (Gill *et a*l., 2006; Rowan-Nash *et al.,* 2019; Virtanen *et al.,* 2023).

If economic questions and costs of DNA- and molecular-based methods are overlooked, the main challenges in targeted mycobiome amplicon sequencing are related to primer selection and sufficient coverage of species, optimal PCR protocol, and the risk of contamination, in addition to which, low fungal abundance compared to non-fungal DNA, such as host DNA or bacteriome, brings challenges in sequencing (Khot & Fredricks, 2009; Tiew *et al.,* 2020). The PCR amplification process might be disturbed by a variety of problems, such as contaminating DNA, amplification inhibition by co-extracted contaminants, formation of artefactual PCR products, and differential amplification of the target genomes (Wintzingerode *et al.,* 1997). Fungal analysis requires rigorous quality control processes in various stages of sample processing, due to the high risk of contamination and fungi being ubiquitous, resulting in increased false positive results (Khot & Fredricks, 2009; Tiew *et al.,* 2020). It has been issued that the results of mycobiome studies show significant variability of both encountered taxa and their relative abundances (Rowan-Nash *et al.,* 2019), and there might be significant variation even among samples from the same individual (Achkar & Fries, 2010). Variability might be

inherent, but also due to methodological aspects. Despite the interest of the scientific community towards mycobiome has increased lately, the history of fungal research is relatively short, lacking homogenous reproducible laboratory workflows and protocols, operative procedures, bioinformatic tools, reference databases, and validation analysis, resulting in difficulties in comparison of the results between studies (Ali *et al.,* 2019; Rowan-Nash *et al.,* 2019; Tiew *et al.,* 2020; Wiesmann *et al.*, 2022; Zoll *et.al.,* 2016). Standardized practices have a key role in adopting and translating the mycobiome investigation into clinical practice.

1.4 Quantitative PCR (qPCR)

A conventional PCR, invented in 1984 by Kary Mullis and coworkers (Kubista *et al.,* 2006), provides qualitative data by enumerating the targeted segment of DNA through an amplification process, and is a basis for a variety of methods in the research of genetic material and microbial communities, such as fungal diversity (Busser *et al.,* 2020; Kubista *et al.,* 2006; OPATHY & Gabaldon, 2019). In traditional end-point PCR, the amplification products are detected by flow-based downstream analysis methods, such as agarose gel electrophoresis. An advancement, a real-time PCR (RT-PCR) or quantitative PCR (qPCR) is a PCR-based method that simultaneously allows the amplification and enumeration of the PCR product. After the invention of PCR, several intentions for qPCR were developed (Heid *et al.,* 1996), until a kinetic real-time monitoring PCR analysis mode, utilizing video clips from the amplification process, was invented in 1993 by Higuchi and coworkers (Higuch *et al.,* 1993). The method was further elaborated in 1996 by Heid and coworkers into an RT-qPCR method with no post-PCR manipulation. Since, it has been a gold standard for molecular diagnostics, such as nucleic acid quantification, detection of pathogens, gene expression analysis, single nucleotide polymorphism analysis, chromosome aberration analysis, and protein detection (Kubista *et al.,* 2006; OPATHY & Gabaldon, 2019; Wilhelm & Pingoud 2003; Zhu *et al.,* 2020).

Some of the benefits of the HT qPCR approach are sensitivity, accuracy, reproducibility, cost-effectiveness, feasibility, and simplicity, as the demands for performing the method are relatively easily accessible. Nevertheless, there are challenges associated with qPCR, including the non-specific binding of fluorophore dyes resulting in problems, such as template-independent duplex dimerization of primers, forming artefactual non-specific primer-dimers (PD), and cross-reactivity with non-targeted DNA (OPATHY & Gabaldon,

2019; Rychlik, 1995, Khot & Fredricks, 2009), or the low abundance of samples, and the risk of contamination (Khot & Fredricks, 2009; Zoll *et.al.,* 2016). For fungal qPCR, there are various approaches targeting different genomic regions or genes, such as 18S rRNA or ITS, complicating the comparability of the results (Wiesmann *et al.*, 2022). ITS is considered to be one of the most promising regions to target fungi (Busser *et al.,* 2020), and common genomic ITS -regions in mycobiome research are ITS1, ITS2, and ITS3 (Rowan-Nash, *al.,* 2019; Wiesmann *et al.*, 2022). ITS1 -region primers have been claimed to detect the most fungal species and genera among human fecal samples, in which they detected members of *Candida, Saccharomyces*, *Debaryomyces,* and *Penicillium*, among other genera (Wiesmann *et al.*, 2022). Nevertheless, ITS2 -region targeting primers are suggested to reach higher accuracy, which is associated with lower variability in size, taxonomic biases, misidentification, and misclassification of fungal species (Ali *et al.,* 2019). In addition to distinct coverages, primers might have variable amplification efficiencies on microbial taxa, leading to differential amplification (Wintzingerode *et al.,* 1997). Indeed, the selection of primers has a major impact on abundances and differences among taxa, as the outcomes of fungal community in a sample might vary significantly depending on the targeted region (Rychlik, 1995; Wiesmann *et al.*, 2022; Wintzingerode *et al.,* 1997).

1.4.1 qPCR assay

Two main results are provided by qPCR: an amplification chart with an amplification threshold for each trace, and a melting profile of an amplification product. The qPCR is based on a fluorescent signal from a fluorescent probe or a dye binding on the double-stranded DNA (dsDNA) (Figure 1) (Heid *et al.,* 1996; Higuchi *et al.,* 1993; Khan *et al.,* 2011; Rinttilä, 2011). The quantification of the target sequence is performed by determining the actual quantity of target DNA amplicons, which are labeled with a fluorescent reporter, of which the most known is SYBR Green (Khan *et al.,* 2011). The measurement of the fluorescent signal at a given cycle enables the real-time quantification of target DNA amplicon copies (Higuchi *et al.,* 1993; Rinttilä, 2011; Thermo Scientific Chemicals, 2024). The standard curve and amplification threshold of each fluorophore enable the determination of the amplification thresholds of unknown samples, and further the estimations of their copy-numbers (CN) representing the number of genomes present, which can be used to estimate their abundances (Bio-Rad, 2006; Heid *et al.,* 1996).

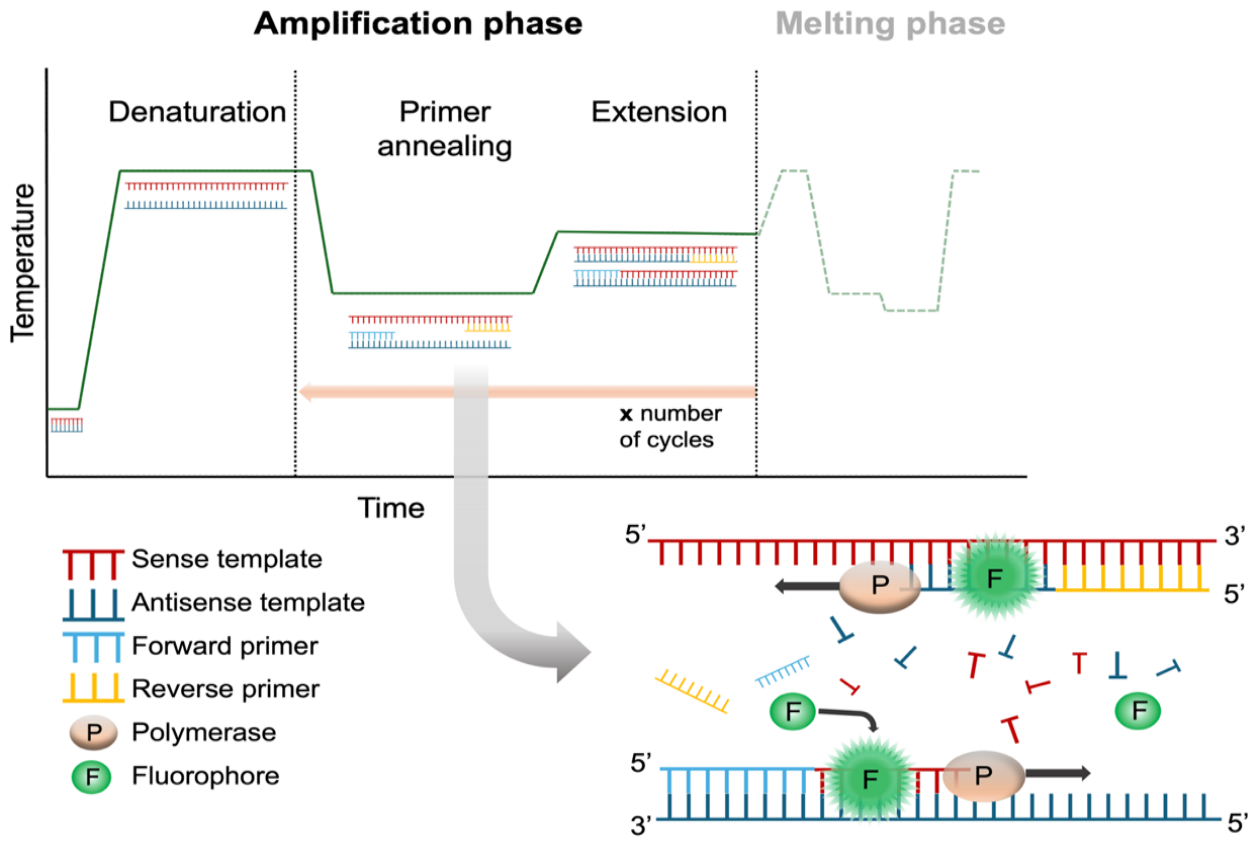

Figure 1. Function of qPCR.
The newly forming antisense strand is prepared according to the sense template, and the new sense strand according to the antisense template. Nucleic acid stains bind the qPCR product, the newly formed dsDNA, resulting in an increase in light emission observed as a signal, which increases in direct proportion to the amplification product (Rinttilä, 2011). The fluorescent signal is monitored to provide information on the amplification process of each template (AAT Bioquest, 2024).

The qPCR assay, adjusted according to the template and the reagents, begins with an initial denaturation and polymerase activation by incubation at high temperature to prevent PD extension and non-specifically annealed primer elongation, which might begin to appear in low temperatures while setting up qPCR (Solis BioDyne). The initial activation step is followed by multiple cycles of the actual polymerization and the amplification of the targeted product, including a short denaturation at high temperature, an annealing step at a given temperature depending on the melting temperature (T_m) of selected primers, and an extension step at the optimal temperature for the polymerase (72°C) (Kubista *et al.,* 2006). The amplification phase is followed by a melting phase, including an equalizing incubation step at a primer-specific temperature to detect the fluorescent data, and the melt curve analysis (MCA), which is performed to ensure the specificity of the amplification products (CFX Maestro, 2023; Kubista *et al.,* 2006, Solis BioDyne).

Reagents required for a qPCR assay include forward and reverse-facing primers, and a buffer with essential components, such as polymerase, magnesium chloride ($MgCl_2$),

deoxyribonucleotide triphosphates (dNTP), and fluorescent dye (Kubista *et al.*, 2006), which are often included in a ready-to-use solution, a Master Mix. A commonly used PCR enhancer, dimethyl sulfoxide (DMSO), may be used in qPCR, especially with GC-rich primers (Chakrabarti & Schutt, 2001). It is a ready-to-use bioreagent, an organosulfur compound (C_2H_6OS), and a polar aprotic solvent that dissolves compounds of different polarity, used to disrupt the formation of secondary structures of the template, enhancing the primer annealing (Thermo Scientific Chemicals, 2024).

1.4.2 Amplification chart

An amplification chart with a quantification cycle (C_q) or threshold cycle (C_T), for each trace is the main result of a qPCR assay (Figure 2). The fluorescence of a fluorophore increases as it interacts with a constantly multiplying dsDNA amplified in the PCR, and the C_q -value represents the quantification cycle upon which a product reaches the amplification threshold, which is the detection limit of the fluorescent signal (Heid et al., 1996). The amplification curve consists of three phases: an initiation phase of moderate amplification below the detection limit, linear to the ground baseline; an exponential growth phase as the amplification threshold is reached and the amplification is constant; and a plateau phase, in which the amplification has reached the maximum yield and reduces due to limiting reagents, and the reaction is decreased until the endpoint of the reaction is eventually reached (Bio-Rad, 2006; Heid et al., 1996; Kubista et al., 2006).

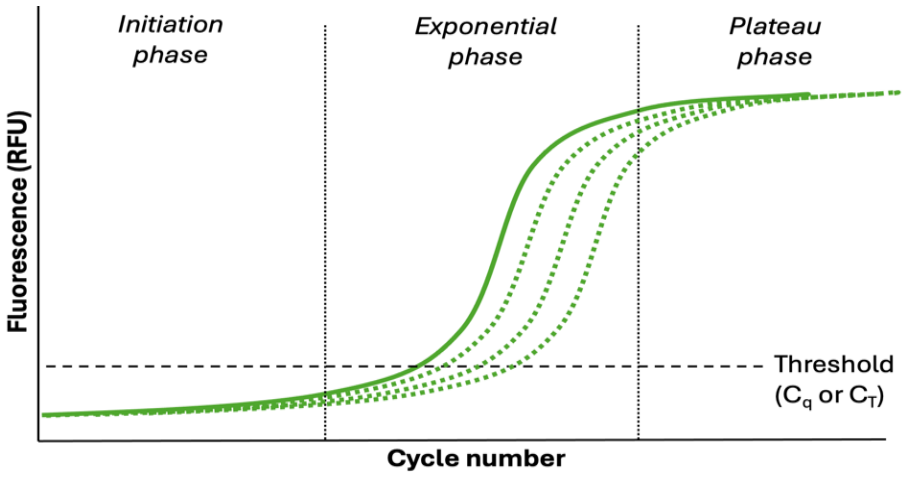

Figure 2. Amplification chart of qPCR.
The amplification chart provides relative fluorescence units (RFU) for each well at every thermal cycle of a qPCR. The software (CFX Maestro, Bio-Rad) calculates a baseline and an amplification threshold for each fluorophore. The baseline is constructed by fitting the best straight line through the fluorescence of each well, and subtracting the best fit data from the background at each cycle. (Bio-Rad, 2006)

The baseline for all fluorescence tracts is a requirement for the construction of a standard curve and the determination of a C_q. A baseline-subtracted curve fit method leaves each C_q as invariant as it uses the centered mean filter to smooth the baseline curve. Baseline-subtracted analysis mode is also used for end-point analysis, in which the relative fluorescence units (RFU) are compared to the negative controls to provide the average RFU values of the last thermal cycle of the PCR. End-point analysis can be used for qualitative analysis, i.e., to define whether the sequence of interest is present or absent in a sample (CFX Maestro, 2023; Heid *et al.,* 1996).

A known PCR product, a standard, may be used to obtain relative quantifications from an unknown target, assuming they amplify with the same efficiency (Heid *et al.,* 1996). A standard curve is generated from the starting quantity (SQ), derived from the C_q -values of the template with known concentrations, and despite it is not required for qPCR, it enables the between-plate normalization, which is pivotal in studies with larger sample sets requiring several parallel qPCR runs (Bio-Rad, 2006; CFX Maestro, 2023). Through interpolation, a standard curve provides known quantities of particles, such as unit mass or CN, which allows the absolute quantity calculation of unknown samples. Furthermore, it is used to assess the performance of the qPCR, as the reaction amplification efficiency (E%) is measured for each fluorophore of standard curve wells (90–105%), and indicates the robustness and reproducibility of an assay, whereas the coefficient of determination (R^2) represents the goodness of fit, which describes how well the standard curve line describes the experimental data (>0.980), and ultimately, whether the qPCR assay is optimized (Bio-Rad, 2006; CFX Maestro, 2023). The optimization of the efficiency of a qPCR is affected by several parameters, such as the reaction conditions (time, temperature), reagents (primers, magnesium, and salt concentration), and template (size, composition, purity) (Heid *et al.,* 1996).

1.4.3 Melt curve analysis

The melting profile of a product is another main result of a qPCR, providing a melt curve, and a melt peak at a specific T_m, at which 50% of the dsDNA has dissociated (Bio-Rad, 2006; Erali, Voelkerding & Wittwer, 2009) (Figure 3). MCA is performed to verify the specificity of an amplified product and is based on the denaturation of the product, resulting in a constant decrease of fluorescence of the dyed dsDNA (Erali *et al.,* 2009).

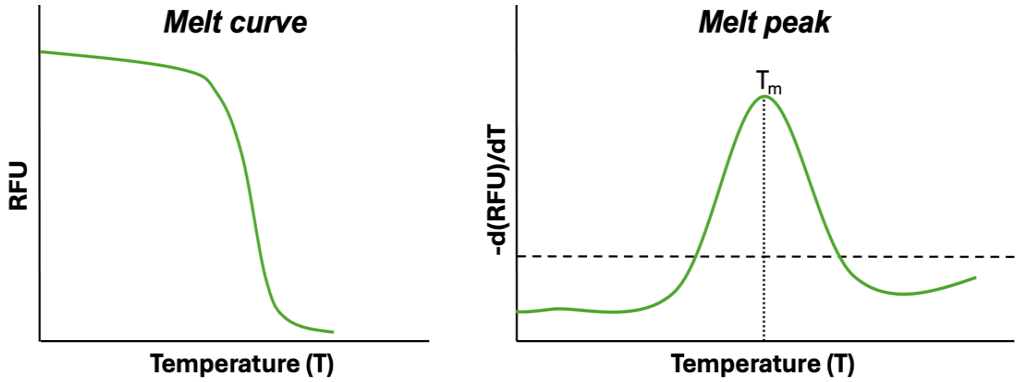

Figure 3. Melting chart of qPCR.
The melting profile is achieved by analyzing the amplification data, i.e., the RFU values collected during the melting phase of the qPCR, as a function of temperature (T). The melt curve demonstrates the real-time relative fluorescence of each fluorophore unit per temperature (RFU/T). The maximum fluorescence is achieved when the two DNA strands are annealed, and decreases gradually as the temperature rises in small increments towards the T_m, eventually dissociating the dsDNA and releasing the fluorophore (Bio-Rad, 2006; CFX Maestro, 2023; Erali et al., 2009).
The melt peak is achieved by plotting a negative first derivative of fluorescence versus the temperature (-d(RFU)/dT), which illustrates the gradually decreasing fluorescence of dsDNA, resulting in the constant slope of a melt curve, followed by a quick change in slope and drastic reduction of the fluorescence as T_m is reached (Bio-Rad, 2006; CFX Maestro, 2023; Erali et al., 2009).

The length and the GC -content of a molecule determine the place of the melt peak, as the cleavage of a long strand of DNA demands a higher temperature (Erali *et al.,* 2009; Rychlik, 1995). Respectively, the duplex stability of a G-C base pairing with three hydrogen bonds is higher, requiring higher dissociation energy than an A-T base pairing with two hydrogen bonds (Rychlik, 1995). MCA is used to identify different products of the qPCR, as the melt peak at the specific T_m is characteristic for each organism, enabling the distinguishing of the target product from others, including nonspecific products, such as PDs (Bio-Rad, 2006), but also characterization of different organisms, such as bacterial pathogens (Huc*et al.,* 2020) or fungal taxa (Thian Lung *et al.,* 2014; Nejad *et al.,* 2020), emphasizing the importance of MCA coupled with amplification results.

