16S rRNA gene profiling: Direct and indirect applications for clinical microbiology

Munckhof, E.H.A. van den

CHAPTER 10

Summary
Nederlandse samenvatting
List of publications
Curriculum vitae
Dankwoord

SUMMARY

The human body is colonised by a vast number of archaea, bacteria, bacteriophages, fungi, protozoa and viruses, collectively referred to as the human microbiota. These microorganisms play a critical role in important physiological processes of the human body. Disturbance of the microbiota has been associated with a wide range of diseases and is therefore the focus of a growing number of research initiatives. The boost in microbiota research resulted from the development of the high-throughput sequencing platforms, which enabled easy detection of a theoretically unlimited number of bacteria based on the 16S ribosomal RNA (rRNA) gene. This gene is present in all bacteria and demonstrates considerable sequence diversity among different bacteria, making it a useful marker for identification of different bacteria. The most widely used approach is based on amplification and subsequently sequencing of the hypervariable regions of the 16S rRNA gene. The collection of obtained sequences is representative of the bacterial community as a whole in terms of membership (those present), and relative abundances, (how many 16S rRNA genes of a member is present compared to the total). This method can also be very valuable for clinical microbiology as it enables detection and identification of a theoretically unlimited number of bacteria present in a specimen and permits (semi)quantitative information about the composition of a microbial community. Although 16S rRNA gene profiling is a very straightforward method, its usefulness may be reduced due to its limited resolution to genus rather than the species level. In this thesis, we aimed to evaluate the clinical usefulness of 16S rRNA gene profiling. In addition, we explored whether clinically relevant cut-off values for interpretation of the sequencing data could be defined. Chapter 1 serves as a general introduction to the human microbiota and its association with disease. In addition, the different high-throughput sequencing methods for identification of bacteria are outlined. Finally, the potentially added value of 16S rRNA gene profiling for the clinical microbiology is explained.

In the first part of this thesis, we focused on using 16S rRNA gene profiling as a direct diagnostic tool for identification of clinically relevant bacteria. In routine clinical microbiology, standard identification of clinically relevant microorganisms from specimens is based on culture of bacteria followed by species identification with matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) technology and antibiotic susceptibility testing of the cultured putative causative microorganism. In **Chapter 2**, we questioned whether a stepwise approach using 16S rRNA gene profiling followed by species-specific real-time quantitative PCRs (qPCRs) and/or culture has the potential to be a more accurate and efficient diagnostic approach than the routine diagnostic approach based on culture. To this end, 62 sputum samples from patients with a suspected lower respiratory tract infection (LRTI) were analysed by the routine culture-based approach, 16S rRNA gene profiling and multiple species-specific qPCRs. We found that 16S rRNA gene

profiling identified 110 potentially pathogenic genera in sputum while 37 pathogens were found using a routine culture-based approach. This difference is caused by the fact that 16S rRNA gene profiling can detect all bacteria present in a specimen while culture is often more specific for a certain class of pathogenic species. Non-pathogenic species are not investigated by routine culture. Since non-pathogenic and pathogenic species can both colonise the respiratory tract without being involved in an infection, it is very important to provide a clinical interpretation of 16S rRNA gene profiling data in measures of relative abundance. This turned out to be a challenge. Furthermore, classification to the genus level is not always sufficient to identify the causative microorganism. Based on these observations, we concluded that 16S rRNA gene profiling provides a more complete characterisation of polymicrobial communities and can be used to identify potentially pathogenic genera in sputum, but only when combined with species-specific qPCR to achieve the needed resolution to the species level. Selective culture was added to the combination of methods to be performed to characterise phenotypic aspects such as the antibiotic susceptibility of a pathogen. Unfortunately, faster, better and cheaper technology for sequencing, which can be expected in the near future, are required before 16S rRNA gene profiling will be adopted by the routine clinical microbiology. Taken together, the stepwise approach using 16S rRNA gene profiling, species-specific qPCRs and culture has the potential to be used in clinical settings for the diagnosis of LRTIs in the near future when challenges of clinical data interpretation, turnaround time and cost of 16S rRNA gene profiling are overcome.

In the second part of this thesis, we focussed on using 16S rRNA gene profiling as an indirect tool for evaluation of diagnostic methods and therapies. In **Chapter 3**, we explored using 16S rRNA gene profiling as an alternative reference test to evaluate existing tests for the diagnosis of bacterial vaginosis (BV). To discriminate between a normal vaginal microbiota and BV, we used a previously reported cut-off value of 47% relative abundance of *Lactobacillus*. *Lactobacillus* dominated vaginal microbiota profiles (with ≥47% relative abundance of *Lactobacillus*) were categorised as normal vaginal microbiota and microbiota profiles with less *Lactobacillus* (<47% relative abundance) with mainly anaerobes as microbiota associated with BV. These definitions were used to categorise 115 vaginal swabs. For each swab, the result was compared with the outcome of five diagnostic methods: (i) a Gram stain scoring system based on the quantitative assessment of three different bacterial morphotypes, called the Nugent score; (ii) Amsel's clinical criteria; (iii) culture and (iv) two commercially available qPCRs. For the current gold standard (Nugent score), we found a specificity (true negative rate) of 100% and a sensitivity (true positive rate) of 64% for the diagnosis of BV using 16S rRNA gene profiling as the reference test. This means that the Nugent score is highly accurate for identification of women without BV but that it misses 36% of the women with BV. Evaluation of the discrepant results between 16S rRNA gene profiling and the Nugent score showed that the unresolved clinical importance of the intermediate

