# CLINICAL UTILITY OF GENOMIC SEQUENCING FOR OCULAR INFLAMMATORY AND INFECTIOUS DISEASES

by

Liying Low

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Academic Unit of Ophthalmology
Institute of Inflammation and Ageing
College of Medical and Dental Sciences
University of Birmingham
January 2021

UNIVERSITY OF
BIRMINGHAM

## University of Birmingham Research Archive

### e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

## Abstract

The human body is thought to be an ecosystem of mutually beneficial interactions between the human cells and the microorganisms that live in and on our bodies. The largest and most diverse microbial community reside in the gut. Collectively, the genomes of these microbes are known as the gut microbiome. The gut microbiome is important in the development and maintenance of a healthy immune system. Additionally, the gut microbiome regulation of secretory immunoglobulin A and polymorphonuclear cells recruitment on the ocular surface contributes to immune protection against ocular infections. Perturbations in the gut microbiome, also known as gut dysbiosis, has been linked to systemic inflammation elsewhere in the body (extraintestinal sites) and may play a role in ocular inflammation and increased susceptibility to ocular infections. The current gold standard for diagnosis of most ocular infections is microbial culture, however, it is time-consuming and approximately 40% of clinically suspected cases are culture-negative.

Next generation sequencing technologies have revolutionised genomic research and has the potential to provide comprehensive microbial and host data with a wide range of translational clinical applications, from microbiome analysis in diseased and healthy states, to rapid identification of pathogens in infectious diseases. However, these sequencing methodologies have yet to be adopted into routine clinical practice, in part due to the highly contextual and bidirectional host-microbiome interactions, as exemplified by the role of the gut microbiome in modulating host immune homeostasis (i.e. commensal versus pathogenic organisms), and the lack of optimised metagenomic protocols for technically-challenging clinical samples such as ocular samples that are often minute and have high background host DNA contamination.

The general hypothesis investigated in this thesis was the contribution of the microbiomes to ocular disease and the potential application of novel sequencing technologies for rapid identification of these microbiomes, both pathogenic and commensals, in ocular inflammatory diseases and infections. I applied Illumina 16S rRNA gene sequencing to characterise the gut microbiome profile of patients with ocular mucous membrane pemphigoid (OcMMP), a devastating mucosal scarring autoimmune condition of unknown cause. I found that patients with OcMMP had lower alpha diversity compared to healthy controls, and that gut dysbiosis correlated with ocular inflammation, and treatment with immunosuppressive therapies. This broad sampling of the gut microbiome provides a framework for the integration of deeper multi-omic datasets to unravel the complex host immune-microbiome interactions in autoimmunity.

Using the novel nanopore sequencing technology, I extended the research to identify specific microbial species causing ocular infections. Microbial keratitis, infection of the cornea, is a leading cause of preventable blindness worldwide. Using a defined mock community and *ex-vivo* porcine eye model, I optimised the sampling method and developed a bioinformatics pipeline for full-length 16S rRNA nanopore sequencing, and was successful in implementing this sequencing workflow directly onto clinical samples of patients presenting with microbial keratitis.

Next, I evaluated the clinical utility of nanopore sequencing in endophthalmitis, a severe, sight-threatening intraocular infection. The sensitivity of metagenomic sequencing in detecting pathogens decreases with high host DNA contamination. Therefore, I examined the

effect of saponification on host DNA depletion by qPCR using murine peripheral blood mononuclear cells spiked with serial dilutions of *E.coli* cells and found that there was a 1500-fold depletion of host DNA with very minimal bacterial DNA loss between saponified and un-saponified samples. I compared the effects of saponification on human DNA depletion by whole genome nanopore sequencing using intraocular fluid biopsy samples from a subset of patients with endophthalmitis and found that saponification increased the proportion of bacterial to human reads. Finally, I performed full-length 16S rRNA nanopore sequencing (16S nanopore), whole genome nanopore sequencing (WGS nanopore) and Illumina whole genome sequencing (Illumina WGS) directly on intraocular fluid biopsy samples of patients with clinically suspected endophthalmitis, showing that whole genome nanopore sequencing performed better than 16S nanopore and Illumina WGS in detecting the predominant organism in endophthalmitis. Taken together, this suggests that with further optimisation, the future use of the portable, real-time nanopore sequencing technology as a point-of-care testing for microbial identification and quantification may be beneficial in ocular infections.

## Dedication

*In loving memory of Carolyn Wallace,*
*a beautiful soul*

## Acknowledgements

I am most grateful to my esteemed supervisors: Miss Saaeha Rauz, for her academic rigour, tireless effort and meticulous attention to detail; Dr Graham Wallace, whose keen intellect, selfless time and care provided the academic freedom to explore new avenues of research; Professor Philip I Murray, for his wise guidance and support; and Professor Nicholas J Loman, whose scientific knowledge provided insightful discussion and critical suggestions on bioinformatics and sequencing.

The work undertaken in this thesis has been made possible through funding from the Fight for Sight Clinical Research Training Fellowship (Reference 1840/1841) and the Royal College of Ophthalmologists Ophthalmology Trainee Research Network Award (Reference: 24RCO3).

I would like to express my appreciation to Professor Russell Van Gelder and his team for the synergistic collaboration during my 3-month research fellowship at the University of Washington, Seattle, USA. I would like to thank Professor Janet Lord for providing institutional support and access to the *Birmingham 1000 Elders* group, and the research team at the Academic Unit of Ophthalmology and Birmingham and Midland Eye Centre (Dr John O'Neill, Dr Kusy Suleiman, the *West Midlands Collaborative Ophthalmology Network for Clinical Effectiveness & Research*) for their technical and logistical assistance. The Medical Research Council-funded Cloud Infrastructure for Big Data Microbial Bioinformatics (CLIMB-BIG-DATA project) provided the computational infrastructure necessary for the storage and processing of the vast amount of genomic data generated from this project – many thanks to Mr Radoslaw Poplawski for administrative support.

My deepest gratitude to the patients and research participants for their enthusiasm and support, which were fundamental to this study. Thank you to Professor Carrie MacEwen, a trusted mentor and role model, whose advice provided clarity at times of need. Special thanks to my parents, in-laws, my husband Dr Jonathan Philip Law and our two daughters Sophie and Ellie for their unconditional love, sacrifices, and encouragement over the past few years.

# Contents listings

## Table of Contents


***Abstract*** ..... ***2***

***Dedication*** ..... ***5***

***Acknowledgements*** ..... ***6***

***Contents listings*** ..... ***7***

***List of figures*** ..... ***10***

***List of tables*** ..... ***10***

***List of supplementary tables*** ..... ***14***

***List of supplementary materials*** ..... ***15***

***List of abbreviations*** ..... ***16***

***Chapter 1 Introduction*** ..... ***21***

**1.1 The structure of the eye** ..... **21**

**1.2 Immune mechanisms of the eye** ..... **22**

1.2.1 The ocular surface ..... 22

1.2.2 Blood-ocular barrier system ..... 24

**1.3 The gut microbiome** ..... **25**

1.3.1 Regulation of the host immune system by the gut microbiota ..... 25

1.3.2 Regulation of the gut microbial profile by the host immune system ..... 26

1.3.3 Links between the gut microbiota and the 'gut-eye' axis ..... 26

**1.4 The ocular surface microbiome** ..... **30**

**1.5 Ocular Mucous Membrane Pemphigoid** ..... **34**

1.5.1 Demographics and clinical presentation ..... 34

1.5.2 Pathogenesis and pathophysiology of OcMMP ..... 36

1.5.3 Diagnosis ..... 41

1.5.4 Treatment strategies ..... 41

**1.6 Ocular infections** ..... **43**

1.6.1 Microbial keratitis ..... 43

1.6.2 Endophthalmitis ..... 45

1.6.3 Conventional clinical microbiology diagnostics and management of ocular infections ..... 48

**1.7 Sequencing methods to evaluate the microbiome and identifying pathogenic microbial genomes in ocular infections** ..... **51**

1.7.1 Current sequencing platforms ..... 51

1.7.2 Targeted amplicon 16S rRNA gene sequencing ..... 55

1.7.3 Untargeted whole genome sequencing ..... 57

1.7.4 Analysis of microbiome data ..... 58

**1.8 Aims and objectives** ..... **63**

***Chapter 2 Gut microbial dysbiosis in ocular mucous membrane pemphigoid*** ..... ***67***

**2.1 Introduction** ..... **67**

**2.2 Methods ..... 68**
2.2.1 Participant enrolment and sample collection ..... 68
2.2.2 DNA extraction from faecal samples, 16S rRNA gene amplification, Illumina MiSeq amplicon sequencing ..... 71
2.2.3 Bioinformatic analysis and statistics ..... 72

**2.3 Results ..... 73**
2.3.1 Differences in alpha-diversity between OcMMP and healthy controls ..... 77
2.3.2 Differences in beta diversity between OcMMP and healthy controls ..... 83
2.3.3 Compositional differences in gut microbiome of OcMMP and healthy controls ..... 84

**2.4 Discussion ..... 92**

***Chapter 3 Evaluation of Full-Length Nanopore 16S Sequencing for Detection of Pathogens in Microbial Keratitis ..... 99***

**3.1 Introduction ..... 99**

**3.2 Materials and Methods ..... 102**
3.2.1 Mock bacterial community ..... 102
3.2.2 Ex vivo porcine model of microbial keratitis ..... 102
3.2.3 Collection method/Swabs ..... 103
3.2.4 DNA extraction ..... 105
3.2.5 16S qRT-PCR for 16S and beta-actin ..... 105
3.2.6 16S gene copy number quantification ..... 106
3.2.7 Nanopore 16S barcoding, sequencing, and bioinformatics ..... 107
3.2.8 Clinical sample collection from patients presenting with microbial keratitis ..... 108
3.2.9 Miseq 16S sequencing and bioinformatics ..... 109
3.2.10 Statistical Analysis ..... 109

**3.3 Results ..... 110**
3.3.1 Evaluation of bioinformatics analysis pipeline on defined mock microbial community ..... 110
3.3.2 Optimising sampling technique for collection of microbial keratitis on ex-vivo porcine model ..... 114
3.3.3 Proof-of-concept clinical study on microbial keratitis patient samples ..... 119

**3.4 Discussion ..... 121**

***Chapter 4 Evaluation of full length 16S and whole genome nanopore sequencing for detection of pathogens in endophthalmitis ..... 129***

**4.1 Introduction ..... 129**

**4.2 Methods ..... 131**
4.2.1 Isolation of murine peripheral blood mononuclear cells ..... 132
4.2.2 Serial dilutions of cultured *Escherichia coli* ..... 132
4.2.3 Host DNA depletion by saponification ..... 133
4.2.4 Patient recruitment and sample collection ..... 134
4.2.5 DNA extraction ..... 134
4.2.6 Quantitative Polymerase Chain Reaction (PCR) ..... 135
4.2.7 Full-length 16S rRNA nanopore sequencing and bioinformatics ..... 136
4.2.8 Nanopore Whole Genome Sequencing (Nanopore WGS) and bioinformatics ..... 138
4.2.9 Illumina Whole Genome Sequencing (Illumina WGS) and bioinformatics ..... 140
4.2.10 Negative controls ..... 142
4.2.11 Definitions ..... 142
4.2.12 Statistical analyses ..... 143

**4.3 Results ..... 144**
4.3.1 Effects of differential centrifugation and saponification on host DNA depletion and bacterial DNA recovery ..... 144

4.3.2 Sequences detected in negative controls ..... 148
4.3.3 Clinical samples ..... 148
4.3.4 Agreement between microbial culture and full-length 16S rRNA nanopore sequencing ..... 150
4.3.5 Agreement between microbial culture and nanopore whole genome sequencing ..... 153
4.3.6 Agreement between culture and Illumina whole genome sequencing (Illumina WGS) ..... 155
**4.4 Discussion ..... 155**
***Chapter 5 Conclusions ..... 163***
**5.1 Future directions ..... 167**
***Bibliography ..... 172***
***Appendices ..... 193***
**Supplementary for Chapter 2 ..... 193**
**Supplementary for Chapter 3 ..... 207**
**Supplementary for Chapter 4 ..... 221**
**Publication arising from this thesis ..... 230**
Low L, *et al*. Evaluation of Full-Length Nanopore 16S Sequencing for Detection of Pathogens in Microbial Keratitis. PeerJ. (in Press) ..... 230

## List of figures

Figure 1.1 Basic anatomy of the eye in the frontal and sagittal planes. ..... 21
Figure 1.2 Diagram of the tear film. ..... 23
Figure 1.3 Gut microbial activation of retina-specific T cells. ..... 27
Figure 1.4 The multi-systemic clinical spectrum of mucous membrane pemphigoid. ..... 34
Figure 1.5 Clinical signs in ocular mucous membrane pemphigoid. ..... 35
Figure 1.6 Structure of the epithelial basement membrane. ..... 36
Figure 1.7 Summary of autoantibody and cellular autoimmune-driven inflammatory and pro-fibrotic mechanisms in OcMMP. ..... 39
Figure 1.8 Immunosuppression strategies for OcMMP. ..... 42
Figure 1.9 Anterior segment photograph of patient presenting with microbial keratitis, showing a large, focal central corneal infiltrate with marked conjunctival injection. ..... 44
Figure 1.10 Anterior segment photograph of patient presenting with post-operative endophthalmitis, showing hypopyon and conjunctival injection. ..... 46
Figure 1.11 Schematic representation of the numerous different media required for culture and susceptibility testing, along with the different primers required for polymerase chain reactions (PCR) in conventional microbiology. ..... 49
Figure 1.12 Overview of Illumina, nanopore and Pacific Biosciences single-molecule, real-time (SMRT) sequencing technologies. ..... 53
Figure 1.13 The structure of the 16S rRNA gene. ..... 55
Figure 2.1 Comparisons in alpha diversity between healthy volunteers and OcMMP. ..... 79
Figure 2.2 Correlations between alpha diversity and Cicatrising Conjunctivitis Clinical Assessment Tool (CCAT) scores. ..... 81
Figure 2.3 Principal Coordinate Analysis (PCoA) of the Yue and Clayton theta distances between faecal samples from OcMMP patients and healthy volunteers. ..... 84
Figure 2.4 Box and whisker plot comparing the relative abundance in phylum-level taxa between patients with ocular mucous membrane pemphigoid (OcMMP) and healthy volunteers. ..... 85
Figure 2.5 Box and whisker plot comparing the relative abundance of taxa that differed significantly at order level (multiple hypothesis testing, adjusted p-values <0.05). ..... 86

Figure 2.6 Box and whisker plot comparing the relative abundance of taxa that differed significantly at family level (multiple hypothesis testing, adjusted p-values <0.05). ..... 87
Figure 2.7 Box and whisker plot comparing the relative abundance of taxa that differed significantly at genus level (multiple hypothesis testing, adjusted p-values <0.05)..... 88
Figure 2.8 Linear discriminant analysis (LDA) effect size (LEfSe) of significantly enriched bacterial taxa in OcMMP and healthy volunteers. ..... 89
Figure 2.9 Comparisons of the relative abundances of bacterial taxa that were significantly enriched in OcMMP. ..... 90
Figure 2.10 Comparisons of the abundances of bacterial taxa that are significantly enriched in healthy volunteers. ..... 91
Figure 3.1 Summary of study workflow. ..... 101
Figure 3.2 Bioinformatics workflow. ..... 108
Figure 3.3 Bar chart comparing the relative abundances of taxa before and after filtering. 110
Figure 3.4 Post-inoculation DNA concentration by collection material. ..... 116
Figure 3.5 Variability in post-inoculation DNA concentrations by collection material ..... 118
Figure 3.6 Comparison of swab collection yield, defined by relative abundances of species normalised by 16S copy number. ..... 119
Figure 4.1 Bar graph comparing the effects of saponification on 16S bacterial recovery by PCR in murine peripheral blood mononuclear cells (PBMC) samples spiked with varying E.coli concentrations, ranging from $10^1$ to $10^5$ colony-forming units (CFU)/ml. ..... 144
Figure 4.2 Bar graph comparing the effects of saponification on murine actin recovery in murine peripheral blood mononuclear cells (PBMC) samples spiked with varying E.coli concentrations, ranging from $10^1$ to $10^5$ colony-forming units (CFU)/ml. ..... 145
Figure 5.1 Potential metagenomic workflow in the management of an unwell patient in the clinical setting. ..... 168

## List of tables

Table 1.1 Summary of gut microbiome studies relating to dry eye diseases. ..... 29
Table 1.2 Summary of findings from studies of the healthy adult ocular surface microbiome. ..... 33
Table 1.3 Comparisons of data type, costs, throughput and portability across the sequencing technologies. ..... 54
Table 1.4 Characteristics of the 16S rRNA gene reference databases. ..... 57
Table 2.1 Demographics of participants ..... 75
Table 2.2 Details of medications/systemic immunotherapy. ..... 76
Table 2.3 Linear regression models of number of OTUs in OcMMP vs. healthy controls ..... 77
Table 2.4 Linear regression models of Shannon Diversity Index in OcMMP vs. healthy controls ..... 77
Table 2.5 Linear regression models of number of OTUs and ocular inflammation score ..... 82
Table 3.1 Summary of collection methods used. ..... 104
Table 3.2 Comparisons of the effects of filtering on taxonomic assignment. ..... 111
Table 3.3 Comparing the effects of using NCBI 16S, RDP and SILVA databases at different levels of clustering (97%, 99% and 100% similarity). ..... 113
Table 3.4 Pre-inoculation average $C_T$ by collection material. ..... 114
Table 3.5 Post-inoculation average $C_T$ and DNA concentrations by collection material. ..... 115
Table 3.6 Variability in post-inoculation DNA concentrations by collection method ..... 117
Table 3.7 Results of culture and 16S rRNA gene sequencing of patient samples aligned to NCBI 16S RefSeq Database at 100% similarity. ..... 120
Table 4.1 Comparisons of the effects of saponification on proportion of human, non-human reads and bacterial load ..... 147
Table 4.2 Demographics of patient participants. ..... 149
Table 4.3 Summary of predominant microorganism detected by culture, 16S rRNA nanopore sequencing, whole genome nanopore and Illumina sequencing. ..... 151
Table 4.4 Comparison of microbial culture, full-length 16S nanopore sequencing, nanopore and Illumina whole genome sequencing (WGS) results for endophthalmitis. ..... 152
Table 4.5 Comparisons of bacterial load, mean alignment breadth, depth and genomic coverage. ..... 154

## List of supplementary tables

Supplementary Table 3.1 Pilot data comparing DNA yield of dry vs. pre-moistened swab at different time points ..... 207
Supplementary Table 3.2 Comparing the effects of using NCBI 16S, RDP and SILVA databases at different levels of clustering (97%, 99% and 100% similarity) on patient samples ..... 215
Supplementary Table 4.1 Sequences detected in negative controls for full-length 16S rRNA nanopore sequencing runs. ..... 221
Supplementary Table 4.2 Sequences detected in negative controls for whole genome nanopore sequencing runs. ..... 222
Supplementary Table 4.3 Sequences detected in negative controls for whole genome Illumina sequencing runs. ..... 224

## List of supplementary materials

Supplementary material 2.1 R script for generating principle coordination analysis plots. 193
Supplementary material 2.2 R script for processing 16S taxonomic data. ..... 195
Supplementary material 3.1 Bioinformatics script written in python for clustering of reference databases using the CD-HIT software. ..... 208
Supplementary material 3.2 Bioinformatics script written on python for read alignment using minimap2. ..... 209
Supplementary material 3.3 Bioinformatics script written in python for read filtering, sorting, and indexing using SAMtools. ..... 210
Supplementary material 3.4 Bioinformatics script written in python for removing the last redundant line on SILVA reference database. ..... 212
Supplementary material 3.5 Python script for dividing the taxonomic assignments into genus and species groupings. ..... 213
Supplementary material 3.6 Python script for calculating relative abundances of taxa from 16S nanopore output. ..... 214
Supplementary material 4.1 Bioinformatics script written in python for alignment and subsequent removal of human reads using minimap2 and SAMtools. ..... 225
Supplementary material 4.2 Bioinformatics script written in python for alignment of reads to reference genomes using minimap2 and SAMtools. ..... 226
Supplementary material 4.3 Python script for calculating breadth of coverage, mean alignment depth and coverage. ..... 228
Supplementary material 4.4 Python script for calculating relative abundances of taxa from Illumina WGS output. ..... 229

# List of abbreviations

| Abbreviation | Meaning |
|---|---|
| 16S rRNA | 16S ribosomal ribonucleic acid |
| 16S rRNA V | 16S ribosomal ribonucleic acid gene variable region |
| AMOVA | analysis of molecular variance |
| AMP | antimicrobial peptides |
| ARVO | Association for Research in Vision and Ophthalmology |
| BD | Becton-Dickinson |
| BH | Benjamini-Hochberg |
| BHI | brain-heart infusion |
| BM | basement membrane |
| BMI | body mass index |
| BP | bullous pemphigoid antigen |
| bp | base pairs |
| BRiSK | biome representational in silico karyotyping |
| CCAT | Cicatrising Conjunctivitis Clinical Assessment Tool |
| CD | cluster designation |
| CFU | colony-forming units |
| conj | conjunctival |
| $C_T$ | cycle threshold |
| CXCL | chemokine (C-X-C) motif ligand |
| DES | dry eye syndrome |
| DIF | direct immunofluorescence |
| DNA | deoxyribonucleic acid |
| dNTPs | deoxynucleoside triphosphates |
| FASTQ | Federal Acquisition STreamlining Act text format data containing quality score |
| Gb | gigabases |
| HC | healthy control |
| HLA | human leukocyte antigen |

| | |
|---|---|
| HOMOVA | homogeneity of molecular variance |
| HSP | heat shock protein |
| HSV | herpes simplex virus |
| IC | impression cytology |
| IFN | interferon |
| Ig | immunoglobulin |
| IIF | indirect immunofluorescence |
| IL | interleukin |
| IOL | intraocular lens |
| IQR | interquartile range |
| IRF | interferon regulatory factor |
| IVT | intravitreal |
| kg | kilogram |
| LB | lysogeny broth |
| LDA | linear discriminant analysis |
| LEfSe | linear discriminant analysis effect size |
| LPS | lipopolysaccharide |
| MHC | major histocompatability complex |
| MK | microbial keratitis |
| ml | milliliter |
| μl | microliter |
| mM | millimolar |
| MMP | mucous membrane pemphigoid |
| MWE | Medical Wire and Equipment |
| MyD88 | myeloid differentiation primary response 88 |
| N/A | not available |
| N50 | sequence length of the shortest contig corresponding to 50% of the total genome length |
| NCBI | National Center for Biotechnology Information |
| NF | nuclear factor |

| | |
|---|---|
| ng | nanogram |
| NHS | National Health Service |
| NK cells | natural killer cells |
| NLR | NOD-like receptor |
| NLRC | nucleotide-binding oligomerisation domain-like receptors (NLR) caspase-associated recruitment domain |
| NLRP | nucleotide-binding oligomerisation domain-like receptors (NLR) pyrin domain |
| NOD | nucleotide-binding oligomerisation domain-containing protein |
| OcMMP | mucous membrane pemphigoid with ocular involvement |
| OD | optical density |
| ONT | Oxford Nanopore Technologies |
| OTU | Operational Taxonomic Unit |
| PacBIO SMRT | Pacific Biosciences Single-Molecule, Real-Time |
| PAMPs | pathogen-associated molecular patterns |
| PBMC | peripheral blood mononuclear cells |
| PBS | phosphate-buffered saline |
| PCoA | principal coordinate analysis |
| PCR | polymerase chain reaction |
| PDGF | platelet-derived growth factor |
| PMH | past medical history |
| PRRs | pattern recognition receptors |
| PSA | polysaccharide A |
| pSS | primary Sjögren's Syndrome |
| PTFE | Polytetrafluoroethylene |
| QC | quality control |
| qPCR | quantitative polymerase chain reaction |
| qRT-PCR | quantitative real-time polymerase chain reaction |
| rDNA | ribosomal deoxyribonucleic acid |
| RDP | Ribosomal Database Project |
| RIPK | receptor-interacting protein kinase |

| | |
|---|---|
| RNA | ribonucleic acid |
| RPM | revolutions per minute |
| RV virus | *Rubivirus* genus |
| SAA | serum amyloid A |
| SAM | Sequence Alignment/Map |
| SCFA | short-chain fatty acids |
| SFB | segmented filamentous bacteria |
| sIgA | secretory immunoglobulin A |
| sp | species |
| spp | several species |
| STUDIOUS | Systemic Gut Microbiota Driving Sight-threatening Inflammatory Ocular Disease |
| Tc cells | T cytotoxic cells |
| TGF | transforming growth factor |
| Th1 cells | type 1 T helper cells |
| Th2 cells | type 2 T helper cells |
| TLR | toll-like receptor |
| TNF | tumour necrosis factor |
| TRIF | Toll/interleukin-1 receptor-domain-containing adapter-inducing interferon-β |
| UCG | uncultured group |
| UK | United Kingdom |
| vs. | versus |
| VZV | varicella zoster virus |
| w/w | weight per weight |
| WGS | whole genome sequencing |
| ZMW | zero-mode waveguides |
| γδ T cells | gamma delta T cells |

# Chapter 1

# Introduction

# Chapter 1 Introduction

## 1.1 The structure of the eye

The normal eyeball measures approximately 70 – 85 mm in circumference and 22 – 27 mm in anteroposterior diameter[1] (Figure 1.1).

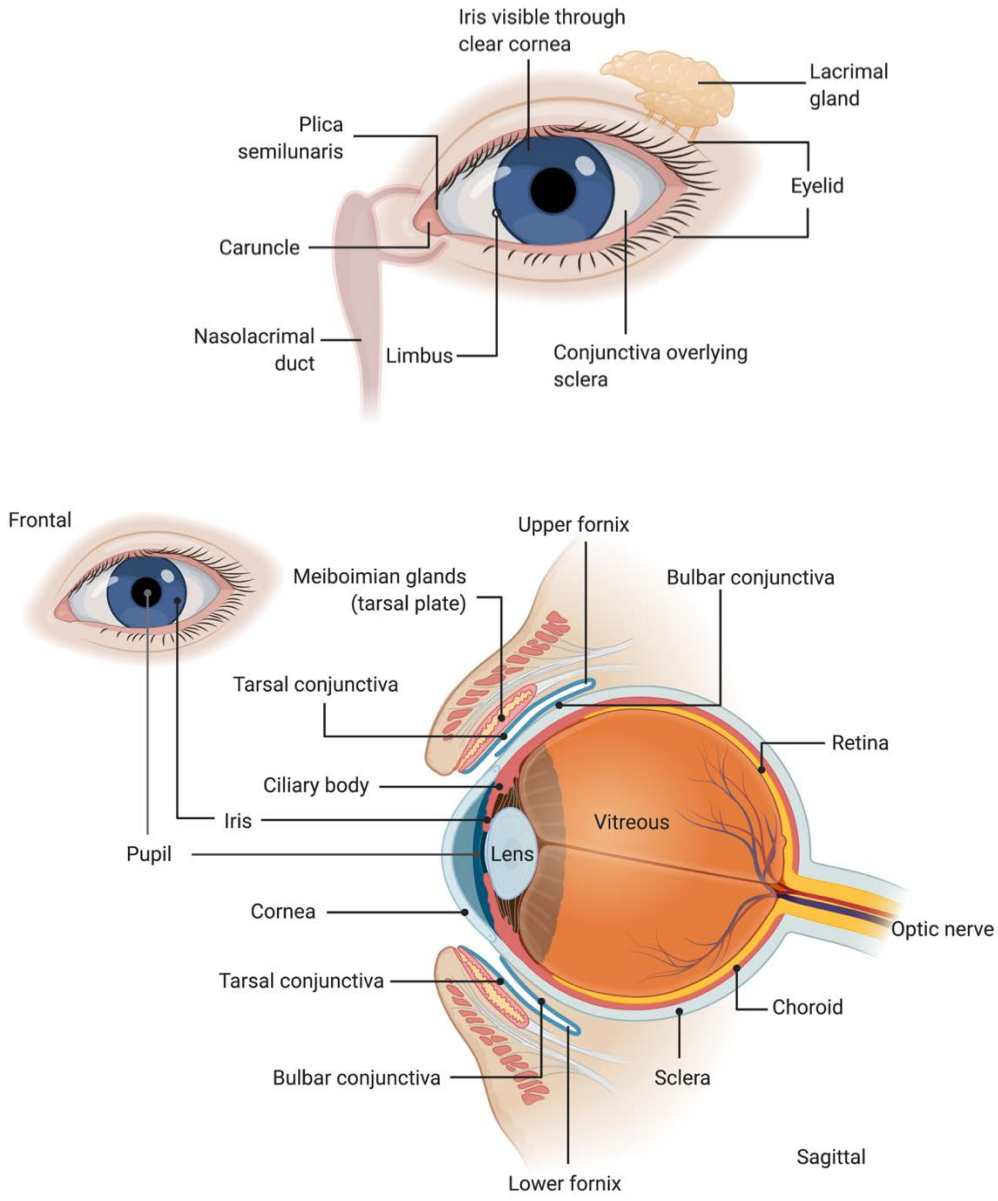



*Figure 1.1 Basic anatomy of the eye in the frontal and sagittal planes.*

*Created with BioRender.com*

The eye, a highly specialised sensory organ, is part of the complex optical system providing vision[2]. First, light rays reflected off the environment or object of interest, is refracted by the clear cornea into the anterior and posterior chambers containing aqueous humour, through the pupil (an opening regulating the amount of light entering the eye and controlled by the iris), refracted by the lens, passes the vitreous humour (4 – 5ml thick, gelatinous fluid accounting for 80% of the eye volume), and reaches the retina. The photoreceptors in the retina transduce the light into electrical signals (action potentials) which are then transmitted by the optic nerve to the brain for visual processing[3].