1.5 Quantitative NGS (qNGS)

The quantitative NGS (qNGS) approach is an integration of the relative microbial profiling method and the quantitative abundance analysis method, thus making it a quantitative microbiome profiling method (Jian, Salonen, Korpela, 2021). It merges cell- or molecular-based quantitative profiling methods, such as flow cytometry or qPCR, and NGS-based amplicon sequencing by combining the number of reads from sequencing data, and the total counts of the target microbes in the same samples achieved from the quantitative method, thus enabling the estimation of absolute abundances of each NGS-detected microbial taxa (Jian *et al.,* 2020) (Figure 4).

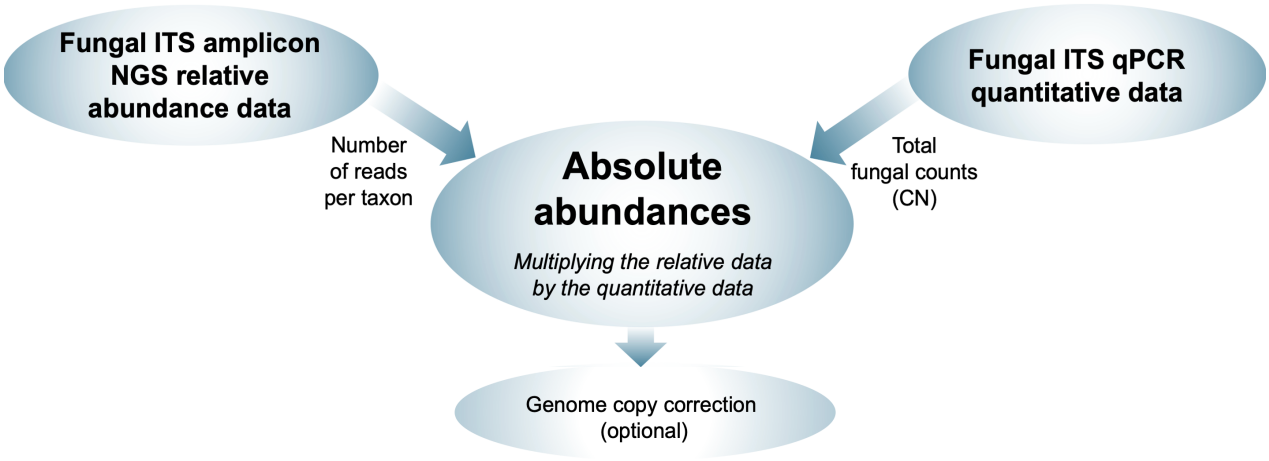

Figure 4. Workflow of the qNGS approach.
Relative abundance data and quantitative data are prerequisites for the estimation of absolute abundances of all detected taxa. An optional correction of genome copies may be done computationally by dividing the absolute abundances by the CN of each NGS-detected taxa (Jian et al., 2020).

When universal primers, targeting all bacteria or fungi in the samples, are used both for qPCR and NGS, the two methods measure the same target organisms, whereas cell-based flow cytometry measures intact cells, some of which might not be captured by the NGS, resulting difficulties in comparability of quantitative and relative abundances (Jian *et al.* 2021). Therefore, qPCR is suggested to give more comparable quantitative microbiome profiling results (Jian *et al.,* 2020). Furthermore, to avoid potential biases, primers used in the qPCR and NGS should optimally target the same area (Jian *et al.,* 2020), such as the ITS1 or ITS2 -region of the fungal genome.

The qNGS was developed to overcome the compositional nature of the normal NGS data, as in the relative abundance NGS data, the shares of different taxa always sum up to 100%. As each change of a given taxon affects all others as well, there is a mutual dependence: if one abundance is increased, others are decreased concurrently. The compositionality of taxonomy may lead to misinterpretations of the community, as the rate of false positive discovery increases in classic statistical methods (Jian *et al.,* 2020). The effect of compositionality is recognized to be extensive both in complex microbial communities, such as intestinal microbiota, but also in communities with low diversity, such as vaginal swab samples (Jian *et al.,* 2020). The qNGS approach eliminates the compositionality of the NGS data, as it provides absolute abundancies of each sample, varying between 0–100%, and the amounts of different taxa in a sample are independent, and not directly affected by the other taxa encountered. Hence, it offers a comprehensive understanding of the microbiome, including interactions and dynamics between the taxa (Jian *et al.,* 2020).

Nevertheless, the qNGS approach does not provide an exact number of cells, but an approximation of the genomic copies and the microbial density, due to copy-number variation (CNV) among different microbial taxa (Jian *et al.,* 2020). The rDNA CNV among different microbial taxa has a major impact on organism function, yet the essence of variation among fungi is not fully known (Lofgren, Uehling, Branco *et al.,* 2019). As there are substantial CNV in fungal rDNA even among closely related taxa, biases in estimating the abundancies in HT-sequencing ITS of rDNA may lead strongly to over- or underestimating actual fungal abundancies (Lofgren *et al.,* 2019). Compared to other microbial groups, such as bacteria or archaea, the total CNV of fungi is greater, although the level of CNV is similar. Regardless there are considerable ranges even between the same fungal species, the estimated CNV of fungal rDNA, thus also ITS CNV, has been suggested to fall between prokaryotes and larger eukaryotes, ranging from tens to a couple of hundreds rDNA (Lofgren *et al.,* 2019). Nevertheless, there is no consensus on the average CN of ITS (Dannemiller, Lang-Yona, Yamamoto *et al.,* 2014).

Previously, the qNGS approach has been used to study bacterial microbiota and environmental fungi. Jian *et al.* (2020) have performed the qNGS approach on human bacterial microbiota, revealing that the estimated absolute abundances and qPCR abundances among bacteria correlated extremely well. Furthermore, Dannemiller *et al.,* (2014) achieved absolute concentration data on environmental fungi by performing a comparison between taxon-specific concentrations achieved from qPCR, and relative abundance data achieved from pyrosequencing. However, the qNGS approach has not been used on human mycobiota.

2 Aims of the study

The quantitative next-generation sequencing (qNGS) approach combines the quantitative polymerase chain reaction (qPCR) and the next-generation sequencing (NGS), enabling the determination of the absolute abundances of microbial taxa. Here, the mycobiota has been investigated by optimizing fungal qPCR on gynecological samples of patients suffering from recurrent pregnancy loss (RPL).

This study had three aims: the first aim was to optimize qPCR for fungi to reliably investigate their absolute amounts in vaginal and endometrial samples derived from women suffering from RPL, and their controls. The second aim was to couple the qPCR data with the previously generated NGS data on fungi, and compare the qPCR and NGS results. The third aim was to investigate if the fungal positivity and quantities were related in the two sample types from the same patient, and if there were differences between RPL-patients and their controls.

Genomic-level knowledge of the microbiota provides possibilities to better understand what influence it has on human health and disease (NIH HMP 2009; Rowan-Nash *et al.,* 2019). Further, combining the mycobiome with the host genetic background might enlighten the understanding of diseases, providing insight into the development of personalized medicine (Zoll *et.al.,* 2016). Regarding the microbiome of the female reproductive tract, mycobiome research might bring new aspects to enhance female reproductive health, and provide cost-effectiveness and robustness to the study of mycobiota as a part of the human microbiota.

3 Materials and methods

3.1 Methods

The methods used in the thesis presented as a flow chart (Figure 5), focus on optimizing qPCR for all fungi and additionally for *C. albicans*. A significant part of the methods included systematic searching of errors, i.e., troubleshooting of the process, as problems were encountered during the optimization work. All the materials, reagents, primers, templates, and machinery used in this thesis are presented in detail in the appendix (Appendix 1: Materials and equipment).

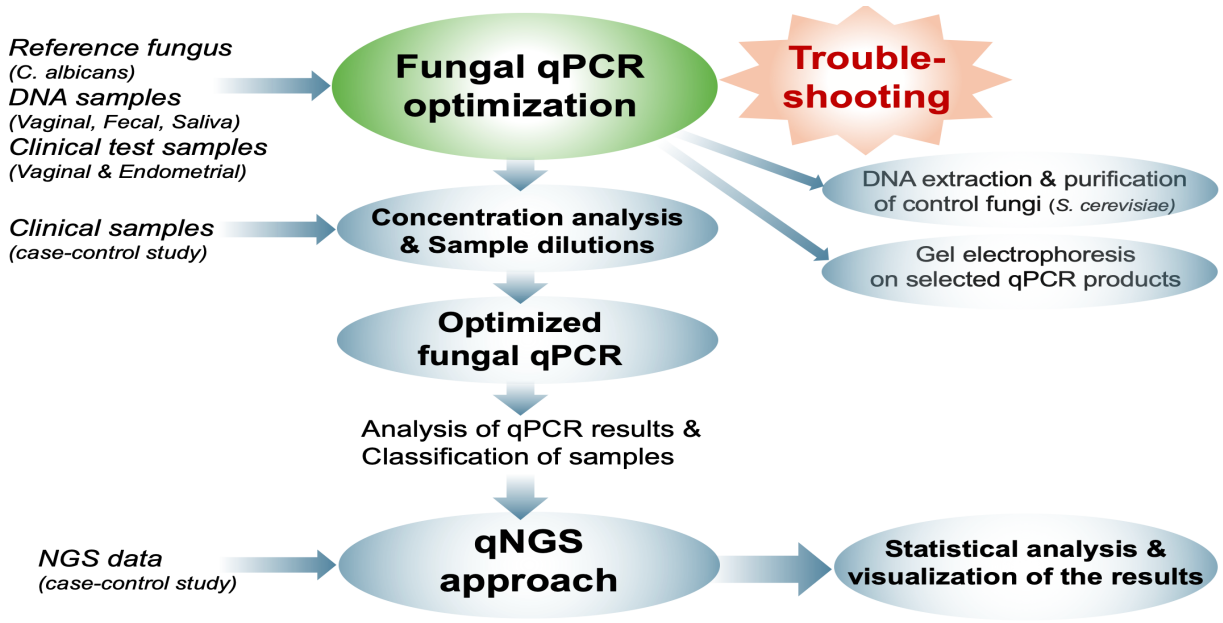

Figure 5. Flow chart of the thesis work.
The fungal qPCR optimization aimed at the qNGS approach, which provides the absolute abundances of the fungi, troubleshooting being a fundamental part of the optimization work. The optimization was performed on DNA samples and was tested on control fungi before continuing to the concentration analysis of the clinical samples with NGS data.

Special attention was given to the general working practices throughout the process. All pipetting and other laboratory work was executed in a laminar flow cabinet (Biowizard Platinum, Kojair ® Tech Oy, Finland). All equipment and materials were cleaned with 70% alcohol, followed by 30-minute UV-cleaning. The entire processing of the reagents and samples was performed on ice (+4° C) to prevent DNA degradation and any premature reactions between the reaction components. To minimize the risk of contamination and degradation, materials were divided into smaller aliquots, including MQ-water, primers, biological samples, and reference fungus cell pellet DNA extracts. Unnecessary opening and freeze-thaw cycles were avoided. In the final qPCR runs for the clinical samples, all reactions were performed using freshly made working reagents from the same stock.

Precautions were taken to ensure the sample replicates were identical in terms of the volume and homogenous distribution of the DNA. The pipettes were dedicated for qPCR work, and recently cleaned and calibrated. The pipette tips (Thermo Fisher Scientific, UK) were filtered, unopened boxes packed by the manufacturer. Standard pipetting technique, i.e., forward pipetting, was committed to sample pipetting due to the limited amount of sample and the elevated risk of over-delivery of the sample, associated with the reverse pipetting method. The pipetting was committed with a multichannel pipette, and with a pipetting technique, following the guidelines '10 steps to improve pipetting accuracy' by Thermo Fisher Scientific. Despite the temperature potentially affecting the pipetting volume through thermal expansion and shrinking, depending on the relative humidity and vapor pressure of the liquid, all pipetting was committed at room temperature.

3.2 Materials and sample preparation

The clinical samples used with the optimized qPCR protocol were derived from the TOIVE study, which is a prospective case-control study (Peuranpää *et al.* 2022), investigating if the composition of endometrial or vaginal microbiota is associated with RPL. The vaginal swab and endometrial pipelle samples were gathered by a gynecologist during diagnostic or treatment procedures. The inclusion and exclusion criteria as well as other characteristics of the RPL patients and controls are found in the original publication (Peuranpää *et al.*, 2022). In this thesis, the total study cohort (N=170) consisted of vaginal swabs and endometrial pipelle samples from RPL-patients (n=46) and from a control group with no history of pregnancy loss, endometriosis, anovulation, or Fallopian tube defects (n=39). The pure culture of *C. albicans* was used as the main reference organism. The samples used for the optimization of the fungal qPCR protocol were previously extracted DNA from vaginal and endometrium samples, and additionally from feces and saliva (Appendix 1). *Saccharomyces cerevisiae* was used as control fungi in qPCR optimization. The final customization of the qPCR protocol was performed on six vaginal swabs and endometrial samples from the study by Peuranpää *et al.,* (2022).

DNA extraction and purification of the clinical samples were performed in November and December of 2020. The process involved mechanical lysis and paramagnetic beads (Virtanen *et al.,* 2019). Samples were prepared for Illumina MiSeq paired-end sequencing of the bacterial 16S rRNA gene and fungal ITS1 -region amplicons. Further information about sample processing is described in the article by Peuranpää *et al.* (2022). Upon

thawing for qPCR analysis, 39 DNA samples (19 vaginal, 20 endometrial) had evaporated almost completely, and were resuscitated by adding 30 µl of ultra-clean molecular biology water (MQ-water) and incubated over two nights at +4° C. DNA extraction and purification of the additional biological test samples used in the qPCR optimization were performed by using a similar protocol used for the clinical samples, involving mechanical lysis and paramagnetic beads to extract fungal DNA, ensuring the maximum release of the cellular contents and DNA liberation (Fredricks, Smith & Meier, 2005; Zhang *et al.,* 2021). The work was performed either manually or automatically by using a beat beating method with Fast Prep 96™ homogenizer instrument (MP Biomechanicals, USA), Vortemp 56-incubator (Labnet International, USA), and a KingFisher Flex purification automate (Thermo Fisher Scientific, Finland) as described by Virtanen *et al.* (2019).

DNA content of the clinical samples was measured with HT Quant-iT assays in a 96-well plate, using dsDNA reagent and kits (Quant-iT ™ PicoGreen® dsDNA Assay Kit, Invitrogen by Thermo Fisher Scientific) and Hidex Sense microplate reader (Brother HL-1212W, Finland). The concentration of the *S. cerevisiae* DNA extracted and purified from the brewer's yeast stock was measured with the Qubit method (Qubit™ dsDNA HS Assay Kit and Qubit 4 fluorometer, Invitrogen by Thermo Fisher Scientific, Singapore). Both Qubit and Quant-iT are fluorescence-based methods analyzing the concentration of the target molecule with the same technology, while the throughput varies, with Quant-iT assay being an HT method, whereas Qubit is performed one sample at a time.

Dilutions of 1:10 (10%) and a further 1:100 (1%) to MQ-water were prepared from the original clinical samples. The sample plates were carefully mixed by vortexing, and briefly centrifuged (140 rcf, 15 sec), before pipetting. The dilutions were re-measured with Quant-iT, and additionally with Qubit, to verify some of the results to ensure that the DNA concentrations between different dilutions of the same sample were correlating. The template volume in fungal qPCR was a constant 5.0 µl (20%) per reaction. To obtain triplicates of each clinical sample, a minimum of 16.0 µl total volume of the final sample concentration was prepared by adding either MQ-water or the original sample to the dilution plates to achieve the desired DNA concentration for the final qPCR reactions.

The reagents and buffers (Appendix 1) were either ready-to-use products from a manufacturer, or autoclaved buffers prepared in the laboratory by the author. The qPCR reagents were primers (Table 1), DMSO (Thermo Fisher Scientific), and ultra-clean MQ-

water (AccuGENE®, Lonza). All primers were obtained straight from a manufacturer (Sigma-Aldrich, Merck) as lyophilized stocks, except for ITS2 -region primers (ITS3/ITS4), which were readymade 50 μM dilutions borrowed from another laboratory (Ventin-Holmberg, R., Folkhälsan). Lyophilized primers were diluted into 1 x TE-buffer to obtain 100 μM dilutions (Sigma-Aldrich, Merck), which were further diluted with MQ-water to obtain 50 μM and 10 μM working dilutions.

Table 1. Primers used in the qPCR.

Reagents:	**Origin**
ITS1: ITS1F forward *5' CTTGGTCATTTAGAGGAAGTAA 3'*	Sigma-Aldrich, Merck stock 1: HA16117389, stock 2: HA16468144 Reference: Gardes & Bruns (1993)
ITS1: ITS2 reverse *5' GCTGCGTTCTTCATCGATGC 3'*	Sigma-Aldrich, Merck stock 1: HA16117388, stock 2: HA16468143 Reference: White, Bruns, Lee, Taylor (1990)
ITS2: ITS3 forward (50μM) *5' GCATCGATGAAGAACGCAGC 3'*	Ventin-Holmberg, R., Folkhälsan Reference: White, Bruns, Lee, Taylor (1990)
ITS2: ITS4 reverse (50μM) *5' TCCTCCGCTTATTGATATGC 3'*	Ventin-Holmberg, R., Folkhälsan Reference: White, Bruns, Lee, Taylor (1990)
***C. albicans* forward** *5' GGGTTTGCTTGAAAGACGGTA 3'*	Sigma-Aldrich, Merck. HA16117386 Reference: Guiver, Levi, Oppenheim (2001)
***C. albicans* reverse** *5' TTGAAGATATACGTGGTGGACGTTA 3'*	Sigma-Aldrich, Merck. HA16117387 Reference: Guiver, Levi, Oppenheim (2001)

3.3 Troubleshooting

A systematic search and correcting of errors were elaborated as problems arose during the optimization, including sudden alteration in both amplification and melting profiles of samples. Compared to previous findings from the same samples with the same parameters, a significant reduction in amplification efficiencies and an elevation of C_q-values were encountered, as well as unexpected melting profiles, and reduced or uncoherent separation between template concentrations among standard series, positive controls, and other sample types. Furthermore, negative controls demonstrated different amplification and melting profiles than before, including a reduction of C_q -values. Notably, all these parameters were originally well-functioning, resulting in expectable amplification and MCA of each sample type, thus suggesting sudden malfunctioning in some or multiple areas of the workflow. The troubleshooting matrix for systematic correcting of errors (Table 2) incorporates a profound investigation of all the elements of the process, suggesting a potential problem and a proposal to solve it. The matrix is divided into four

stages representing both the chronological order and the severity of the precautions: basic, intermediate, and intense action, and the detection of the potential source of error.

Table 2. The troubleshooting matrix for systematic identification and correcting of errors.