category was mainly responsible for the poor sensitivity of the current gold standard. For all five diagnostic methods, we found a specificity of at least 92%, suggesting that ≥47% relative abundance of *Lactobacillus* was highly accurate for identification of women without BV. However, the sensitivities of the five diagnostic methods ranged between 39% and 81%. Discrepancy analysis showed that at least two methods were in agreement with 16S rRNA gene profiling in 81% of the swabs categorised as microbiota associated with BV. The microbiota of the remaining swabs was characterised by a relatively high abundance of bacteria weakly associated with BV and/or not commonly found in the vagina. Based on these data, we consider 16S rRNA gene profiling as a good alternative for the current gold standard to evaluate BV diagnostic methods. Currently, the definition of <47% relative abundance of *Lactobacillus* and mainly anaerobes is sufficient to identify women with BV, but more research is required to identify the role of specific bacteria in the vaginal microbiota.

Next, we changed our focus from the evaluation of diagnostic methods to the evaluation of treatments. We discussed the usefulness of 16S rRNA gene profiling as a tool to determine the impact of treatment on polymicrobial communities. First, we focussed on the treatment of BV because up to 40% of the patients treated for BV return to their physician with persistent symptoms of abnormal vaginal discharge. Since the diagnosis of BV is challenging due to the subjective scoring of the Nugent test, the wide spectrum of anaerobes associated with BV and the limited number of bacteria that can be effectively cultivated or targeted by qPCR, misdiagnosis rather than treatment failure might be the cause of the persistent symptoms. In **Chapter 4**, we determined the best algorithm for the diagnosis of BV based on clinical symptoms and signs, bedside and/or laboratory test findings. To achieve this, tree-based classification analysis was performed on the clinical data and bedside test results obtained from 56 women with abnormal vaginal discharge. Laboratory tests were compared. We found that the best algorithm was to screen with two bedside tests, and if positive, to confirm with qPCR (sensitivity 94%; specificity 97%). Secondly, we performed retrospective analysis to determine whether implementation of the developed algorithm would have reduced the number of patients that returned to their physician with persistent symptoms. We found that 30% of the persistent symptoms were caused by misdiagnosis and another 30% by the occurrence of a different infection or treatment failure. For both analyses, 16S rRNA gene profiling and other reference tests were needed to diagnose BV as well as fungal and protozoan infections that may cause abnormal vaginal discharge. These data show that 16S rRNA gene profiling is not sufficient as stand alone test.

The next step is to understand why treatment failure occurs in some women with BV. In **Chapter 5**, we performed comprehensive analysis on the 16S rRNA gene profiling data obtained during two subsequent hospital visits from 21 women with clinically diagnosed BV. We observed two bacterial community types before and after antibiotic treatment with significantly different bacterial diversity. The community state types identified in women before treatment were driven by *Lactobacillus*, *Gardnerella* and *Atopobium* or by

Gardnerella, *Atopobium*, *Prevotella* and *Sneathia* while the community state types identified after treatment were either driven by *Lactobacillus* or by multiple bacteria. Unfortunately, we found no association between the community state types before or after treatment and the clinical outcome defined as: no, or persistent symptoms. It might be the case that treatment failure involves bacterial strains with high virulence potential, which cannot be differentiated from strains with low virulence potential based on the 16S rRNA gene. Taken together, the 16S rRNA gene profiling contributes to better understanding as to why women return to their physician with persistent complaints of abnormal vaginal discharge but lack resolution to discriminate between virulent and non-virulent bacterial strains.