The eye consists of three layers – from the outer to inner layers:

- External fibrous tunic composed of sclera and cornea
- Middle, vascular pigmented tunic, also known as uvea, composed of choroid, ciliary body, and iris
- Inner, sensory neural tunic composed of the retina

## 1.2 Immune mechanisms of the eye

### 1.2.1 The ocular surface

The ocular surface system comprises of the cornea, conjunctiva, lacrimal and accessory glands, Meibomian glands, tear film, the nasolacrimal duct and the eyelids[4]. The meibomian glands produce the superficial lipid layer of the tear film that prevents surface evaporation[4]. The lacrimal and accessory glands secrete water, immunoglobulin A, and other antimicrobial proteins such as lysozyme, defensins and lactoferrin, forming the intermediate aqueous layer

of the tear film[4]. The goblet cells within the corneal and conjunctival epithelia secrete hydrophilic mucins[4]. The components of the tear film is shown in Figure 1.2. Eye-associated lymphoid tissue, or immune follicles that are analogous to the gut-associated isolated lymphoid follicles, are present within the conjunctiva and develop soon after birth[5].

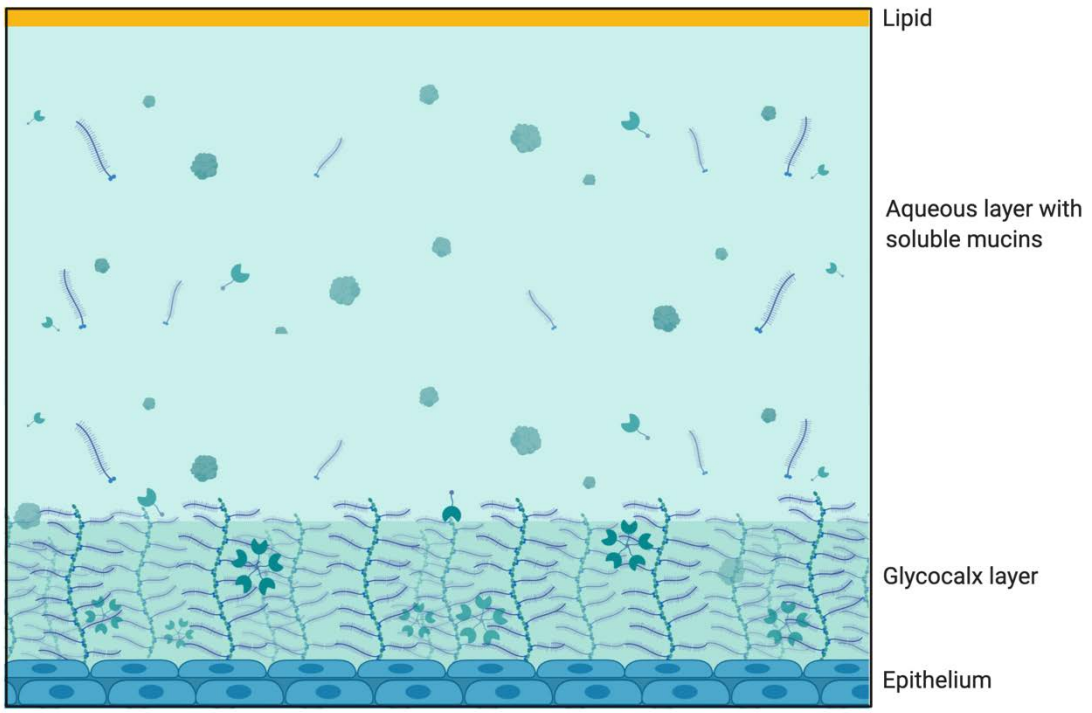



*Figure 1.2 Diagram of the tear film.*

*Created with BioRender.com*

Resident cells of the both the innate and adaptive immune system, such as neutrophils, eosinophils, macrophages, natural killer cells, Langerhans cells, γδ T cells, T helper 1 (Th1) and Th2 cells, are found on the ocular surface[6]. Together, they play a vital role in protecting and maintaining the smooth, wet refractive surface of the cornea. The cornea is the clear window in front of the eye compromising of intricately arranged collagen fibrils that are vital for optical transparency and visual neuro-processing. When damaged, the avascular cornea loses transparency leading to visual impairment.

### 1.2.2 Blood-ocular barrier system

The eye is an immune-privileged site, with the blood-ocular barrier system (consisting of the blood-aqueous and the blood-retinal barrier) protecting the ocular antigens from invasion by infectious pathogens or from circulating immune cells in an immunocompetent host[7]. The blood-aqueous barrier, located at the anterior portion of the eye, is formed by the tight junctions of the non-pigmented ciliary epithelium and the endothelium of the iris vasculature[7]. Tolerance to antigens in the anterior chamber is also mediated by the anterior chamber-associated immune deviation, whereby there is downregulation of the T helper 1 and T helper 2-mediated immune responses to intraocular antigens[8]. The blood-retina barrier is composed of both physical and immunologic barriers. The physical barrier is formed by the tight junctions of the retinal capillary endothelium and the retinal pigment epithelium. The immunologic barrier is formed by the retinal neurovascular unit (astrocytes, Müller cells, and pericytes) and the retinal pigment epithelium secretion of immunomodulatory factors (IL-10, macrophage migration inhibitory factor, and complement regulatory proteins) and conversion of naïve T cells into regulatory T cells[7,9]. Together, these mechanisms of ocular immune-privilege play an important role in maintaining the function and structural integrity of ocular tissues, and preserving optical clarity by reducing the potentially damaging collateral effects of excessive inflammation to ocular tissues, which have limited capacity for regeneration[10]. However, this comes at a cost of delayed clearance of certain pathogens, such as *Toxoplasma gondii* and herpes simplex virus, incomplete immune tolerance to ocular antigens and unchecked tumour growth, such as ocular melanoma[11].

## 1.3 The gut microbiome

The human body consists of symbiotic interactions between microorganisms, otherwise known as microbiota, that reside in and on our bodies. The largest and most diverse microbial communities reside in the colon of the human gastrointestinal system, collectively the genomes of these microbiota are known as the gut microbiome[12]. The gut microbiota is important in the regulation of energy homeostasis, metabolism of drugs and modulation of the immune system, either through direct contact with the mucosal surfaces of the lining of the gut or through the production of gut microbial metabolites[12,13].

### 1.3.1 Regulation of the host immune system by the gut microbiota

Influences of the gut microbiota on the host innate immune system mainly rely on metabolites. Gut microbial tryptophan metabolites affect the innate lymphoid cells and reduce mucosal inflammation through interleukin(IL)-22 dependent aryl hyrdrocarbon receptor signalling[14]. Short chain fatty acids, such as butyrate, a by-product of starch fermentation by anaerobic bacteria in the gut, exerts anti-inflammatory properties by inhibiting histone deacetylases, thereby downregulating macrophage secretion of nitric oxide, IL6, and IL-12, which are pro-inflammatory mediators[15]. In terms of the influences of the gut microbiota on the adaptive immune system, segmented filamentous bacteria (SFB) adherent to epithelial cells of the terminal ileum have been shown to induce the accumulation of Th17 cells in the lamina propria and the increase production of IL-22, serum amyloid A (SAA) and IL-17[16]. Colonisation of Bacteroides fragilis, through the zwitterionic capsular polysaccharide, PSA, induces CD4+ T cell proliferation, balances Th1/Th2 cytokine production and directs regulatory T cell development[17].

### 1.3.2 Regulation of the gut microbial profile by the host immune system

The gut microbiota composition is in turn influenced by the host immune system[13]. Antimicrobial peptides (AMP) produced in the gut is capable of altering the gut microbial colonisation[18]. Microbial peptidoglycan recognition by nucleotide-binding oligomerisation domain-containing protein 2 (NOD2) in lamina propia, signals through kinase receptor-interacting protein 2 (RIPK2) and nuclear factor-κβ (NF- κβ), and induces the production of antimicrobial peptides and mucin[19]. Flagellin and lipoproteins of microbes stimulate toll-like receptor 5 (TLR5) in CD103+ dendritic and luminal epithelial cells, leading to the secretion of IL-23 by dendritic cells, maturation of innate lymphoid cell, secreting IL-22, thereby, enhancing the production of AMPs by luminal epithelial cells[20]. Microbial metabolites also activate NOD-, leucin-rich repeat-, and pyrin domain-containing 6 (NLRP6) inflammasome to secrete IL-18 and AMPs[21]. T follicular helper cells produce secretory IgA, whilst invariant natural killer T cells, through CD1d, secrete IL-4 and IL-10 to regulate the composition of the gut microbiota[22].

### 1.3.3 Links between the gut microbiota and the 'gut-eye' axis

The gut microbial components and their metabolites have the capacity to modulate both the local and systemic immunity at extra-intestinal sites, such as the brain, lung, liver, skin and endocrine organs[13,23]. Emerging evidence indicates that there is potential for the gut microbiota to modulate the ocular immunity, which has traditionally been classified as an 'immune-privileged site'[24,25]. Mouse models of spontaneous experimental autoimmune uveitis have shown that exposure to the gut commensal microbiota activates retina-specific

autoreactive T cells through their T cell receptors in the lamina propria of the intestine, and leads to development of spontaneous uveitis[26]. Depletion of the gut microbiota by broad-spectrum antibiotics reduces Th17 cells in the lamina propria of the intestine, and attenuates uveitis (Figure 1.3)[26].

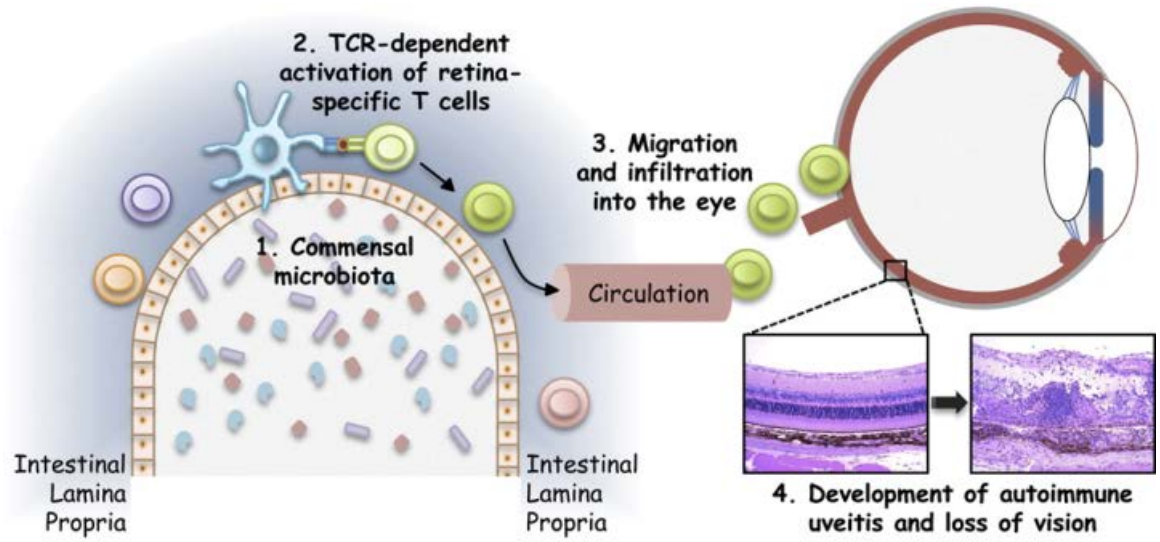



*Figure 1.3 Gut microbial activation of retina-specific T cells.*

*Commensal gut microbiota activates retina-specific T cells in a T-cell receptor (TCR)-dependent manner. Auto-reactive T cells migrate and infiltrate into the eye leading to the development of autoimmune uveitis and loss of vision. Image reproduced from Horai et al[26]. Elsevier user license.*

Studies of the gut microbiome in uveitis patient cohorts have reported enrichment of Enterobacteriaceae, and *Fusobacterium*, and lower levels of abundance of *Prevotella* compared to healthy controls[27]. Similarly, the gut microbiome is altered in Vogt-Koyanagi-Harada disease, and is restored with immunosuppressive treatment[28]. In neovascular AMD mouse model, high-fat diet alters the Bacteroides to Firmicutes ratio, reduces the gut barrier function, and increases intestinal permeability to AMPs into the systemic circulation, triggering activation of pattern recognition receptors and inflammation mediated through

elevated cytokine levels of IL-1β, IL-6, TNF-α and vascular endothelial growth factor A that contribute to pathological neovascularisation[29]. Enrichment of *Oscillibacter, Anaerotruncus, Eubacterium* and *Ruminococcus* were reported in stool samples of patients with neovascular age-related macular degeneration, along with upregulation of glutamate degeneration, L-alanine fermentation, and arginine biosynthesis, and down-regulation of fatty acid elongation pathway[30]. Patients with glaucoma have higher oral bacterial load in compared to healthy controls[31]. Low dose LPS administered to mouse model of glaucoma (DBA/2J) upregulates Toll-like receptor 4 signalling and complement pathway, activating microglia in the optic nerve and retina, leading to axonal degeneration and loss of neurons[31].

The gut microbiome, in addition to host genetics, has been shown to be an important modulator of the ocular surface homeostasis[25]. Using animal studies, the tear secretory IgA (sIgA) levels were reduced following oral antibiotic intake in specific pathogen free Swiss Webster mice but not in C57BL/6N mice, and reconstitution of the gut microbiome in germ-free Swiss Webster mice with *Bacteroides acidifaciens* restored sIgA levels on the ocular surface[32]. Findings of the gut microbiome profile studies in patients with immune-mediated dry eyes are summarised in Table 1.1

| Authors | Country | Sample size | Microbiome analysis technique | Alpha diversity | Findings | | Correlation with disease |
|---|---|---|---|---|---|---|---|
| | | | | | Increased | Decreased | |
| de Paiva, 2016 | USA | 10 (pSS) HMP (HC) | 16S rRNA V4 | Reduced in pSS | pSS vs HC: *Pseudobutyrivibrio, Escherichia/ Shigella, Blautia, and Streptococcus* | *Bacteroides, Parabacteroides, Faecalibacterium, and Prevotella* | Number of OTUs negatively correlates with disease severity |
| Moon, 2020 | South Korea | 10 (pSS), 14 (DES), 12 (HC) | 16S rRNA V3 - V4 | No difference | pSS vs HC: Bacteroides (phylum), *Alistipes, Odoribacter, Prevotella, Veillonella*<br><br>DES vs HC: *Veillonella* | *Actinobacteria, Agathobacter, Bifidobacterium, Blautia, Dorea, Bifidobacterium longum, Eubacterium hallii Subdoligranulum* | |
| Mendez, 2020 | USA | 13 (SS), 8 (DES) | 16S rRNA V4 - V5 | No difference | SS vs DES: Actinobacteria, Bacteroidetes, Proteobacteria (phylum), Actinomycetaceae, Akkermanciaceae, Coriobacteriaceae, Eggerthellaceae, Eubacteriaceae, Lactobacillaceae (family), *Megasphaera, Parabacteroides, Prevotella* (genus) | Firmicutes (phylum), Clostridiales (order), Lachnospiraceae, Ruminococcaceae (family), *Bacteroides, Faecalibacterium, Veillonella* (genus) | |

*Table 1.1 Summary of gut microbiome studies relating to dry eye diseases.*

Abbreviation: primary Sjögren's Syndrome, pSS; Human Microbiome Project, HMP; dry eye syndrome, DES; healthy control, HC; operational taxonomic unit, OTU.

## 1.4 The ocular surface microbiome

The ocular surface was initially thought of as 'sterile', with the physical protective mechanism of the blink reflex, presence of ocular resident immune cells and antimicrobial proteins and immunoglobulins in the tear film providing a hostile environment for microorganisms[33]. Using modern culture techniques, Perkins and associates were able to isolate coagulase-negative *Staphylococcus* and *Propionibacterium acnes* from 87 of 96 conjunctival swabs of healthy individuals[34]. With the advent of molecular sequencing technologies, a much broader repertoire of microbial sequences have been found on the ocular surface. The findings of these studies are summarised in Table 1.2.

The existence of a core microbiome, defined as "the set of genes present in a given habitat in all or a vast majority of humans *[sic]*", remains uncertain[12]. In ophthalmology, the debate arises about whether the presence of the organisms at the ocular surface are truly stable 'residents' or merely transient 'passersby'. To address this issue, Zegans and Van Gelder have proposed a framework for studying the ocular microbiome - "A true ocular surface microbiota would be characterized by the persistence of a stable consortium of viable microorganisms on the ocular surface over time." Ozkan and associates evaluated the temporal stability of the ocular surface by sampling the conjunctiva of 45 healthy adults at baseline, 1 month and 3 months. There were no culturable or detectable bacterial taxa (by 16S rRNA sequencing) present in all participants at all three time points, or in all participants at any time point. However, *Corynebacterium* and *Streptococcus* and 26 other bacterial taxa were consistently present at all three time points of sampling in at least one or more participant[35]. Similarly, Doan and associates have reported a predominance of *Corynebacterium*, *Streptococci, Propionibacteria* and torque teno virus in the ocular surface microbiome, which was distinct

from the facial skin and buccal mucosa bacterial community[36]. Previous quantitative studies have shown that the abundance of bacteria on the ocular surface are much fewer in magnitude compared to other mucosal sites, such as the oral or gut mucosa[36,37]. Taken together, this suggests that although the healthy ocular surface is not sterile, it is paucibacterial and its constituents have high interindividual variability.

The physiological and functional role of the human ocular surface microbiome has yet to be determined. However, in a proof-of-concept murine study, St Leger and associates showed that the commensal *Corynebacterium mastitidis* was able to colonise the murine ocular surface and induce interleukin 17 production from conjunctival γδ T cells, leading to increased recruitment of neutrophils to the conjunctiva and release of antimicrobial S100A8-S100A9 dimer in tears, thus protecting the eye from pathogenic *Pseudomonas aeruginosa* and *Candida albicans* infection[5]. In another study by Kugadas and associates, they found that germ-free Swiss Webster mice had higher susceptibility to *Pseudomonas aeruginosa*-induced infection compared to specific pathogen free mice[38]. Treatment with oral antibiotics to decrease the number of gut microbiota in specific pathogen free mice resulted in higher bacterial burden and susceptibility to *Pseudomonas aeruginosa*[38]. Polymorphonuclear leukocytes derived from germ-free mice and oral antibiotic treated-specific pathogen free mice were shown to have delayed maturation of neutrophils and reduced bactericidal capacity to *Pseudomonas aeruginosa PAO1* and *PA14* compared to polymorphonuclear cells derived from non-antibiotic treated specific pathogen free mice[38]. Reconstitution of the germ free mice with mouse- and human-derived microbiota decreased the susceptibility to *Pseudomonas aeruginosa* infection[38]. Monocolonising germ free mice with a single topical application of coagulase negative *Staphylococcus* sp. promotes IL-1beta-driven responses and

protected against *Pseudomonas aeruginosa* infection[38]. Similarly, Zaidi and associates have shown that germ-free mice had delayed neutrophil recruitment to *Staphylococcus aureus* infected corneas compared to wild type mice due to deficiency in intercellular adhesion molecule 1-dependent polymorphonuclear leukocyte recruitment to the cornea[39]. This suggests that both the gut and commensal ocular surface microbiome may contribute to the immune homeostasis of the eye.

Ocular surface microbiome alterations in immune-mediated dry eye disease is undetermined[40,41]. Graham and associates have shown that patients with dry eyes had higher numbers of bacterial colony forming units on culture compared to healthy, and was inversely correlated with conjunctival goblet cell density[40]. However, their finding was not replicated when using PCR to detect 16S rRNA gene. Although de Paiva and associates were able to detect decreased diversity in the gut microbiome of patients with Sjögren's syndrome keratoconjunctivitis compared to healthy controls, which correlated with systemic disease severity, they were not able to detect any significant differences in the ocular surface microbiome [37,40]. This suggests that the gut microbiome, which contains the largest amount of commensal bacteria in the body, is more likely to be affected by perturbations in the systemic immunity compared to the low-biomass, paucibacterial ocular surface microbiome.

| Authors [REF] (Year) | Country | Sample size | Type of sample | Microbiome analysis technique | Findings |
|---|---|---|---|---|---|
| Perkins[34] (1975) | USA | 96 | Conj cotton swab | Culture | *Staphylococcus epidermidis, Streptococcus sp., Micrococcus, Staphylococcus aureus, Gram negative rods, Corynebacterium sp., Bacillus sp., Propionibacterium acnes, Peptostreptococcus sp., Lactobacillus sp., Clostridium sp., Eubacterium sp.* |
| Graham[40] (2007) | Northern Ireland | 57 | Conj cotton swab & IC (cellulose acetate filters) | Culture, broad-range 16S rDNA PCR, dye-terminator Sanger sequencing | Coagulase negative *Staphylococcus, S. epidermidis, Corynebacterium sp., Propionibacterium sp., Rhodococcus erythropolis, Klebsiella sp., Erwinia sp.* |
| Dong[42] (2011) | USA | 4 | Conj cotton swab | V3-V4 16S rRNA gene Roche 454 sequencing | Phyla: Proteobacteria, Actinobacteria, Firmicutes; Genera: *Pseudomonas, Bradyrhizobium, Propionibacterium, Acinetobacter, Corynebacterium, Brevundimonas, Sphyngomonas, Staphylococcus, Streptococcus* |
| Lee[43] (2012) | South Korea | 4 | Eyelash & tear sample | V1-V3 16S rRNA gene Roche 454 sequencing | Phyla: Actinobacteria, Proteobacteria, Firmicutes, Cyanobacteria, and Bacteroidetes |
| Doan[36] (2016) | USA | 107 | Conj calcium alginate swab | Culture, BRiSK, V3-V4 Illumina sequencing | Coagulase-negative *Staphylococci, Diphtheroids, Propionibacteria,, Streptococci,* torque teno virus |
| de Paiva[37] (2016) | USA | 8 | IC (EyePrim) | V1-V3 16S rRNA gene Roche 454 sequencing | Phyla: Firmicutes, Actinobacteria, Proteobacteria, and Bacteroidetes |
| Ozkan[35] (2017) | Australia | 45 | Conj flock swabs | V4 16S rRNA Illumina MiSeq | Culture: Phyla: Firmicutes, Actinobacteria. Genera: Staphylococcus, Propionibacterium, Micrococcus; Sequencing: Phyla: Firmicutes, Actinobacteria & Proteobacteria. Genera: *Corynebacterium, Streptococcus, Anaerococcus, Sphingomonas and Acinetobacter* |
| Wen[44] (2017) | China | 90 | IC (MF membrane filter) | WGS Illumina HiSeq | *Propionibacterium acnes, Staphylococcus epidermidis, Escherichia coli, Micrococcus luteus, Onchrobactrum anthropic, Acidovorax sp., Actinobacter baumanii, Pseudomonas aeruginosa, Staphylococcus haemolyticus,* other sp. |
| Shimizu[41] (2019) | Japan | 11 | Conj cotton swab | Culture | Culture: *Staphylococcus epidermidis, Propionibacterium acnes* |

*Table 1.2 Summary of findings from studies of the healthy adult ocular surface microbiome.*

*Abbreviations: biome representational in silico karyotyping, BRiSK; conjunctival, conj; impression cytology, IC; Millipore mixed cellulose ester membranes, MF; whole genome sequencing, WGS.*

## 1.5 Ocular Mucous Membrane Pemphigoid

### 1.5.1 Demographics and clinical presentation

Mucous membrane pemphigoid (MMP) is a rare, orphan, life-threatening autoimmune scarring disease, involving all mucosal sites, such as the ocular, oral, nasopharyngeal, anogenital, tracheal, oesophageal mucosa, and skin[45–47] (Figure 1.4). Mucous membrane pemphigoid involving the ocular surface (OcMMP) is the commonest cause of cicatrising conjunctivitis in the United Kingdom, accounting for 61% of newly diagnosed cases, with an estimated minimum incidence of 0.8 per million population[48]. OcMMP mainly affects the ages of 60 – 80 years, and is likely to be more common in Caucasians than in Indian Asians or Chinese[49].

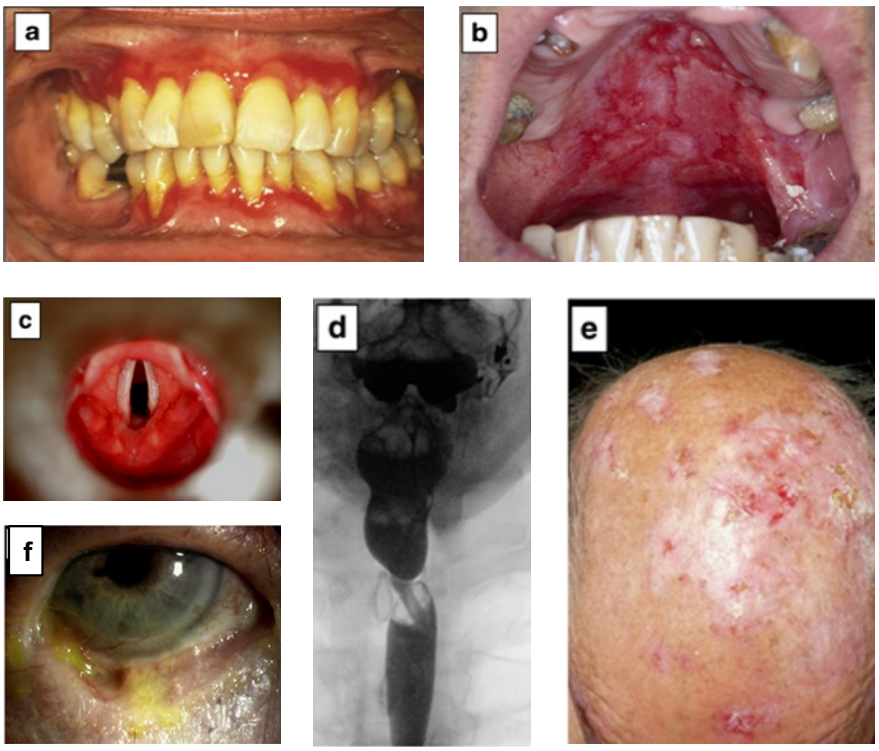



*Figure 1.4 The multi-systemic clinical spectrum of mucous membrane pemphigoid.*

*Involvement of the oral mucosa (Panels a and b. gingival and palatal inflammation), involvement of the supraglottic mucosa (Panel c. supraglottic inflammation), involvement of the gastrointestinal tract (Panel d. oesophageal stenosis), involvement of the skin (Panel e. scalp ulceration), involvement of the ocular mucosa (Panel f. conjunctival inflammation). Images reproduced from Dart[46]. CC BY-NC-ND 4.0.*

Patients typically present with red eye or persistent conjunctivitis not responding to topical treatment, or entropion and trichiasis, progressing to persistent corneal epithelial defect, corneal vascularisation, scarring and corneal blindness. In a small subset of patients there is a rapid progression from initial presentation of acute conjunctivitis and/or limbitis (stem cell failure) to severe conjunctival overgrowth, ocular surface failure and keratinisation leading to corneal blindness[46,50]. Common clinical signs seen in OcMMP are shown in Figure 1.5.