	Potential problem	Proposal to solve the problem
1	**Pipetting error**	Re-run of the samples with the same parameters
	Degradation and/or contamination	New aliquots of reagents (10 mM primers); New aliquots of template (controls, standard)
	Primer failure	Cross-examination and comparison of the problematic primers with different primer types (e.g., ITS1 vs. ITS2)
2	**Working practices and contamination**	Special attention to sterility; Ultra-cleaning of the laminar flow cabinet with chlorine; Profound cleaning of the pipettes; New boxes of pipette tips, tubes, strips, etc.
	Degradation and/or contamination	New aliquots of reagents: 50 mM primers, qPCR Mix, MQ-water; New aliquots of template (samples, controls, standard)
	Machinery failure	Comparison of two different qPCR machines; Ensuring the function of the fridge and freezer (24h T measurements).
	Amplification failure	Ensure qPCR products with gel electrophoresis
3	**Working practices**	Special attention to working order and protocols; Pausing the laboratory working
	Environmental failure	Another room and a new laminar flow cabinet; No cross-examination or comparison of old materials or samples
	Material failure	Calibration and cleaning of pipettes; Replacement of pipette tips, tubes, strips, etc. into new sterile
	Reagent failure	New bottle of ultra-clean clean MQ-water; Preparation of fresh buffers; New stocks of lyophilized primers (diluted into fresh buffers); New sterile tube of DMSO; New tube of qPCR mix (another LOT)
	Template failure	New pure culture of the standard; New negative control; New clinical test samples from the original sample plate; Preparation of new sample dilutions and concentration measurement
4	**Detecting of the potential source of error**	Comparison of the old and new primers
		Comparison of the old and new DMSO
		Comparison of the old and new negative control
		Comparison of the old and new pure culture of standard
		Comparison of the old and new test samples and dilutions

The first stage of troubleshooting involved basic error detection, starting from the most possible sources of problems and double-checking of the results. Further precautions were taken, concentrating on the possible degradation and contamination of materials, and cross-examination of the variables with identical conditions. The second stage concentrated on working practices, risks of contamination, and the environment. The third and most resource-consuming and demanding stage was constructed from the complete replacement of all the materials, and of the fundamental assessment of the laboratory practices, methods, and protocols. The purpose of the last steps was to evaluate the possible source of error afterwards, when the protocol was already working well. The machinery, pipettes, and environment were not included in the fourth phase due to the high risk of possible contamination.

As a part of the troubleshooting procedure, some of the qPCR products from vaginal swabs and endometrial test samples were visualized on electrophoresis in an agarose gel to ensure successful polymerization. The qPCR product was not purified from reagents but used as such instead. Electrophoresis of amplification products was implemented with Power Pac 200 (Bio-Rad, USA) in a thick 2.5% agarose gel (Agarose, Molecular grade Bioline), suitable for small products such as ITS PCR products, with 5.0 µl of Midori green dye (Advance DNA stain Midori Green, Nippon Genetics Europe GmbH by BIOTOP), and 5.0 µl of FastRuler ™ low range (50-1500 bp) ready-to-use DNA ladder (FastRuler ™ Low Range DNA Ladder, ready-to-use, Thermo Fisher Scientific). Gels were loaded with 4.0 µl of impurified amplification product together with 1.0µl of loading buffer (6x DNA Loading Dye, Thermo Scientific, Thermo Fisher). Running time was 15 min on 120 V and 2.0 A followed by 40 min on 100 V and 2.0 A. Gels were visualized on a Gel Doc™ XR+ UV transilluminator (Bio-Rad, USA).

3.4 Fungal qPCR optimization

The fungal DNA in samples was amplified and quantified with qPCR, in 96-format. The quantification was carried out using a Bio-Rad real-time thermal cycler (Bio-Rad CFX96™ Optics Module Real-Time System C1000 Touch™ Thermal Cycler, Singapore) with Bio-Rad CFX Maestro 1.1–4.1.2433.1219 software, and with HOT FIREPol® EvaGreen® qPCR Mix Plus (no ROX) 5x (Solis BioDyne, Estonia) as a fluorophore reagent, which is an optimized ready-to-use solution for qPCR assays and an analogy for SYBR Green (Khan *et*

al., 2011). The plates were 96-well semi-skirted clear PCR plates with standard profile (4titude® Biotop, UK).

The workflow of each of the qPCR runs was executed in identical order (Table 3). Between-plate normalization was obtained by identical standard curves produced from *C. albicans* pure culture, as well as the positive control (5.0 ng), which originated from a different stock of *C. albicans* than the standard series. MQ-water (5.0 µl) and human blood DNA (5.0 ng) were used as negative non-template control. The optimized fungal qPCR protocol was applied on clinical sample qPCR runs (eight plates), made within one week.

Table 3. Workflow of the qPCR.

	Working step
1	Preparation of the Master Mix in the Eppendorf tube (excluding the qPCR Mix)
2	Pipetting of MQ-water onto the plate for standard series
3	Pipetting of negative controls onto the plate: 5.0µl of MQ-water & human DNA (1.0ng/µl)
4	Pipetting of positive controls onto the plate: 5.0µl of C. albicans (1.0ng/µl)
5	Pipetting of the samples onto the plate: 5.0µl of each sample (2.0ng/µl)
6	Preparation of the standard series onto a strip
7	Pipetting of the standard series onto the plate: 1.0µl of C. albicans (10.00–0.001ng/µl)
8	Adding the qPCR reagent to the Master Mix prepared earlier
9	Pipetting of the Master Mix onto the plate (20.0µl/well)
10	Sealing of the plate with a qPCR-seal cover
11	Vortexing and spinning the plate (140rcf, 1 min)
12	Putting the plate into the qPCR machine & starting the run

The fungal qPCR optimization was committed by cross-examination of different variables, which consisted of qPCR cycling conditions, Master Mix elements, and template elements (Table 4). The optimization focused on *C. albicans* and gynecological samples, whereas other template types were used as supportive elements of the optimization. The work commenced by testing the *C. albicans* pure culture with different combinations of reagents and continued by elaborating the optimization by adding variables. The main interest of this thesis was in primer type and primer volume, template type, and template amount, which were further examined with DMSO and different qPCR conditions.

Table 4. Fungal qPCR optimization matrix.

Cycling conditions	Master Mix elements			Template elements	
Annealing T (°C)	***Primer type (region)***	***Primer c μM (μl)***	***DMSO μl (%)***	***Type (DNA)***	***Amount (ng)***
52 to 62	ITS1 ITS2	0.8 (1.0) 2.0 (2.5)	none (0) 0.50 (2.0) 0.60 (2.4) 0.75 (3.0)	*C. albicans* *(S. cerevisiae)* (Fecal) (Saliva) Vaginal Endometrial	0.001 0.01 0.10 1.00 2.50 5.00 10.00 25.00 50.00

General qPCR cycling conditions, such as the initial qPCR cycling protocol, as well as the scan mode for a reporter dye, fluorescein amidites (FAM), were elaborated upon the guidelines of the qPCR Master Mix reagent EvaGreen® (Solis BioDyne, Estonia). The annealing temperature was tested for each concentration of primers and DMSO, but not for all template types and amounts due to the limited amount of sample material. The optimal temperature for each tested variable was obtained by adding an annealing temperature gradient (Table 4). The melting conditions were set to match the amplification conditions, and the increment modifications were selected to give the maximum specificity of the melting conditions of each sample.

The examined elements of qPCR Master Mix included primer type and concentration, and DMSO. Two universal fungal ITS-primers, ITS1- and ITS2 -region coding primers (Table 1), with two different amounts: 0.8 μM (4%) and 2.0 μM (10%), corresponding to 1.00 μl and 2.50 μl of 10 μM forward and reverse primer dilution per reaction. ITS1- and ITS2 -region coding primers were tested on a *C. albicans* template, biological test samples, and *S. cerevisiae*. In addition, *C. albicans* -specific primers were used, not in the actual optimization of fungal qPCR but to verify the efficient polymerization of the *C. albicans* stock used as a standard. DMSO was examined by ranging typically used concentrations per reaction: none (0%); 0.50 μl (2.0%); 0.60 μl (2.4%) and 0.75 μl (3.0%). As template volume and qPCR Mix reagent constituted the constant 10.0 μl of the 25.0μl total volume, the remaining volume was achieved by adding MQ-water a total volume of 9.25–13.0 μl, according to the volume of primers and DMSO in the reaction.

The template elements consisted of the template type and amount. The template volume was a constant 5.0 µl (20%) per reaction throughout the examinations, but the template amount varied from 0.001 ng up to 50.00 ng per reaction (Table 4). Template types were two separate pure cultures of *C. albicans* and different types of test samples, and extracted DNA from fecal, saliva, vaginal, and endometrial samples, of which the focus was on gynecological samples. Spiking of additional biological test samples with the known positive template was used to verify the functioning of the reaction, by adding 1.0 ng of *C. albicans* pure culture to 4.0 ng of vaginal and fecal template. Furthermore, *S. cerevisiae* was studied, to ensure the functioning of the DNA extraction and purification protocol.

3.5 Fungal qPCR quality and data analysis

The optimized fungal qPCR protocol for gynecological samples was performed on clinical samples. The analysis and selection of the fungal qPCR data were performed visually and statistically by analyzing standard deviations and end-point values of the amplification C_q -values and MCA of each sample replicate. The fungal qPCR data was classified into three categories: positive, negative, and intermediate or inconclusive results. The classification was performed according to the amplification (C_q -value < lowest standard), MCA (T_m > 80 °C), and additionally to end-point analysis. To enlighten the relationships between sample types and patient groups, further statistical analysis and visualization of the qPCR data was performed, and later used in the qNGS approach.

Settings for the evaluation of the quality of each qPCR run were adjusted according to the recommendations of the thermal cycler software (Bio-Rad, CFX Maestro, 2023). Reaction efficiency limits were 90–110%, and replicate group C_q -value standard deviation was set to be 0.2 or less. The replicate variation, i.e., the standard deviation of the replicate group C_q -value, was the primary criterion for the exclusion of a sample. MCA was used to decide outliers in varying triplicates. A result was considered as an outlier according to the replicate variation, but also if the melt curve was significantly distinct from the other two replicates by shape or T_m.

If only one sample replicate was chosen due to standard deviation variation, but the rejected ones were supportive as their melting profiles, the remaining sample was classified to be either clear positive or negative. If the melting profiles of the triplicates were different, the remaining sample was considered inconclusive, regardless of the clarity of the melt curve. The sample replicate was always classified as inconclusive if there was no

support from other samples. If the triplicates were similar and supported each other by their amplification and melting profiles, but only one was chosen due to replicate variation, the chosen one was the one with the lowest C_q -value. Thereby, the positivity of a given sample was more significant than the negativity. End-point analysis was used additionally. If the result from end-point analysis was against the suggestion received from manual analysis, or there was variation between sample replicates in terms of end-point results, the sample was classified as inconclusive. The categorized qPCR results were compared to relative sequencing data to evaluate the correlation between the two methods.

The final qPCR data was statistically analyzed and visualized with R software (version 2024.4.2.764, Posit Team 2024), with packages: *pacman* (Rinker & Kurkiewicz, 2017), *remotes* (Csárdi *et al.,* 2024), *psych* (Revelle, 2024), *ggplot2* (Wickham, 2016), *Munsell* (Wickham, 2024), *BiocManager* (Morgan & Ramos, 2024), *phyloseq* (McMurdie & Holmes, 2013), *stringi* (Gagolewski, 2022), and *farver* (Pedersen, 2022).

The relationships between sample types (vaginal and endometrial) and patient groups (RPL patients and control group) were analyzed with nonparametric Wilcoxon exact tests by comparing the differences of the mean C_q -values, and computing approximations and probability of a sample from one group to be greater than a sample from another group. Wilcoxon test was chosen as the data consisted of two groups with independent observations, and was not normally distributed. A paired Wilcoxon signed rank exact test was performed to compare the differences between the sample types among each patient. Wilcoxon rank sum exact test, also known as the Mann-Whitney test, was performed unpaired to compare the two sample groups in terms of sample types. To compare fungal positivity and quantities in gynecological samples of RPL-patients and their controls, and to investigate their relationship, the results were visualized by plotting their C_q -values.

To assess the effect of the read counts, i.e. the number of sequences in each NGS sample data on the success of qPCR amplification, statistical correlation tests were performed by comparing Illumina MiSeq total read count of each sample to the mean C_q -values of fungal qPCR replicates. As the data was continuous with independent nonparametric variables, correlation tests were performed by Spearman's rank methods without correction adjustment, and with "pairwise" mode to include solely complete cases with both data types from each sample.

3.6 Fungal qNGS approach

The qNGS absolute abundancies were achieved by multiplying the total fungal abundance from qPCR by the relative abundance of each fungal taxa derived from MiSeq NGS data. The qPCR starting quantity (SQ) values were converted into ITS CN, which provides the number of genomic copies (DNA molecules) per given mass of the sample. The *C. albicans* genome was used as a reference to calculate the default fungal CN, which was used to convert SQ values of all the other fungal taxa into CNs.

The mass of the *C. albicans* genome (m) was calculated by multiplying *C. albicans* genome size (n) (14.851 Mb), an average molecular weight (M) of a dsDNA (660 g/mol) (Stephenson, 2003), and Avogadro's number (N_A): $m = n * M * N_A$. The ITS CN of *C. albicans* ($CN_{C.alb}$) was derived by dividing the mass of the genomic DNA (m_{gDNA}) used in standard series by the mass of the genome (m): $CN_{C.alb} = \frac{m_{gDNA}}{m}$. To obtain the CN values of all the other fungal taxa, their SQ values were converted into CN by multiplying all the SQ mean values included in the analysis with a CN of 1.00 ng of *C. albicans*: $CN_{sample} = CN_{C.alb} * SQ_{sample}$.

Further analysis and combining of the absolute CN values and relative NGS data were performed with R (version 2024.4.2.764, Posit Team 2024) *pacman* (Rinker & Kurkiewicz, 2017), and *remote* (Csárdi *et al.*, 2024) packages. The minimum cut-off value of the qNGS was set to the limit of the lowest standard ITS1 CN to ensure the positivity of the sample and to remove values close to negative. The visualization of the data was performed according to the results to provide the most comprehensible view of the taxa and their abundances.

4 Results

4.1 Troubleshooting

Due to a major part of the thesis work being spent on troubleshooting the poorly performing qPCR assay, some of the troubleshooting results are presented. As the first stage of the troubleshooting matrix (Table 2), involving basic error-detecting methods, offered no answers to the problems, the troubleshooting continued by cross-examination of the primers with identical conditions. The amplification of different types of templates was problematic, including the additional biological samples (fecal, vaginal, endometrial), and the *C. albicans* reference fungus (Figure 6).

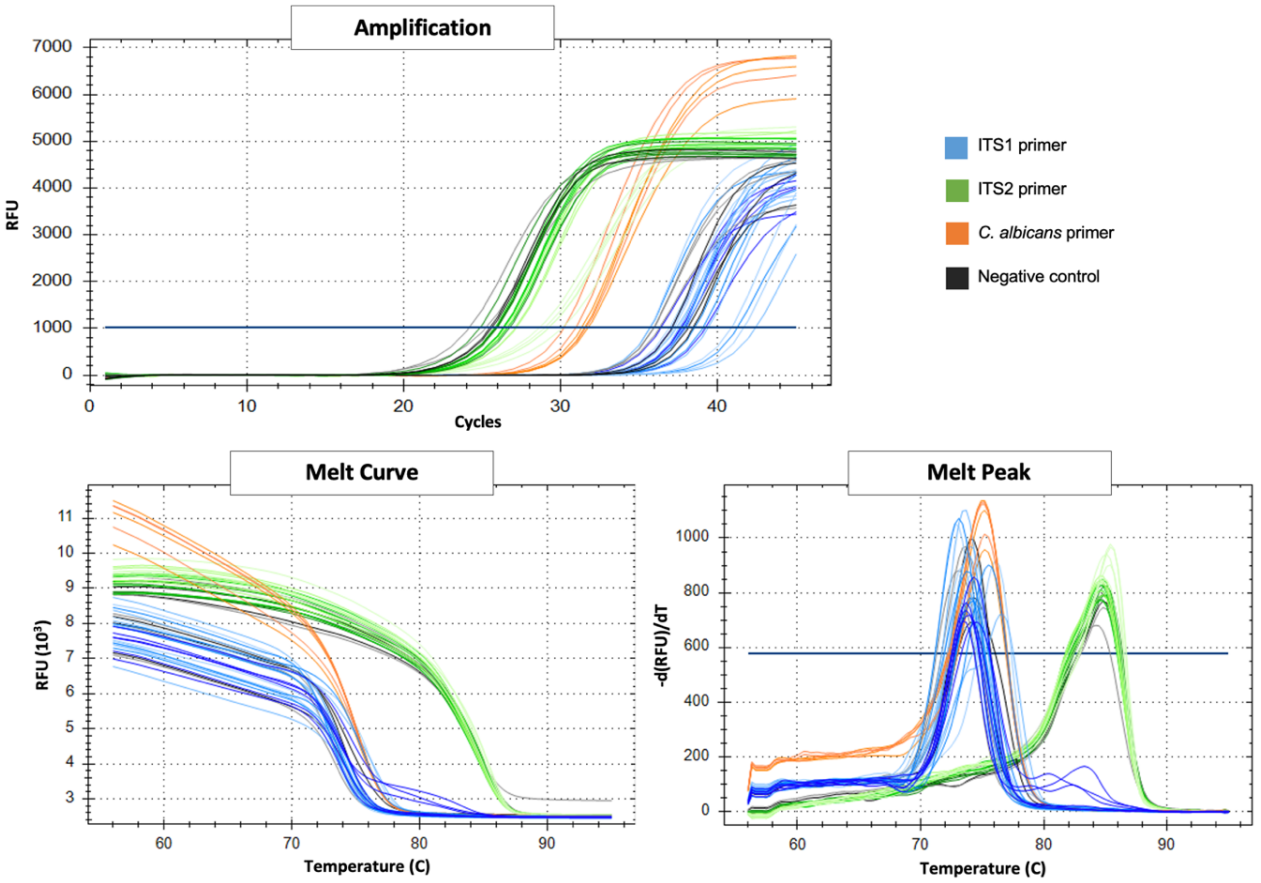

Figure 6. qPCR amplification and MCA of different primer pairs on *C. albicans* template.
Different profiles of a partly functional ITS2 -primer and an inefficient ITS1 and *C. albicans* -specific -primers result in different amplification curves and melt profiles. *C. albicans* template concentrations are represented as darker and lighter tones of green and blue, and should be well separated in the amplification chart. The amplification of all primers should start around cycles 20–35, the melt curves should be similar, and the melt peaks should be around the expected area of products (>80°C).
The amplification of *C. albicans* with ITS1 -region primers started late (C_q >34), and their melt curves and melt peaks were entirely around the PD -region. The *C. albicans* specific primers acted moderately (C_q 30), but their melting profiles were not acceptable. The ITS2 -region primers amplified early (C_q 24-30), and their melt peaks were around the expected area of products, but there was no clear separation between the different template dilutions. The negative controls were entirely among other samples.

The results, achieved from the first stages of the troubleshooting matrix, guided towards the intermediate troubleshooting, in which the fresh working dilutions from stronger primer stocks, nor new aliquots of other reagents or templates resulted in any solutions to the failure of qPCR. The temperature of the freezer was stable during the measured 24 hours. Both tested qPCR machines gave similar results. After accomplishing each step of the third intensive troubleshooting step, including a pause of more than a month of laboratory work, the problems in qPCR runs were evaded, as all parameters of the qPCR were functioning as expected, enabling the final optimization of the protocol. Regardless of the troubleshooting, no reason could be pinpointed to be the source of poor qPCR performance. There were no significant differences between the old and new MQ-water, nor with primer stocks or DMSOs. The possibility of qPCR mix being the source of the problem was low, as it is replaced constantly, every 1–2 assays. There were no clear differences between the new and the older, originally well-worked samples and negative controls, nor between the new stock of *C. albicans* pure culture and the first pure culture and dilutions of it.

4.2 DNA concentrations

An additional challenge in the thesis work in parallel to the non-optimal performance of the qPCR assay was that the patient sample DNA concentration results were partly inconsistent, as the different dilutions of the samples (10% and 1% dilutions) did not completely correlate together, nor to the undiluted original sample concentrations. Several measurements with Quant-iT were committed: the original samples were mostly over the detection limit of the assay, while the lighter 1% dilutions generally were below it, whereas the 10% dilution plate gave the most reliable concentrations and had the strongest correlation between different assays. The 10% dilution plates of both sample types were used as a base to prepare the desired final sample dilution plates for optimized qPCR.