In **Chapter 6**, we further explored the use of 16S rRNA gene profiling to determine the impact of treatment on the microbiota. We focussed on how 16S rRNA gene profiling can be applied in clinical trials to assess the effectiveness of new drugs in patients with atopic dermatitis (AD). For these clinical trials, the skin microbiota is an interesting biomarker since AD is associated with *S. aureus* colonisation and reduced microbial diversity. The microbiota composition of healthy skin may vary significantly between humans (inter-individual variation) and within a human over time (intra-individual variability) due to host and environmental factors, such as antibiotic exposure, hygiene and lifestyle. Importantly, skin affected by AD is likely to have larger inter- and intra-patient variability of the skin microbiota, implying the need for frequent sampling when evaluating the impact of treatment on the affected skin microbiota. Accordingly, we analysed 16S rRNA gene profiling, culture and qPCR data obtained from skin swabs collected weekly from 20 patients with mild to moderate AD. To measure the inter- and intra-patient variability over a period of 42 days, the standard deviation of the mean was calculated for microbial diversity, relative abundance of *Staphylococcus* spp. and *S. aureus* concentration. For microbial diversity of lesional skin, we observed a high inter-patient variability (36-46% vs 16-28%) and a wide range of intra-patient variability (7-92% vs 4-29%) compared to unaffected skin. For the relative abundance of *Staphylococcus* spp. and *S. aureus* concentration of affected skin comparable results were found. In addition, we were able to define three patient groups with different microbiological phenotypes. We categorized these patients according to their microbial diversity of skin microbiota, the relative abundance of the genus *Staphylococcus*/*S. aureus* and microbiota and temporal variability. Groups I and II could be described as high relative abundance of *S. aureus*, low microbial diversity and either microbiologically stable, or unstable over time, respectively. In contrast, the affected skin microbiota of group III was closely related to their unaffected skin microbiota with a low relative abundance of *S. aureus* and high microbial diversity. Based on these observations, we concluded that microbiological phenotype stratification with 16S rRNA gene profiling and frequent sampling is required to analyse the microbiota in clinical AD trials.

In the third part of this thesis, we used 16S rRNA gene profiling as a research tool to study the link between microbiota and disease. Previously, we explored whether 16S rRNA gene profiling added value for clinical microbiology as a direct or indirect diagnostic tool. To address this, studies focussing on different diseases were performed, including BV and AD. Currently, microbiota research looking for a link between other diseases and the human microbiota is increasing exponentially. However, we believe more basic research is needed before microbiota findings can be safely and widely applied into the clinical setting. To illustrate this, we used 16S rRNA gene profiling as a research tool to study the link between the microbiota and two different medical conditions. In **Chapter 7**, we explored whether specific nasal and/or oropharyngeal microbiota profiles can be associated with increased age and RTIs in adults. To achieve this, 16S rRNA gene profiling was performed on nasal and oropharyngeal swabs of 152 controls and 152 patients with an upper or lower respiratory tract infection. We defined eight nasal and nine oropharyngeal microbiota clusters. Unfortunately, we were unable to clarify why the elderly are more susceptible to RTIs. Unexpectedly, we showed that nasal microbiota dominated by the genus *Moraxella* (presumably *Moraxella nonliquefaciens*) is associated with respiratory health in the elderly population. The nasal microbiota cluster dominated by the genus *Moraxella* was significantly more prevalent in the healthy elderly population ($p = 0.002$) compared to the healthy middle aged adults, and it was significantly less prevalent in the elderly with a LRTI ($p = 0.001$) compared to the healthy elderly population. However, some caution is necessary when translating this research based on 16S rRNA gene profiling into the clinic. We collected data at one timepoint. This means that we cannot distinguish whether *M. nonliquefaciens* was less prevalent in elderly patients due to its protective nature or that the infection changed the microbiota. Furthermore, fundamental research to explore the protective properties of *M. nonliquefaciens* is needed.

In **Chapter 8**, we explored whether the balanopreputial and urine microbiota are associated with a chronic inflammatory, scarring dermatosis associated with penile cancer, called male genital lichen sclerosus (MGLSc). Accordingly, 16S rRNA gene profiling was performed on balanopreputial swabs and urine of 40 controls and 40 men with MGLSc. For the balanopreputial sac, we observed a difference in microbiota profiles between the men with MGLSc and controls. The relative abundance of the genus *Finegoldia* was decreased in men with MGLSc (median relative abundance of 9% vs 28%) while the relative abundance of the genus *Prevotella* was increased (median relative abundance of 20% vs 4%). Both the prevalence (50% vs 15%) and relative abundance (median relative abundance of 4% vs 0%) for the genus *Fusobacterium* were increased in men with MGLSc. *Fusobacterium* spp., particularly *Fusobacterium nucleatum*, have been associated with inflammatory diseases such as periodontitis and inflammatory bowel disease and some cancers such as colorectal and

oral squamous cell carcinoma. Longitudinal data is lacking to establish whether dysbiosis of the balanopreputial sac microbiota is the cause or consequence of the disease. Furthermore, involvement of balanopreputial sac dysbiosis in the aetiology and pathogenesis of MGLSc, with a specific role for *F. nucleatum*, should be confirmed by reproduction of the disease *In vitro* or in animal models. Thereafter, interventional research can be performed to develop treatment for specific diseases.

Finally, in **Chapter 9** we evaluated the results of the studies described in this thesis and discussed their implications for future research. In this final chapter, we highlighted the clinical application for which 16S rRNA gene profiling should be considered and the corresponding data interpretation challenges. Furthermore, we speculated about the usefulness of whole metagenomic shotgun sequencing for clinical microbiology compared to 16S rRNA gene profiling. The clinical value of 16S rRNA gene profiling will probably become clear in the forthcoming years.