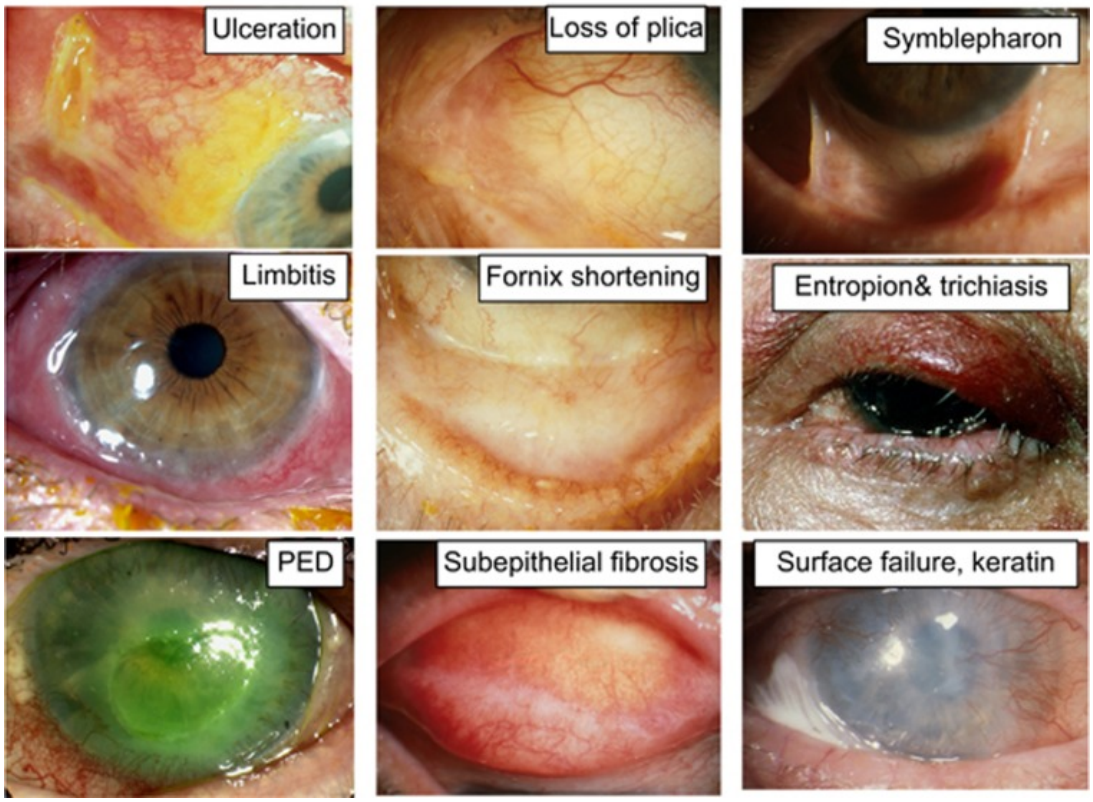



*Figure 1.5 Clinical signs in ocular mucous membrane pemphigoid.*

*From top left: conjunctival ulceration, limbitis, persistent corneal epithelial defect (PED), loss of plica, fornix foreshortening, subepithelial fibrosis, symblepharon, entropion (inversion of the eyelid margin) and trichiasis (abnormally positioned eyelashes), and ocular surface failure leading to keratinisation, opacification and vascularisation of the cornea and corneal blindness. Images reproduced from Dart[46]. CC BY-NC-ND 4.0.*

### 1.5.2 Pathogenesis and pathophysiology of OcMMP

The pathogenesis of OcMMP involves a combination of both genetic and environmental factors leading to the loss of tolerance to conjunctival epithelial basement membrane proteins, and generation of autoreactive T cell and/or autoantibodies that are central both to the autoantibody (IgG or IgA) and cellular autoimmune responses[46]. The target antigens in OcMMP are BP180[51,52], BP230, laminin 332[53,54], α6β4 integrin[55,56], and type VII collagen[57] (Figure 1.6).

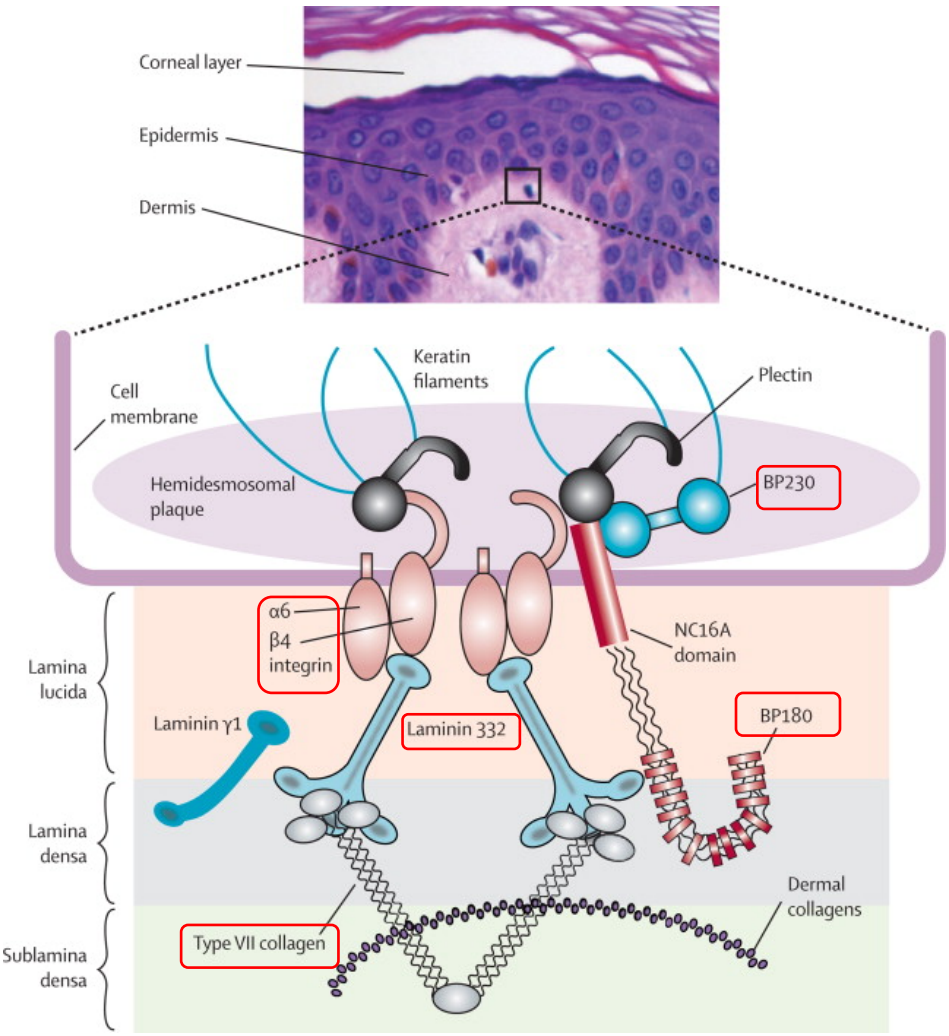



*Figure 1.6 Structure of the epithelial basement membrane.*

*Target antigens in MMP are highlighted in the red boxes. Image reproduced from Schmidt[58]. Elsevier user license.*

The pathophysiology of ocular MMP involves two main processes: (1) inflammation and (2) scarring, both of which are not mutually exclusive[46,58] . A potential autoimmune inflammatory response is initiated through the presentation of self-antigens, such as proteins from the basement membrane of the conjunctiva, oral mucosa or skin, on the MHC (class II region of DQB1)[59] of antigen presenting cells to T cell receptors in genetically susceptible individuals, thereby initiating B cell activation, proliferation, and generation of IgG or IgA autoantibodies against basement membrane proteins[46]. Another suggested hypothesis for the generation of these autoantibodies is through molecular mimicry secondary to gut microbiome dysbiosis[60]. The binding of autoantibodies to the basement membrane activates the complement cascade, and results in an acute inflammatory response - the recruitment of inflammatory effector cells (macrophages, neutrophils, CD8-$T_C$ cell, CD4-$T_H$ cell, and plasma cells), the release of cytokines and growth factors (interleukins (IL)-2, IL-5, IL-13, IFN$\gamma$, TNF$\alpha$) and expansion of CD4 $T_H$ subset 1 and subset 2 cells leading to chronic inflammatory response[46,61–65].

In the absence of autoantibodies, the profibrotic inflammatory response is thought to be initiated by a cell-mediated response by autoreactive T lymphocytes against the basement membrane proteins, overexpression of macrophage-derived IL-4, platelet derived growth factor and transforming growth factor $\beta$, leading to the generation of "persistently profibrotic fibroblasts", and production of extracellular matrix[46,66,67]. The aldehyde dehydrogenase/ retinoic acid metabolic pathway, which is mediated through paracrine and autocrine effects of dendritic cells, has also been shown to activate the profibrotic phenotype of fibroblasts in patients with OcMMP, both during acute inflammation and quiescence[68,69]. Inhibition of the acetaldehyde dehydrogenase enzyme with disulfiram in vitro reverts OcMMP fibroblasts to a

normal phenotype[68]. The putative mechanisms underpinning inflammation and profibrotic scarring processes in OcMMP is summarised in Figure 1.7.

The loss of tolerance to basement membrane proteins is associated with enhanced susceptibility in patients with the major histocompatibility complex (MHC) class II allele, HLA-DQB*0301 and HLA-DR4[59,70,71]. However, in a familial case study of a pair of female monozygotic twins that share the same HLA haplotypes, only one of the twin developed OcMMP, whilst the other cotwin and her children were unaffected. This suggests that pathogenesis of OcMMP could be multi-gene and associated with other environmental factors[56].

There is mounting evidence that the gut microbiome plays a critical role in the development and maturation of the immune system[72]. Alteration in the gut microbiome (gut dysbiosis) has been linked to ocular surface inflammation in a murine model of desiccating stress, but whether this change is a cause or effect of the disease process has yet to be determined[37]. The gut microbiome can be shaped by the host genotype, in particular the genes encoding for the human leukocyte antigen (HLA) molecules which may confer genetic susceptibility to autoimmunity[73]. Elhusseiny and colleagues have recently reported an interesting case of a patient with refractory OcMMP and ulcerative colitis despite being on aggressive immunosuppressive therapy, who had long-term remission of ocular surface inflammation following colectomy[60]. Taken together, this suggests that there might be a link between the gut microbiome and ocular inflammation in patients with OcMMP.

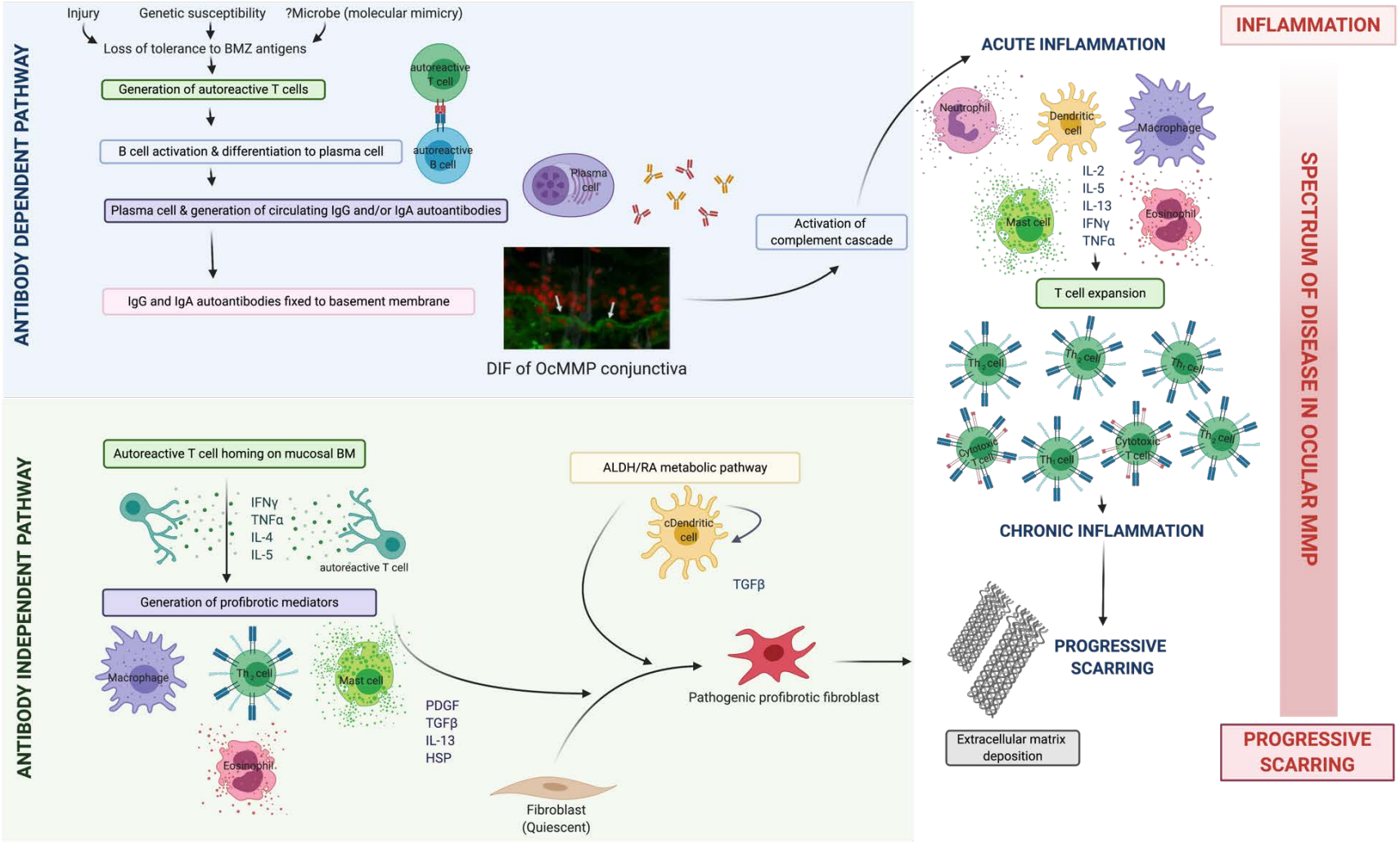



*Figure 1.7 Summary of autoantibody and cellular autoimmune-driven inflammatory and pro-fibrotic mechanisms in OcMMP.*

*(Figure 1.7, cont.) The pathogenesis of OcMMP involves a combination of both genetic and environmental factors leading to the loss of tolerance to conjunctival epithelial basement membrane proteins, generation of autoreactive T cell and/or autoantibodies that are central both to the autoantibody (IgG or IgA) and cellular autoimmune responses (antibody independent). Antibody may also be produced locally by plasma cells found in mucosal tissue sites. Binding to the conjunctival mucosal basement membrane results in complement fixation followed by an acute inflammatory response leading to influx of neutrophils, macrophages, dendritic cells, plasma cells and both cytotoxic and helper T cells. Release of profibrotic mediators (TGFβ in acute disease but not chronic, PDGF, HSP, TNFα, IL13) trigger fibroblasts and deposition of extracellular matrix. Accumulation of neutrophil collagenases, gelatinase, matrix metallo proteineases and myeloperoxidase lead to persistence of profibrotic fibroblasts and progressive scarring in eyes that have clinical control of inflammation. In addition, antibody independent pathways involving aldehyde dehydrogenase (ALDH)/retinoic acid (RA) metabolism in both human MMP fibroblasts and murine dendritic cells have been shown to drive a profibrotic conjunctival phenotype and progressive scarring.*

*Abbreviations: basement membrane zone, BMZ; immunoglobulin, Ig; direct immunofluorescence, DIF; ocular mucous membrane pemphigoid, OcMMP; interleukin, IL; TNF$\alpha$; T helper subset 1,Th1; T helper subset 2, Th2; platelet-derived growth factor, PDGF; transforming growth factor beta, TGFβ; heat shock protein, HSP; extracellular matrix, ECM. Created with BioRender.com*

### 1.5.3 Diagnosis

The First International Consensus on Mucous Membrane Pemphigoid guidance recommends that a direct immunofluorescence (DIF) biopsy is required for a diagnosis of OcMMP[47], however, this presents a long-standing diagnostic conundrum for DIF-negative patients who have typical features of progressive scarring in the absence of other causes such as Stevens-Johnson syndrome, toxic epidermal necrolysis or drug-induced cicatrising conjunctivitis[46,74]. Nevertheless, DIF-negative MMP patients are recognised in the ocular cohort as having a differing, and possibly more severe disease course. Therefore, proposals for a new diagnostic criteria for OcMMP has been made to include patients with negative DIF but positive indirect immunofluorescence (IIF) (i.e. presence of circulating autoantibodies in the serum) and patients with negative immunopathology (i.e. negative DIF and negative IIF) but have a typical phenotype of progressive cicatrising conjunctivitis in the absence of other causes[45,46,48].

### 1.5.4 Treatment strategies

Systemic immunomodulation is required to control inflammation as current commercially-available topical therapies are often ineffective[46]. The 'step-ladder' or 'step up' and 'step down' approach to the use of systemic immunotherapy has been advocated to control inflammation and minimise side-effects from medication, and is summarised in Figure 1.8. Low-dose oral tetracycline (doxycycline, initially 100mg then 50mg/day maintenance dose) is also used for its anti-inflammatory properties, which include inhibition of metalloproteinases, chemiotaxis, and pro-inflammatory cytokines (IL-6, IL-8, TNF$\alpha$)[75–77]. However, in many cases, subclinical scarring can still progress despite apparent control of inflammation. Once scarring has developed, at present, there are no clinical therapies available to reverse the process[46].

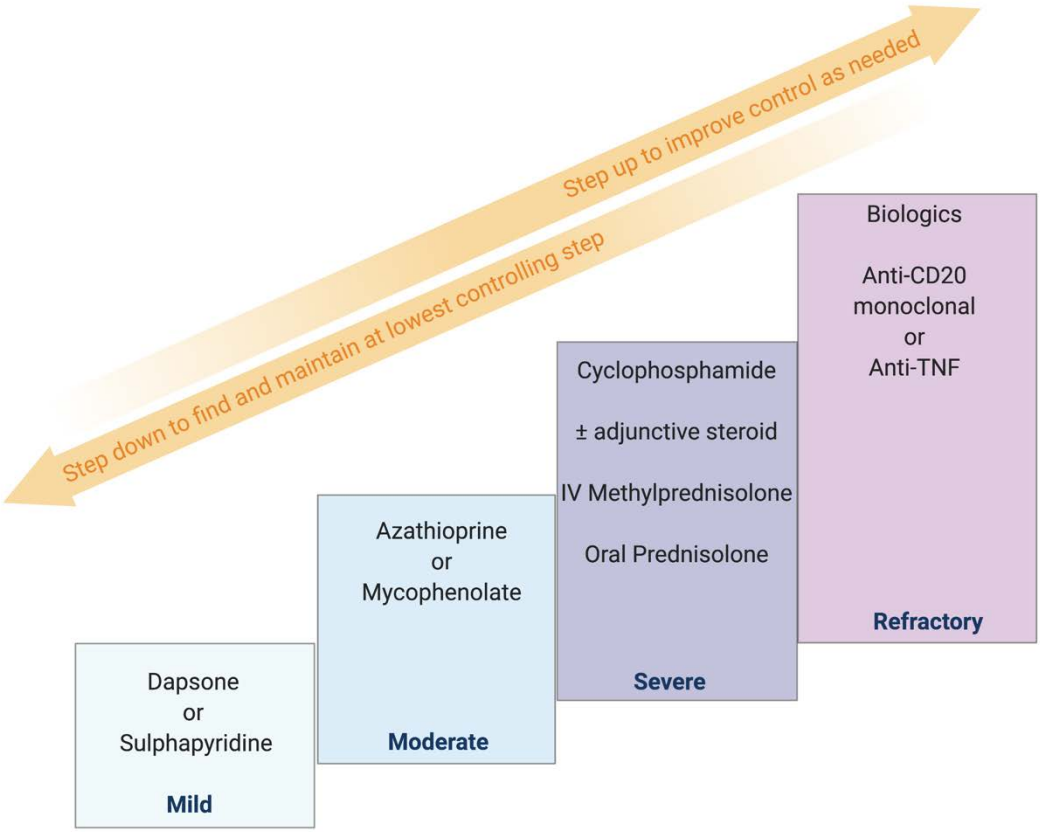



*Figure 1.8 Immunosuppression strategies for OcMMP.*

*'Step ladder' approach to use of immunosuppression (based on Rauz et al and Dart[45,46,78]). Patients start treatment at the step most appropriate to the initial severity of their disease (mild, moderate, or severe) and may step down therapy based on control of inflammation. For mild inflammation, dapsone or sulphapyridine is used. For persistent disease, azathioprine, or mycophenolate (if intolerant to azathioprine) may be added or substituted. For severe inflammatory disease, cyclophosphamide is used with adjuvant steroids. For refractory disease, biologics such as anti-CD20 monoclonal immunoglobulins (Rituximab) or anti-tumour necrosis factor (anti-TNF) may be used. Created with BioRender.com*

## 1.6 Ocular infections

### 1.6.1 Microbial keratitis

The World Health Organisation estimates 15 million people are visually impaired worldwide from corneal blindness secondary to a range of underlying pathological processes including infections (trachoma, onchocerciasis, bacterial, herpetic), trauma, nutritional deficiencies and inflammatory diseases[79]. Often under-reported, the incidence approximates 1.5-2.0 million cases of unilateral blindness per annum and includes cases of microbial keratitis (MK)[79]. Microbial keratitis is a term that describes acute corneal infection or abscess that forms the most common cause of acute medical ophthalmology emergency admission within the UK and is a leading cause of preventable blindness[80,81]. Worldwide incidence, prevalence and causative organisms have geographical variance such as an incidence of 799 per 100,000 in Nepal compared to an estimated UK incidence of 40-52 per 100,000 per year, and fungal species are more common in tropical or subtropical climes versus Gram positive or negative bacterial species predominating in temperate areas[82–84].

Patients typically present to the emergency department with painful loss of vision associated with discharge (**Figure 1.9**). The most significant risk factor is contact lens wear together with trauma/injury, previous eye surgery, topical or systemic immunosuppression and neuropathic cornea[85]. During the acute episode, many patients are reviewed daily in Emergency Departments, whilst those with the severest eye infections are admitted to specialist ophthalmology wards for intensive eye care. Data from NHS Digital estimate 2,617 admissions per year for microbial keratitis and each NHS hospital admission is associated with a significant direct and indirect cost burden[81].

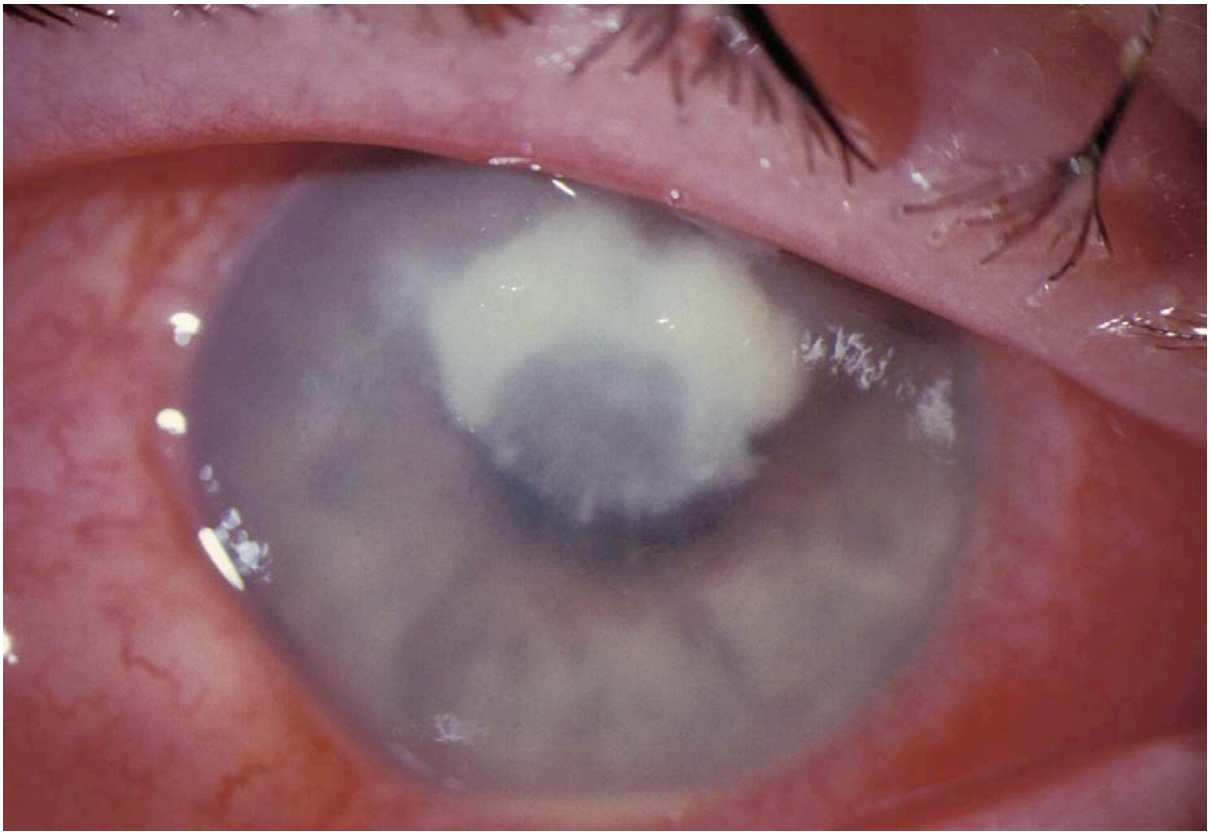


*Figure 1.9 Anterior segment photograph of patient presenting with microbial keratitis, showing a large, focal central corneal infiltrate with marked conjunctival injection.*

*Courtesy of P.I. Murray.*

The immunopathogenesis of ocular infection is largely dependent upon virulence factors of the offending pathogen, such as secretion of toxins or proteases, biofilm formation, and host factors, including host genetic background and age[86]. Broadly, recognition of molecular structures on pathogens (pathogen-associated molecular patterns, PAMPs) by pattern recognition receptors (PRR) such as toll-like receptor (TLR), N-formyl methionyl receptor, mannose-binding lectin, activates the complement cascade (opsonisation, phagocytosis and formation of membrane attack complex enabling lysis of the microbe), influx of polymorphonuclear leucocytes, natural killer cells, macrophages, γδ T cells, dendritic cell maturation, release of pro-inflammatory cytokines (IL-1, IL-6, IL-8, IL-12, IL-17, interferon-α), macrophage inflammatory proteins and monocyte chemotactic proteins[87,88]. Upregulation of

corneal surface inflammasomes involving intracellular nucleotide-binding oligomerisation domain-like receptors (NLR) pyrin domain 3 (NLRP3) and caspase-associated recruitment domain 4 (NLRC 4) activates caspase-1 and subsequent release of IL-1β and IL-18 pro-inflammatory cytokines in microbial keratitis[87]. Polymorphonuclear cell and fibroblast secretion of TNFα, reactive oxygen metabolites, vascular endothelial growth factor and matrix metalloproteinase-9 may also inadvertently contribute to the development of corneal vascularisation, protein degradation and scarring[89]. Treatment strategies for bacterial microbial keratitis involves sterilisation of infection, with antimicrobial therapy guided by regional epidemiology of pathogens and their resistance profiles, followed by adjuvant use of steroids or anti-collagenases to dampen inflammation[81,89]. Newer treatment modality using collagen cross-linking, which utilises the antimicrobial effects of ultraviolet-A light with photochemically activated riboflavin, for the treatment of bacterial and fungal keratitis is currently been investigated[89].

### 1.6.2 Endophthalmitis

Endophthalmitis is a severe, sight-threatening intraocular inflammation and is one of the most feared complications for any ocular surgical procedures[90]. It most commonly occurs secondary to infectious pathogens entering the eye, either exogenously through a breach in the globe during surgery or post-traumatic events, or endogenously, stemming from systemic infection elsewhere in the body[86,90]. A leading cause of endophthalmitis is following intraocular procedures, with incidence rates ranging from 1.2 to 5 per 1000 cases post-cataract surgery and 0.7 to 13 per 10000 cases following intravitreal injections worldwide[90–96].

The microbiology spectrum is dependent upon aetiology with coagulase negative *Staphylococci, Staphylococcus aureus*, *Streptococci*, *Propionibacterium acnes* and Enterobacteriaceae predominantly implicated in acute post-operative cases; *Bacillus cereus*, fungi, Streptococci, *Clostridium* sp., and *Microsporidium* sp. in post-traumatic cases; *Candidia albicans*, *Aspergillus*, *Staphylococcus*, and Enterobacteriaceae in endogenous endophthalmitis[90,95,97–100]. Potential risk factors for developing post-procedural endophthalmitis include history of ocular surface disease, diabetes mellitus, and immunocompromised and older age[90,100]. Patients typically present with painful loss of vision associated with hypopyon and vitritis (**Figure 1.10**)[97]. Despite aggressive medical and surgical interventions, endophthalmitis generally results in permanent partial or total sight loss. Failure in treatment is attributed to the rise in antimicrobial resistance strains from widespread use of broad-spectrum antibiotics due to the lack of accurate diagnostics tools to rapidly identify these causative pathogens[101,102].