The DNA concentrations of the six clinical test samples, upon which the optimization work of the qPCR protocol was ultimately adjusted, were verified by re-measuring them with Qubit. The vaginal swab concentrations were: 4.3 ng/µl, 1.0 ng/µl, and 6.1 ng/µl, and endometrial samples were: 11.2 ng/µl, 5.1 ng/µl, and 9.8 ng/µl. The replicates of *S. cerevisiae* DNA, measured with Qubit, were consistent: 48.3; 52.0; 55.0, and 55.0 ng/µl, with a mean value of 52.6 ng/µl and a median of 53.5 ng/µl. The DNA concentrations of *C. albicans* reference fungus, nor the additional biological test samples (fecal, saliva) used

in the preliminary optimization of the qPCR protocol were readymade stocks in the laboratory and not re-measured.

4.3 Gel electrophoresis

The electrophoresis gels (Figure 7) were used as a secondary detection method to the qPCR amplicons and compared to the melting profiles of respective samples. Some of the samples had fungal amplification, resulting in a visible band in the middle parts of the gel (150–400 bp), and a melting peak around the expected area of products. The majority did not amplify efficiently, resulting in a band in the lower area of the gel (50 bp) and a melt peak around the PD -region. The peaks not reaching the melting temperature did not appear at the gel either, whereas samples with double bands on the gel had a double peak melt curve as well. Hence, there was a good concordance between the qPCR and agarose gel results. None of the endometrial qPCR products were visible in the gel as clear bands, suggesting solely fungi-negative samples or an unsuccessful qPCR.

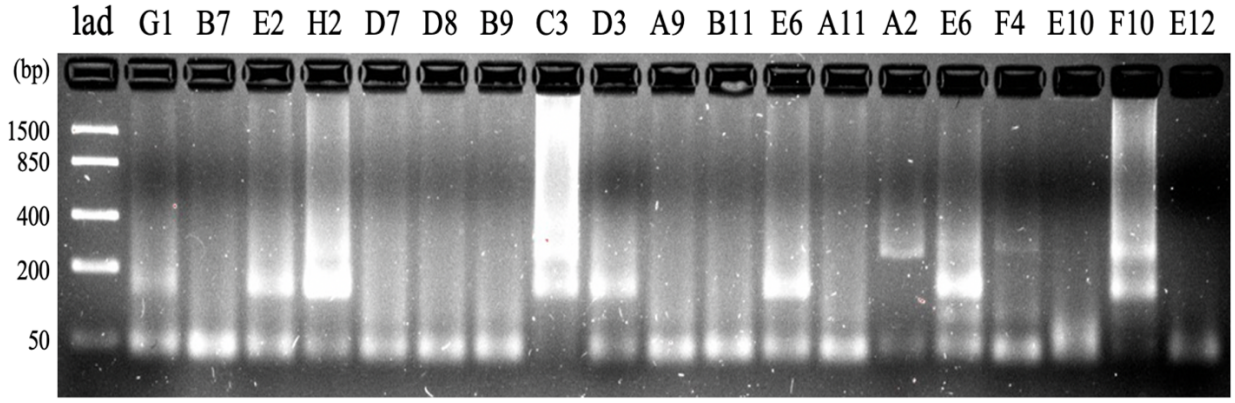

Figure 7. Vaginal swab qPCR products on agarose gel.
Fungal qPCR products are visible at the gel as white bands, and the length of a product can be scaled to a ladder: a clear band between 150–400 bp represents an amplified product.

4.4 Fungal qPCR optimization

A fungal qPCR protocol for gynecological samples was optimized. The optimization matrix (Table 5) summarizes the most efficient combination of each variable, further discussed in more detail.

Table 5. Fungal qPCR optimization results.

Cycling conditions	Master Mix elements			Template elements	
Annealing T (°C)	***Primer type (region)***	***Primer c µM (µl)***	***DMSO µl (%)***	***Type (DNA)***	***Amount (ng)***
52 to 62	ITS1 ITS2	0.8 (1.0) 2.0 (2.5)	none (0) 0.50 (2.0) 0.60 (2.4) 0.75 (3.0)	*C. albicans* *(S. cerevisiae)* (Fecal) (Saliva) Vaginal Endometrial	0.001 0.01 0.10 1.00 2.50 5.00 10.00 25.00 50.00
→ 60.0° C	***→ ITS1***	***→ 0.8 µM***	***→ none***	***C. albicans Vaginal Endometrial***	***→ 10.00 ng***

4.4.1 Cycling conditions

The qPCR thermal cycling conditions were adjusted as presented in Table 6. The initial denaturation was set three minutes longer than was recommended by the guidelines of the Solis BioDyne qPCR Master Mix reagent. The number of polymerization cycles was elevated four cycles above the recommendation. The melt curve was set according to the annealing temperature. The annealing temperature close to 60.0°C resulted in the most efficient amplification in each run, with all primer volumes, and both with and without DMSO. Efficiencies started to decrease in temperatures below 55.0°C and over 60.0°C.

Table 6. Adjusted qPCR thermal cycling protocol.

	Step	Temperature (°C)	Time (min:sec)
1	Initial denaturation and polymerase activation	95	15:00
2	Denaturation	95	0:30
3	Annealing + plate read	60	0:30
4	Extension	72	0:30
5	44 cycles of steps 2–4	-	-
6	Incubation (equalizing)	95	1:00
7	Detection (fluorescent data)	60	1:00
8	Melt curve + plate read	56 to 95, increment 0.3	0:01

4.4.2 Master Mix elements

Three elements of Master Mix were investigated, including primer type, primer concentration, and DMSO. Different primers were tested using DNA from pure cultures of fungi. The ITS1 -region primers were the most efficient (Figure 8) with the *C. albicans* template, and the difference to the ITS2 -region primers was 4–5 amplification cycles on average, corresponding to a >10-fold higher sensitivity.

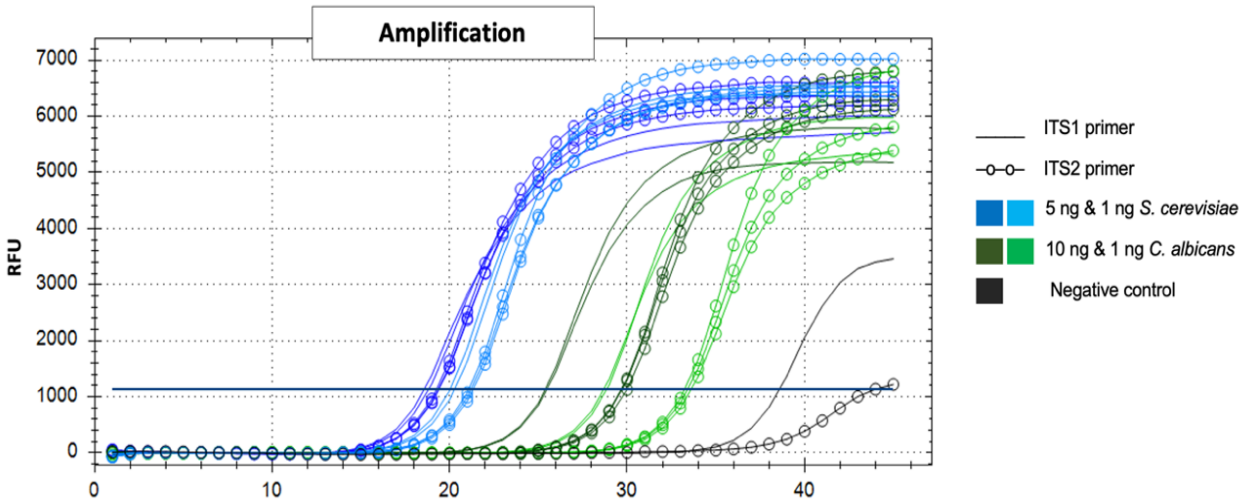

Figure 8. The different amplification of ITS1 and ITS2 -region primers.
The different templates and primers provide different amplification efficiencies, corresponding to the C_q-value, i.e., the number of cycles the sample reaches the amplification threshold.

The difference was not as clear with other tested samples, such as with *S. cerevisiae*, where the difference was 1–2 thermal cycles on average. Melting profiles of the two primer regions were not alike due to their varying T_m, the ITS1 -region corresponding to 83°C and the ITS2 -region to 87°C on average. The *C. albicans* -specific primers amplified the *C. albicans* template well, yet not as efficiently as the ITS -region primers, with C_q -values 1.4–2 thermal cycles higher than ITS1 -region primers. Melting temperatures were similar, 83.4–84.0° C, with all primers, but with *C. albicans* -specific primers the height of the melt peak was lower, corresponding to less efficient amplification. The amplification of *S. cerevisiae* with ITS -primers was more efficient than the amplification of *C. albicans*, and both ITS -primers gave comparable results, as both the amplification and the MCA had moderate differences. The ITS1 -region primers were selected as the most efficient, thus further represented results are based on those.

A higher concentration (2.0 μM) of primer per reaction resulted in more efficient amplification among all the samples, including negative controls. The lower concentration

(0.8 µM) separated the samples more efficiently by their template amount, as the difference between amplification thermal cycles of the highest and the lowest standard series template amounts was greater (Table 7). The separative effect was particularly clear with the standard series and negative controls.

Table 7. Effect of primer concentration on C_q -values.

Primer concentration (µM)	**C_q -value (mean)**		
	10.0 ng template	**0.1 ng template**	**Thermal cycle difference**
2.0	23.4	28.0	**4.6**
0.8	23.8	30.0	**6.2**

The higher concentration generally increased the dimerization of primers, respectively decreasing or entirely removing the melt peak in the expected area of PCR products (Figure 9:A). Nevertheless, in some cases the effect of higher primer concentration was the opposite, enhancing the correct melt peak and lowering the PDs respectively (Figure 9:B). There was inconsistency especially with the endometrial samples, whereas among vaginal swab samples, the lower concentration resulted in higher melt peaks in the expected area.

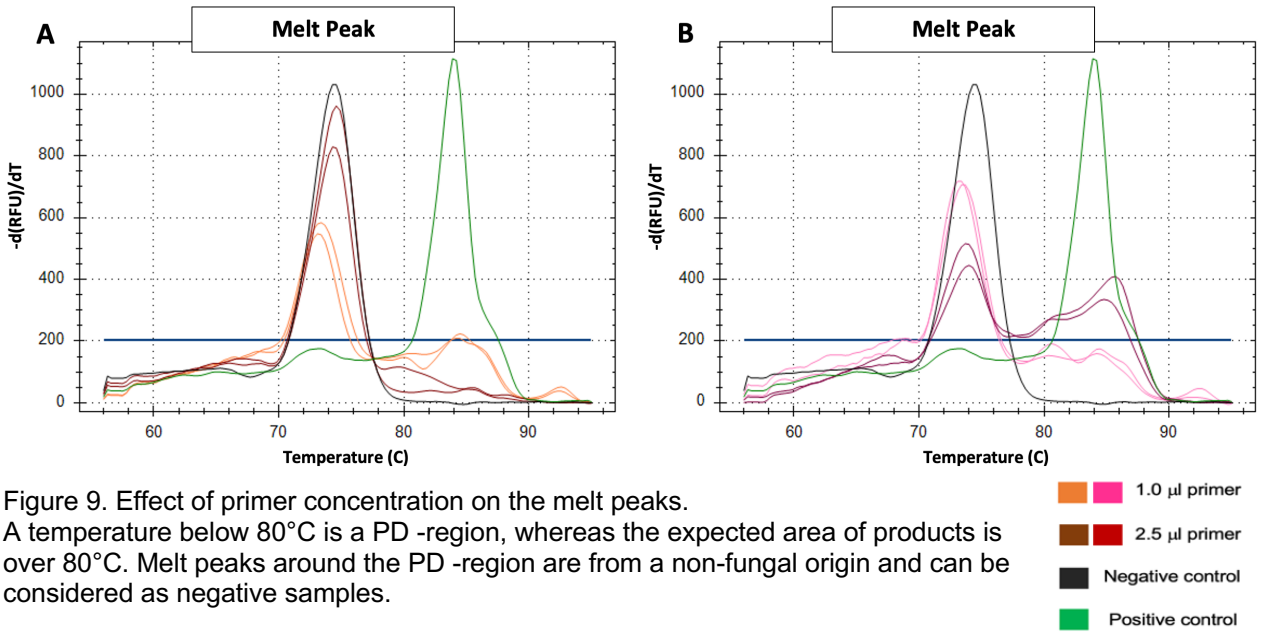

Figure 9. Effect of primer concentration on the melt peaks.
A temperature below 80°C is a PD -region, whereas the expected area of products is over 80°C. Melt peaks around the PD -region are from a non-fungal origin and can be considered as negative samples.

The effects of DMSO were controversial, as it both enhanced and reduced the efficiency of the reaction. The results were stable with *C. albicans* DNA from pure culture but fluctuated with other tested samples. When DMSO was involved, the amplification was the most efficient with low annealing temperatures (54.0–55.0°C). However, lower temperatures

reduced the melt peaks in the supposed area of products with all dilutions of DMSO, whereas high annealing temperature (>60°C) enhanced it. Furthermore, DMSO lowered the melting temperature throughout the examinations by 0.5–2 degrees, depending on the amount of DMSO used per reaction (Figure 10).

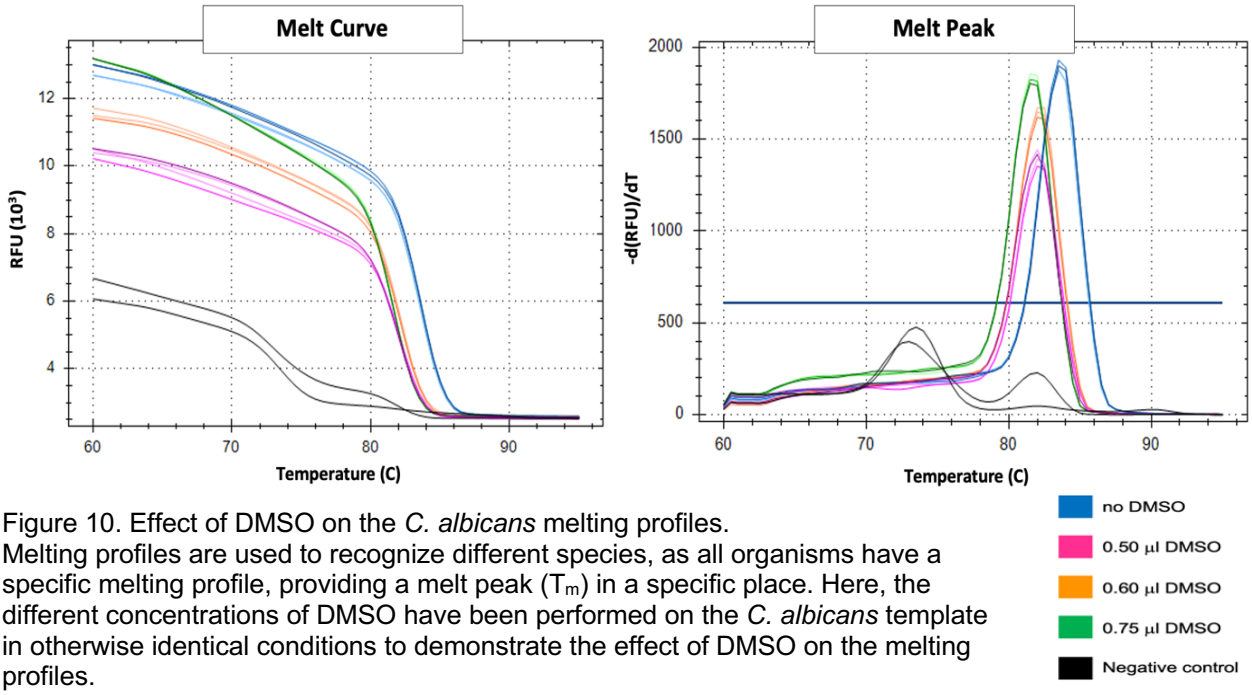

Figure 10. Effect of DMSO on the *C. albicans* melting profiles.
Melting profiles are used to recognize different species, as all organisms have a specific melting profile, providing a melt peak (T_m) in a specific place. Here, the different concentrations of DMSO have been performed on the *C. albicans* template in otherwise identical conditions to demonstrate the effect of DMSO on the melting profiles.

Despite DMSO improved the amplification of endometrial test samples by lowering the C_q-value by 1–3 thermal cycles, it increased mainly the PDs. With vaginal test samples, DMSO reduced the efficiency by elevating the C_q-value by 2–6 thermal cycles. Furthermore, DMSO enhanced the amplification in negative controls, decreasing the C_q-value, observed in MCA as PDs.

4.4.3 Template elements

The aimed amount of the DNA per reaction was 10 ng. The low (<1 ng) template amounts resulted in low E% in every sample type, and with every primer or DMSO combination, as they amplified late (C_q>34), close to negative controls. The difference between low and intermediate (1–10 ng) template amounts was particularly clear in melting profiles, where lower template amounts mostly increased the PDs, and higher template amounts developed melt peaks around the expected area of products. The amplification curves of low template amounts were overlapping and had high replicate variation. Higher amounts (>10 ng) of template resulted in more efficient amplification within every sample type. Within gynecological samples the difference between the highest and the intermediate

template amounts was moderate: for the vaginal samples with the highest dilutions of 50.00 ng/reaction the mean C_q -value was 32.8, and with 25.00 ng/reaction it was 32.5. For the intermediate 10.00 ng/reaction, the mean C_q -value was 33.3, and for the 5.00 ng/reaction, it was 34.5. The highest dilutions elevated the amplification only with the *C. albicans* positive samples, as with the highest template amount their C_q -values were 24–26, whereas with the intermediate template amount their C_q -values were 28–31. Instead, there was no difference in the amplification, if the sample was *C. albicans* negative. With endometrial samples, the results were comparable to the vaginal swab samples, yet more inconsistent.

The vaginal test samples were amplified either well (mean C_q 27.9), or with low E% (mean C_q 34.3), reflecting their fungal positivity. The saliva DNA samples did not amplify efficiently in any combination of reagents, with C_q -values of 33–35 (mean C_q 34.2). The amplification of fecal DNA samples was more efficient than saliva samples and most of the vaginal samples, with C_q -values of 29–34 (mean C_q 31.5). The spiking of samples with *C. albicans* resulted in a significant elevation of amplification within samples that were not amplifying well otherwise: the difference was 4–7 thermal cycles, and the C_q -value was elevated from around negative control into an area of a *C. albicans* of 1.00 ng. The spiking of the already fungal-positive sample did not cause an elevation of amplification.

4.5 Standard, controls, and thresholds

The 10-fold standard series produced from *C. albicans* pure culture originally included five dilutions (0.001–10.000 ng) of the template per reaction, corresponding to an ITS1 -region CN from 10^2 (614.375) to 10^{-2} (0.061) CN/µl of genomic copies in each dilution (Table 8).

Table 8. Standard series qPCR results.

Dilution	template (ng)	CN/µl	average C_q
10^2	10.00	614,375	24.2
10^1	1.00	61,438	27.6
10^0	0.10	6,144	31.0
10^{-1}	0.01	0,614	33.9
10^{-2}	0.001	0,061	35–36
neg	0	0	35.4

The lightest 0.001 ng dilution, being entirely indistinguishable from the negative control, was removed from the analysis due to its inaccuracy, which led to a rise of E%. Eventually,

the standard curve included four dilutions from 10^{2} to 10^{-1} of the *C. albicans* template (Figure 11).

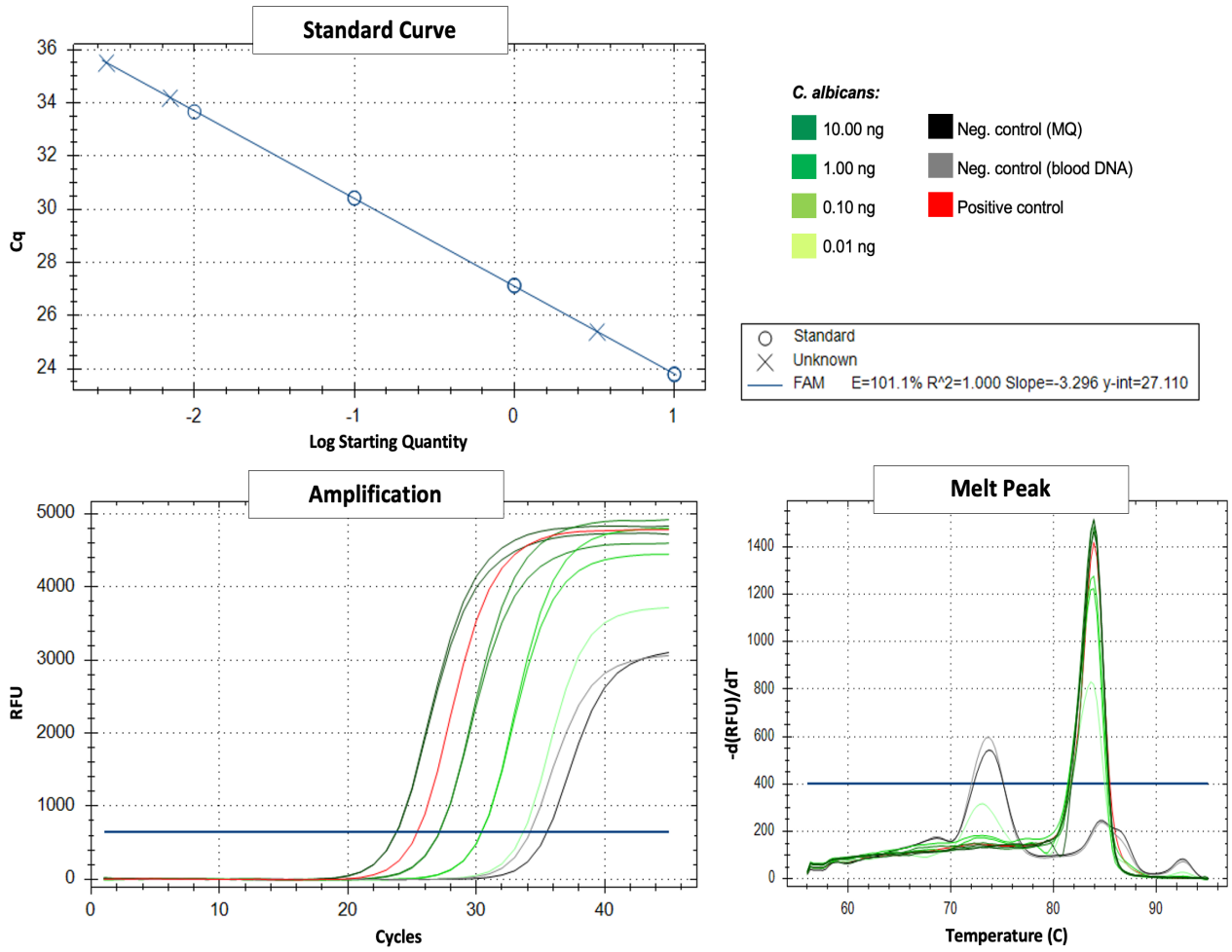

Figure 11. Standard curve of *C. albicans.*
A standard curve is generated by plotting the C_q of a template with known concentration dilutions against the logarithm of the SQ, which is an initial CN of a template, derived from the C_q -values (Bio-Rad, 2006; CFX Maestro, 2023). The standard curve is derived from the amplification curves of *C. albicans*, which are gradually changing according to the dilution. Negative controls may amplify after the lightest standard dilution. Melt peaks of the standard are expected to be at the area of T_m of *C. albicans*, whereas negative controls may have melt peaks at the PD area.

The *C. albicans* positive control was functioning predictably throughout the examinations, generating results corresponding to the standard series dilutions between 10.0 ng and 1.0 ng. There was no significant amplification in either of the negative controls or T_m at the region of the expected product. There were elevated melt curves around the area of expected products, as well as melting temperatures in the PD -region, indicating primer pairing occurring while no template was available. The between-plate normalization and the standard curves of the eight qPCR assays were evaluated by their E% and R^2 (Table 9). The amplification threshold (FAM (RFU)) was set automatically. The T_m threshold was realigned manually according to the negative control melt peak (+50 units of -d(RFU)/dT).

Table 9. qPCR thresholds for standards and controls.

	VAGINAL SWAB				ENDOMETRIAL				median	mean
plate	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1–8**	**1–8**
E%	101	100	101	106	100	106	104	102	**102**	**103**
R^2	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	**1,00**	**1,00**
Amplification FAM (RFU)	650	864	909	728	666	787	792	755	**787**	**786**
T_m threshold -d(RFU)/dT	350	220	380	350	330	355	310	355	**350**	**329**
Neg.melt peak -d(RFU)/dT	300	170	330	300	280	305	260	305	**300**	**279**

4.6 Fungal qPCR

An optimized fungal qPCR protocol for gynecological samples was performed on clinical samples, and a comparison of fungal positivity and quantities in the two sample types and patient groups was performed. The qPCR replicate variation of the clinical samples was relatively high, which resulted in the elimination of sample replicates as part of the quality control. For the included samples, the between-plate variation was moderate in terms of replicate variation, as well as the difference between the two sample types (Table 10), with greater variation among endometrial samples.

Table 10. Included replicates of the clinical samples.

	Plate	triplicate	duplicate	single
VAGINAL SWAB n (%)	**1**	4	12	7
	2	4	10	10
	3	3	9	11
	4	2	9	5
	1-4	**13 (15.1)**	**40 (46.5)**	**33 (38.4)**
ENDOMETRIAL n (%)	**5**	4	14	6
	6	1	8	15
	7	2	16	6
	8	1	10	2
	5-8	**8 (9.4)**	**48 (56.5)**	**29 (34.1)**
ALL n (%)	**1-8**	**21 (12.3)**	**88 (51.5)**	**62 (36.3)**

The C_q -values for the included fungal qPCR replicates of the clinical samples were high, as most samples were around 34–35 (Figure 12), which was the area of negative controls, suggesting low amplification efficiency in general, low number of fungal positivity among samples, or both.

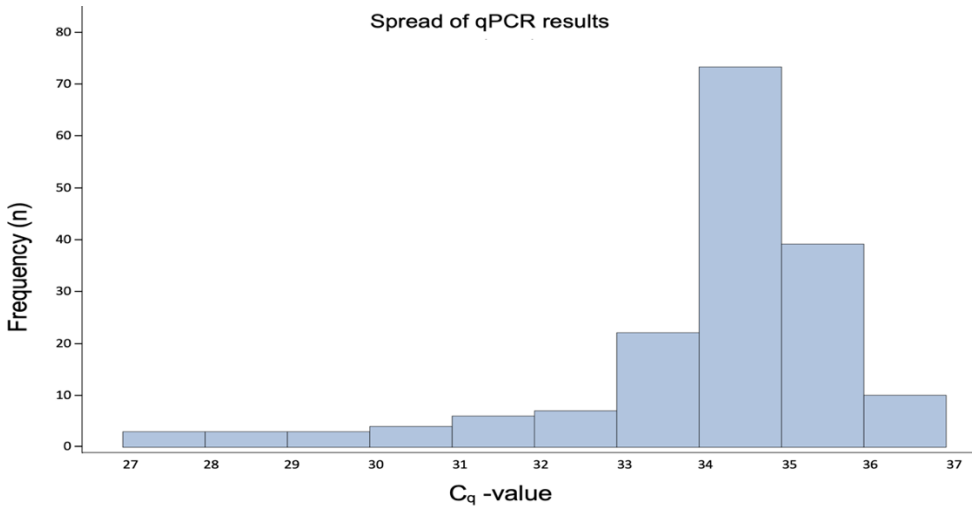

Figure 12. Spread of C_q -values of the clinical samples.

Visualizing the C_q -values (Figure 13), overall, the RPL-patients had more fungal amplification than the control group, yet vaginal samples of the control group had the lowest C_q -values of all sample groups and types. Although most of the samples had high C_q -values around 33–36, there are single samples with lower C_q -values, especially among vaginal samples. The RPL-patients had more fungal amplification in the endometrial samples, and the control group in the vaginal samples.

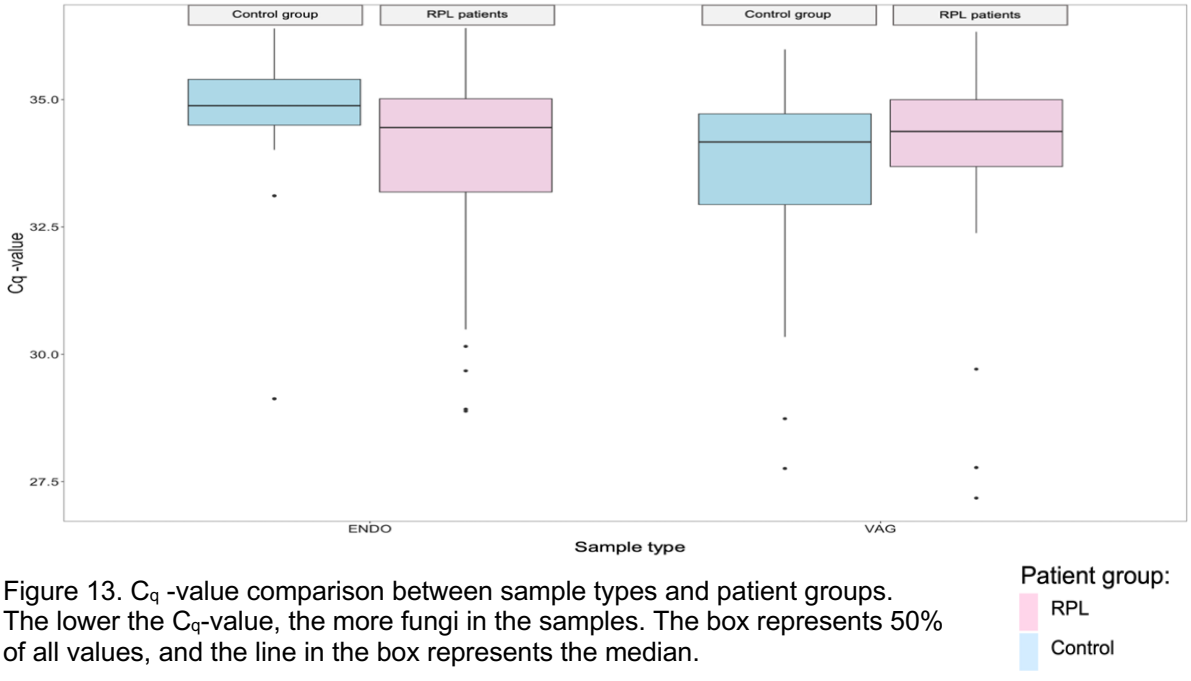

Figure 13. C_q -value comparison between sample types and patient groups. The lower the C_q-value, the more fungi in the samples. The box represents 50% of all values, and the line in the box represents the median.

No statistical significance (p<0,05), measured with Wilcoxon correlation tests (Table 11), was observed between the two sample types of RPL-patients, nor between vaginal samples of the two patient groups. The RPL-patients had more fungal amplification in the endometrium than in the vaginal samples, but there was no statistical significance. The control group had significantly more fungal amplification in the vaginal samples (p<0,001), whereas their endometrium samples were mostly fungi-negative. No correlation was encountered in the vaginal samples among patient groups. The endometrium samples of the RPL-patients had significantly more fungal amplification (p=0.011), whereas the controls were mostly negative.

Table 11. Wilcoxon rank sum test results of the fungal qPCR data.

	Variables	Method	V / W	p-value
RPL	vaginal & endometrial	paired	586 (V)	0.626
CONTROL	vaginal & endometrial	paired	135 (V)	< 0.001
VAG	RPL & control	unpaired	1074 (W)	0.120
ENDO	RPL & control	unpaired	610 (W)	0.011

The correlation between the qPCR and NGS methods was evaluated by comparing the relative sequencing data and qPCR data (Table 12). The fungal qPCR data was classified into three categories by visually comparing the qPCR amplification, MCA, and end-point analyses. Almost one-third of all the results remained inconclusive due to the disagreement of the three criteria, thus out-ruled from the comparison, diminishing the portion of the qPCR results compared to MiSeq results. The qPCR results categorized as positive or negative were classified as either matching or mismatching to the MiSeq reference result (pos or neg), with percentages of the match or mismatch respectively.

Table 12. qPCR curve and MiSeq relative data comparison.

	ALL n (%)			VAGINAL n (%)			ENDOMETRIAL n (%)		
	TOT	**pos**	**neg**	**TOT**	**pos**	**neg**	**TOT**	**pos**	**neg**
MiSeq	170 (100)	106 (62.4)	64 (37.6)	85 (100)	68 (80.0)	17 (20.0)	85 (100)	38 (44.7)	47 (55.3)
qPCR	121 (71.2)	67 (39.4)	54 (31.8)	66 (77.6)	42 (49.4)	24 (28.2)	55 (64.7)	25 (29.4)	30 (35.3)
match	**68 (40.0)**	44 (41.5)	24 (37.5)	**44 (51.8)**	35 (51.5)	9 (52.9)	**24 (28.2)**	9 (23.7)	15 (31.9)
mismatch	**53 (31.2)**	23 (21.7)	30 (46.9)	**22 (25.9)**	7 (10.3)	15 (88.2)	**31 (36.5)**	16 (42.1)	15 (31.9)
inconclusive	**49 (28.8)**	-	-	**19 (22.4)**	-	-	**30 (35.3)**	-	-

The Spearman's rank overall correlation (r) between the high read count of sequencing data and early amplification of the samples was -0.19 (Table 13), i.e., a high read count corresponded to a successful qPCR. The correlation results (t-value), as well as the probability (p-value) of all samples and vaginal swab samples, were statistically significant, whereas there was no statistical significance in the endometrial samples. The correlations were negative, as low C_q -values are a sign of early amplification of a product, a result of a successful qPCR run.

Table 13. Spearman's rank correlation test of qPCR amplification and NGS read count.

	n	correlation (r)	t-value	p-value	standard error
ALL	170	-0.194	2.562	0.011	0.076
VAG	85	-0.222	2.074	0.041	0.107
ENDO	85	-0.076	0.692	0.491	0.109

4.7 Fungal qNGS approach

The fungal qPCR results from clinical samples were coupled with existing NGS data to provide qNGS results. The qNGS approach provided the absolute abundances of 102 samples, divided nearly equally between the two patient groups (Table 14). In the histogram representing the qNGS absolute abundances of all the samples (Figure 14), there were great variances in the absolute abundances of fungi among samples, as the majority remained very low, and some had high absolute abundance. The absolute abundances were converted into a logarithmic scale, which made the extreme values more manageable and enabled a better evaluation of the data, but also made the large portion of samples at detection limit bulking the plot. A filtration of the results was performed to remove samples with very low abundances by setting a cut-off value to the limit of the lowest value of the qPCR standard curve (Table 8), hence the positivity of samples beneath it could be questioned. After the filtration, 42 samples in total were included in the qNGS approach (Figure 15 & Figure 16). The division between patient groups changed to slightly favor RPL-patients over controls, and the sample group division increased accordingly.

Table 14. Breakdown of samples in the qNGS approach to sample groups and types.

	TOT	RPL	CONTROL	VAG	ENDO
before filtration n (%)	**102 (100)**	53 (52,0)	49 (48,0)	65 (63,7)	37 (36,3)
after filtration n (%)	**42 (100)**	25 (59,5)	17 (40,5)	31 (73,8)	11 (26,2)

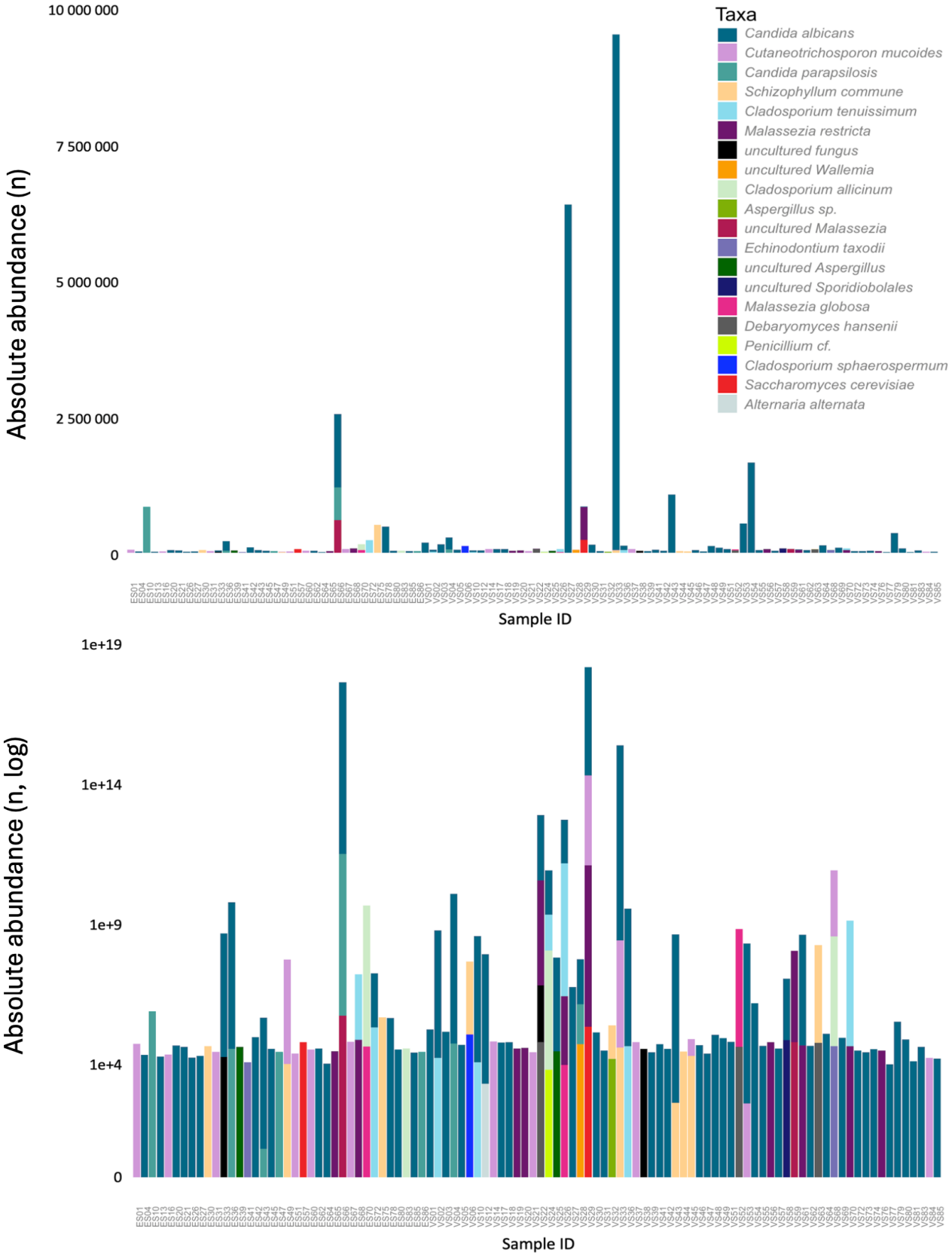

Figure 14. Absolute abundances of all samples.
The qNGS approach provided results in 102 samples, represented in real scale according to their CN (upper histogram). The differences in the heights of the bars are large, as the histogram is scaled by the most abundant samples. The fungal taxa are more recognizable after converting the abundances into a logarithmic scale (lower histogram).