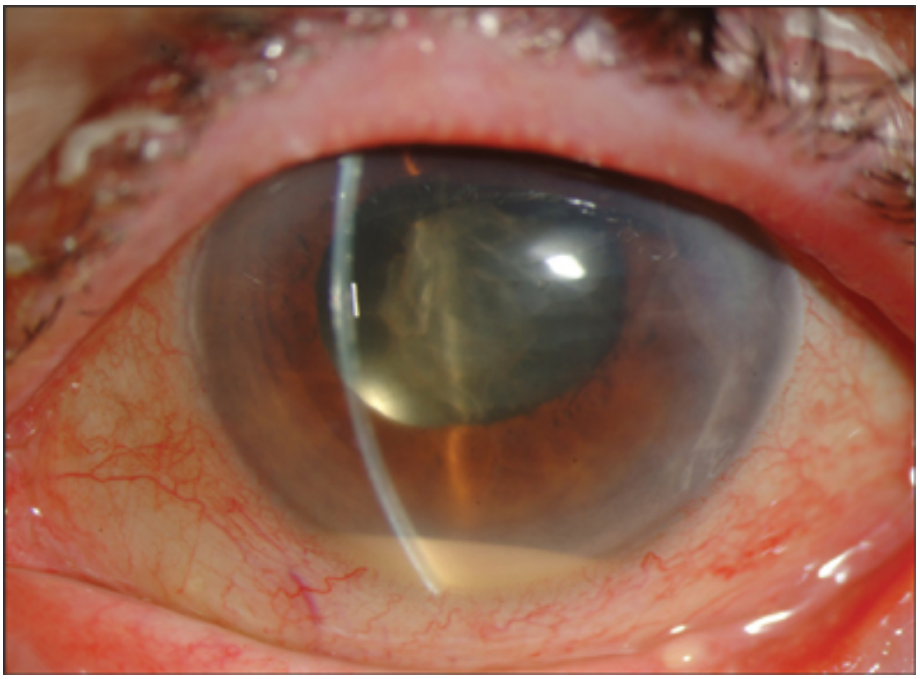


*Figure 1.10 Anterior segment photograph of patient presenting with post-operative endophthalmitis, showing hypopyon and conjunctival injection.*

*Courtesy of P. I. Murray.*

In endophthalmitis associated with microbial infection, the recognition of pathogen associated molecular patterns by pattern recognition receptors, such as toll-like receptors, activate signalling through adaptor proteins [myeloid differentiation primary response 88 (MyD88) and Toll/interleukin-1 receptor-domain-containing adapter-inducing interferon-β (TRIF)], upregulating nuclear factor-κβ and interferon regulatory factor 3 (IRF-3)-mediated expression of proinflammatory mediators, such as TNFα, IL-1β, IL-6, chemokine (C-X-C) motif ligand-1 (CXCL1), in the ocular cells (Muller cells, microglia, retinal pigment epithelium)[86,103]. The permeability of the blood-retina barrier is increased to allow the migration of effector cells into the posterior segment of the eye[86]. Polymorphonuclear cell activation leads to further release of cytokines with unintended bystander ocular tissue destruction, such as gliosis and endothelial cell damage[103]. Apoptosis, or programmed cell death, mediated through Bax and Fas proteins, help to regulate inflammation upon pathogen elimination, thereby limiting damage to ocular tissue[86]. A key component in the initial clinical management of exogenous bacterial endophthalmitis is intravitreal administration of antibiotics, most commonly broad-spectrum vancomycin for Gram-positive coverage and ceftazidime for Gram-negative coverage[86,91]. The use of systemic antibiotics with good ocular penetration such as ciprofloxacin and moxifloxacin has been also reported[91].

### 1.6.3 Conventional clinical microbiology diagnostics and management of ocular infections

Currently, conventional microbial diagnostic methods are based on culturing of samples on different media for species identification, followed by antimicrobial susceptibility testing (Figure 1.11). Ocular samples taken for microscopy, culture and sensitivity for detection of underlying aetiological organisms provides the mainstay investigation together with molecular diagnostics using polymerase chain reaction (PCR) where underlying herpes viral aetiology is also suspected[80].

In addition, the sensitivity and specificity yields using conventional culture-based techniques for the diagnosis of ocular infections is poor, ranging from 36-41% as most causative pathogens are not easily cultured in-vitro[104–107]. Time to identification of organism is around 24-72 hours but due to the micro-size of the inoculum, it may be several days before the causative organism is known and 7 days or more before full sensitivities are available. During this period, generic broad-spectrum antimicrobial therapeutic protocols are implemented based upon an extrapolation of likely organisms and culture sensitivities derived from local institutional or regional microbiology datasets[80,108]. Antimicrobial therapy is refined in the minority of cases when cultures and sensitivities are known[80]. This approach may reduce the effectiveness of ophthalmically important antimicrobials leading to evolution of resistant species[109,110].

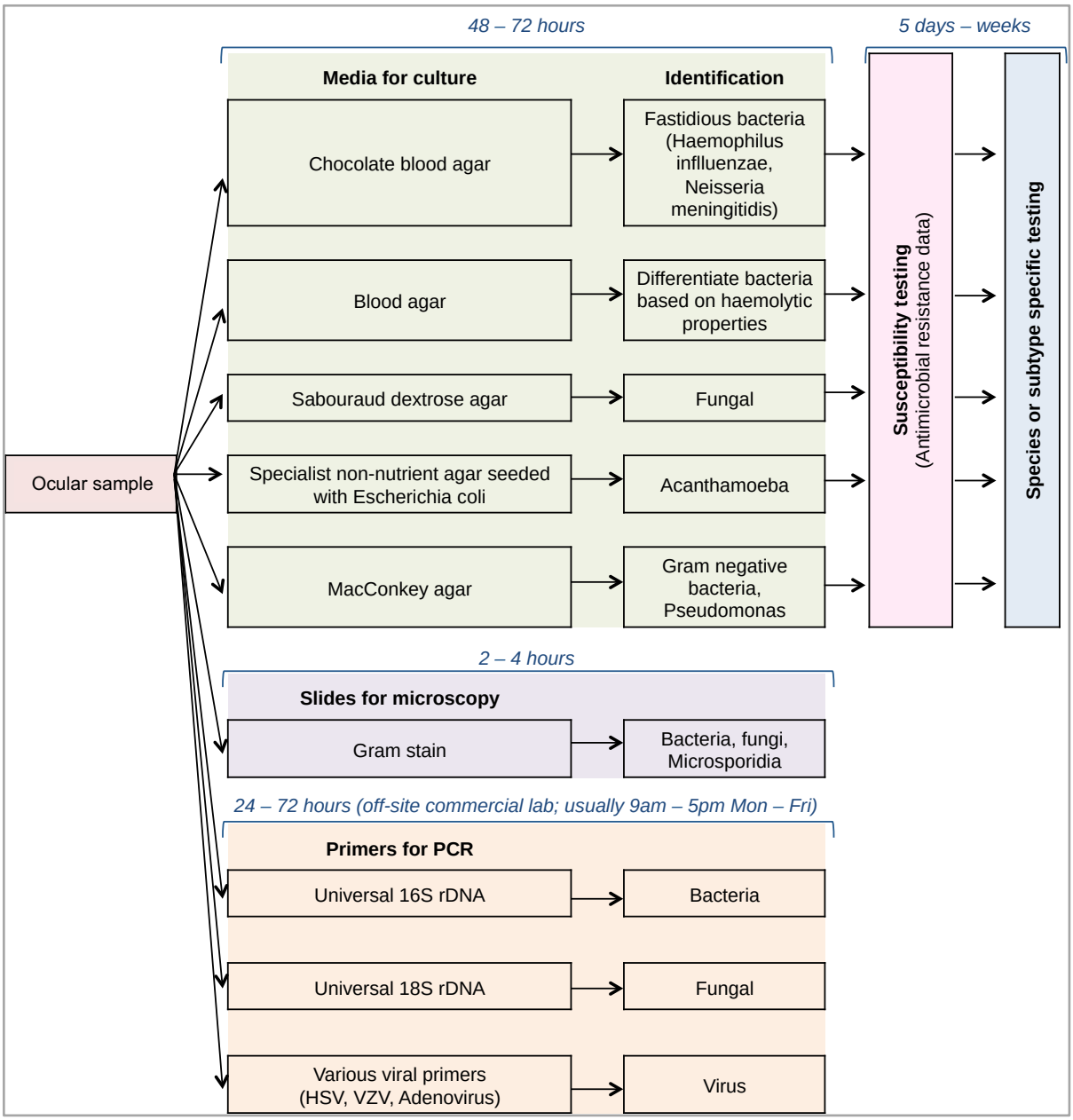



*Figure 1.11 Schematic representation of the numerous different media required for culture and susceptibility testing, along with the different primers required for polymerase chain reactions (PCR) in conventional microbiology.*

*Note: Not shown in the diagram is the use of fastidious anaerobe agar/media for culturing anaerobes and cysteine-lactose-electrolyte deficient media for culturing actinomycetes.*

*Abbreviations: rDNA, ribosomal deoxyribonucleic acid; HSV, herpes simplex virus; VZV, varicella zoster virus.*

Prognosis is dependent upon the virulence of the pathogen and efficacy of early treatment in the first 3 days[107]. Delayed diagnosis and failure in treatment could result in corneal perforation, emergency tectonic corneal transplantation and loss of eye (evisceration) in

microbial keratitis, and retinal detachment and evisceration in cases of endophthalmitis[111]. Virulence of the organisms is attributed to the rise in antimicrobial resistance strains from widespread use of broad-spectrum antibiotics stemming from the lack of accurate diagnostic tools to rapidly identify these causative pathogens[101,108,112]. Non-culture methods have entered clinical practice including PCR and in vivo laser scanning confocal microscopy[113,114]. These technologies are limited and do not give information beyond species identification, require pre-testing assumption of the pathogen and target only a few known, common pathogens. PCR testing of pathogen DNA from the eye is usually commercially driven. The service is largely 9-5pm Monday to Friday, impacting on an increased delay for out-of-hours patients. The technique has a high false-positive rate and emerging antimicrobial susceptibility patterns cannot be determined[33,82,114,115].

## 1.7 Sequencing methods to evaluate the microbiome and identifying pathogenic microbial genomes in ocular infections

### 1.7.1 Current sequencing platforms

Technological revolutions in sequencing platforms and its applications have driven advances in the discovery of microbial pathogenic variants and microbiome studies[116]. In recent years, whole genome sequencing has enabled the discovery of unexpected, novel pathogens, such as *Vittaforma corneae* in infectious conjunctivitis[117] and Torque teno virus in culture-negative endophthalmitis[118].

Deep metagenomic sequencing techniques, performed predominantly on Illumina sequencing platforms, have also enabled phylogenetic analysis of the temporal and geographic origin of ocular infection[106,119]. However, the prohibitively high running costs of these large sequencing platforms mean that these sequencing facilities are only available in select centres, and samples would have to pooled for batch processing resulting in delayed turnaround time[105,107,120]. Figure 1.12 and Table 1.3 summarises the commercially available sequencing platforms – Illumina, nanopore and Pacific Biosciences Single-Molecule Real-Time (SMRT) sequencing.

Major advantages of using nanopore sequencing compared to the other sequencing technologies include portability (pocket-sized MinION sequencer), relatively low start-up cost and rapid workflow, with less than an hour from sample preparation to data analysis[121–125]. Nanopore sequencing has been utilised for on-the-field, real-time genomic surveillance of

Ebola[126] and Zika[127,128] viruses. In addition, nanopore sequencing has also been used for pathogen detection in proof-of-concept clinical studies of sepsis[129], lower respiratory tract[130,131], urinary tract[132], and prosthetic joint infections[133]. However, there is relatively high error rates in nanopore sequencing (~95% raw read accuracy for nanopore)[134], but strategies to reduce nanopore sequencing error rate are constantly evolving with improvements to the pore chemistry and basecalling software[134].

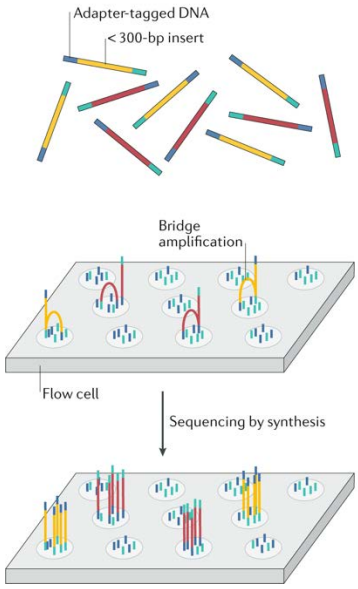
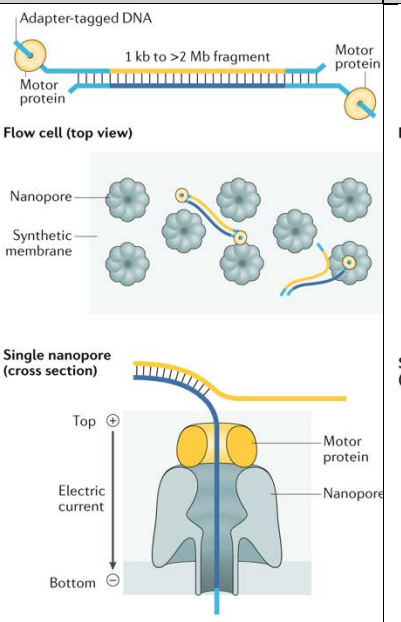
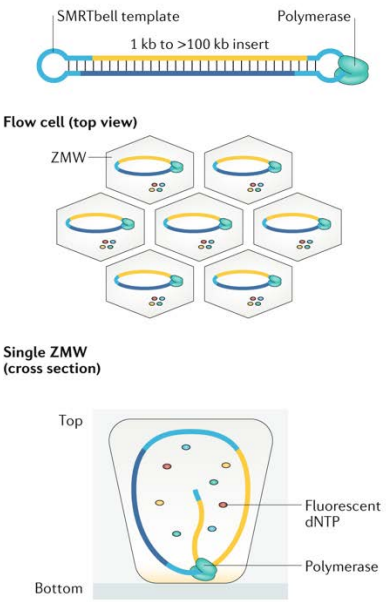
| Illumina short-read sequencing | Nanopore sequencing | Pacific Biosciences SMRT sequencing |
|---|---|---|
|  |  |  |
| DNA fragment is tagged with adapters complimentary to the oligonucleotide coating the flow cell, allowing hybridisation to flow cell surface. Clusters are generated through bridge amplification by polymerases. Reverse DNA strands are cleaved and leaving only the forward strand. Synthesis by sequencing begins whereby incorporation of fluorescently-labelled deoxynucleoside triphosphates (dNTPs) are detected via fluorescence signal from laser excitation on the DNA strand. | DNA fragment is tagged with adapters linked to tethering and motor proteins that inserts to the opening of the nanopores embedded onto a synthetic membrane. Motor proteins unwind the double-stranded DNA. Electrical current is applied, driving negatively-charged DNA through the pore. Changes in the electrical current within the pore as the nucleotide passes through the pore are recorded as 'squiggles' and used to identify the DNA sequence. | DNA fragment is ligated to hairpin adapters, forming a circular SMRTbell, bound to a DNA polymerase and loaded onto zero-mode waveguides (ZMW) chambers. As polymerase incorporates fluorescently-labelled dNTP, light pulse emitted from the bottom of the well (20 – 30nm) excites the fluorophore and is detected by a camera. Changes in the fluorescence corresponding to each DNA base is recorded within each ZMW and is used to identify the DNA sequence. |

*Figure 1.12 Overview of Illumina, nanopore and Pacific Biosciences single-molecule, real-time (SMRT) sequencing technologies.*

*Images reproduced from Logsdon[135]. Springer Nature license number: 974501209042, Dec 2020.*

| Sequencing platforms | Illumina | Oxford Nanopore | Pacific Biosciences Single-Molecule Real-Time |
|---|---|---|---|
| **Method** | Sequencing by synthesis | Nanopore sequencing | Sequencing by synthesis |
| **Data type** | Short-read<br>Single-end, paired-end | Long, ultralong read | Continuous long read<br>High fidelity read |
| **Read length (N50, maximum, platform)** | 75 bp, 500 bp (NextSeq)<br>50 bp, 250 bp (NovaSeq) | 10000 bp, >1500 kbp (MinION, GridION, PromethION) | 10000 bp, > 200kbp (Sequel II) |
| **Read accuracy** | >99.9% | 87 – 98% | 87 – 99% |
| **Runtime** | 24 – 48 hours | 0.5 – 48 hours | 10 – 30 hours |
| **Estimated instrument cost†** | £75,000 (MiSeq)<br>£185,000 (NextSeq)<br>£800,000 (NovaSeq) | £1,000 (MinION)<br>£40,000 (GridION)<br>£150,000 (PromethION) | £600,000 (Sequel II) |
| **Estimated cost per Gb†** | £10 - 60 | £20 - £500 | £13 - £500 |
| **Throughput per flow cell** | 10 Gb (MiSeq)<br>120 Gb (NextSeq)<br>>3000 Gb (NovaSeq) | 2 – 20 Gb (MinION, GridION)<br>50 – 100 Gb (PromethION) | 15 – 100 Gb (Sequel II) |
| **Portability** | Bench top (MiSeq, NextSeq)<br>Large >600kg (NovaSeq) | Portable, pocket-size (MinION)<br>Benchtop (GridION, PromethION) | Large >600kg (Sequel II) |

*Table 1.3 Comparisons of data type, costs, throughput and portability across the sequencing technologies.*

*†Costs at year 2020*

*Abbreviations: N50, sequence length of the shortest contig covering 50% of the total genome length; base pairs, bp; gigabases, Gb, kilograms, kg.*

Newer, high-throughput sequencing approaches can be broadly classified into two major options: targeted amplicon sequencing (selective amplification of specific genetic region of interest such as 16S rRNA in bacteria or 18S rRNA in fungi) and metagenomic sequencing (untargeted amplification of all genomic DNA)[82].

### 1.7.2 Targeted amplicon 16S rRNA gene sequencing

Targeted amplicon sequencing of marker genes using conserved, housekeeping regions of the genome interspersed with variable regions have been utilised to infer phylogenetic links and microbial taxonomy[136,137]. The universal 16S ribosomal RNA (rRNA) gene sequence in bacteria, which is approximately 1550 base pairs long, composed of a highly conserved region interspersed with nine variable regions (V1-9), is the most commonly used marker gene for assessing bacterial profiles (Figure 1.13)[136,138,139].

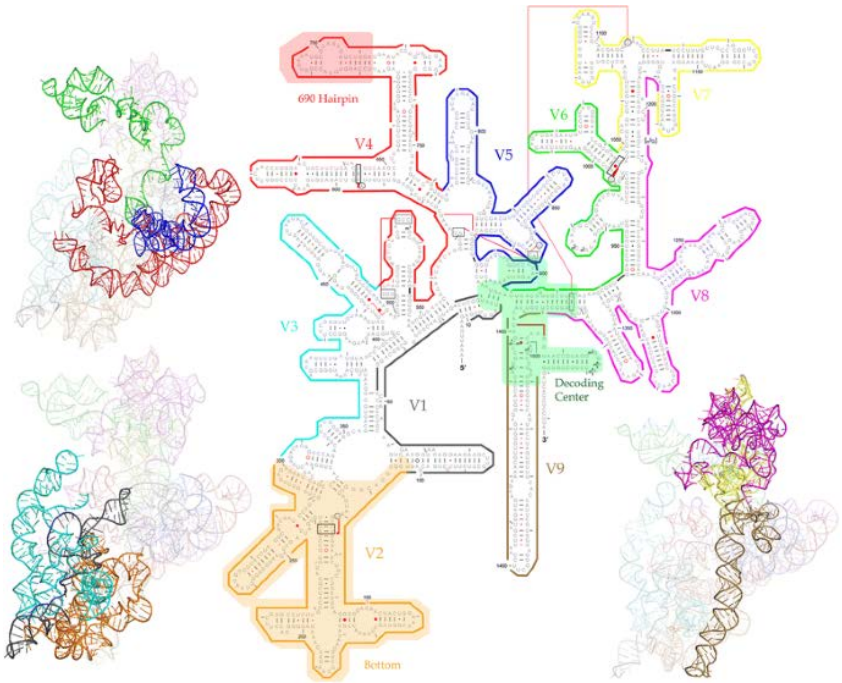



*Figure 1.13 The structure of the 16S rRNA gene.*

*Variable regions (V1 – V9) are highlighted with colour on the 2D and corresponding 3D structures. Image reproduced from Yang[140]. CC BY 4.0*

Initial microbiome analyses on the US National Institutes of Health-funded Human Microbiome Project were performed using the 16S rRNA gene sequencing[12]. Amplification of the 16S rRNA region by PCR reduces the background host contamination and by only sequencing a smaller part of the genome, this drastically reduces the sequencing depth requirements and cost. The vast majority of clinical studies have only sequenced part of the 16S gene, ranging from the single variable region of V4, V6 to three variable regions V1-3 or V3-5, because the widely used Illumina sequencing platform is only capable of producing short reads of less than 500 bases[141,142]. This short-read sequencing further compounds the taxonomic resolution of 16S rRNA sequencing, which is typically limited to genus-level resolution[139]. Choice of the hypervariable region affects the taxonomic assignment, for example the V4 hypervariable region providing better whole bacterial diversity in human microbiome studies whereas the V1-V2 hypervariable region is more specific for skin microbiota profiling[143]. With the introduction of long-read sequencing technologies, such as the Pacific Bioscience (PacBio) circular consensus sequencing and Oxford Nanopore Technologies (ONT) sequencing platforms, sequencing the full-length of the 16S rRNA gene, along with all V1 – V9 regions, is now feasible[141,144]. Several studies have shown that full-length 16S rRNA reads provides better taxonomic resolution compared to reads that only target a certain region of the 16S rRNA gene[144,145]. The choice of 16S reference database also impacts upon the taxonomic assignment[145–147]. For example, the expanded Human Oral Microbiome Database contains references to microbes specifically from the aerodigestive tract whilst the National Center for Biotechnology Information (NCBI) 16S RefSeq, the Ribosomal Database Project (RDP) and the SILVA rRNA databases include taxa from all sources (human and non-human hosts) and the environment[148–151]. Differences in the characteristics of the three databases are summarised in Table 1.4.

| Databases | NCBI 16S RefSeq[148] | RDP[149] | SILVA[150] |
|---|---|---|---|
| **Number of entries** | 21,314 | 3,356,809 | 2,090,668 |
| **Description** | Manually curated 16S ribosomal rRNA sequences | Aligned and annotated 16S ribosomal rRNA sequences (and 28S rRNA fungi sequences) | Manually curated 16S small and large subunit rRNAs (16S prokaryotes and 18S eukaryotes) |
| **Domains/Type materials** | Bacteria, archaea | Bacteria, archaea, eukarya | Bacteria, archaea and eukarya |
| **Sequence validation steps** | Trimming of long sequences, removal of low quality sequences, vector screening, removal of terminal Ns, chimera check, corrected or added intron annotation | Chimera detection | Alignment quality score above 50, sequence length above 1,200 bases for Bacteria and Eukarya, and above 900 for bases for Archaea |

*Table 1.4 Characteristics of the 16S rRNA gene reference databases.*

*Characteristics of the 16S rRNA gene reference databases. Abbreviations: National Center for Biotechnology Information 16S Reference Sequences, NCBI 16S RefSeq; Ribosomal Database Project, RDP.*

Targeted sequencing involving universal PCR amplification of the specific gene of interest, such as the 16S ribosomal RNA gene for bacteria, has the propensity for high false discovery rates[152], provides limited taxonomic resolution to the genus level, and does not capture the metagenome of other organisms, such as virus, fungi and protists[136,153–155].

### 1.7.3 Untargeted whole genome sequencing

Untargeted, 'shotgun', whole ('meta-') genome sequencing profiles the entire genomic content of the sample and is able to provide deeper taxonomic resolution, to the species and strain level, along with the predicting the functional potential of the microbial community[156]. However, a key challenge in the implementation of 'shotgun' whole genome sequencing on clinical samples is the high host background which can lead to overwhelmingly greater

amounts of host reads and very few microbial reads, thereby decreasing the sensitivity of WGS in identifying the potential pathogen[157]. Several techniques to selectively enrich for microbial reads have been developed, such as the use of antibodies targeting the methyl-CpG motifs that are present in human/vertebrate DNA but lacking in most microbial DNA and differential lysis of host cells using chaotropic buffers followed by degradation of released background DNA with DNase[158–160]. Charalampous and associates have optimised a saponin-based host DNA depletion protocol for nanopore sequencing on 41 clinical respiratory samples of patients with suspected lower respiratory tract infections and reported up to 10,000 fold depletion or 99.9% of human DNA[130]. Firstly, differential centrifugation was used to pellet the host and microbial cells. Then, saponin was used to permeabilise host cell membrane by solubilising cholesterol with subsequent addition of high concentrations of NaCl to deliver an osmotic shock to lyse the damaged host cells[130,161]. Finally, DNase was added to deplete the released host DNA[130,158]. The respiratory samples in study conducted by Charalampous and associates were derived from sputum, endotracheal secretions and bronchoalveolar lavage, and therefore, had larger starting volumes of 400µl[130]. However, these host depletion techniques have not been tested on ocular clinical samples that are of much smaller in volume.

### 1.7.4 Analysis of microbiome data

Alpha diversity is a measure of the richness (number of taxonomic units) and/or evenness (relative abundances of the taxonomic units) of an ecological community[162]. Samples are rarefied to a common sequencing depth to control for sampling effort. Richness of a community may be calculated through the observed number of operational taxonomic units (OTUs) or Chao1 estimates. OTU designation is based on clustering sequences at a set

similarity threshold, such as 97% similarity for genus-level or 99% similarity for species-level taxonomy. Newer denoising algorithms have also been developed for marker gene analysis which incorporates the identification of amplicon sequence variants[163]. Chao 1 richness estimate is calculated using the following formula[164]:

$$S_{chao1} = S_{obs} + \frac{n_1(n_1 - 1)}{2(n_2 + 1)}$$

Where $S_{chao1}$ is the estimated richness, $S_{obs}$ is the observed number of OTUs, $n_1$ is the number of OTUs with only one sequence and $n_2$ is the number of OTUs with only two sequences. Chao1 is an estimate of the minimum richness in a community.

The measure of the diversity within a community is a combination of the richness and evenness, and can be calculated using Simpson or Shannon diversity indexes.

Shannon diversity index is calculated using the following formula[164]:

$$\mathrm{H}_{Shannon} = -\sum_{i=1}^{S_{obs}} \frac{n_i}{N} \ln \frac{ni}{N}$$

Where $S_{obs}$ is the number of observed OTUs, $n_i$ is the number of individuals in OTU *i*, and N is the total number of individuals in the community. The larger the H value, the higher the diversity.

Simpson diversity index is calculated using the following formula[164]:

$$D_{\text{simpson}} = \frac{\sum_{i=1}^{S_{obs}} n_i \, (n_i - 1)}{N \, (N - 1)}$$

Where $S_{obs}$ is the number of observed OTUs, $n_i$ is the number of individuals in *i*th OTU, and *N* is the total number of individuals in the community. The larger the D value, the lower the diversity. Simpson diversity index can be affected by the unevenness of the 16S copy number, therefore, Shannon Diversity Index has been used in this thesis.

Beta diversity is a measure of the overlap of multiple communities to each other. Beta diversity indices include:

- Jaccard index, which looks at the membership of the communities and calculates the overlap in the communities. However, this measurement is sensitive to under- and uneven sampling[164].
    - $D_{Jaccard} = \frac{S_{AB}}{S_A + S_B - S_{AB}}$
    - Where $S_{AB}$ is the number of shared OTUs between communities A and B, $S_B$ is the number of OTUs in community B, and $S_A$ is the number of OTUs in community A.
- Bray-Curtis dissimilarity index measures the structure dissimilarity between two communities[164].
    - $D_{Bray-Curtis} = 1 - 2\frac{\sum min(S_{A,i}, S_{B,i})}{\sum S_{A,i} + \sum S_{B,i}}$
    - Where $S_{A,i}$ is the number of individuals in the *i*th OTU in community A, and $S_{B,i}$ is the number of individuals i the *i*th OTU in community B
- Yue-Clayton theta similarity coefficient measures the structure dissimilarity between two communities[164].
    - $D_{\theta YC} = 1 - \frac{\sum_{i=1}^{S_T} a_i b_i}{\sum_{i=1}^{S_T} (a_i - b_i)^2 + \sum_{i=1}^{S_T} a_i b_i}$

  - Where $S_T$ is the total number of OTUs in communities A and B, $a_i$ is the relative abundance of OTU *i* in community A and $b_i$ is the relative abundance of OTU *i* in community B.

- UniFrac method, which is based on the phylogenetic tree shared branch length shared or unique between two communities. Unweighted UniFrac only measures the sequence distance between two communities, whilst weighted UniFrac takes into account the relative abundance of the OTU[164].

The beta diversity indices result in a distance matrix for ordination using principal coordinate analysis for visualisation of the data, and statistical tests such as analysis of molecular variance (AMOVA) and homogeneity of molecular variance (HOMOVA) are used to test for clustering in the data[165]. AMOVA determines if the genetic diversity within two or more populations is greater than their pooled genetic diversity whilst homogeneity of molecular variance (HOMOVA) determines whether the amount of genetic diversity in each population is significantly different[164].

AMOVA is used widely to estimate F-statistics, which decomposes the genetic variance into within-individual, within-population and among-populations, and detect population clustering in genetic datasets (Meirmans, 2018). AMOVA statistic tests the null hypothesis that there is no significant difference in the genetic diversity among the populations and is calculated by the following equations[166]:

$$SS_{within} = \frac{1}{n}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_{ij}^2 \epsilon_{ij}$$

$$SS_{total} = \frac{1}{N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_{ij}^2$$

$$SS_{among} = SS_{total} - SS_{within}$$

$$MS_{among} = \frac{SS_{among}}{a-1}$$

$$MS_{within} = \frac{SS_{within}}{N-a}$$

$$F_{statistic} = \frac{MS_{among}}{MS_{within}}$$

Where n is the number of samples per group/population, a is the number of groups/populations, N is the number of samples, dij is the distance between samples i and j, and εij. is 1 when i and j are from the same group/population and 0 when they are from different groups/populations. SS is sum of squares, MS is mean of squares.

HOMOVA (Bartlett's test of homogeneity of variance) is used to test the hypothesis that the genetic diversity within the populations is homogenous, and is calculated by[166]:

$$B = \frac{(N-P)\ln\left(\frac{SS_W}{N-P}\right) - \sum_{i=1}^{P}(N_i-1)\ln\left(\frac{SS_{Wi}}{N_i-1}\right)}{1+\frac{1}{3(P-1)}\left(\sum_{i=1}^{P}\frac{1}{N_i-1} - \frac{1}{N-P}\right)}$$

Where N is the total number of samples in the study, P is the number of groups/populations, Ni is the number of samples in group/population i, and SSWi is the amount of sum of squared deviations contributed by group i.

## 1.8 Aims and objectives

The gut microbiome is important for the development and maintenance of a healthy immune system and contributes to immune protection against ocular infections. Perturbations in the gut microbiome, also known as gut dysbiosis, has been linked to systemic inflammation elsewhere in the body (extraintestinal sites) and may play a role in ocular inflammation and increased susceptibility to ocular infections. Next generation sequencing technologies are capable of providing comprehensive microbial data directly from clinical samples. Targeted and untargeted metagenomic sequencing approaches may have a wide range of clinical applications spanning from microbiome analysis in diseased and healthy states, to rapid identification of pathogens in infectious diseases[18,157].

*Gut microbiome analysis in Ocular Mucous Membrane Pemphigoid using Illumina 16S rRNA gene sequencing*

The gut microbiome, initially characterised using 16S rRNA gene sequencing in the Human Microbiome Project, plays a critical role in the development of the immune system[12,13]. Alterations in the gut microbiome have been implicated in various autoimmune conditions and have been linked with ocular surface inflammation in murine models[13,37]. Mucous membrane pemphigoid is a multi-systemic autoimmune scarring disease involving all mucosal sites, including the ocular surface and gastrointestinal tract[46]. The gut microbiome profile of patients with mucous membrane pemphigoid and its relation to ocular inflammation is currently unknown. Therefore, I have undertaken a broad sampling of the gut microbiome in ocular mucous membrane pemphigoid using 16S rRNA Illumina sequencing to provide a framework for future integration with multi-omic datasets.

*Pathogen identification in ocular infections using nanopore 16S rRNA and whole genome sequencing*

Nanopore long-read sequencing, a new and emerging sequencing technology, enables real-time data analysis and is portable, therefore is ideal for rapid detection of pathogen in acute ocular infectious diseases, such as microbial keratitis and endophthalmitis, whereby prognosis is dependent upon early detection and elimination of causative pathogen[82,157]. To date, there has been no published study rigorously examining the application of nanopore sequencing in ophthalmic infectious diseases, which require optimisation as the ocular samples are small in volume and have high background host to pathogen DNA[106]. Therefore, I will describe the collection methods, host DNA depletion technique, and bioinformatics pipeline which have been optimised for the detection of pathogens directly from ocular samples.

The general hypothesis investigated in this thesis was the contribution of the microbiomes to ocular disease and the potential application of novel sequencing technologies for rapid identification of these microbiomes, both pathogenic and commensals, in ocular inflammatory diseases and infections.

Specific hypothesis:

Chapter 2: The gut microbiome profile in OcMMP is different compared to healthy controls.

Chapter 3: Full-length 16S nanopore sequencing is able to identify causative pathogen directly from clinical sample of microbial keratitis.

Chapter 3: Metagenomic and full-length 16S nanopore sequencing is able to identify causative pathogen directly from clinical sample of endophthalmitis.

The aims and objectives of this thesis are to: a) characterise the gut microbiome profile in ocular mucous membrane pemphigoid using 16S rRNA short-read sequencing, b) optimise

collection method and bioinformatics workflow for full-length nanopore 16S sequencing for detection of pathogens in microbial keratitis, and c) evaluate the clinical utility of whole genome nanopore sequencing for detection of pathogens in endophthalmitis.

# CHAPTER 2

# Gut microbial dysbiosis in Ocular Mucous Membrane Pemphigoid

# Chapter 2 Gut microbial dysbiosis in ocular mucous membrane pemphigoid

## 2.1 Introduction

Mucous membrane pemphigoid (MMP) is a rare, orphan, life-threatening autoimmune scarring disease, involving all mucosal sites, such as the ocular, oral, nasopharyngeal, anogenital, tracheal, oesophageal mucosa, and skin[45–47]. MMP with ocular involvement (OcMMP) is the commonest cause of cicatrising conjunctivitis in the United Kingdom, accounting for 61% of newly diagnosed cases, with an estimated minimum incidence of 0.8 per million population[48]. The pathogenesis of OcMMP involves a combination of both genetic and environmental factors leading to the loss of tolerance to conjunctival epithelial basement membrane proteins, and generation of autoreactive T cell and/or autoantibodies to basement membrane proteins that are central to the disease process[46]. The mechanisms on how and where the T cells first become reactive to self-antigens in OcMMP is unknown.

There is mounting evidence that the gut microbiome plays a critical role in the development and maturation of the immune system[72]. Alteration in the gut microbiome (gut dysbiosis) has been linked to ocular surface inflammation in a murine model of desiccating stress, but whether this change is a cause or effect of the disease process has yet to be determined[37].

In a murine model of spontaneous uveitis (R161H mice), Caspi and associates showed that activation of retina-specific T cell receptor was dependent on the gut microbiota[26]. Elhusseiny and associates have recently reported an interesting case of a patient with refractory OcMMP and ulcerative colitis despite being on aggressive immunosuppressive therapy, who had long-term remission of ocular surface inflammation following colectomy[60]. Taken together, this

suggests that there might be a link between the gut microbiome and ocular inflammation in patients with OcMMP.

The aims of this study were to 1) compare the gut microbiome profiles of OcMMP patients with age- and sex-matched healthy controls, and to 2) examine the relationship between the gut microbiome diversity and ocular inflammation in OcMMP.

## 2.2 Methods

### 2.2.1 Participant enrolment and sample collection

The study was approved by the London Bridge Research Ethics Committee (Systemic Gut Microbiota Driving Sight-threatening Inflammatory Ocular Disease (STUDIOUS), reference: 17/LO/0062, IRAS project ID: 140601)) and written informed consent was obtained from all participants.

Patients diagnosed with OcMMP attending the Ocular Surface Disease clinics at the Birmingham and Midland Eye Centre, Birmingham, United Kingdom were invited to participate in the study. There is no preceding microbiome study in OcMMP in the existing literature review to perform a reliable sample size calculation. Hence, the maximum feasible sample size was used, whereby all OcMMP patients attending the tertiary referral ocular surface disease clinic at Birmingham and Midland Eye Centre who fulfilled the inclusion and exclusion criteria of the study were approached and invited to participate in the study. Outcome measures : Alpha (number of OTUs, Shannon Diversity Index) and beta-diversity

[analysis of molecular variance (AMOVA) and homogeneity of molecular variance (HOMOVA)linear discriminate analysis of effect size].

The diagnosis of OcMMP was based on clinical findings typical of OcMMP [subepithelial fibrosis, loss of plica, symblepharon, limbitis, forniceal foreshortening, symblepharon formation with or without a history of intermittent or persistent conjunctival inflammation (acute or chronic)] after exclusion of other causes of cicatrising conjunctivitis, regardless of direct immunofluorescence biopsy results[74]. Direct immunofluorescence (DIF) biopsy result was deemed positive based on the presence of IgG, IgA and/or C3 deposits in the epithelial basement membrane zone[47] but a positive test was not essential for the diagnosis of OcMMP[48].

A validated clinical assessment tool for cicatrising conjunctivitis was used to measure the inflammation (disease activity), scarring (disease progression) and morbidity (the sight-threatening effects of the disease) in patients with OcMMP at time of recruitment to the study[50]. Inflammation was scored based on the degree of bulbar hyperaemia and limbitis. Bulbar hyperaemia was assessed under normal light (daylight) and compared against the standardised grading scale panel of photographs of patients with cicatrising conjunctivitis provided in the clinical assessment tool[50]. Each quadrant of the bulbar conjunctiva (superotemporal, inferotemporal, inferonasal, superonasal) was graded using a 5-point scale, range from 0 to 4, giving a maximum score of 16 for each eye. Composite score for inflammation was calculated by transposing the maximum score for each eye on a percentage scale of 0 to 100. The combined scarring and morbidity was scored based on the extent of symblepharon, corneal vascularisation and corneal opacity. The absence or presence of

symblepharon was determined by direct observation of the patient. The extent of symblepharon was graded as absent (score of 0), if less than half horizontal involvement by symblepharon of horizontal fornix intercanthal distance (score of 1), if more than or equal to half horizontal involvement by symblepharon of horizontal fornix intercanthal distance (score of 2), with the patient looking in the opposite direction of the fornix being measured. Corneal vascularisation and opacity was assessed on the slit lamp. The central cornea was defined as the central 4mm of the cornea delineated by a slit-beam scale. The peripheral cornea was defined as the area of cornea outside the central 4mm zone. Each peripheral corneal quadrant was scored as positive with a score of 1 for the involvement by vessels or opacity separately. For corneal vascularisation, a score of 1 was given to each peripheral quadrant involved, and if the central cornea was involved, a score of 1 was given, giving a maximum score of 5. For corneal opacification, a score of 1 was given to each peripheral quadrant involved, and an additional score of 5 if the central cornea was involved, giving a maximum score of 9. The composite score for combined scarring and morbidity was calculated by transposing the maximum score for each eye on a percentage scale of 0 to 100.

Immunosuppressive treatment for OcMMP followed a “step-ladder” approach guided by disease activity as outlined by Rauz and associates[45,78]. Mild cases of OcMMP were treated with Dapsone (25-50mg twice a day) and/or low dose oral tetracycline (doxycycline, 50mg once a day) for its anti-inflammatory properties[46]. Patients with persistent disease, or those with mild to moderate inflammation, either azathioprine (1 – 2.5mg/kg/day) or mycophenolate mofetil (500 – 1000 mg twice a day) was initiated. For patients with severe inflammation, either continuous oral cyclophosphamide (1 – 2 mg/kg/day) and adjuvant prednisolone (1 mg/kg/day) or pulsed oral or intravenous cyclophosphamide with

intravenous methylprednisolone was delivered to induce rapid remission before step-down to less toxic therapy[167]. For patients with refractory disease, intravenous anti-CD20 monoclonal antibody was employed[45,168].

Healthy volunteers were identified from visitors or staff working at the Birmingham and Midland Eye Centre, or from the University of Birmingham *1000 Elders project*, and invited to participate in the study. Exclusion criteria for both OcMMP and healthy controls were as follows: no history of bowel surgery, inflammatory bowel syndrome, or systemic malignancy. Participants were provided with a faecal sample collection kit comprising: clean nitrile gloves, sterile faecal sample collection container containing 97% ethanol, re-sealable plastic bag with absorbent pad, disposable paper collection sheet to cover over the toilet seat, an instruction leaflet and pre-paid postal package. Participants collected a single faecal sample produced at any time of day. Collection containers were placed in a pre-paid postal package and delivered via First class Royal Mail® services at room temperature. Upon receipt of samples, aliquots were frozen at -80$^{o}$C until DNA extraction.

### 2.2.2 DNA extraction from faecal samples, 16S rRNA gene amplification, Illumina MiSeq amplicon sequencing

Faecal DNA was extracted using DNeasy PowerLyzer Powersoil Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA yield was evaluated by Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). The 16S rRNA V4 region was amplified by PCR and sequenced on the Illumina MiSeq platform (Illumina Inc, San Diego, California, USA) using the dual-indexed 2 x 150 bp paired end protocol modified from Kozich et al[169]. Briefly, 16S

rRNA gene libraries were constructed from the genomic faecal DNA using primers to amplify the V4 region (250bp) using the primers (16Sf: GTGCCAGCMGCCGCGGTAA, 16Sr: GGACTACHVGGGTWTCTAAT). Each PCR reaction consisted of 0.5μl Phusion Polymerase (Thermo Fisher Scientific, Waltham, MA, USA), 1.5μl DMSO (Thermo Fisher Scientific, Waltham, MA, USA), 1μl each of V4 primer at 10μM, and 1μl DNA template (1ng). PCR conditions consisted of an initial 30 seconds at 98°C denaturation step, followed by 28 cycles of 10 seconds denaturation at 98°C, 30 seconds annealing cycle at 60°C, and 10 seconds extension cycle at 72°C, with final extension for 10 minutes at 72°C and kept at 4°C. Amplicon clean-up was done using 0.85x of Agencourt AMPure XP magnetic purification beads (Beckman Coulter, Indianapolis, USA). The concentration of the 16S library pools was determined using Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and the library pool was diluted to 2nM. Sequencing was performed on the Illumina Miseq platform using the Miseq Reagent Kit V2 500 cycle reagent kit (Illumina Inc, San Diego, California, USA). Negative (blank, no DNA template) and positive controls (aliquots of faecal sample from one healthy, 'generous' donor) were used for DNA extraction, 16S amplification and Illumina MiSeq sequencing.

### 2.2.3 Bioinformatic analysis and statistics

Sequencing reads were processed using bioinformatics pipeline available in the mothur v.1.44.1 software[164] – assembly of contigs between read pairs, alignment to reference region, trimming of sequence ends to the same alignment coordinates, preclustering to denoise sequences within each sample, screening for chimeras using UCHIME, classification against the SILVA reference files release 138[150] using a naïve Bayesian classifier, split into groups at the level of order, and assignment to operational taxonomic units (OTUs) at a level of 3%

dissimilarity. Alpha diversity was measured using by calculating the Shannon index and observed number of OTUs using mothur[164]. Principal coordinate analysis (PCoA) plots were created after calculating for distance matrices, demonstrating the distance between samples in two dimensions – the further the distances, the greater the difference in the microbiome composition. Comparisons were made using Yue and Clayton theta distances between samples, taking into account the relative abundance of shared and non-shared OTUs – the larger Yue and Clayton theta distance, the more dissimilar the bacterial populations between the samples. The analysis of molecular variance (AMOVA) and homogeneity of molecular variance (HOMOVA) were conducted within mothur[164]. Samples were examined for differentially abundant bacterial taxa using linear discriminant analysis of effect size (LEfSe) within mothur[170]. Statistical testing for all 16S rRNA OTU data (Kruskal-Wallis test) were adjusted for multiple hypothesis testing using the Benjamini and Hochberg method in R software v 4.0.2[171] (R scripts available in Supplementary material 2.1 and 2.2). Continuous data were summarised using medians and interquartile ranges. Statistical analyses (Mann-Whitney, Spearman non-parametric correlation) were performed using GraphPad Prism v.8.0.0 (GraphPad Software, San Diego, California, USA) whilst linear regressions were performed using SPSS v.26.0 (IBM Corp., Armonk, NY) with p-value of less than 0.05 deemed to be indicative of significance throughout.

## 2.3 Results

Faecal samples were collected from 50 OcMMP patients and 40 healthy volunteers. The demographics were comparable between the OcMMP and healthy cohorts, except that the OcMMP cohort had a higher BMI ($p<0.001$). A majority of the OcMMP patients had positive direct immunofluorescence (DIF) biopsy results (68%), with a median duration of disease of

33 months [Interquartile range (IQR): 11 – 98 months]. None of the OcMMP patients reported active extraocular symptoms. Participant characteristics are shown in Table 2.1.

A median of 94,965 sequences were generated per sample, with an IQR of 74,230 to 123,647 sequences. Samples were rarefied to 25,000 sequences. Three samples had sequences lower than 25,000 and were therefore excluded from the bioinformatic analyses (2 healthy controls, and 1 OcMMP). The average coverage was 99.7%.

| | MMP (n=50) | HC (n=40) |
|---|---|---|
| **Age (years)** | 71 (66 - 88) | 72 (65 - 76) |
| **Sex (Female, %)** | 31 (62%) | 20 (50%) |
| **Ethnicity (White, %)** | 48 (96%) | 39 (98%) |
| **BMI (kg/m$^2$)** | 29.1 (25.7 - 32.7) | 23.6 (21.9 - 26.0) |
| **Proton Pump Inhibitor Use** | 5 (10%) | 4 (10%) |
| **Duration of disease (months)** | 33 (11 – 98) | N/A |
| **Positive immunofluorescence status** | 34 (68%) | N/A |
| **Cicatrising conjunctivitis clinical assessment**[†] | | |
| Active inflammation | 14 (28%) | |
| Combined scarring & morbidity score ≥50% | 6 (12%) | |
| **Tetracycline use** | | |
| Never | 25 (50%) | N/A |
| Previously used tetracycline, now stopped | 7 (14%) | N/A |
| Currently using tetracycline | 18 (36%) | N/A |
| **Systemic Immunotherapy** | | |
| Never | 13 (26%) | N/A |
| Previously used immunotherapy, now stopped | 10 (20%) | N/A |
| Currently using immunotherapy | 27 (54%) | N/A |

*Table 2.1 Demographics of participants*

*[†]Cicatrising conjunctivitis clinical assessment scores reported for the worse eye. Active inflammation is defined as total inflammation score above 25% at time of faecal sample collection. The combined scarring and morbidity score assesses the lower and upper fornix symblepharon, corneal vascularisation and opacity. Data reported as median and interquartile range, or as n (%). Abbreviations: body mass index, BMI; direct immunofluorescence, DIF.*

| | Currently taking medication | Previously used, now stopped | Never used |
|---|---|---|---|
| **Azathioprine** | 6 (12%) | 7 (14%) | 37 (74%) |
| **Mycophenolate** | 14 (28%) | 11 (22%) | 25 (50%) |
| **Cyclophosphamide** | 2 (4%) | 10 (20%) | 38 (76%) |
| **Biologics** | 0 (0%) | 2 (4%) | 48 (96%) |
| **Systemic steroids** | 7 (14%) | 16 (32%) | 27 (54%) |
| **Dapsone** | 2 (4%) | 4 (8%) | 44 (88%) |

*Table 2.2 Details of medications/systemic immunotherapy.*

### 2.3.1 Differences in alpha-diversity between OcMMP and healthy controls

To determine any differences in the microbial diversity within each sample, observed OTUs and Shannon Index, measures of alpha-diversity, were assessed. After adjusting for participant factors in the multivariable model (age, gender, BMI, diet, proton pump inhibitor use), OcMMP cohort was found to be significantly independently associated with lower number of OTUs, with adjusted number of OTUs of 115 (95% CI: 73-157, p<0.001) and lower Shannon Diversity Index, with an adjusted Shannon Diversity Index 0.70 (95% CI: 0.35-1.04, p<0.001) lower than in healthy controls (Table 2.3 and Table 2.4).

| Variables | Univariable Models Coefficient (95% CI) | p-Value | Multivariable Model Coefficient (95% CI) | p-Value |
|---|---|---|---|---|
| **Cohort (OcMMP)** | -97.27 (-132.72, -61.83) | **<0.001** | -115.13 (-156.76, -73.49) | **<0.001** |
| **Age (per decade)** | 11.74 (-8.64, 32.12) | 0.26 | 10.92 (-9.24, 31.08) | 0.28 |
| **Gender (Female)** | -5.28 (-46.74, 36.18) | 0.80 | 14.91 (-25.95, 55.77) | 0.47 |
| **BMI (per 5kg/m$^2$)** | -13.35 (-30.89, 4.20) | 0.134 | -4.26 (-21.86, 13.35) | 0.63 |
| **Diet (Standard diet)** | 20.84 (-20.78, 62.47) | 0.32 | 56.90 (-6.88, 120.68) | 0.08 |
| **PPI use (Yes)** | -16.75(-87.35, 53.86) | 0.64 | 5.83 (-58.38, 70.04) | 0.86 |

*Table 2.3 Linear regression models of number of OTUs in OcMMP vs. healthy controls*

*Results are from linear regression models. The coefficient represents the increase in the number of OTUs per the stated number of units increase for continuous variables, or for the stated category relative to the reference category for nominal variables. Bold p-values are significant at p<0.05.*

| Variables | Univariable models Coefficient (95% CI) | p-Value | Multivariable models Coefficient (95% CI) | p-Value |
|---|---|---|---|---|
| **Cohort (OcMMP)** | -0.55 (-0.85, -0.26) | **<0.001** | -0.70 (-1.04, -0.35) | **<0.001** |
| **Age (per decade)** | 0.11 (-0.05, 0.26) | 0.17 | 0.09 (-0.08, 0.26) | 0.28 |
| **Gender (Female)** | 0.01 (-0.31, 0.32) | 0.97 | 0.18 (-0.17, 0.52) | 0.31 |
| **BMI (per 5kg/m$^2$)** | -0.04 (-0.18, 0.10) | 0.56 | 0.02 (-0.13, 1.65) | 0.79 |
| **Diet (Standard diet)** | 0.17 (-0.14, 0.49) | 0.28 | 0.40 (-0.13, 0.93) | 0.14 |
| **PPI use (Yes)** | 0.15 (-0.39, 0.69) | 0.57 | 0.31 (-0.23, 0.84) | 0.26 |

*Table 2.4 Linear regression models of Shannon Diversity Index in OcMMP vs. healthy controls*

*Results are from linear regression models. The coefficient represents the increase in the number of OTUs per the stated number of units increase for continuous variables, or for the stated category relative to the reference category for nominal variables. Bold p-values are significant at p<0.05.*

To determine if systemic treatment or clinical disease in OcMMP patients could be contributing to the alpha-diversity differences seen in the gut microbiome, observed OTUs and Shannon index parameters were assessed within the OcMMP cohort. Patients who were currently on systemic immunotherapy had significantly lower observed OTUs compared to those who had never used or had previously used systemic immunotherapy (Kruskal-Wallis statistic: 6.8, p-value: 0.03), however, there was no significant difference in the Shannon Index for these patients [**Figure 2.1, Panel B and D**]. No significant alpha-diversity differences were detected between OcMMP patients who had never used, previously used or currently on long-term, low-dose oral tetracycline [**Figure 2.1, Panel A and C**]. No significant alpha-diversity differences, as quantified by the number of OTUs (p=0.88) or Shannon Index (p=0.47), were detected between OcMMP patients who had never used, previously used or were currently using Dapsone.

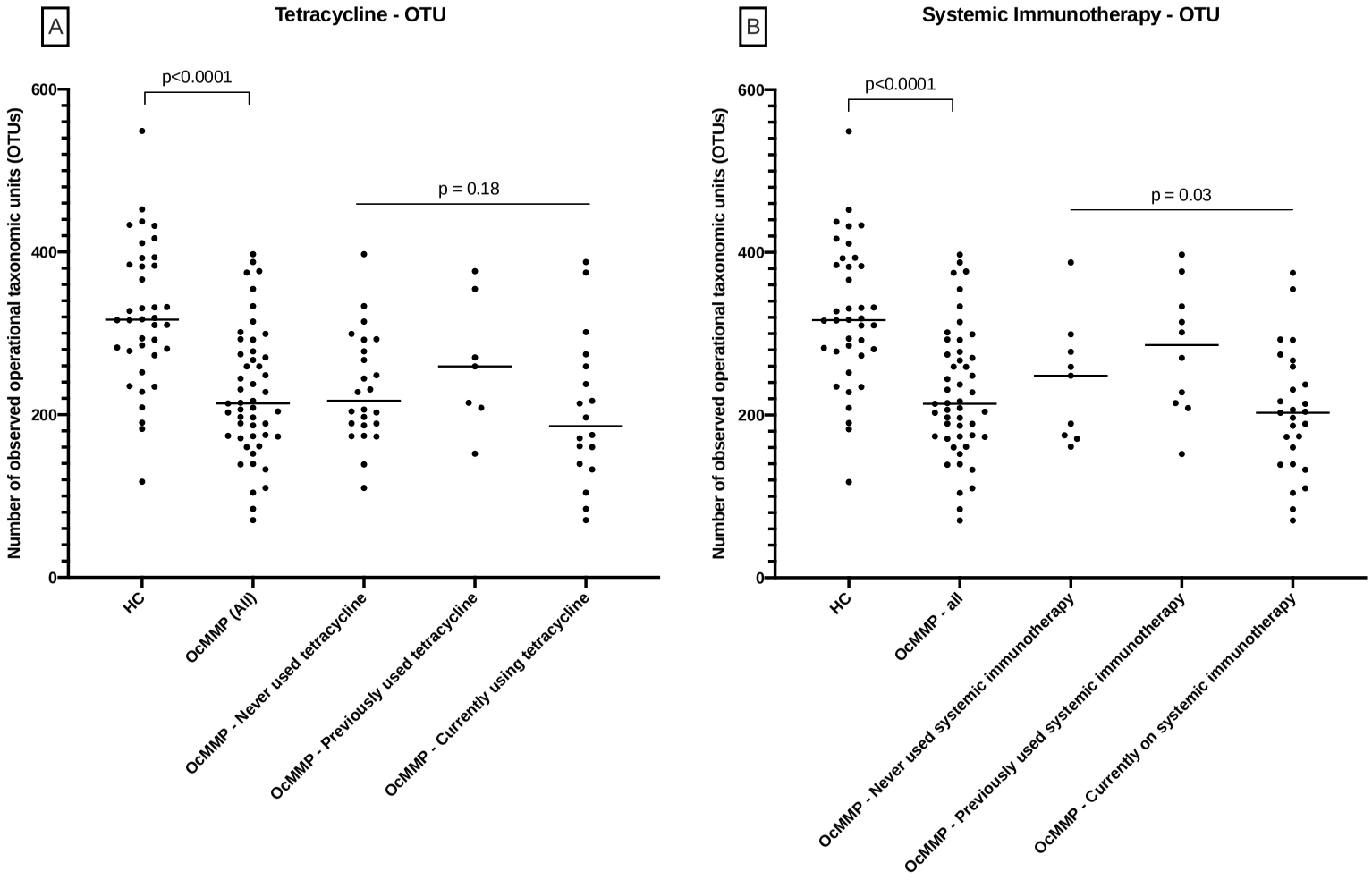



*Figure 2.1 Comparisons in alpha diversity between healthy volunteers and OcMMP.*

*Comparisons in observed operational taxonomic units (OTUs) and Shannon Index between healthy controls and OcMMP, with subsets of OcMMP patients in relation to their use of low-dose oral tetracycline (Panel A & C), and use of systemic immunotherapy (Panel B & D).*

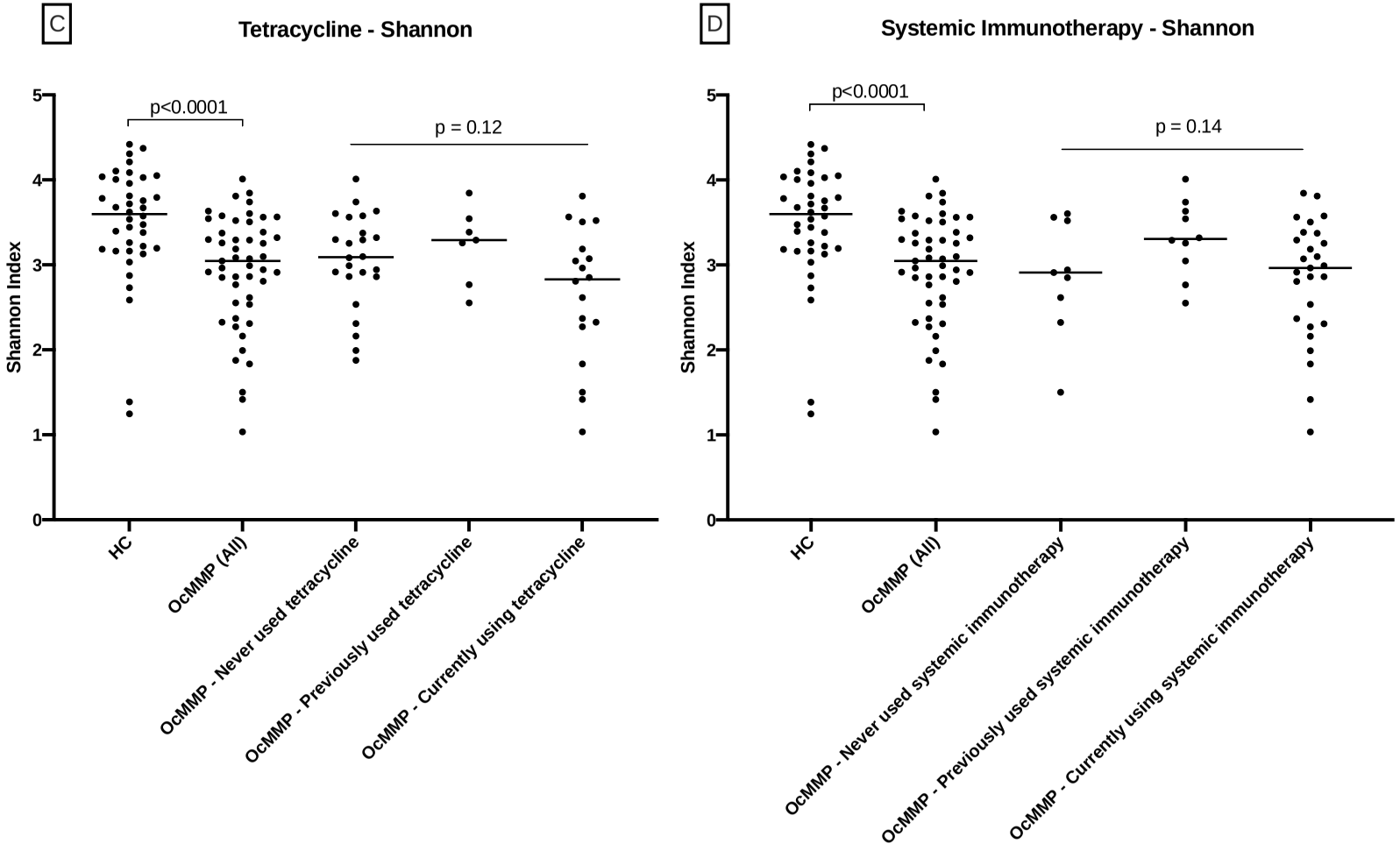



*Figure 2.1 (cont). Comparisons in observed operational taxonomic units (OTUs) and Shannon Index between healthy controls and OcMMP, with subsets of OcMMP patients in relation to their use of low-dose oral tetracycline (Panel A & C), and use of systemic immunotherapy (Panel B & D).*

Reduced number of observed OTUs were weakly correlated with inflammation ($R^2$: 0.1, p= 0.03). However, there was no significant correlation between Shannon Index and inflammation (**Figure 2.2, Panels A & B**). No significant correlation was detected between alpha-diversity measures and the cicatrising conjunctivitis clinical assessment scores of scarring and morbidity (**Figure 2.2, Panels C & D**). Alpha-diversity was not significantly associated with duration of disease or DIF biopsy results.