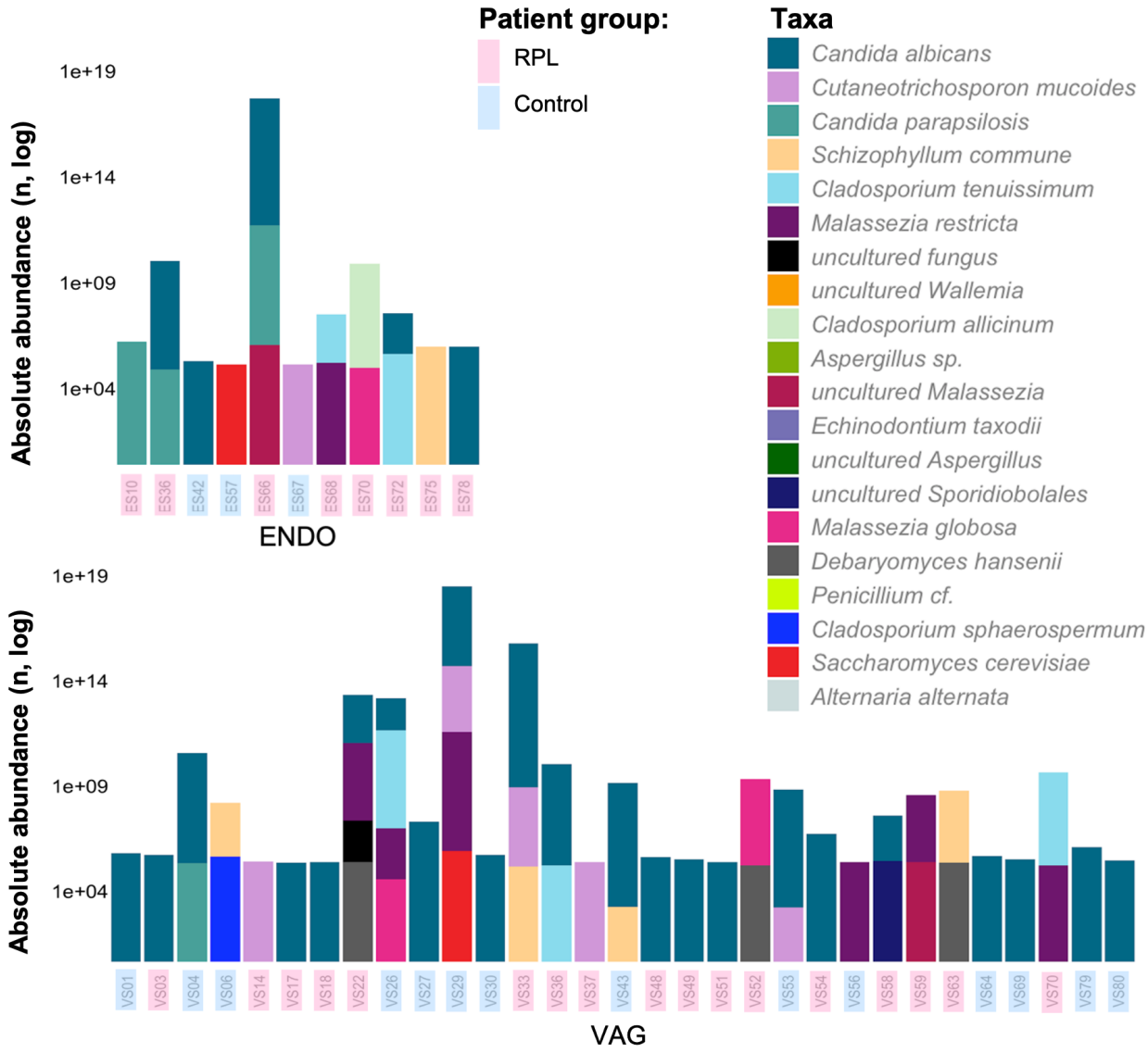

Figure 15. Absolute abundances of filtered samples in logarithmic scale.
The log conversion enlightens the fungal taxa of each sample the most profoundly, revealing more taxa per sample than the non-log scale or relative abundance scale. However, the logarithmic scale decreases the height of the top bar taxa, increasing the lower bar taxa, making it challenging to compare the different samples.

The 42 filtrated samples were selected for the qNGS approach representing the absolute abundancies in a log scale (Figure 15), which revealed all the taxa encountered from each sample, not visible in the non-logarithmic representations, including the two endometrium samples (ES68, ES72) and 10 vaginal swab samples (VS06, VS22, VS26, VS29, VS33, VS43, VS53, VS58, VS59, VS63). Instead, the absolute abundancies represented according to their CN directly (Figure 16) provided the most realistic view of the taxa of each sample in relation to other samples and compared to the relative abundance of the taxa. Four samples had significantly higher CN than the rest of the samples, which resulted in drastically cutting the bars with a CN over 1,000,000 to fit the scale.

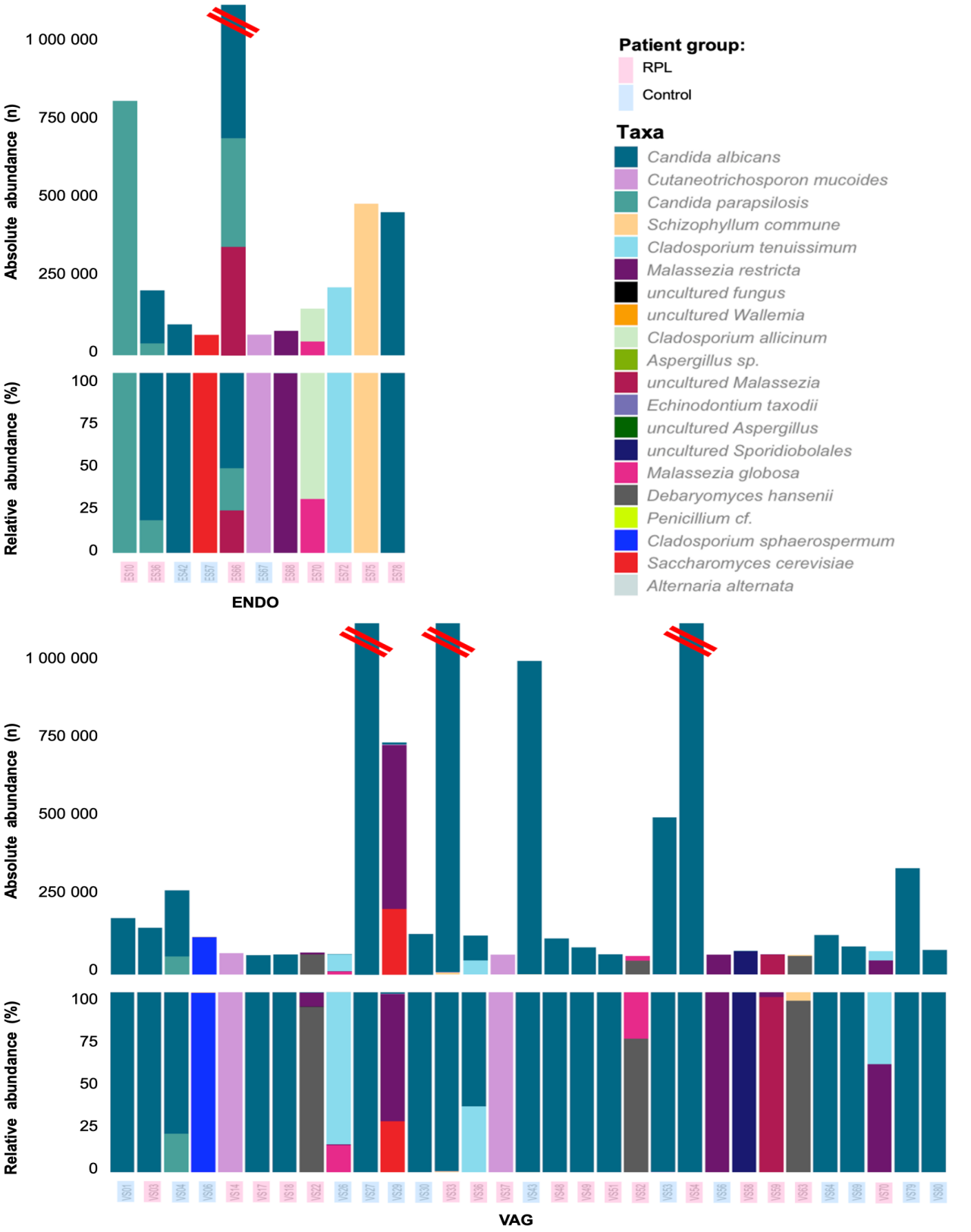

Figure 16. qNGS approach of selected samples.
Classic relative abundances (%) of taxon encountered both in NGS and qPCR always sum up to 100%, whereas the qNGS approach eliminates the compositionality of NGS data, estimating the absolute abundances (n) in genomic CN of fungal taxa identified by sequencing.
Represented samples had a clear positive value, as the cut-off value was set on the limit of the lowest qPCR standard. Four samples have been cut (red lines) to fit the scale, as they had significantly higher CN (n > 1,000,000) than the others. The qNGS approach helps to evaluate the differences between the samples, and enlightens the fungal taxa of each sample more realistically.

The number of the fungal taxon in the sample correlated directly to the absolute abundances. There were 23 samples containing a single taxon, and 19 samples consisting of 2–4 taxa, 13 of which belonged to the RPL-patient group and 6 to the controls (Figure 14). Each of these also had absolute abundances higher than the highest sample with a single taxon (~10^7). Within vaginal samples, the highest absolute abundances belonged to both sample groups, as within the 10 highest absolute abundances value there were five RPL-patient samples and five control group samples. Within endometrial samples, the control group samples had the lowest absolute abundance values throughout the samples.

There were 69 fungal taxa observations in the qNGS approach (Table 15), consisting of 14 fungal taxa, of which *C. albicans* was the prevalent taxon throughout the samples. It was present in 28 samples, divided relatively evenly among patient groups, but unevenly among sample types. Furthermore, uncultured *Wallemia*, uncultured *Aspergillus* and other *Aspergillus*, *Echinodontium taxodii*, *Penicillium*, and *Alternaria Alternaria* were encountered with low abundances. From all the observations, 75% of the species were encountered from vaginal samples, and 25% from the endometrium samples. The observations were more common among RPL-patients (61%) than the controls (39%).

Table 15. Fungal species encountered from the qNGS.

	TOT		**RPL**		**Control**	
	VAG	**ENDO**	VAG	ENDO	VAG	ENDO
Candida albicans	**23**	**5**	11	4	12	1
Malassezia restricta	**6**	**1**	3	1	3	-
Cutaneotrichosporon mucoide	**5**	**1**	3	-	2	1
Schizophyllum commune	**4**	**1**	2	1	2	-
Cladosporium tenuissimum	**3**	**2**	2	2	1	-
Candida parapsilosis	**1**	**3**	-	3	1	-
Debaryomyces hansenii	**3**	**-**	3	-	-	-
Malassezia globosa	**2**	**1**	1	1	1	-
Uncultured Malassezia	**1**	**1**	1	1	-	-
Saccharomyces cerevisiae	**1**	**1**	-	-	1	1
Uncultured Sporidioblales	**1**	**-**	1	-	-	-
Cladosporium allicinum	**-**	**1**	-	1	-	-
Cladosporium sphaerospermum	**1**	**-**	-	-	1	-
Uncultured fungus	**1**	**-**	1	-	-	-
All (%)	**52 (75)**	**17 (25)**	**28**	**14**	**24**	**3**
	69 (100)		**42 (61)**		**27 (39)**	

Among the achieved 42 qNGS results there were two patients with sample pairs (Figure 17) belonging to both sample type groups, whereas the other 38 samples were either vaginal or endometrial samples. Both sample pairs belonged to the RPL-patient group and consisted of two fungal taxa: patient 1 harbored *C. albicans* in both vagina and endometrial sample, whereas the second taxa varied in the two samples type, as *C. tenuissimum* was present in the vaginal sample and *C. parapsilosis* in the endometrial sample. Patient 2 harbored different taxa in both sample types: *Malassezia restricta* and *C. tenuissimum* in the vaginal sample, while *Malassezia globose* and *C. allicinum* were predominant in the endometrial sample. According to these two patient cases the taxonomy was not alike in the two sample types of a patient.

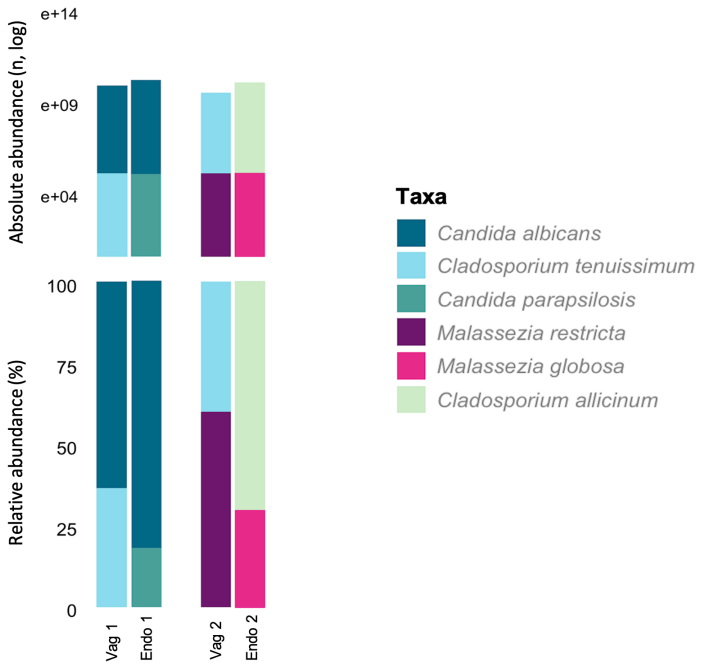

Figure 17. qNGS sample pairs.
The fungal entity might vary in the two sample types of the patient.

5 Discussion

5.1 qPCR and qNGS findings

This thesis intended to reliably investigate the absolute amounts of mycobiome in gynecological samples, employing the qNGS approach, after optimizing fungal qPCR for vaginal and endometrial samples. The aim was to investigate whether the fungal positivity and quantities were related in the two sample types from the same patient, and whether there were differences between the patient groups. The optimized fungal qPCR was used on clinical samples, resulting in a successful fungal qNGS approach, providing estimates of absolute abundances of fungal taxa in gynecological samples, in vaginal swab samples in particular. The qNGS approach prevents the risk of misinterpretation and enables the comparison of relative and absolute abundances of the taxa in each sample, in a relationship with other samples.

The fungal qPCR results showed statistically significant differences between the patient groups and sample types. Control group members had a significant difference in the abundance of fungi in the vagina and the endometrium, according to which the vaginal mycobiota was abundant, and the endometrium hardly harbored fungi. Instead, there was no statistically significant difference in the sample types among the RPL-patients, who had more abundant fungal amplification in the endometrium samples compared to controls. Furthermore, the endometrial samples were a distinguishing factor, demonstrating a statistical difference, whereas the vaginal samples were more similar among both groups.

The qNGS approach encountered 14 fungal taxa, *C. albicans* being the prevalent taxon throughout the samples, corresponding to the previous findings of *Candida* species being the most prevalent colonizer of the vagina (Achkar & Fries, 2010; Bradford & Ravel, 2017; Drell *et al.,* 2013; Guo *et al.,* 2012; Ma *et al.* 2020; Rowan-Nash *et al.,* 2019; Virtanen *et al.,* 2023). Another *Candida* species*, C. parapsilosis*, an emerging human pathogen causing invasive infections (Trofa *et al.,* 2008), was the second most prevalent taxa among endometrial samples. The second common fungal taxon in all samples, *Malassezia restricta*, belongs to the most common colonizer of skin (Rowan-Nash *et al.,* 2019; Virtanen *et al.,* 2023). The third prevalent taxa, *Cutaneotrichosporon mucoide*, is a species causing superficial infections (Capoor *et al.,* 2015). *Schizophyllum commune* is an environmental mushroom (Mahajan, 2022), occasionally causing nasal and superficial infections in humans (Premamalini *et al.,* 2011), as well as *Cladosporium tenuissimum*

(Nasiri-Jahrodi *et al.,* 2023). The low amount of sample pairs did not allow the comparison of the fungal taxonomy of both sample types in one patient. Thus, a wider number of sample pairs would be a requirement to make any further conclusions.

Previously, issues in analyzing other than *Candida* species in gynecological samples have been reported, and the portion of uncultured fungi has been considerable. In the pioneering study by Drell *et al.,* (2013) more than 85% of the samples lacked the taxonomic specification lower than a kingdom, with a relative abundance of 38.6%, whereas Virtanen *et al.,* (2023) encountered uncultured fungi in 66% of samples, with a 9.5% total abundance. Furthermore, air-born contaminants and highly sensitive methods, including contamination despite stringent conditions, primer efficiency-related problems, or biases due to bioinformatic or sequencing artifacts have been described (Drell *et al.,* 2013). In the study with a bacterial qNGS approach by Jian *et al.,* (2020), the estimated absolute abundances from sequencing, and qPCR abundances among bacteria have been described to correlate nearly perfectly. However, microbial profiles achieved from cell- or molecular-based quantitative and relative microbial profiling methods are expected to differ due to varying microbial densities and challenges in the enumeration and quantification of microbial samples (Jian *et al.* 2021). In this study, the correspondence was not high, when comparing the categorized samples to the MiSeq reference data, as less than half of all the samples matched. Furthermore, the qPCR result categorization into positive, negative, or inconclusive could not be done robustly, as nearly one-third of the qPCR results fell into the inconclusive category. Compared to the previously generated NGS data, the fungal qPCR failed to amplify the fungi-positive samples in nearly half of the cases, whereas some of the fungi-negative NGS samples amplified well with qPCR. Additional fungi were amplified in some of the samples that were fungi-negative according to MiSeq reference data, which might be due to contamination or errors in sequencing. Similar discrepancies have been achieved from other qNGS studies as well, in which qPCR has been more sensitive than NGS (Dannemiller *et al.,* 2014). For our results, it can also be concluded that the fungal NGS, involving several qPCR steps during the library preparation, seems to likely contain some false positives.

Furthermore, the DNA extraction and purification methods have an impact on the findings and may alter the fungal community observed, as the cell walls and capsules lyse differently (Fredricks *et al.,* 2005; Huseyin *et al.,* 2017), potentially resulting in the over-expression of fungal taxa that lyse more easily, such as *Candida* (Lehtoranta *et al.,* 2021).

Notably, these methods are often optimized for bacterial DNA, which might not be ideal for fungi (Lehtoranta *et al.,* 2021). According to Huseyin *et al.* (2017), the qPCR and NGS samples derived from the same extracts, generated by using a mechanical lysis as part of the DNA extraction, were shown to efficiently break fungal cell walls. This method has been widely used in microbiome studies, thus allowing mycobiome research from the stored DNA originally used for bacteriome research (Huseyin *et al.* 2017). Regardless of whether this study included false-positive qPCR or NGS results due to DNA extraction and purification methods, contamination, untargeted non-specific binding, or false-negatives due to problems in qPCR protocol (Khot & Fredricks, 2009), or if they proposed additional information to the sequence data, the results between qPCR and MiSeq were partly uncoherent. In this light, both PCR-based study methods of endometrial samples with low microbial biomass would require further optimization and validation, as their qPCR results were more often mismatched than matched to the MiSeq reference data.