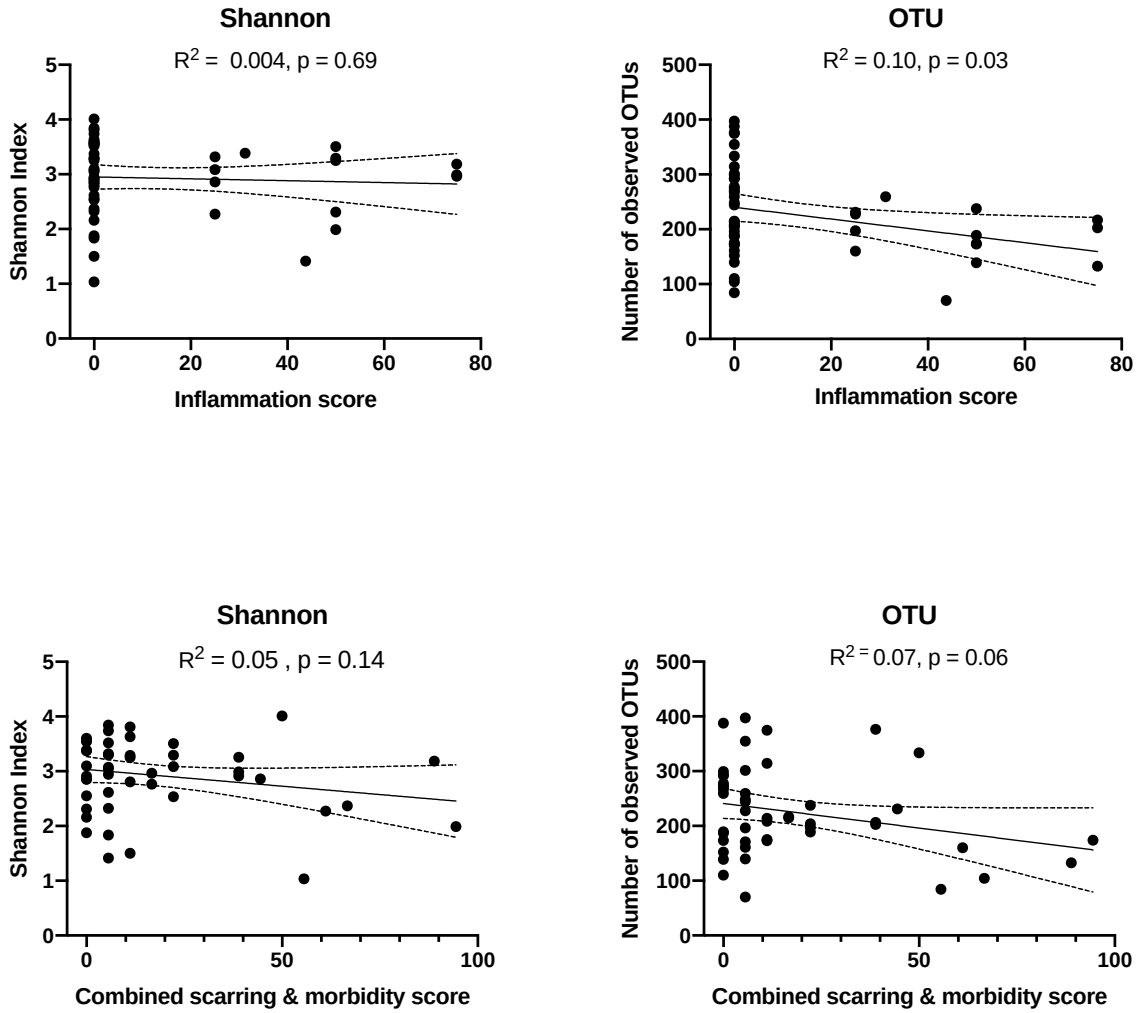



*Figure 2.2 Correlations between alpha diversity and Cicatrising Conjunctivitis Clinical Assessment Tool (CCAT) scores.*

*Correlations between cicatrising conjunctivitis clinical assessment scores of inflammation (Panel A & B) and scarring and morbidity (Panels C & D) in the worse eye. A higher clinical assessment score represents a more severe clinical disease. Abbreviations: operational taxonomic units (OTUs).*

Associations between factors and alpha-diversity (number of OTUs) were assessed within the OcMMP cohort. On univariable analysis, the number of OTUs was found to be significantly correlated with both the ocular inflammation score (p=0.03) and the current use of systemic immunotherapy (p=0.02). However, on multivariable analysis, which additionally adjusted for other factors, none of the factors considered were identified as significant independent predictors of the number of OTUs, including ocular inflammation score (p=0.21) and the current use of systemic immunotherapy (p=0.11) [Table 2.5]. The lack of statistical significance of these two factors in the multivariable model may be partly a consequence of the relatively small sample size, resulting in low statistical power. However, multicollinearity may also have played a role, given that there is a significant correlation between the ocular inflammation score and current use of systemic immunotherapy (Spearman 0.36, p = 0.01).

| Variables | Univariable model Coefficient (95% CI) | p-Value | Multivariable model Coefficient (95% CI) | p-Value |
|---|---|---|---|---|
| Inflammation score | -1.08 (-2.02, -0.14) | **0.03** | -0.84 (-2.17, 0.49) | 0.21 |
| Age (per decade) | 20.44 (-1.07, 41.96) | 0.06 | 18.22 (-7.22, 43.65) | 0.15 |
| Gender (Female) | 20.28 (-26.99, 67.55) | 0.39 | 18.06 (-31.05, 67.18) | 0.46 |
| Ethnicity (White) | -68.78 (-139.20, 1.65) | 0.06 | -74.55 (-158.08, 8.98) | 0.08 |
| BMI (per 5kg/$m^2$) | -0.26 (-4.44, 3.92) | 0.90 | 1.51 (-24.30, 27.30) | 0.91 |
| Diet (Standard diet) | 30.68 (-22.48, 83.84) | 0.25 | 16.23 (-40.31, 72.78) | 0.56 |
| PPI use (Yes) | -34.35 (-109.55, 40.86) | 0.36 | -3.53 (-88.94, 81.87) | 0.93 |
| Duration of disease (per decade) | 3.48 (-1.00, 7.96) | 0.13 | -0.86 (-7.17, 5.45) | 0.78 |
| Immunofluorescence status (Positive) | 17.58 (-32.61, 67.78) | 0.48 | -6.10 (-61.36, 48.16) | 0.82 |
| Tetracycline use (current) | -35.57 (-82.05, 10.92) | 0.13 | -42.81 (-91.66, 6.05) | 0.08 |
| Dapsone use (current) | -26.34 (-142.15, 89.48) | 0.65 | -48.27 (-169.87, 73.33) | 0.43 |
| Systemic Immunotherapy use (current) | -54.17 (-98.55, -9.80) | **0.02** | -45.54 (-101.18, 10.10) | 0.11 |

*Table 2.5 Linear regression models of number of OTUs and ocular inflammation score*

*Results are from linear regression models. The coefficient represents the increase in the number of OTUs per the stated number of units increase for continuous variables, or for the stated category relative to the reference category for nominal variables. Bold p-values are significant at p<0.05.*

2.3.2 Differences in beta diversity between OcMMP and healthy controls

To determine the differences in overall microbial community structure between OcMMP and healthy controls, Yue and Clayton theta distances were used to plot the ordination in the PCoA (**Figure 2.3**). There were significant differences in the clustering within the OcMMP and healthy control ordinations (analysis of molecular variance (AMOVA) F statistics (Fs): 2.4, p = 0.001), however, there was no significant difference in the variation (homogeneity of molecular variance (HOMOVA) B value: 0.03, p = 0.26). No significant differences were detected in the clustering within OcMMP patients based on treatment with tetracycline (AMOVA Fs: 0.73, p = 0.88) or on treatment with systemic immunosuppression (AMOVA Fs: 0.85, p = 0.68). There were no significant differences in the clustering within OcMMP patients based on the DIF biopsy results (AMOVA Fs: 1.04, p = 0.37).

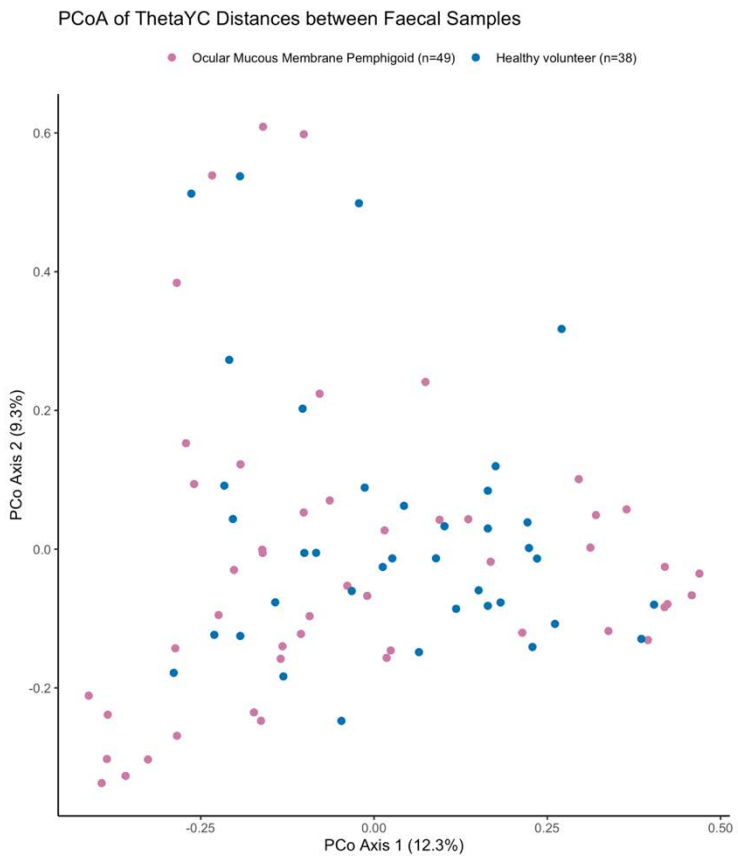


*Figure 2.3 Principal Coordinate Analysis (PCoA) of the Yue and Clayton theta distances between faecal samples from OcMMP patients and healthy volunteers.*

### 2.3.3 Compositional differences in gut microbiome of OcMMP and healthy controls

To assess the differences in microbial taxonomic abundances between OcMMP and healthy controls, OTU-assigned sequence reads were compared at various levels from phylum to genus. At the phylum level, the gut microbiome of both groups were dominated by Firmicutes [median relative abundance in OcMMP: 56.7%, IQR: 40.0 - 71.9% vs HC: 54.6% (45.9 – 66.7%)] and Bacteroidota/Bacteroides [OcMMP: 20.2% (4.6 – 36.5%) vs HC: 23.2% (14.5 – 40.7%)] followed by Proteobacteria [OcMMP: 4.8% (1.4 – 10%) vs HC: 2.7% (1.1 – 4.6%)], Actinobacteriota [OcMMP: 2.7% (0.5 – 8.1%) vs HC: 2.2% (0.9 – 6.3%)] and Desulfobacterota. [MMP: 0.3% (0.1 – 0.8%) vs HC: 0.5% (0.1 – 1.0%)] (**Figure 2.4**). There were no significant differences in the gut microbiome composition of OcMMP and HC in the phylum, class and order levels. There were significantly altered proportions of microbial taxa at the order (n=13) (**Figure 2.5**), family (n=18) (**Figure 2.6**) and genus (n=40) (**Figure 2.7)** levels, with Benjamini-Hochberg (BH) adjusted p-values of less than 0.05.

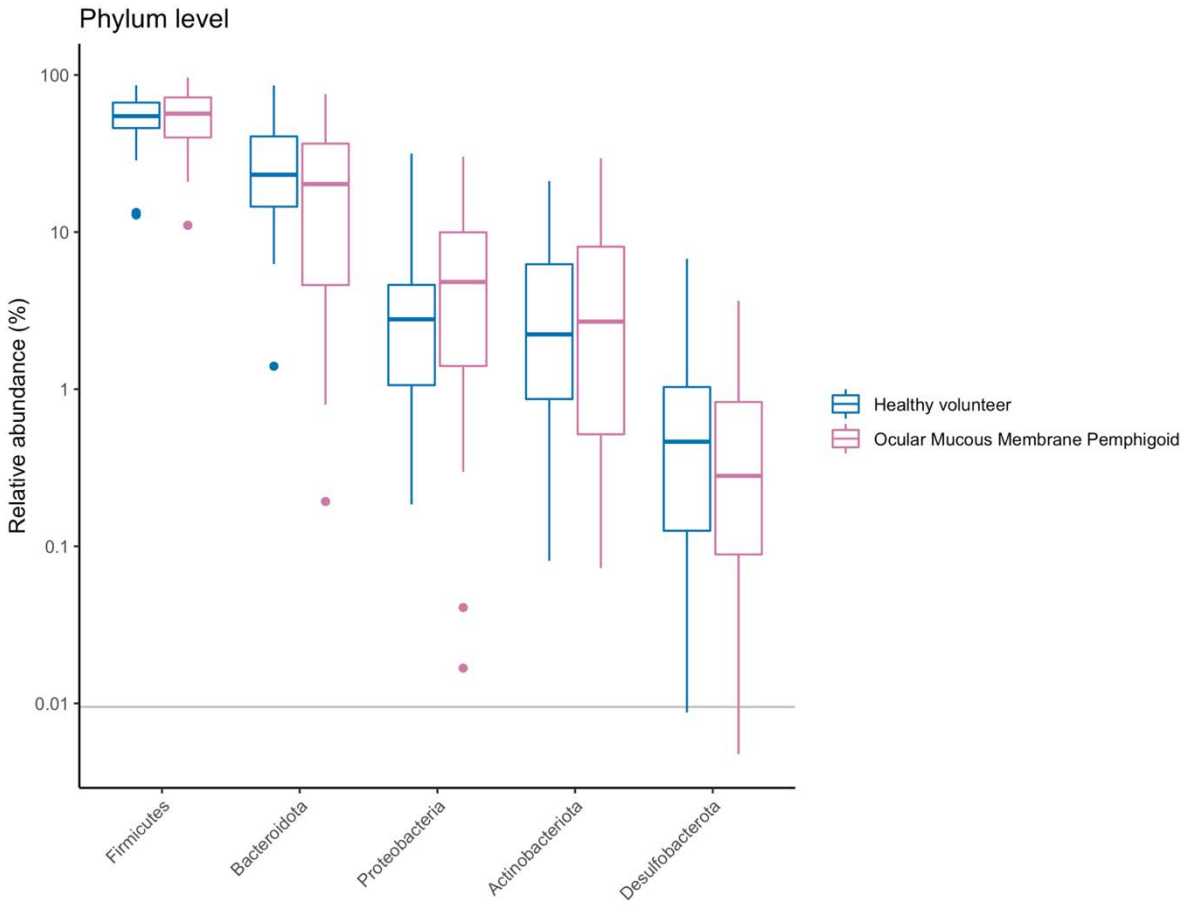



*Figure 2.4 Box and whisker plot comparing the relative abundance in phylum-level taxa between patients with ocular mucous membrane pemphigoid (OcMMP) and healthy volunteers.*

*There were no significant difference in the taxa at phylum level between the OcMMP and healthy volunteers.*

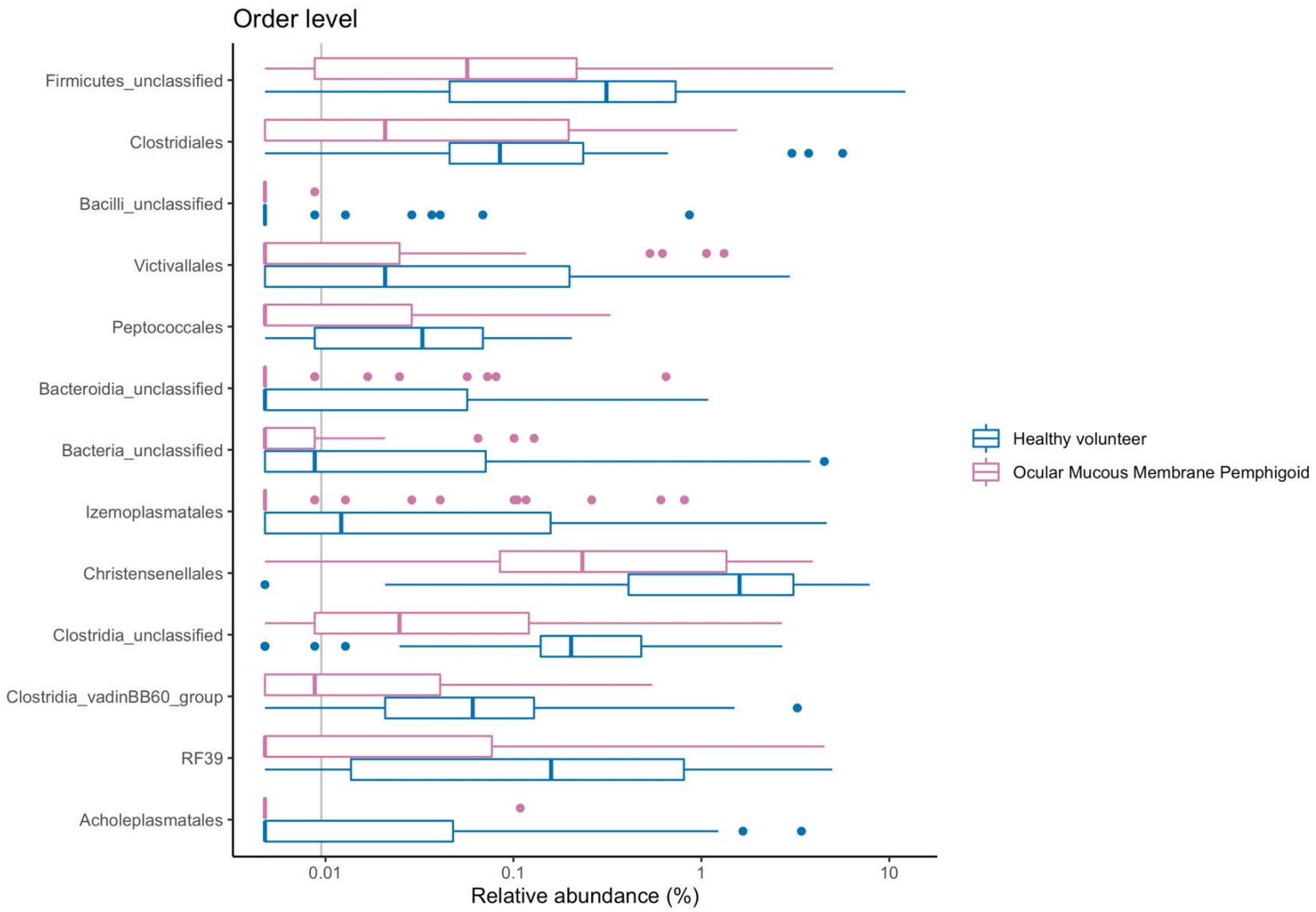



*Figure 2.5 Box and whisker plot comparing the relative abundance of taxa that differed significantly at order level (multiple hypothesis testing, adjusted p-values <0.05).*

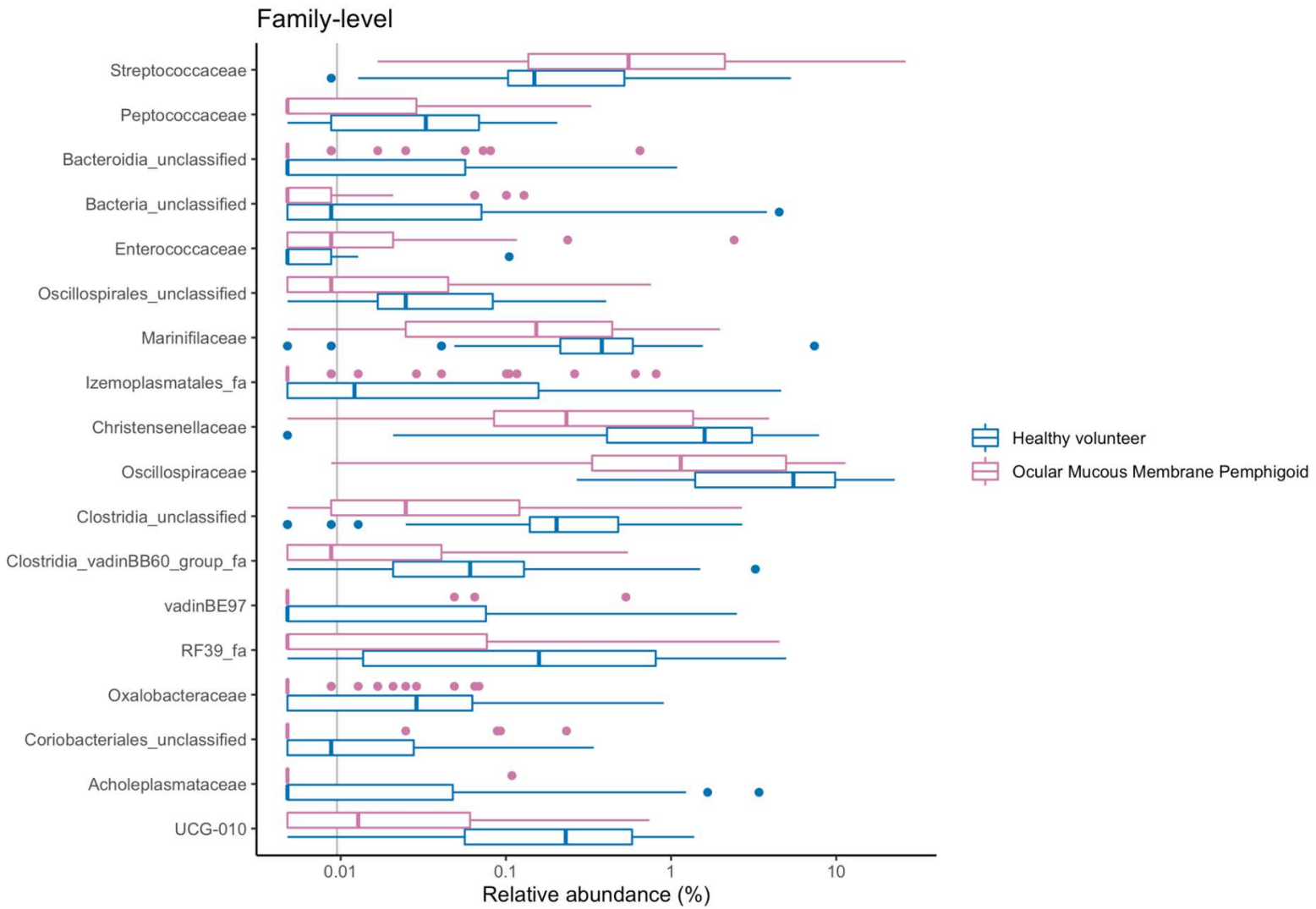



*Figure 2.6 Box and whisker plot comparing the relative abundance of taxa that differed significantly at family level (multiple hypothesis testing, adjusted p-values <0.05).*

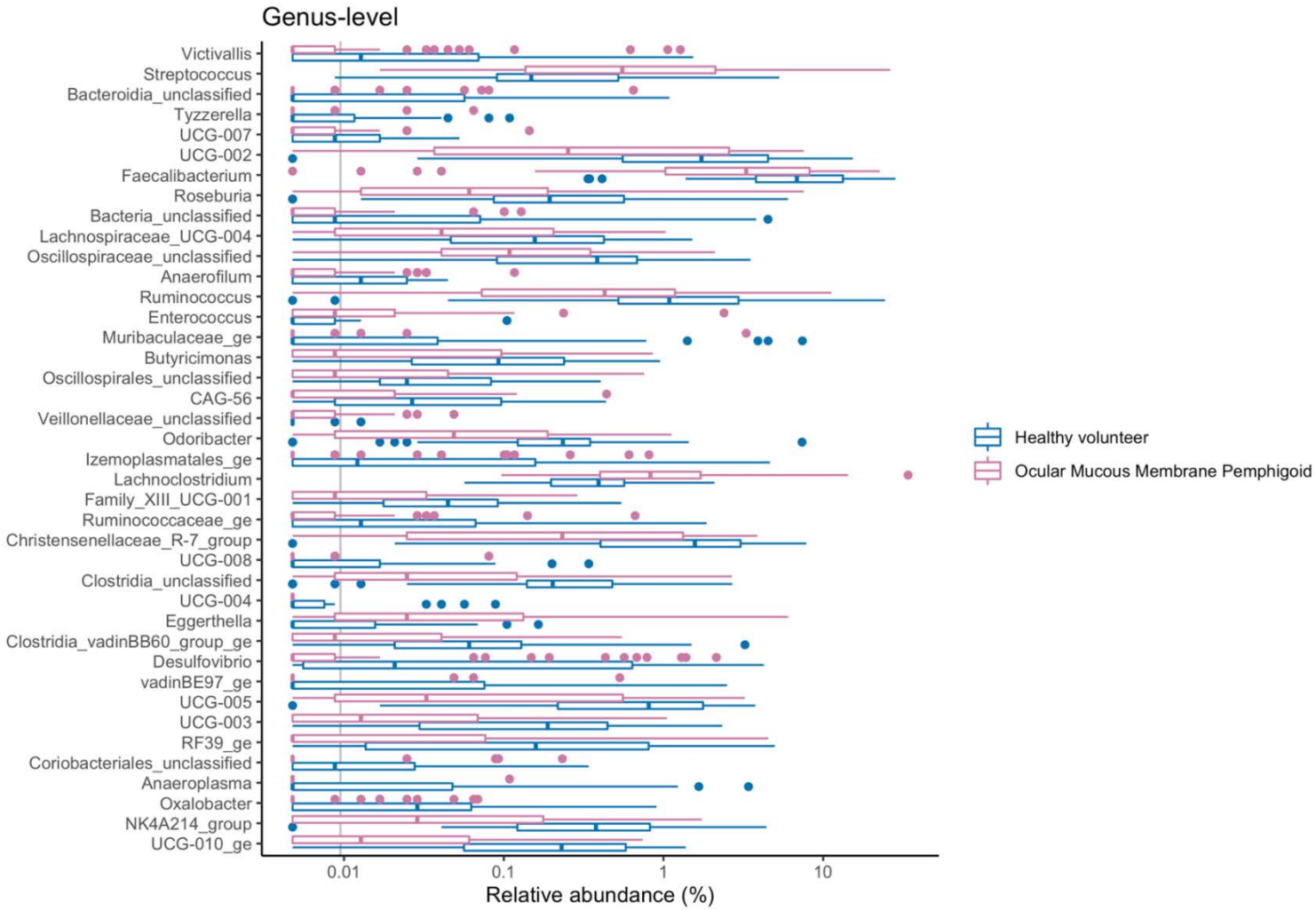



*Figure 2.7 Box and whisker plot comparing the relative abundance of taxa that differed significantly at genus level (multiple hypothesis testing, adjusted p-values <0.05).*

The gut microbial dysbiosis between OcMMP and healthy control were further analysed by linear discriminant analysis (LDA) effect size (LEfSe) method and revealed significant differences in the abundances of 15 bacterial taxa (**Figure 2.8**). The LDA scores indicated that *Streptococcus*, *Lachnoclostridium*, and *Eggerthella* were enriched in OcMMP patients ($Log_{10}$ LDA score > 2, BH-adjusted p-value <0.05) [**Figure 2.9**] whilst *Oxalobacter*, *Clostridia uncultured genus-level group (UCG) 014*, *Christensenellaceae R-7 group,* UCG 002, 003, 005, NK4A214 group and butyrate-producing bacteria such as *Ruminococcus*, *Lachnospiraceae*, *Oscillospiraceae*, *Coprococcus*, *Roseburia* were enriched in healthy controls ($Log_{10}$ LDA score < -2, BH-adjusted p <0.05) [**Figure 2.10**].