The methods for filtration and representation of the qNGS data are essential and may change the entire interpretation of the results. A significant portion of the samples were out-ruled from the analysis due to their negative, or close to negative qPCR values, which altered the division between the patient groups and sample types, favoring the RPL-patients and vaginal samples. The filtered data might have been worth investigating as well, and additionally, the qNGS approach could have been performed with different exclusion criteria, such as all qPCR samples without any selection of replicates, or solely positive qPCR results, regardless of their replicate variation or support from the other replicates. Thus, this would have reduced the reliability of the results, as fungal qPCR on the type of samples used in this thesis seemed to be very error-prone, requiring manual selection. However, a manual assortment of any type increases the risk of errors and wrong interpretations, which is a considerable result-affecting factor.

Furthermore, as the distribution of the qNGS data was high, a scaling was performed. The filtration and logarithmic conversion of the samples enabled the recognition of the taxa more specifically, revealing taxa in several samples that otherwise remained invisible. A log transformation is useful in results assuming normal distribution (Jian *et al.,* 2020). However, the data of this study was not normally distributed, hence the log transformation altered the proportions of different taxa: the top part of the bar was decreased, and the lower parts were increased in size. Thus, the log scale might be confusing, possibly leading to wrong interpretations, regardless of it providing a more profound interpretation of the

fungal taxa encountered in a sample. In this light, the non-log representation directly according to the CN would be preferable, in which the highest parts of the bars with extremely high abundances were cut, thus avoiding the demand for a very high abundancy scale of the y-axis of the histogram. Nevertheless, this approach decreased the taxonomic comparability of the samples in the sense that the difference between the bars of the histogram was too large to recognize the low-abundant taxa.

To improve comparability, the qPCR between-plate normalization was performed with a pure culture of *C. albicans*, used as the main reference organism. In addition to being the prevalent taxa in vaginal samples, it is a model yeast (Fredricks *et al.,* 2005), a well-known fungal species, providing high amplification efficiencies and predictable multiplication, thus making it suitable to use as a standard. The C_q -values of each standard dilution were in line with results achieved in other publications about the standard series of *C. albicans* (Busser *et al.,* 2020). The effect of NGS sequencing species level read counts of all the fungi on successful qPCR was evaluated statistically. A high read count corresponded to a successful amplification with statistical significance, which emphasizes the importance of template amount in qPCR optimization. However, among the endometrial samples, there was no correlation, which supports the hypothesis of suboptimal fungal qPCR protocol for endometrial samples, resulting in low E% in qPCR, and thus requiring still further optimization.

5.2 Strengths and limitations

There are several reliability-affecting factors, such as the rate of reproducibility, and the consistency and coherence of the results. Furthermore, there are limitations and strengths in this thesis, some of which might be both reliability decreasing and increasing factors, such as complications during optimization, and the troubleshooting resulting from them. This thesis utilized clinical patient data, underlining the importance of research ethics. Here, the author did not have access to any detail on patient data that could have enabled identification, such as name or date of birth. All the data was coded according to the patient group and sample type.

Sample handling and storage, including their stability with time, is one of the key challenges in mycobiome sequencing (Tiew *et al.,* 2020). The clinical samples were stored at -20° C for several years, with no data on their freeze-thaw cycles, and had potentially suffered from degradation but also evaporation, leading to possible concentration of the

samples. DNA concentrations of the original samples were measured several times to different dilutions of the sample to gain accuracy, as the results were inconsistent and non-correlative regarding different dilutions of the same sample. The inaccuracy in concentration analysis regarding the template amount was evident, decreasing the reliability of the results, thus being a limitation of this thesis. In addition, it raises some concerns regarding the reliability of the concentration assays in general, as usually there would be only one measurement of the concentration. Here, the confusing results led to both elongation of the process, and problems with the entire optimization work, as the amount of template was not necessarily the desired or even known.

The fluctuation in concentrations affected also the optimization process through the six clinical test samples used in the final optimization. Assuming the sample handling and storage affected the concentration, the DNA concentrations of recently extracted *S. cerevisiae* were evaluated, and confirmed consistent, suggesting the extraction and purification of the brewer's yeast stock were successful. Notably, these DNA samples were fresh, whereas the clinical samples, used in the optimization of the qPCR protocol and the qNGS approach, were years old. The re-measured results of the six clinical test samples were partly comparable to their expected concentrations, as half of them were similar, but the remaining three samples ranged from one-fifth to double the expected. However, the concentrations were confirmed after the optimization of the protocol, as their actual concentrations were not questioned until the inconsistent results from the clinical patient samples, obtained after the optimization of the qPCR protocol. The confirmation of the concentration of all the samples should have been performed before starting the optimization of the qPCR, which would have increased the reliability of this study.

The clinical test samples used in the optimization of the qPCR protocol were randomly picked and, by chance, one out of six was *C. albicans* positive, whereas the other five were low-abundancy or fungi-negative samples, resulting in low amplification, thus bringing great challenges to the process. Regardless of the test sample randomization increased the reliability, as the sample take truly illustrated the sample take, it might have been beneficial to use the information from fungal ITS1-amplicon Illumina NGS data, to choose an optimal set of samples from both sample types, including samples with no fungi, variant fungi, and solely *C. albicans*. Despite the larger number of samples would have extended and complicated the setup of the qPCR runs, it would probably have provided simpler, straightforward results and eventually improved the analysis, increasing the reliability.

Regardless the template amount was one of the most important variables investigated in this thesis, the importance of the portion of the DNA as such is not entirely clear. The total DNA concentration of each sample was provided, not the amount of fungal DNA in particular, thus including also non-fungal DNA such as bacterial or host DNA. This might have caused errors in terms of template amount optimization (Khot & Fredricks, 2009; Tiew *et al.,* 2020), as the fungal DNA amount might have been negligible among the entire DNA content. In some studies, this issue has been solved by applying a specific volume of the sample, regardless of the concentration (Virtanen *et al.,* 2023). Furthermore, the low-abundant samples might have caused further issues regarding the sample replicates, as another limitation of this study was that more than one-third of all samples were not studied as triplicates. The replicate variation was low with standard series and controls, indicating high accuracy in working methods, such as the pipetting technique. However, it was high in clinical samples, resulting in a rejection of replicates. A low amount of fungal DNA in the template might have resulted in uneven sample distribution between sample replicates. The low-abundant samples were usually those with the highest variation, as they did not amplify efficiently, resulting in high C_q -values and low melt peaks. Efficient amplification was associated with the samples of low replicate variation, resulting in the acceptance of all three replicates. The risk of uneven sample distribution was acknowledged, and precautions were taken to prevent it. Nevertheless, they were not sufficient for the low-abundant samples, supporting the hypothesis that the method requires further optimization. Hence, the sensitivity of the qPCR and the protocol used in this thesis was not enough to recognize the fungal positivity or negativity of all the samples, resulting in inefficient amplification in general.

Despite an optimized qPCR for clinical samples being performed, unpredictability and instability of the successful qPCR runs were reliability-decreasing factors and a limitation of this thesis. As the qPCR runs started to give unexpected and uncoherent results in the middle of the optimization process, special arrangements were committed to solve the problems, and a systematic error detection was launched, including a profound investigation of all the reagents, materials, and working habits. Regardless of the systematic troubleshooting, no explaining factor could be pinpointed for the sudden unsuccessful qPCR runs. One possible reason might have been the presence of PCR inhibitors in some of the reagents or equipment used, although they remained undetected, if present. Ultimately, only time and a pause in laboratory working seemed to help, or as suggested, perhaps it was just '*the position of the moon*'. Regarding the timing of the

complications, if would have occurred at the beginning of the laboratory work, they might have had a major impact on the process, possibly leading to a reconsideration concerning the fungal qPCR. Likewise, the problems might have potentially emerged during the final qPCR runs on unique clinical samples with limited amounts, leading to the restricted possibilities of reproducibility, which could be considered a limitation of this thesis. Despite the entire troubleshooting procedure being a time and resource-consuming process, not having unexpected problems, the troubleshooting would have been left out, which can be seen as an important part of the optimization process as well as a reliability-increasing factor and a strength of this thesis.

Additional reliability-increasing factors and strengths of this study are systematic and careful working methods, which were polished due to the complications, as the demand for repeated assays resulted in established routines in different phases of the laboratory work and an active mindset of constant error-avoiding and detecting. Lastly, the willingness to perform the work with high ethics, deep concentration, and success, which arose from the author's interest and the importance of the subject of this thesis, may be considered an essential strength of this study.

5.3 Prospects

The fungal qPCR protocol for gynecological samples was concluded from the information available, and should not be seen as a finalized protocol, and requires further examination. Greater variance in sample replicates and a larger portion of inconclusive samples, a lower correlation between qPCR and MiSeq data, and a lower portion of endometrial samples in the qNGS approach compared to vaginal samples reflect remaining challenges in the qPCR protocol of the endometrial samples and the demand for further optimization. In addition, there would be further analytical and statistical work to elaborate on the data achieved from this thesis, such as the clinical significance and relationship between different taxa encountered from the samples. Due to the CNV of each species, an additional step of genome copy correction could have been performed but was not executed due to it being a potential source of further errors, as there are substantial variations in CNs of rDNA among different fungal taxa (Lofgren *et al.,* 2019). Furthermore, additional background information from the clinical patients, such as the history of infections and the use of antibiotics, could have been combined with the qPCR and qNGS results. To continue the

research on RPL -patients and their fungal profiles, the data could be further compared to healthy pregnant women.

Additional methodologic suggestions for the future would be the volume-based approach by comparing them to those with measured DNA content; Investigation of the effect of non-fungal DNA among the sample, including the optimization of DNA extraction and purification methods for fungi, and a comparison of older samples and recently extracted and purified fresh samples; Experimenting additional primers, or a combination of them, including further investigation of ITS2 -region coding primers, although the selection of ITS1 -region primers was supported, as they were used in the Illumina MiSeq paired-end sequencing of the clinical samples, and would thereby result in a more comparable qNGS approach (Jian *et al.,* 2020); Testing another qPCR reagent; and Switching into digital qPCR and combining the results. Concluding the prospects, the systematic error searching, the troubleshooting matrix, might be worth elaborating, thus the problems that occurred during the optimization work would not need to be explained by variants, such as '*the position of the moon*'.

6 Acknowledgments

The thesis was performed in Docent Anne Salonen's laboratory in Meilahti campus, University of Helsinki, and has been funded by Ilkka Kalliala (Helsinki HUS, HUCS). The clinical samples were from the TOIVE study by Pirkko Peuranpää e*t al.,* (2022), and the genomic DNA of *C. albicans* strains used as a reference fungus, and saliva DNA samples were kindly provided by Dr. Pirjo Pärnänen (University of Helsinki, Helsinki HUS, HUCS, Head and Neck Center). The fecal samples were from the HELMi study. The R-scripts used for the qNGS approach were originally from Schahzad Saquib, TOIVE study. I would like to thank you all for enabling my study.

Special thanks to you, Anne and Rebecka, my supervisors from who I got both autonomy and help when needed, who encouraged me through the process, and who understood my time limitations due to being a mother for small children, and a part-time worker as a midwife, which is another path of my professional career. Thank you, Schahzad Saquib and Roosa Jokela, who walked me through the statistical analysis of the results with R software; Tuomas Heini, who helped me with all the small and bigger problems that occurred in the laboratory phase of my thesis; TOIVE-team members, from whom I got the patient materials of this thesis; and the entire HUMi laboratory members, with whom it was refreshing to take a cup of coffee. Finally, thank you, Mom, for taking care of my little one when he was still too young for daycare, enabling me to start this work, and Tommi, my love and my rock, for supporting me in every manner throughout the process. This thesis is dedicated to you, Pihla, Pauli, and Aaro, the lights of my life.

References

AAT Bioquest. (2024). Real-Time PCR (qPCR). Read 15.8.2024. https://www.aatbio.com/catalog/real-time-pcr-qpcr

Achkar, J.M., Fries, B.C. (2010) Candida infections of the genitourinary tract. *Clin Microbiol Rev,* 23(2), 253–273. 10.1128/CMR.00076-09

Ali, N.A.B.M., Mac Aogain, M., Morales, R.F., Tiew, P.Y., Chotirmall, S.H. (2019). Optimisation and benchmarking of targeted amplicon sequencing for mycobiome analysis of respiratory specimens. *Int J Mol Sci*, 20(20), 4991. 10.3390/ijms20204991

Anderson, I.C., Parkin, P.I (2007). Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *Journal of Microbiological Methods*, 68(2), 248–253. 10.1016/j.mimet.2006.08.005

Bio-Rad, CFX Maestro. (2023). The qPCR software guidelines. CFX Maestro, 1.1–4.1.2433.1219.

Bio-Rad. (2006). Real-Time PCR Applications Guide. Bio-Rad Laboratories, Inc. Read 15.8.2024. https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_5279.pdf

Boskey, E.R., Telsch, K.M., Whaley, K.J., Moench, T.R., Cone, R.A. (1999). Acid production by vaginal flora in vitro is consistent with the rate and extent of vaginal acidification. *Infection and immunity*, 67(10), 5170–5175. https://doi.org/10.1128/IAI.67.10.5170-5175.1999

Bradford, L. L., Ravel, J. (2017). The vaginal mycobiome: A contemporary perspective on fungi in women's health and diseases. *Virulence*, *8*(3), 342–351. https://doi.org/10.1080/21505594.2016.1237332

Busser, F.D., Coelho, V.C., Fonseca, C., Barbaro Del Negro, G.M., Shikanai-Yasuda, M.A., Lopes, M.H., Magri, M.M.C., Teixeira de Freitas, V.L. (2020). A Real Time PCR strategy for the detection and quantification of *Candida albicans* in human blood. *Rev inst Med Trop Sao Paulo*, 62, e9 https://doi.org/10.1590/S1678-9946202062009

Capoor, M.R., Agarwal, S., Yadav, S., Saxena, A.K., Ramesh, V. (2015). Trichosporon mucoides causing onychomycosis in an immunocompetent patient. *International journal of dermatology*, *54*(6), 704–707. https://doi.org/10.1111/ijd.12157

Chakrabarti, R., Schutt, C.E. (2001). The enhancement of PCR amplification by low molecular-weight sulfones. *Gene*, 274(1–2), 293–298. https://doi.org/10.1016/S0378-1119(01)00621-7

Contijoch, E.J., Britton, G.J., Yang, C., Mogno, I., Li, Z., Ng, R., Llewellyn, S.R., Hira, S., Johnson, C., Rabinowitz, K.M., Barkan, R., Dotan, I., Hirten, R.P., Fu, S.C., Luo, Y., Yang, N., Luong, T., Labrias, P.R., Lira, S., Peter, I., Faith, J.J. (2019). Gut microbiota density influences host physiology and is shaped by host and microbial factors. *eLife*, *8*, e40553. https://doi.org/10.7554/eLife.40553

Csárdi, G., Hester, J., Wickham, H., Chang, W., Morgan, M., Tenenbaum, D. (2024). remotes: R Package Installation from Remote. Repositories, Including 'GitHub'. R package version 2.5.0. https://CRAN.R-project.org/package=remotes

da Matta, D.A., Souza, A.C.R., Colombo, A.L. (2017). Revisiting Species Distribution and Antifungal Susceptibility of Candida Bloodstream Isolates from Latin American Medical Centers. *J Fungi (Basel)*, 3(2), 24. 10.3390/jof3020024

Dannemiller, K.C., Lang-Yona, N., Yamamoto, N., Rudich, Y., Peccia, J. (2014). Combining real-time PCR and next-generation DNA sequencing to provide quantitative comparisons of fungal aerosol populations. *Atmospheric Environment,* 84, 113–21. https://doi.org/10.1016/j.atmosenv.2013.11.036

Drell, T., Lillsaar, T., Tummeleht, L., Simm, J., Aaspõllu, A., Väin, E., Saarma, I., Salumets, A., Donders, G.G.G., Metsis, M. (2013). Characterization of the Vaginal Micro- and Mycobiome in Asymptomatic Reproductive-Age Estonian Women. *PLOS ONE,* 8(1), e54379. https://doi.org/10.1371/journal.pone.0054379

Erali, M., Voelkerding, K.V., Wittwer, C.T. (2008). High resolution melting applications for clinical laboratory medicine. *Experimental and molecular pathology*, *85*(1), 50–58. https://doi.org/10.1016/j.yexmp.2008.03.012

Fredricks, D.N., Smith, C., Meier, A. (2005). Comparison of Six DNA Extraction Methods for Recovery of Fungal DNA as Assessed by Quantitative PCR. *Journal of Clinical Microbiology,* 43. https://doi.org/10.1128/jcm.43.10.5122-5128.2005

Gagolewski, M. (2022). stringi: Fast and portable character string processing in R. Journal of Statistical Software, 103(2), 1–59. https://doi.org/10.18637/jss.v103.i02

Gardes, M., Bruns, T.D. (1993). ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology*, 2(2), 113–8. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-294X.1993.tb00005.x.