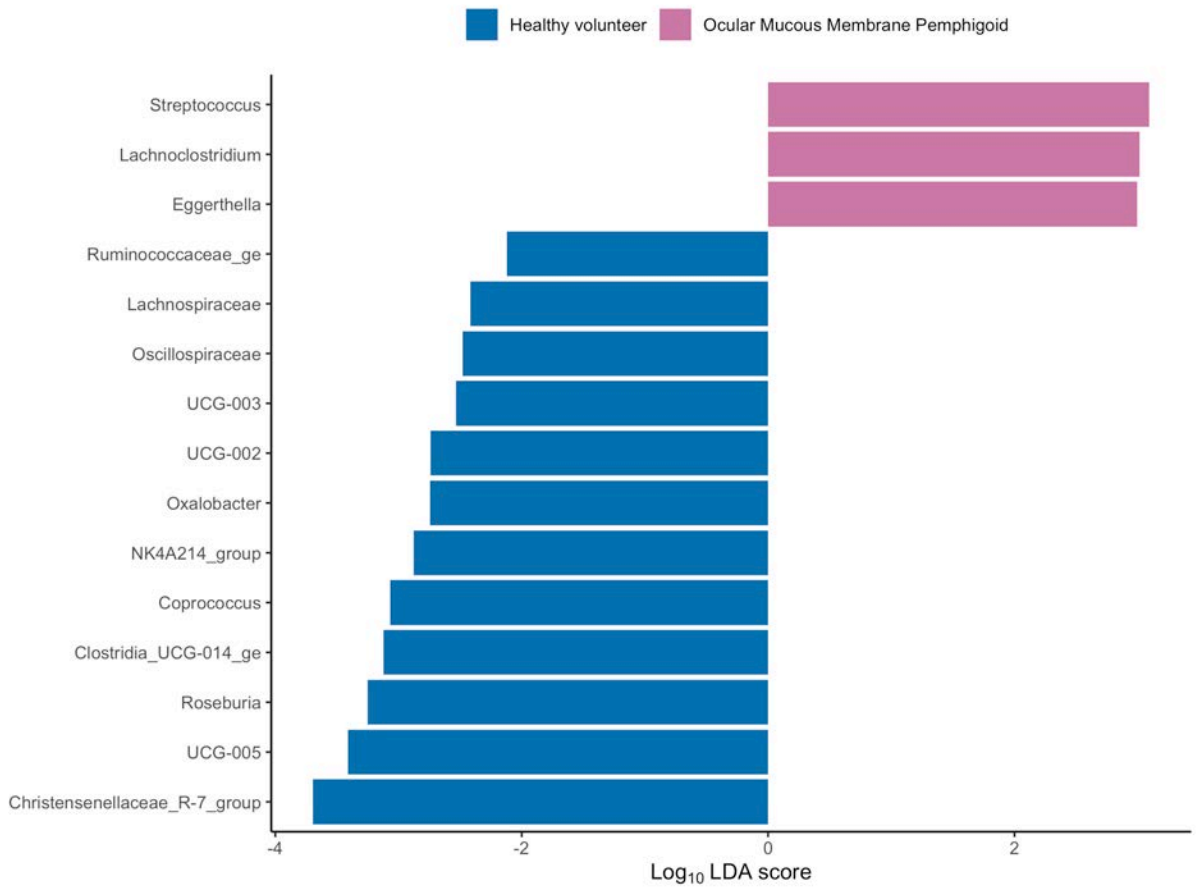



*Figure 2.8 Linear discriminant analysis (LDA) effect size (LEfSe) of significantly enriched bacterial taxa in OcMMP and healthy volunteers.*

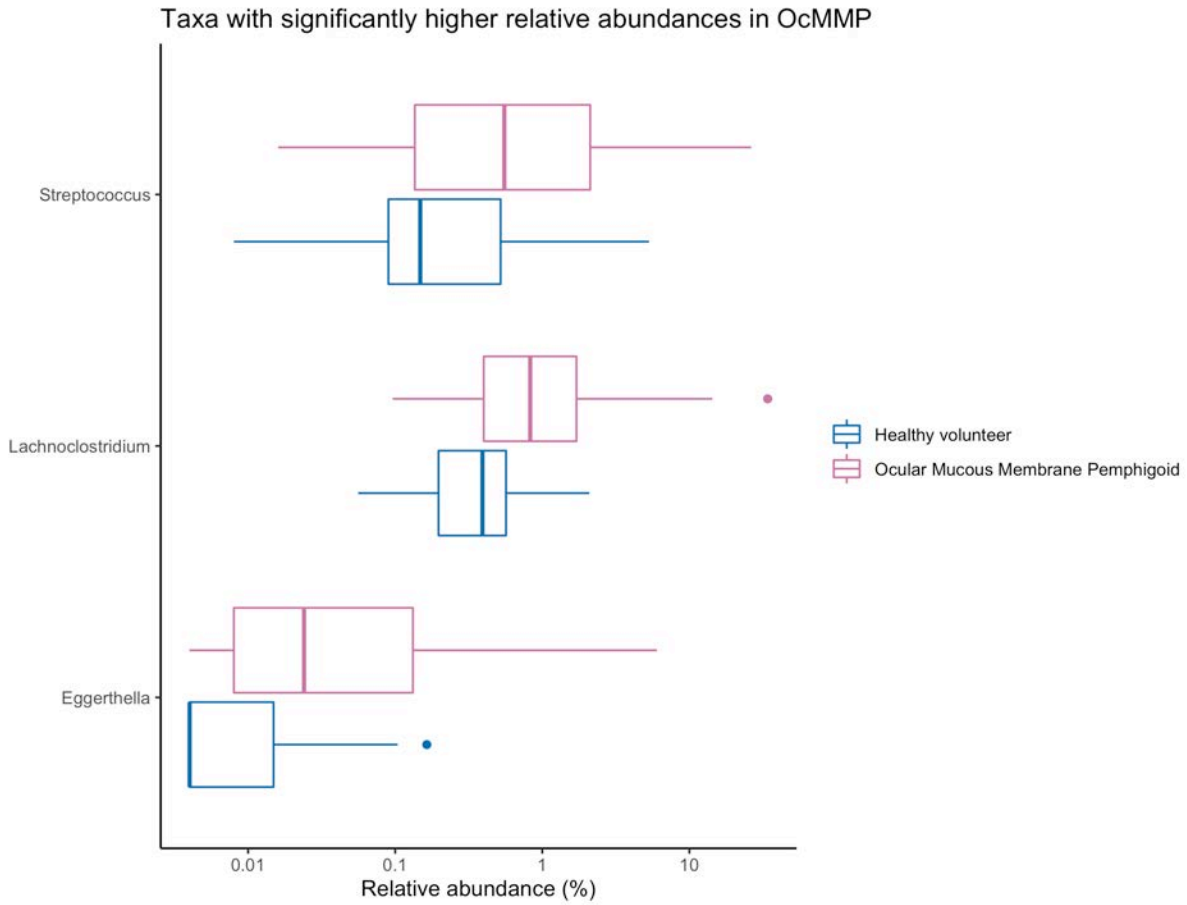



*Figure 2.9 Comparisons of the relative abundances of bacterial taxa that were significantly enriched in OcMMP.*

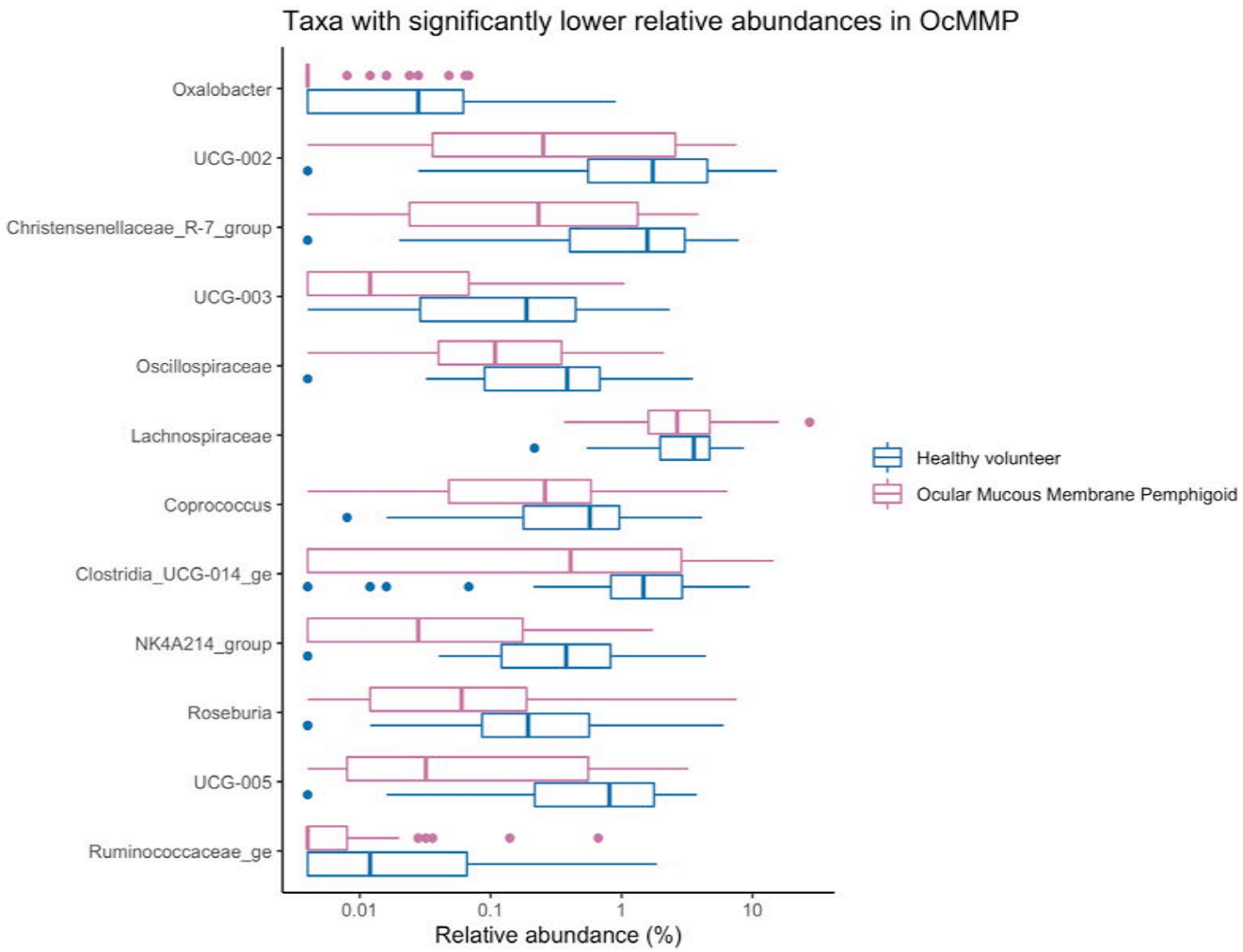



*Figure 2.10 Comparisons of the abundances of bacterial taxa that are significantly enriched in healthy volunteers.*

## 2.4 Discussion

This study is the first to compare the gut microbiome profile between patients with OcMMP and healthy controls using 16S rRNA gene sequencing. Patients with OcMMP were found to have lower alpha-diversity and altered gut bacterial composition compared with healthy controls. The reduction in the number of observed OTUs in the gut microbiome was associated with ocular surface inflammation scores using the CCAT scale and the use of systemic immunotherapy. There were no differences in the gut microbiome profile of OcMMP patients with DIF biopsy positive compared to those with DIF biopsy negative results. Taken together, these results suggest that OcMMP is associated with gut dysbiosis and correlates with ocular inflammation and the use of systemic immunotherapies.

The predominant taxa at the phylum level observed in both the OcMMP and healthy control group in this study was *Firmicutes*, followed by *Bacteroidota*, *Proteobacteria*, and *Actinobacteria*, is similar to other publications on large-scale White population cohorts[172–174]. In this study, patients with OcMMP were observed to have lower alpha-diversity compared to healthy volunteers, in concordance with previously reported studies in Sjögren's Syndrome and uveitis[37,175]. Several bacterial taxa identified as being altered in OcMMP in this study have been associated with autoimmunity. The gut microbiome in patients with OcMMP were enriched for the genera of *Streptococcus*, *Eggerthella*, and *Lachnoclostridium*. Chakravarthy and associates performed 16S rRNA sequencing on the V3-V4 region of faecal samples from 13 patients with uveitis and 13 healthy controls in a small south Indian population, and have shown increased abundance of *Streptococcus*, and depletion of *Lachnospiraceae*, *Ruminococcaceae* in their uveitis patient cohort[175]. Enrichment of *Streptococcus* in the gut microbiome has also been reported in Sjögren's Syndrome patients[37]. The relative and

absolute levels of *Streptococcus* spp in the gut microbiome correlated with higher osteoarthritis-related knee pain in a large, population-based Rotterdam cohort of over 1427 participants and was further validated in an independent cohort of 867 participants[172]. Although unable to establish causality, the authors postulated that the metabolites or membrane vesicles containing endotoxin produced by the gastrointestinal luminal *Streptococcus* spp. may be responsible in eliciting extra-intestinal inflammation in the synovial lining[172]. Given that the relative abundances of each taxa that were significantly altered in the OcMMP patient cohort accounted for less than 10% of the total gut microbiome, it is likely that a group of microbes sharing common metabolic or functional profiles, rather than a single predominant microbe, is associated with OcMMP. Patients with OcMMP were also found to have decreased relative abundances of butyrate-producing bacteria such as *Ruminococcus*, *Lachnospiraceae*, *Oscillospiraceae*, *Coprococcus*, *Roseburia*, similar to previous gut microbiome studies in patients with Behçet's Disease and uveitis[175,176]. Butyrate, a short-chain fatty acid by-product of the large intestinal microbial fermentation, have been shown to induce and regulate the differentiation of T regulatory cells[177]. Preliminary data by Hernandez and associates confirmed the presence and expression of butyrate receptors (G protein-coupled receptors 41, 43 and 109B) and sodium-coupled monocarboxylate transporter 1 (an electrogenic sodium- and chloride-dependent sodium-coupled solute transporter for butyrate) on the corneal and conjunctival epithelial cells[178]. Treatment with butyrate blunted inflammatory responses in conjunctival and corneal epithelial cells, and bone marrow dendritic cells *ex-vivo,* suggesting a potential role of butyrate in modulating the ocular surface homeostasis[178].

In this study, gut dysbiosis was weakly correlated with ocular inflammation in OcMMP, suggesting the gut microbiome could be driver of inflammation. It is also plausible that patients with more severe ocular inflammation could either had treatment with more aggressive systemic immunotherapies or had subclinical multi-systemic mucosal inflammation involving the intestinal system, thus affecting their gut microbiome. A prospective multi-disciplinary systemic screening study done by Ong and associates showed that a high proportion of OcMMP patients had asymptomatic disease at extra-ocular sites, including the gastrointestinal tract[74]. Seminal work from de Paiva and associates demonstrated that Sjögren's Syndrome patients with the most severe keratoconjunctivitis sicca had the lowest diversity in the gut microbiome, and were further confirmed by Moon and colleagues[179], however ocular surface microbiome was not significantly different from healthy controls[37]. This could be partly explained by the paucibacterial nature of the ocular surface secondary to the innate antimicrobial compounds found in tears and mechanical elimination through the blinking reflex or the lack of technical methods to sufficiently collect, extract and sequence the ocular surface microbiome[36].

The link between the gut-eye axis remains unclear. Alterations in the gut microbiome could play a role in the pathogenesis of OcMMP or the immune responses in OcMMP pathophysiology on the mucosal surface could be driving the changes in the gut microbiome profile in OcMMP. Another potential confounder is the effect of treatment with systemic low-dose tetracycline and/or systemic immunosuppression on the gut microbiome in patients with OcMMP. There were no significant differences in the number of observed OTUs between patients who had never taken, previously used or currently using tetracycline. The tetracycline (doxycycline) maintenance dose, used primarily for its anti-

inflammatory properties, in this OcMMP patient cohort was 50mg daily, which is much lower than the usual 200mg – 100mg dose required for treating acute infections[180]. The cumulative effect on the gut microbiome from long-term use of tetracycline, a bona-fide bacteriostatic antibiotic, at sub-therapeutic dose is still unclear. Studies have shown that antibiotic bacteriostatic/bactericidal activity is dose-dependent and that the gut microbiome is a reservoir for antibiotic-resistance genes[181,182]. There was significantly greater diversity (observed OTUs) in the gut microbiome of patients who had never been on systemic immunotherapy compared to those currently on treatment. This suggests that systemic inflammation influences the gut microbiome and those individuals who do not require immunotherapy may have less severe disease. Immunosuppressive therapies have been shown to change the gut microbial composition[183,184], and in turn the gut microbiota may affect response to immunotherapy by modulating the drug metabolism or effectiveness of lymphocyte function[185,186]. Given the correlation between systemic immunotherapy use and ocular inflammation score, it is difficult to isolate the effects of these two factors in the current study. The small sample size also limits the ability to assess interaction effects between medications, as dividing the cohort into subgroups based on all combinations of the use/non-use of two (or more) medications, the within-group sample size would be insufficient to give reliable estimates of effect sizes. In addition, the healthy control group was not on any immunosuppression therapies or tetracycline or dapsone, hence it would not be meaningful to include these patients in this analysis. A larger study would ideally be required to give a sufficient sample size to perform either matched-cases or subgroup analyses, to further interrogate the impacts of these factors. Inclusion of another group of patients on immunosuppression for other conditions (i.e. disease control) may in theory enable treatment-based comparisons of the gut microbiome profile,

however, it is very difficult to source and match patients on similar immunosuppressive therapies (type of treatment, threshold for starting or switching treatment, duration, dose, disease response to treatment, and combination with other drugs). It would also be difficult to ascertain if changes in the gut microbiome were indeed affected by the treatment or through the disease process itself (i.e. systemic vs. local gut inflammation). There are large intra- and interhost variations in the gut microbiome even amongst healthy individuals, as described by Vujkovic-Cvijin and associates[187]. To address this, longitudinal studies (i.e. follow-up of participants pre- and post-treatment) and larger sample size of 400 – 500 subjects per cohort may help to account of confounding host-associated variations in the gut microbiome.

There was no observable difference in the gut microbiome profile of patients with DIF biopsy positive and negative results. Given that DIF biopsy positivity requires the presence of circulating autoantibodies, it is likely that non-autoantibody driven components of inflammation, such as autoreactive T cells, could either contribute to or be affected by gut dysbiosis[46].

Using 16S rRNA gene sequencing, the current study is the first to characterise the gut microbiome in OcMMP, a very rare and orphan disease, in a relatively substantial number of patients (n=50) and correlating gut dysbiosis with clinical disease. There are several limitations to this study. This is a cross-sectional, observational study, therefore the distinction between causation or effect is inconclusive. Medications that are used in the treatment of OcMMP may have impact on the gut microbiome. Future longitudinal studies with larger patient cohort of varying stages of clinical activity and morbidity would be

required to investigate the relationship between the changes in the gut microbiome and clinical parameters, along with treatment-associated gut microbiome alterations. The gut microbiome was profiled using 16S rRNA gene sequencing of faecal samples. Although 16S sequencing provides useful bacterial taxonomic information, metagenomic, metatranscriptomic sequencing and metabolic profiling of both the gut (representative of the luminal microbiota) and mucosal samples along with host inflammatory markers may reveal changes in the gut bacterial, viral, and fungi microbiome in relation to the host genetics and inflammatory status. There might be other lifestyle and environmental confounding factors that have not been accounted for in this study, such as alcohol intake, smoking status, and quality of bowel movement. However, this study likely reflects the current real-world challenges in determining the gut microbiome profile in a very rare disease cohort with complex clinical heterogeneity.

In summary, this study shows that OcMMP patients have gut dysbiosis and alterations in the gut microbiome correlated with disease activity. These preliminary correlative data will provide the framework for future deep phenotyping and integration with multi-omic datasets, such as metatranscriptonomic, host genetics and metabolomics, and causative studies on the role of the gut microbiome in OcMMP. Modulation of the gut microbiome could potentially improve the efficacy and reduce the side-effects of immunosuppressive treatments, as well as prolonging remission[186,188,189]. Further work on investigating the prognostic utility of gut microbiome profile in OcMMP and improving the characterisation of the ocular surface microbiome is required to provide insights into the gut-eye inflammatory axis in disease and health.

# Chapter 3

# Evaluation of Full-Length Nanopore 16S Sequencing for Detection of Pathogens in Microbial Keratitis

# Chapter 3 Evaluation of Full-Length Nanopore 16S Sequencing for Detection of Pathogens in Microbial Keratitis

## 3.1 Introduction

Microbial keratitis is a leading cause of preventable blindness worldwide and is the most common cause of acute medical ophthalmology admission[190]. Bacterial keratitis accounts for the majority of the cases in the Western hemisphere, with a preponderance for Gram-positive organisms such as *Staphylococcus aureus* and *Streptococcus pneumoniae*, and to a lesser extent Gram-negative organisms such as *Pseudomonas aeruginosa*, and *Klebsiella pneumoniae*[84,191–193]. The current gold standard for diagnosis of bacterial keratitis is by conventional sampling and culture techniques, which is time-consuming (48 hours – 5 days) and has poor sensitivity of around 60%[85,105]. Culture-independent molecular techniques such as polymerase chain reaction (PCR) requires priori knowledge of likely pathogenic micro-organism and only a limited number of tests can be multiplexed in a PCR assay[82,114,194]. Newer high throughput sequencing, such as Illumina short-read sequencing, have enabled discovery of novel pathogens from a single sample, however, the prohibitively high running costs of these large sequencing platforms mean that they are only available in select centres, and samples are generally pooled for batch processing, resulting in delayed turnaround time[82,105].

Several challenges have hindered the adoption of sequencing technologies in routine clinical practice for microbial keratitis, including sample collection methods that are frequently ineffective[195] compounded by the low abundance of pathogens[82], high contamination from background host DNA or laboratory reagents[196], and lack of standardisation in sequencing and bioinformatics processing methods[157]. Sampling and DNA extraction methodology

significantly impact upon the downstream sequencing data in ocular samples with low biomass[197,198]. Microbial cells within the sample must be sufficiently lysed to liberate its DNA content, and this is particularly challenging for thick-walled microorganisms whereby mechanical disruption method such as bead-beating in conjunction with heat, chemical or enzymatic treatment is required in the DNA extraction protocol[199,200]. Corneal clinical samples have high host DNA content, usually constituting more than 90% of the sequences, with relatively low abundance of pathogen DNA, and therefore requiring greater depth of sequencing and costs[82,117]. Targeted amplicon sequencing of the 16S rRNA gene, a marker gene that is highly conserved among bacteria and archea kingdoms, might provide a feasible cost-effective solution for assessing bacterial taxa in microbial keratitis.

The aim of this study was to evaluate the potential of full-length 16S nanopore sequencing directly from clinical samples of bacterial microbial keratitis, focussing on the optimisation of the sampling method and bioinformatics pipeline. Firstly, using a defined mock microbial community, I evaluated and benchmarked my bioinformatics analysis pipeline for taxonomic assignment. Then, I optimised the sample collection using an ex vivo porcine model of microbial keratitis. Lastly, I performed a proof-of-concept study in which I applied the sampling technique that provided the highest DNA recovery and the optimised bioinformatics pipeline to prospectively collected samples from patients with suspected microbial keratitis, comparing the nanopore sequencing results to standard microbiology culture methods **(Figure 3.1)**.

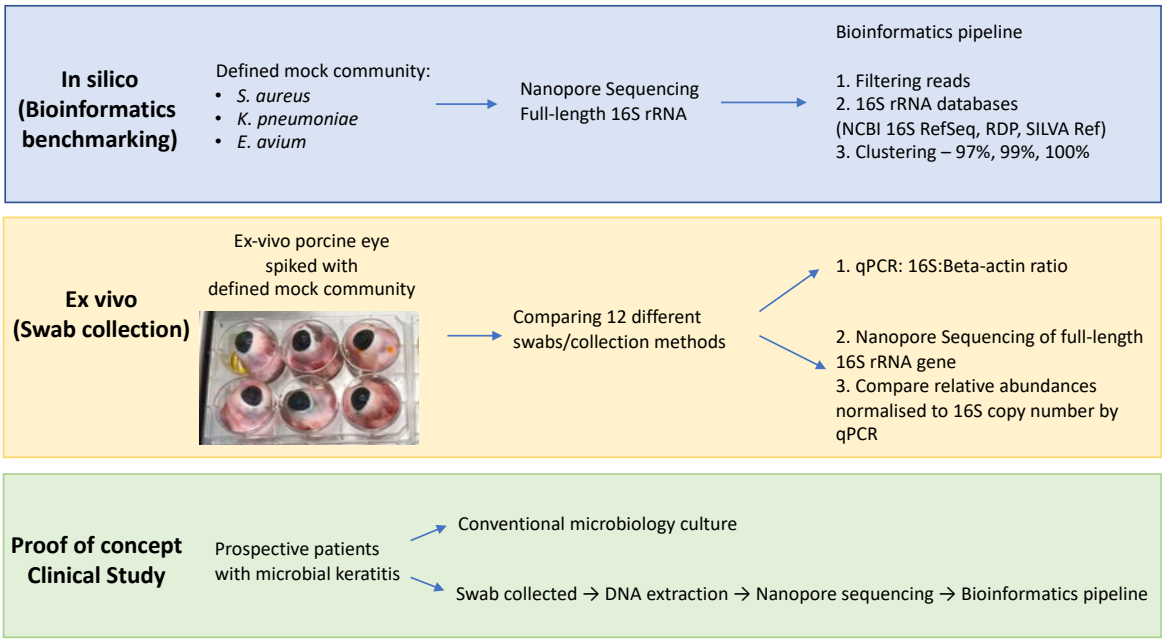



*Figure 3.1 Summary of study workflow.*

*Firstly, using a defined mock community containing* S. aureus*,* K. pneumoniae*, and* E. avium*, the full-length 16S rRNA bioinformatics pipeline was optimised by comparing the effects of filtering, choice of 16S rRNA database (NCBI 16S RefSeq, RDP and SILVA), and clustering of references reads at 97%, 99% and 100% similarity on bacterial taxonomic assignment. Secondly, using ex-vivo porcine eye model spiked with defined mock community (to simulate microbial keratitis), 12 different swabs/collection methods were used to sample the cornea and comparisons of the DNA yield, 16S to beta-actin ratio by qPCR, and the relative abundances of retrieved bacterial taxa by 16S nanopore sequencing were made. Lastly, the optimised sampling method and bioinformatics pipeline was applied directly to clinical samples of patients (n=2) presenting with microbial keratitis in a proof-of-concept study. Abbreviations:* S. aureus, Staphylococcus aureus; K. pneumoniae, Klebsiella pneumoniae; E. avium, Enterococcus avium; *National Center for Biotechnology Information 16S Reference Sequences, NCBI 16S RefSeq; Ribosomal Database Project, RDP.*

## 3.2 Materials and Methods

### 3.2.1 Mock bacterial community

Bacterial species representative of the spectrum of microbial keratitis causative Gram-positive and Gram-negative pathogens [*Staphylococcus aureus, Klebsiella pneumoniae*, and *Enterococcus avium* (previously classified as group D *Streptococcus)*] were used to define the mock bacterial community. *Klebsiella pneumoniae* and *Staphylococcus aureus* were grown overnight at 37$^{o}$C in lysogeny broth (LB) respectively whilst *Enterococcus avium* was grown in Brain Heart Infusion (BHI) broth. Negative controls consisted of the LB and BHI broth without any inoculum, respectively. Overnight cultures were diluted into fresh medium to an optical density 600nm ($OD_{600}$) of 0.05 and incubated at 37$^{o}$C with aeration. The culture $OD_{600}$ was measured every 30 minutes for 5 hours using spectrophotometer (Ultrospec™ 2100 pro, Amersham Biosciences, UK). For enumeration of bacteria, cultures were plated onto agar plates (*Enterococcus avium,* blood agar; *Staphylococcus aureus* and *Klebsiella pneumoniae*, LB agar respectively) and incubated overnight at 37$^{o}$C. The colony-forming units (CFUs) for each species were enumerated the following day. Bacterial growth curves were plotted and the mid-exponential growth phases of the samples were taken. The mock bacterial community consisted of $1x10^{5}$ CFU/ml each of *Klebsiella pneumoniae*, *Enterococcus avium* and *Staphylococcus aureus*.