Gill, S.R., Pop, M., DeBoy, R.T., Eckburg, P.B., Turnbaugh, P.J., Buck S. Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., Nelson, K.E. (2006) Metagenomic Analysis of the Human Distal Gut Microbiome. *Science,* 312(5778), 1355–1359. 10.1126/science.1124234

Godoy-Vitorino, F., Romaguera, J., Zhao, C., Vargas-Robles, D., Ortiz-Morales, G., Vázquez-Sánchez, F., Sanchez-Vázquez, M., de la Garza-Casillas, M., Martinez-Ferrer, M., White, J.R., Bittinger, K., Dominguez-Bello, M.G., Blaser, M.J. (2018). Cervicovaginal Fungi and Bacteria Associated With Cervical Intraepithelial Neoplasia and High-Risk Human Papillomavirus Infections in a Hispanic Population. *Frontiers in microbiology*, 9, 2533. https://doi.org/10.3389/fmicb.2018.02533

Guiver, M., Levi, K., Oppenheim, B.A. (2001). Rapid identification of candida species by TaqMan PCR. *Journal of Clinical Pathology*, 54(5), 362–366. https://doi.org/10.1136/jcp.54.5.362

Guo, R., Zheng, N., Lu, H., Yin, H., Yao, J., Chen, Y. (2012). Increased diversity of fungal flora in the vagina of patients with recurrent vaginal candidiasis and allergic rhinitis. *Microb Ecol*, 64(4), 918–27. 10.1007/s00248-012-0084-0

Haahr, T., Zacho, J., Bräuner, M., Shathmigha, K., Skov Jensen, J., Humaidan, P. (2019). Reproductive outcome of patients undergoing in vitro fertilisation treatment and diagnosed with bacterial vaginosis or abnormal vaginal microbiota: a systematic PRISMA review and meta-analysis. *BJOG: an international journal of obstetrics and gynaecology*, 126(2), 200–207. https://doi.org/10.1111/1471-0528.15178

Han, Y., Liu, Z., Chen, T. (2021). Role of Vaginal Microbiota Dysbiosis in Gynecological Diseases and the Potential Interventions. *Frontiers in Microbiology,* 12, 643422. https://doi.org/10.3389/fmicb.2021.643422

Heid, C.A., Stevens, J., Livak, K.J., Williams, P.M. (1996). Real time quantitative PCR. *Genome Research*, 6(10), 986–994. https://doi-org.libproxy.helsinki.fi/10.1101/gr.6.10.986

Higuchi, R., Fockler, C., Dollinger, G., Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Bio/technology (Nature Publishing Company)*, 11(9), 1026–1030. https://doi-org.libproxy.helsinki.fi/10.1038/nbt0993-1026

Hu, L., Han, B., Tong, Q., Xiao, H., Cao, D. (2020). Detection of Eight Respiratory Bacterial Pathogens Based on Multiplex Real-Time PCR with Fluorescence Melting Curve Analysis. *The Canadian journal of infectious diseases & medical microbiology*. https://doi.org/10.1155/2020/2697230

Huseyin, C.E., Rubio, R. C., O'Sullivan, O., Cotter, P. D., Scanlan, P.D. (2017). The Fungal Frontier: A Comparative Analysis of Methods Used in the Study of the Human Gut Mycobiome. *Frontiers in microbiology*, *8*, 1432. https://doi.org/10.3389/fmicb.2017.01432

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., Korpela, K. (2020). Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS ONE,* 15(1), e0227285. https://doi.org/10.1371/journal.pone.0227285

Jian, C., Salonen, A., Korpela, K. (2021). Commentary: How to Count Our Microbes? The Effect of Different Quantitative Microbiome Profiling Approaches. *Frontiers in cellular and infection microbiology*, *11*, 627910. https://doi.org/10.3389/fcimb.2021.627910

Khan, S.A., Sung, K., Nawaz, M.S. (2011). Detection of aacA-aphD, qacEδ1, marA, floR, and tetA genes from multidrug-resistant bacteria: Comparative analysis of real-time multiplex PCR assays using EvaGreen® and SYBR® Green I dyes. *Molecular and Cellular Probes,* 25(2-3), 78-86. https://doi.org/10.1016/j.mcp.2011.01.004

Khot, P. D., Fredricks, D. N. (2009). PCR-based diagnosis of human fungal infections. *Expert review of anti-infective therapy*, 7(10), 1201–1221. https://doi.org/10.1586/eri.09.104

Kubista, M., Andrade, J.M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., Sindelka, R., Sjöback, R., Sjögreen, B., Strömbom, L., Ståhlberg, A., Zoric, N. (2006). The real-time polymerase chain reaction. *Molecular aspects of medicine*, 27(2-3), 95–125. https://doi.org/10.1016/j.mam.2005.12.007

Lehtoranta, L., Hibberd, A.A., Yeung, N., Laitila, A., Maukonen, J., Ouwehand, A.C. (2021). Characterization of vaginal fungal communities in healthy women and women with bacterial vaginosis (BV); a pilot study. *Microbial pathogenesis*, 161(Pt A), 105055. https://doi.org/10.1016/j.micpath.2021.105055

Liu, F., Yang, S., Yang, Z., Zhou, P., Peng, T., Yin, J., Ye, Z., Shan, H., Yu, Y., Li, R. (2022). An Altered Microbiota in the Lower and Upper Female Reproductive Tract of Women with Recurrent Spontaneous Abortion. *Microbiology Spectrum,* 10, e00462–22. https://doi.org/10.1128/spectrum.00462-22

Lofgren, L.A., Uehling, J.K., Branco, S., Bruns, T.D., Martin, F., Kennedy, P.G. (2019). Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Molecular Ecology*, *28*(4), 721–730. https://doi.org/10.1111/mec.14995

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. *Nat Commun,* 11, 940. https://doi.org/10.1038/s41467-020-14677-3

Mahajan, M. (2022). Schizophyllum commune. *Emerging Infectious Diseases*, 28(3), 725. https://doi.org/10.3201/eid2803.211051

McMurdie, P.J. Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE,* 8(4), e61217. http://dx.plos.org/10.1371/journal.pone.0061217

McQueen, D.B., Maniar, K.P., Hutchinson, A., Confino, R., Bernardi, L., Pavone, M.E. (2021). Redefining chronic endometritis: the importance of endometrial stromal changes. *Fertility and Sterility,* 116(3), 855–61. 10.1016/j.fertnstert.2021.04.036

Mitchell, C.M., Haick, A., Nkwopara, E., Garcia, R., Rendi, M., Agnew, K., Fredricks, D.N., Eschenbach, D. (2015). Colonization of the upper genital tract by vaginal bacterial species in nonpregnant women. *American Journal of Obstetrics and Gynecology*. 212(5), 611.e1–611.e9. https://doi.org/10.1016/j.ajog.2014.11.043

Morgan, M., Ramos, M. (2024). BiocManager: Access the Bioconductor Project Package Repository. R package version 1.30.24. https://CRAN.R-project.org/package=BiocManager

Nasiri-Jahrodi, A., Sheikholeslami, F.M., Barati, M. (2023). Cladosporium tenuissimum-induced sinusitis in a woman with immune-deficiency disorder. *Brazilian journal of microbiology,* 54(2), 637–643. https://doi.org/10.1007/s42770-023-00978-4

Nejad, E.E., Almani, P.G.N., Mohammadi, M.A., Salari, S. (2020). Molecular identification of *Candida* isolates by Real-time PCR-high-resolution melting analysis and investigation of the genetic diversity of *Candida* species. *J Clin lab analysis,* 34(10), e23444. https://doi.org/10.1002/jcla.23444

NIH HMP. (2009). Working Group; Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J.A., Bonazzi, V., McEwen, J.E., Wetterstrand, K.A., Deal, C., Baker, C.C., Di Francesco, V., Howcroft, T.K., Karp, R.W., Lunsford, R.D., Wellington, C.R., Belachew, T., Wright, M., Giblin, C., David, H., Mills, M., Salomon, R., Mullins, C., Akolkar, B., Begg, L., Davis, C., Grandison, L., Humble, M., Khalsa, J ., Little, A.R., Peavy, H., Pontzer, C., Portnoy, M., Sayre, M.H., Starke-Reed, P., Zakhari, S., Read, J., Watson, B., Guyer, M. The NIH Human Microbiome Project. *Genome Res*, 19(12), 2317–23. 10.1101/gr.096651.109

Nilsson, R.H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., Tedersoo, L. (2019). Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews. Microbiology*, *17*(2), 95–109. https://doi.org/10.1038/s41579-018-0116-y

OPATHY Consortium, Gabaldon T. (2019). Recent trends in molecular diagnostics of yeast infections: from PCR to NGS. *FEMS Microbiol Rev*, 43(5), 517–47. 10.1093/femsre/fuz015

Pedersen, T., Nicolae, B., François, R. (2022). farver: High Performance Colour Space Manipulation. R package version 2.1.1, https://CRAN.R-project.org/package=farver

Premamalini, T., Ambujavalli, B.T., Anitha, S., Somu, L., Kindo, A.J. (2011). Schizophyllum commune a causative agent of fungal sinusitis: a case report. *Case reports in infectious diseases*, 2011, 821259. https://doi.org/10.1155/2011/821259

Peuranpää, P., Holster, T., Saquib, S., Kalliala, I., Tiitinen, A., Salonen, A., Hautamäki, H. (2022) Female reproductive tract microbiota and recurrent pregnancy loss: a nested case-control study. *RBMO*, 45(5). https://doi.org/10.1016/j.rbmo.2022.06.008

Posit Team (2024). RStudio: Integrated Development Environment for R. Version 2024.4.2.764. Posit Software, PBC, Boston, MA. http://www.posit.co/

R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Revelle, W. (2024). psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.4.6. https://CRAN.R-project.org/package=psych

Rinker, T. W. Kurkiewicz, D. (2017). pacman: Package Management for R version 0.5.0. Buffalo, New York. http://github.com/trinker/pacman

Rinttilä. T. (2011). REAL-TIME PCR – A Molecular Approach to Investigate the Role of Intestinal Microbiota in the Pathophysiology of Irritable Bowel Syndrome. Academic Dissertation. Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki. https://helda.helsinki.fi/server/api/core/bitstreams/eadfa430-d406-4a54-a259-4e79746c7c91/content

Rowan-Nash, A.D., Korry, B.J., Mylonakis, E., Belenky, P. (2019). Cross-Domain and Viral Interactions in the Microbiome. *Microbiology and Molecular Biology Reviews,* 83(1). https://doi.org/10.1128/mmbr.00044-18

Rychlik, W. (1995). Selection of primers for polymerase chain reaction. *Mol Biotechnol* 3, 129–134 https://doi-org.libproxy.helsinki.fi/10.1007/BF02789108

Stephenson, F.H. (2003). Quantitation of Nucleic Acids, Calculations for Molecular Biology and Biotechnology, *Academic Press,* 2003, 90–108. https://doi.org/10.1016/B978-012665751-7/50046-9

Stop Neglecting Fungi. (2017). Editorial, Nature Microbiology. *Nature Microbiology,* 2: 17120. https://doi.org/10.1038/nmicrobiol.2017.120

Thermo Fisher Scientific. (2004). 10 steps to improve pipetting accuracy. Proper pipetting technique. The fundamentals of pipetting. Thermo Fisher Scientific. Read 8.5.2024. https://www.thermofisher.com/fi/en/home/life-science/lab-plasticware-supplies/lab-plasticware-supplies-learning-center/lab-plasticware-supplies-resource-library/fundamentals-of-pipetting/proper-pipetting-techniques/10-steps-to-improve-pipetting-accuracy.html

Thermo Scientific Chemicals. (2024). Dimethyl sulfoxide, Bioreagent, Thermo Scientific Chemicals. Thermo Fisher Scientific. Read: 1.8.2024. https://www.thermofisher.com/order/catalog/product/J66650.AK

Thian Lung, T.L., Pei Pei, C., Kee Peng, N., Heng Fong, S. (2014). Detection of medically important Candida species by absolute quantification Real-time polymerase chain reaction. *Jundishapur Journal of Microbiology,* 8(1), e14940. https://doi.org/10.5812/jjm.14940

Tiew, P.Y., Mac Aogain, M., Ali, N.A.B.M., Thng, K.X., Goh, K., Lau, K.J.X., Chotirmall, S.H. (2020). The Mycobiome in Health and Disease: Emerging Concepts, Methodologies and Challenges. *Mycopathologia,* 185, 207–231. https://doi.org/10.1007/s11046-019-00413-z

Trofa, D., Gácser, A., Nosanchuk, J.D. (2008). Candida parapsilosis, an emerging fungal pathogen. *Clinical microbiology reviews*, 21(4), 606–625. https://doi.org/10.1128/CMR.00013-08

Virtanen, S., Rantsi, T., Virtanen, A., Kervinen, K., Nieminen, P., Kalliala, I., Salonen, A. (2019). Vaginal Microbiota Composition Correlates Between Pap Smear Microscopy and Next Generation Sequencing and Associates to Socioeconomic Status. *Scientific Reports,* 9,7750. https://doi.org/10.1038/s41598-019-44157-8

Virtanen, S., Saqib, S., Kanerva, T., Nieminen, P., Kalliala, I., Salonen, A. (2023). Metagenome-validated Parallel Amplicon Sequencing and Text Mining-based Annotations for Simultaneous Profiling of Bacteria and Fungi: Vaginal Microbiota and Mycobiota in Healthy Women. *Research Square*. 18 July 2023, PREPRINT (Version 1). https://doi.org/10.21203/rs.3.rs-3166913/v1

White, T.J., Bruns, T., Lee, S., Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A guide to methods and applications/Academic Press, Inc.* 315–22. https://doi.org/10.1016/B978-0-12-372180-8.50042-1

Wickham, C. (2024). munsell: Utilities for Using Munsell Colours. R package version 0.5.1. https://CRAN.R-project.org/package=munsell

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wiesmann, C., Lehr, K., Kupcinskas, J., Vilchez-Vargas, R., Link, A. (2022). Primers matter: Influence of the primer selection on human fungal detection using high throughput sequencing. *Gut Microbes*, 14(1) https://doi.org/10.1080/19490976.2022.2110638

Wilhelm, J., Pingoud, A. (2003), Real-Time Polymerase Chain Reaction. *Chem Bio Chem,* 4(11), 1120-1128. https://doi-org.libproxy.helsinki.fi/10.1002/cbic.200300662

Wintzingerode, F.V., Göbel, U.B., Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis, *FEMS Microbiology Reviews*, 21(3), 213–229. https://doi.org/10.1111/j.1574-6976.1997.tb00351.x

Zhang, B., Brock, M., Arana, C., Dende, C., van Oers, N.S., Hooper, L.V., Raj, P. (2021). Impact of Bead-Beating Intensity on the Genus- and Species-Level Characterization of the Gut Microbiome Using Amplicon and Complete 16S rRNA Gene Sequencing. *Frontiers in cellular and infection microbiology*, 11, 678522. https://doi.org/10.3389/fcimb.2021.678522

Zhu, H., Zhang, H., Xu, Y., Laššáková, S., Korabečná, M., & Neužil, P. (2020). PCR past, present and future. *BioTechniques*, *69*(4), 317–325. https://doi.org/10.2144/btn-2020-0057

Zoll, J., Snelders, E., Verweij, P.E., Melchers, W.J. (2016). Next-Generation Sequencing in the Mycology Lab. *Curr Fungal Infect Rep*, 10, 37–42. 10.1007/s12281-016-0253-6

Appendices

Appendix 1. Materials and equipment

Substance	Origin / LOT/ Batch # /information
Reagents:	
AccuGENE® Molecular Biology Water	Lonza LOT: 22MB025
100% DMSO	Thermo Scientific, Thermo Fisher LOT: 00691653
HOT FIREPol® EvaGreen® qPCR Mix Plus (no ROX) 5x	Solis BioDyne, Estonia LOT: 08251680.3
Primers: (name, direction, region, sequence 5'-3')	
ITS1F forward (ITS1 -region) 5' CTTGGTCATTTAGAGGAAGTAA 3'	Sigma-Aldrich, Merck stock 1: HA16117389 stock 2: HA16468144
ITS2 reverse (ITS1 -region) 5' GCTGCGTTCTTCATCGATGC 3'	Sigma-Aldrich, Merck stock 1: HA16117388 stock 2: HA16468143
ITS3 forward (ITS2 -region) (50mM) 5' GCATCGATGAAGAACGCAGC 3'	Ventin-Holmberg, R., Folkhälsan
ITS4 reverse (ITS2 -region) (50mM) 5' CTTGGTCATTTAGAGGAAGTAA 3'	Ventin-Holmberg, R., Folkhälsan
C. albicans forward 5' GGGTTTGCTTGAAAGACGGTA 3'	Sigma-Aldrich, Merck HA16117386
C. albicans reverse 5' TTGAAGATATACGTGGTGGACGTTA 3'	Sigma-Aldrich, Merck HA16117387
Electrophoresis:	
Agarose, Molecular grade	Bioline LOT: ES520-B053850
Advance DNA stain Midori Green	Nippon Genetics Europe GmbH by BIOTOP LOT 422MG13013
6x DNA Loading Dye	Thermo Fisher Scientific LOT: 2793361 REF: R0611
FastRuler ™ Low Range DNA Ladder, ready-to-use	Thermo Fisher Scientific Snellman, L., HLA Lokki
Fluorophores:	
Quant-iT™ PicoGreen™ dsDNA Assay Kit (20X TE buffer, DNA standard, reagent)	Invitrogen, Thermo Fisher Scientific LOT: 2652720, REFP7589
Qubit ™ dsDNA HS Assay Kit (Buffer, reagent, Standard #1 (0 ng/μl), Standard #2 (10 ng/μl))	Invitrogen, Thermo Fisher Scientific LOT: 2064448, REF Q32854
Templates	
Candida albicans (325.8 ng/μl) #1	Pärnänen, P., HUS, Head and Neck Center, 2.11.2022
Candida albicans (401.9 ng/μl) #2	Pärnänen, P., HUS, Head and Neck Center, 2.11.2022
human DNA (107.0 ng/μl) #1	Snellman, L., HLA Lokki (30.11.2022)
human DNA (50 ng/μl) #2	Snellman, L., HLA Lokki
Vagina swab DNA: V1, V2, V3, V4, V5 (5 ng/μl)	TOIVE-project by Virtanen, S. & Glazer-Livson, S. (23.03.2017)
Endometrial DNA test samples: T1A (219.4 ng/μl); T2C (94.7 ng/μl); T3A (205.6 ng/μl)	TOIVE-project by Virtanen, S. & Glazer-Livson, S. (2017), extracted by Kanerva, T. (27.11.2020)
Fecal DNA: F1, F2, F3, F4, F5 (5 ng/ μl)	Salonen, A., Human Early Life Microbiota (HELMi) - study

Saliva DNA: S1 (52.2 ng/μl), S2 (382.0 ng/μl)	Pärnänen, P. HUS, Head and Neck Center, 2.11.2022
Baker's yeast block (*Saccharomyces cerevisiae*)	bought 11.1.23 from a supermarket
Solutions, Buffers, Reagents	
buffer 1 x TE, pH 8.00	19.10.20 ED
buffer 1 x TE, pH 8.00	MN
DNA extraction solutions:	
10M Sodium acetate, pH 5.2	24.9.20 ED
1M Tris-HCl, pH 8.0	21.9.22 AP
5M NaCl	3.5.22 EL
0.5M EDTA, pH 8.0	8.11.22 EH
20% SDS	4.1.23 EH
RBB Lysis buffer	14.11.22 EH
Isopropanol	16.9.22 AP
70% EtOH	12.1.23 MN
1 x TE, pH 8.0	12.1.23 MN
Devices and equipment:	
qPCR thermal cycler (primary)	Bio-Rad CFX96™ Optics Module Real-Time System C1000 Touch ™ Thermal Cycler, Singapore Bio-Rad Maestro 1.1-4.1.2433.1219
qPCR thermal cycler (secondary)	Bio-Rad CFX96™ Real-Time System C1000 Touch ™ Thermal Cycler, Singapore Bio-Rad Maestro 1.1-3.1.1517.0823
Quant-iT™ plate reader	Hidex Sense Microplate Reader, Type 425-301, Finland. Software_0.5.35.0
Qubit™ fluorometer	Qubit™ 4 fluorometer, Invitrogen, Thermo Fisher Scientific, Singapore. Program: 1x ds DNA High Sensitivity
Electrophoresis	Bio-Rad Power Pac 200, Bio-Rad, USA
Molecular Imager®	Gel Doc ™ XR+, Image Lab ™ Software, Bio-Rad, USA
DNA homogenizer	Fast Prep 96 ™ MP Biomechanicals, Inc., USA
Incubator -shaker	Vortemp 56 , Labnet International, USA
Centrifuges	Mikro 200R (+4° C) zentrifugen, Hettich, Germany
Purification automate	KingFisher Flex, Type 711, REF: 5400630, Thermo Fisher Scientific, Finland
Spinner	Micro centrifuge, LLG® Labware, Germany
Plate centrifuge	Axygen®, Mexico
Vortex	Scientific Industries ™ Vortex-genie® 2, Scientific Industries, USA
Laminar flow cabinet	Biowizard Platinum, Kojair ® Tech Oy, Finland
Pipettes	Finnpipette ® F2 (0.2–2 ml; 2–10 ml; 20–200 ml; 100–1000 ml), Thermo Fisher Scientific Finland. Multichannel: Sartorius Biohit Proline Plus (10–100 ml), Sartorius Biohit mLINE (0–10 ml), Sartorius, Germany
Pipette tips	Fisher brand Sure One, Fisher Scientific, Thermo Fisher Scientific, UK
Plates and plate cover	96-well semi-skirted clear PCR plates, standard profile, qPCR seal, 4titude®, Biotop, UK