### 3.2.2 Ex vivo porcine model of microbial keratitis

Freshly enucleated porcine eyes (*Sus scrofa domestica*) were obtained as a by-product of the meat industry and transported to the laboratory under storage at 4$^{o}$C. Each eye was disinfected with Povidone-iodine 10% w/w for 1 minute, followed by two rinses of sterile 0.9%

Sodium Chloride for 1 minute, and placed in an individual chamber of a sterile 6-well culture plate (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany). Using stereoscopic surgical loupes, a 4mm trephine punch (Acu-Punch®, Acuderm, Fort Lauderdale, USA) was used to create a single central anterior stromal corneal lesion (with debridement of the central 4mm). The area was sampled with the respective swabs/collection methods pre-inoculation, as a background control. Each eye was then inoculated with 20 µL of $1x10^5$ CFU/ml each of the mock community (*Enterococcus avium*, *Staphylococcus aureus*, and *Klebsiella pneumoniae*). The amount of inoculum was determined according to the estimated density of bacteria encountered in vivo in clinical infections[195]. The area was re-sampled with the respective collection methods after 30 minutes to prevent bacterial overgrowth, and placed immediately into a ZR BashingBead™ Lysis Tube containing 750 µl of DNA Shield™ (Zymo Research, Irvine, CA, USA) and stored at -80$^{o}$C until DNA extraction.

### 3.2.3 Collection method/Swabs

The different collection methods evaluated are Isohelix™ SK-2S, Sugi® Eyespear, Cotton, Rayon, Dryswab™, Hydraflock, Albumin-coated, Purflock, Purfoam, Polyester swabs, 21-gauge needle, and Polytetrafluoroethylene (PTFE) membranes **(Table 3.1)**. Sampling order was varied. A total of 60 ex-vivo porcine eye models were used. Each sampling replicate was performed on a separate eye. Swabs were not pre-moistened, as a previous pilot study that I have conducted showed that dry swabs provided higher DNA yield, compared to pre-moistened swabs **(Supplementary Table 3.1)**.

| Name | Manufacturer | Composition of Bud | Shaft | Catalogue number |
|---|---|---|---|---|
| Isohelix™ SK-2S | Isohelix, Kent, UK | Viscose Rayon | - | SK-2S |
| Sugi® Eyespear | Kettenbach GmbH & Co. KG, Escehnburg, Germany | Cotton & Cellulose | - | 30901 |
| MW1021 Dryswab™ Rayon | MWE Medical Wire, Corsham, UK | Rayon | Plastic | MW1021 |
| MW1041 Cotton | MWE Medical Wire, Corsham, UK | Cotton | Wood | MW1041 |
| MW1021D Dryswab™ Polyester | MWE Medical Wire, Corsham, UK | Polyester | Plastic | MW1021D |
| MW821 Dryswab™ Flock | MWE Medical Wire, Corsham, UK | Purflock® | Plastic | MW831 |
| MW100 Fine tip Dryswab™ Rayon | MWE Medical Wire, Corsham, UK | Rayon | Plastic | MW100 |
| MW130 Hospiswab™ Albumin | MWE Medical Wire, Corsham, UK | Albumin coated | Wood | MW130 |
| MW840 Hydraflock® Plastic | MWE Medical Wire, Corsham, UK | Hydraflock® | Plastic | MW840 |
| MW946 Sigma Swab® Purfoam | MWE Medical Wire, Corsham, UK | Purfoam | Plastic | MW946 |
| BD Needle 21G | Becton-Dickinson (BD) & Co, New Jersey, USA | Stainless Steel | - | BD301155 |
| Biopore® PTFE 6mm | Merck & Co, New Jersey, USA | PTFE | - | BGCM00010 |
| Biopore® PTFE 4mm | Merck & Co, New Jersey, USA | PTFE | - | BGCM00010 |

*Table 3.1 Summary of collection methods used.*

*Abbreviations: Becton-Dickinson, BD; Medical Wire and Equipment, MWE; Polytetrafluoroethylene, PTFE.*

### 3.2.4 DNA extraction

DNA from the mock community, conjunctival swabs, and negative controls were extracted using ZymoBIOMICS DNA Miniprep kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. Negative control DNA extraction was performed on reagents without a DNA template and also on sterile 0.9% Sodium Chloride that was used to clean the porcine eyes, whilst positive control DNA extraction was performed on 20 µL of $1x10^5$ CFU/ml each of *Enterococcus avium*, *Staphylococcus aureus*, and *Klebsiella pneumoniae* inserted directly into the lysis tube containing 750 µl of DNA Shield™. DNA concentration was determined fluorometrically using a Qubit dsDNA high-sensitivity assay (Thermo Fisher Scientific, Waltham, MA, USA).

### 3.2.5 16S qRT-PCR for 16S and beta-actin

16S qRT-PCR for 16S and beta-actin genes was done to compare the host to microbial DNA ratio across the different sampling methods.

The primers used to amplify the 16S rRNA gene were: Forward primer 341F 5'-CCTACGGGAGGCAGCAG-3', and reverse primer 534R 5'-ATTACCGCGGCTGCTGGCA-3'. These primers are complementary to the conserved regions in the 16S rRNA gene, nucleotide positions 290 - 484 in Escherichia coli, producing a fragment of 195 bp.

The primers used to amplify the beta-actin *Sus scrofa* gene were: Forward primer 5'-CCAAGCCTGGACTACCTCCT-3', and reverse primer 5'-AAACCTGGAGAGGTTCACCG-3'. These primers were complementary to the *Sus scrofa* actin beta (ACTB) transcript RNA gene, spanning the nucleotide positions 1348 – 1535, producing an amplicon length of 188 bp.

Primers were synthesized by Eurofins Genomics (Ebersberg, Germany). The final PCR mix contained 0.4 µl each of forward and reverse primers (total concentration of 0.4 nmol each), 5 µl of SYBR Green Master Mix (TaKaRa BioTech Corporation, Dalian, China), 3.2 µl of UltraPure™ DNase/RNase-Free distilled water (ThermoFisher Scientific, Waltham, USA) and 1 µl of unamplified genomic DNA, giving a final reaction volume of 10 µl. All samples were performed in triplicate. RT-PCR was performed using the Roche LightCycler®480 Instrument (Roche Diagnostics, Meyland, France) on the following programme: 1 cycle of 30 seconds at 95°C, 40 cycles of 10 seconds at 95 °C, 30 seconds at 59 °C, and 20 seconds at 78 °C.

### 3.2.6 16S gene copy number quantification

The genomic 16S copy number was quantified using Eubacteria 16S Ribosomal gene genesig® Standard kit (Primerdesign™ Ltd, Camberley, UK). The final PCR mix contained 10 µl of PrecisionPLUS 2X qPCR Master Mix, 1 µl of Eubacteria/probe mix and 5 µl of unamplified genomic DNA, and 4 µl of RNase/DNase free water giving a final reaction volume of 15 µl. Samples were divided and performed in triplicates. For the standard curve, the positive control template (16S Eubacteria) was serially diluted 10-fold to obtain copy numbers ranging from 2 x $10^5$ to 2 copy/µL. Quantitative PCR was performed with the Roche LightCycler®480 Instrument (Roche Diagnostics, Meyland, France) on the following programme: 1 cycle of 2 minutes at 95°C, 50 cycles of 10 seconds at 95°C, and 1 minute at 60°C. The $C_T$ number of experimental samples were interpolated against the standard curve to calculate the corresponding 16S copy number.

### 3.2.7 Nanopore 16S barcoding, sequencing, and bioinformatics

The full-length 16S rRNA genes (16S sequencing primers: 27F-AGAGTTTGATCMTGGCTCAG; 1492R-CGGTTACCTTGTTACGACTT) were amplified and samples with amplicons above 1nM were subsequently sequenced on GridION [Oxford Nanopore Technologies (ONT), Oxford, UK] using the R9 flow cell (ONT, FLO-MIN106), as per manufacturer's protocol. Raw sequence reads were basecalled using ONT's MinKNOW software Guppy v.3.2.4 with the R9.4 high accuracy model. Raw reads were demultiplexed using qcat version 1.0.1. Sequence summary and read length histograms were generated using NanoPlot version 1.30.1. 16S rRNA databases were obtained from National Center for Biotechnology Information (NCBI) 16S RefSeq[148] (Nucleotide search details: 33175 [BioProject] or 33317 [BioProject]); the Ribosomal Database Project (RDP)[149] release 11, update 5; and the SILVA rRNA database project[150] version 132 repositories respectively. Reference sequences from the 16S rRNA databases were then clustered into 97%, 99% and 100% similarity thresholds using heuristic clustering method (greedy incremental clustering algorithm) on CD-HIT[201] version 4.8.1 (commands: -c 0.97, 0.99, 1.0, -M 62900, -d 250). I used Minimap2[202] version 2.12-r87 (commands: -K 100M, -ax map-ont) to align the demultiplexed reads to the respective 16S databases and processed the resulting files using the Sequence Alignment/Map (SAM)tools[203] version 1.9 [commands: samtools view -b -F 2308 (to remove unmapped, non-primary and supplementary reads), samtools sort, samtools index, samtools idxstats]. Bioinformatics workflow is summarised in **Figure 3.2**. Bioinformatics scripts that I have written for batch processing is available in Supplementary material 3.1 – 3.6.

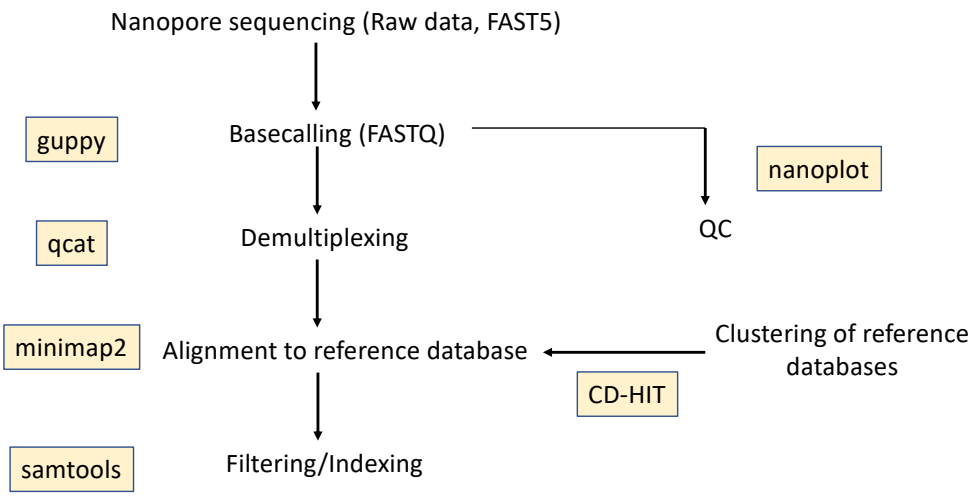



*Figure 3.2 Bioinformatics workflow.*

*Raw data from nanopore sequencing were basecalled with guppy software and demultiplexed using qcat software. Demultiplexed reads were then aligned to reference databases with minimap2 software. SAMtools software was used for read filtering, sorting and indexing. Basic quality control data visualisation was done on nanoplot software. Reference databases were clustered at 97%, 99%, 100% similarity with cd-hit software. Abbreviations: Federal Acquisition STreamlining Act text format data containing quality score, FASTQ; quality control, QC.*

### 3.2.8 Clinical sample collection from patients presenting with microbial keratitis

Patients (n=2) presenting to the Birmingham and Midland Eye Centre, Birmingham, West Midlands, UK with suspected microbial keratitis affecting one eye were recruited to the study. Informed written consent was obtained. The research followed the tenets of the Declaration of Helsinki and was approved by the Health Research Authority Ethics Committee [Rapid Diagnosis of Ocular Infections (RADAR); Reference: 11/EM/0274].

Standard clinical microbiology corneal scrape culture results were compared with nanopore sequencing. For each patient, corneal scrapes were taken for standard clinical microbiology as per routine clinical practice[204]. Corneal scrapes were taken at the base and edge of the lesion, inoculated onto blood, chocolate and Sabaraud's agar plates, and processed at the

local clinical microbiology laboratory [The agar plates were incubated for 5 days – blood and chocolate (6% $CO_2$ at 37$^oC$); Sabourand's (air at 30$^oC$)]. The collection method with the highest DNA recovery in the ex vivo porcine study was used to collect corneal, conjunctival and negative control samples from the patients. Swabs were taken from the affected cornea and unaffected contralateral conjunctiva, together with a negative control 'air swab' of the examination room at the time point of participant sampling, to exclude contamination. Swabs were placed immediately into a ZR BashingBead™ Lysis Tube containing 750ul of DNA Shield™ (Zymo Research, Irvine, CA, USA) and stored at -80$^oC$ until DNA extraction, as described above. Sequence data were compared with the corresponding microbiology culture results of the hospital corneal scrapes. In cases where there was no concordance between culture and nanopore sequencing results, Illumina MiSeq 16S rRNA V4 sequencing was performed.

### 3.2.9 Miseq 16S sequencing and bioinformatics

The V4 variable region of the 16S rRNA gene was amplified from DNA extracts using primer sets (515F: GTGCCAGCMGCCGCGGTAA; and 806R: GGACTACHVGGGTWTCTAAT) and sequenced on the Illumina MiSeq platform[169]. Raw data were filtered and analysed with Mothur and QIIME software packages, as previously described[169,205].

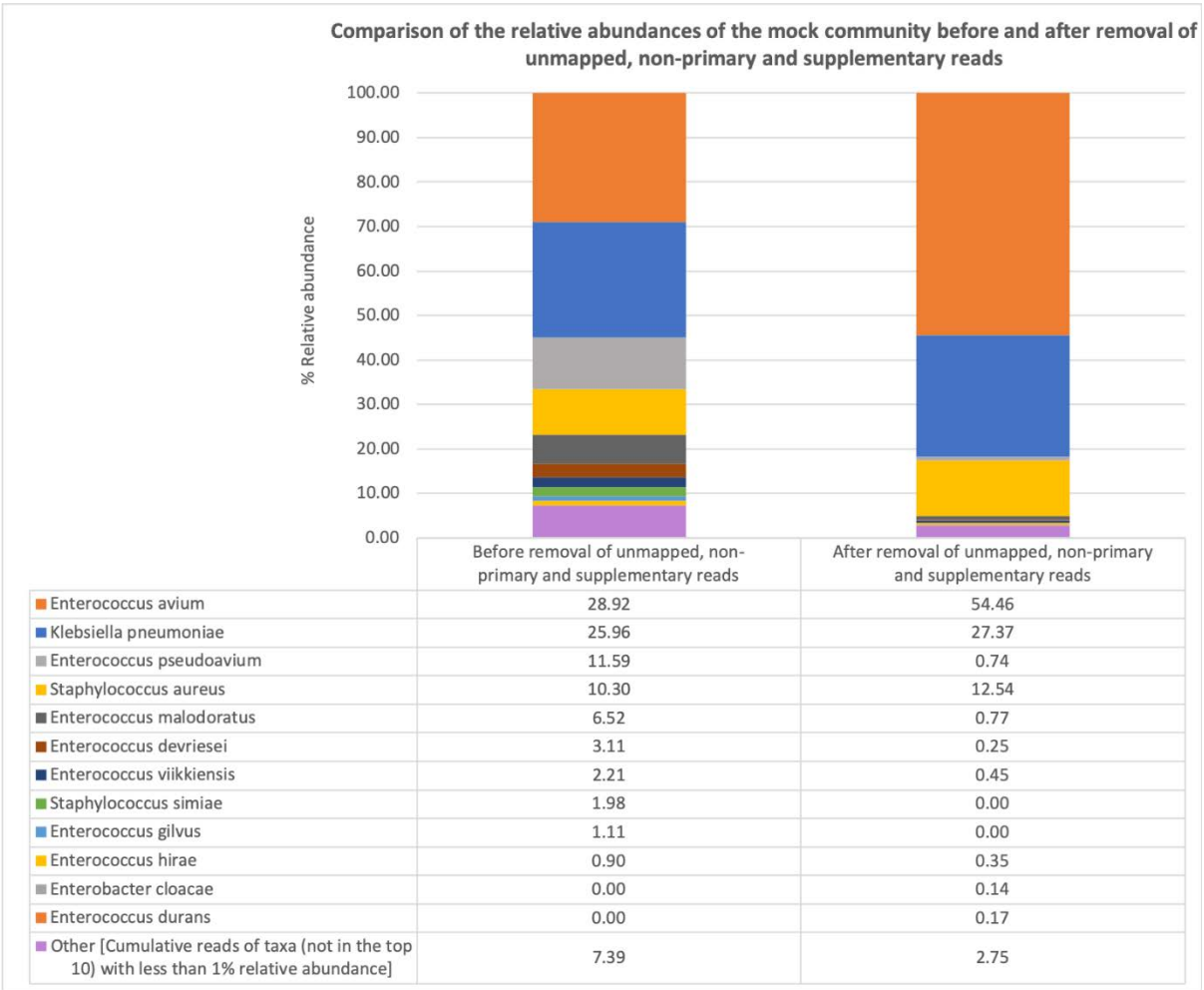
### 3.2.10 Statistical Analysis

Comparisons of DNA concentrations and $C_T$ across the collection materials were performed using Kruskal-Wallis tests. The mean ranks for the DNA concentrations were then used to sort the collection material in order, and data were summarised using medians and ranges. All analyses were performed using IBM SPSS 22 (IBM Corp. Armonk, NY), with $p<0.05$ deemed to be indicative of statistical significance throughout.

## 3.3 Results

3.3.1 Evaluation of bioinformatics analysis pipeline on defined mock microbial community

To evaluate the different methods of performing taxonomic assignment for the most reliable bacterial taxonomic assignment on 16S rRNA nanopore sequencing reads, I compared the effects of removal of unmapped, non-primary and supplementary reads on the taxonomic identification of the mock community **(Table 3.2, Figure 3.3)**. Removal of unmapped, non-primary and supplementary reads provided a more accurate taxonomic assignment of the mock community.




| | Before removal of unmapped, non-primary and supplementary reads | After removal of unmapped, non-primary and supplementary reads |
|---|---|---|
| Enterococcus avium | 28.92 | 54.46 |
| Klebsiella pneumoniae | 25.96 | 27.37 |
| Enterococcus pseudoavium | 11.59 | 0.74 |
| Staphylococcus aureus | 10.30 | 12.54 |
| Enterococcus malodoratus | 6.52 | 0.77 |
| Enterococcus devriesei | 3.11 | 0.25 |
| Enterococcus viikkiensis | 2.21 | 0.45 |
| Staphylococcus simiae | 1.98 | 0.00 |
| Enterococcus gilvus | 1.11 | 0.00 |
| Enterococcus hirae | 0.90 | 0.35 |
| Enterobacter cloacae | 0.00 | 0.14 |
| Enterococcus durans | 0.00 | 0.17 |
| Other [Cumulative reads of taxa (not in the top 10) with less than 1% relative abundance] | 7.39 | 2.75 |

*Figure 3.3 Bar chart comparing the relative abundances of taxa before and after filtering.*

| Without removal of unmapped, non-primary and supplementary reads | | |
|---|---|---|
| Total alignments | 705919 | |
| **Species** | **#Mapped** | **Relative abundance** |
| *Enterococcus avium** | 204171 | 28.92 |
| *Klebsiella pneumoniae** | 183274 | 25.96 |
| *Enterococcus pseudoavium* | 81810 | 11.59 |
| *Staphylococcus aureus** | 72691 | 10.30 |
| *Enterococcus malodoratus* | 46019 | 6.52 |
| *Enterococcus devriesei* | 21978 | 3.11 |
| *Enterococcus viikkiensis* | 15614 | 2.21 |
| *Staphylococcus simiae* | 13980 | 1.98 |
| *Enterococcus gilvus* | 7862 | 1.11 |
| *Enterococcus hirae* | 6380 | 0.90 |
| | | |
| **Genus** | | |
| *Enterococcus** | 409683 | 58.04 |
| *Klebsiella** | 187581 | 26.57 |
| *Staphylococcus** | 90775 | 12.86 |
| *Serratia* | 3078 | 0.44 |
| *Enterobacter* | 3011 | 0.43 |
| *Pectobacterium* | 1238 | 0.18 |
| *Streptococcus* | 763 | 0.11 |
| *Citrobacter* | 653 | 0.09 |
| *Bacillus* | 601 | 0.09 |
| *Lactobacillus* | 579 | 0.08 |

| With removal of unmapped, non-primary and supplementary reads | | |
|---|---|---|
| Total alignments | 117827 | |
| **Species** | **#Mapped** | **Relative abundance** |
| *Enterococcus avium** | 64172 | 54.46 |
| *Klebsiella pneumoniae** | 32248 | 27.37 |
| *Staphylococcus aureus** | 14772 | 12.54 |
| *Enterococcus malodoratus* | 913 | 0.77 |
| *Enterococcus pseudoavium* | 868 | 0.74 |
| *Enterococcus viikkiensis* | 535 | 0.45 |
| *Enterococcus hirae* | 410 | 0.35 |
| *Enterococcus devriesei* | 294 | 0.25 |
| *Enterococcus durans* | 203 | 0.17 |
| *Enterobacter cloacae* | 170 | 0.14 |
| | | |
| **Genus** | | |
| *Enterococcus** | 68665 | 58.28 |
| *Klebsiella** | 32482 | 27.57 |
| *Staphylococcus** | 15184 | 12.89 |
| *Enterobacter* | 214 | 0.18 |
| *Serratia* | 165 | 0.14 |
| *Pectobacterium* | 124 | 0.11 |
| *Streptococcus* | 83 | 0.07 |
| *Citrobacter* | 76 | 0.06 |
| *Bacillus* | 73 | 0.06 |
| *Lactobacillus* | 51 | 0.04 |

*Table 3.2 Comparisons of the effects of filtering on taxonomic assignment.*

*Comparison of the effects of removal of unmapped, non-primary and supplementary reads on the taxonomic identification of the mock community. *denotes mock microbial community in the positive control sample – E.avium, S.aureus and K.pneumoniae.*

Clustering of reference databases by a threshold of sequence similarity may reduce search time and make mapping faster[206]. To determine the effects of clustering, the reference sequences of three different 16S databases were reviewed (the National Center for Biotechnology Information Reference Sequence, NCBI 16S RefSeq; the Ribosomal Database Project, RDP; and the SILVA ribosomal RNA gene database) **(Table 3.3)**. For all three databases, clustering at a threshold of 99% and above provided more accurate operational taxonomic unit assignment (OTU) compared to clustering at 97% similarity. The NCBI 16S RefSeq database with reference sequences clustered at 100% similarity provided the most accurate OTU assignment, with more than 95% of the reads corresponding to the mock community. Based on these results, sequencing reads were aligned to the NCBI 16S RefSeq database at 100% similarity (reference database sequences clustered at 100% similarity) and removed any unmapped, non-primary or secondary alignments.

| NCBI 16S RefSeq | | | | RDP | | | | SILVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of reads | Cluster | #Mapped | Relative abundance | Number of reads | Cluster | #Mapped | Relative abundance | Number of reads | Cluster | #Mapped | Relative abundance |
| 482900 | **97% similarity** | | | 483062 | **97% similarity** | | | 481985 | **97% similarity** | | |
| | *Enterococcus hirae* | 183027 | 37.90 | | *Enterobacter soli* | 173164 | 35.85 | | *Klebsiella quasipneumoniae* | 114637 | 23.78 |
| | *Enterobacter soli* | 172739 | 35.77 | | *Enterococcus faecium* | 123393 | 25.54 | | *Enterococcus faecium* | 109825 | 22.79 |
| | *Staphylococcus aureus** | 99582 | 20.62 | | *Staphylococcus aureus** | 99695 | 20.64 | | *Staphylococcus aureus** | 99153 | 20.57 |
| | *Salmonella enterica* | 7266 | 1.50 | | *Enterococcus asini* | 21726 | 4.50 | | *Klebsiella pneumoniae** | 54508 | 11.31 |
| | *Enterococcus casseliflavus* | 2761 | 0.57 | | *Enterococcus phoeniculicola* | 19122 | 3.96 | | *Enterococcus casseliflavus* | 34290 | 7.11 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 483211 | **99% similarity** | | | 483374 | **99% similarity** | | | 482392 | **99% similarity** | | |
| | *Klebsiella pneumoniae** | 184328 | 38.15 | | *Klebsiella pneumoniae** | 187465 | 38.78 | | *Klebsiella pneumoniae** | 164929 | 34.19 |
| | *Enterococcus avium** | 173319 | 35.87 | | *Enterococcus avium** | 182655 | 37.79 | | *Enterococcus avium** | 109569 | 22.71 |
| | *Staphylococcus aureus** | 97455 | 20.17 | | *Staphylococcus aureus** | 96923 | 20.05 | | *Staphylococcus aureus** | 98477 | 20.41 |
| | *Enterococcus hirae* | 4975 | 1.03 | | *Enterococcus malodoratus* | 1281 | 0.27 | | *Enterococcus devriesei* | 39961 | 8.28 |
| | *Enterococcus malodoratus* | 4401 | 0.91 | | *Enterobacter cloacae* | 1235 | 0.26 | | *Enterococcus raffinosus* | 24276 | 5.03 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 483382 | **100% similarity** | | | 483432 | **100% similarity** | | | 482599 | **100% similarity** | | |
| | *Klebsiella pneumoniae** | 187745 | 38.84 | | *Klebsiella pneumoniae** | 179200 | 37.07 | | *Klebsiella pneumoniae** | 184252 | 38.18 |
| | *Enterococcus avium** | 175497 | 36.31 | | *Enterococcus avium** | 176822 | 36.58 | | *Enterococcus avium** | 112163 | 23.24 |
| | *Staphylococcus aureus** | 97505 | 20.17 | | *Staphylococcus argenteus* | 90126 | 18.64 | | *Staphylococcus aureus** | 98461 | 20.40 |
| | *Enterococcus pseudoavium* | 3758 | 0.78 | | *Staphylococcus aureus** | 8510 | 1.76 | | *human gut* | 55245 | 11.45 |
| | *Enterococcus malodoratus* | 3279 | 0.68 | | *Klebsiella quasipneumoniae* | 7614 | 1.57 | | *Enterococcus gilvus* | 4491 | 0.93 |

*Table 3.3 Comparing the effects of using NCBI 16S, RDP and SILVA databases at different levels of clustering (97%, 99% and 100% similarity).*

**denotes mock microbial community in the positive control sample* – E.avium, S.aureus *and* K.pneumoniae.

### 3.3.2 Optimising sampling technique for collection of microbial keratitis on ex-vivo porcine model

To optimise the sampling technique on ex-vivo porcine model of microbial keratitis, comparisons were made between the different collection materials, both pre- and post-inoculation. Pre-inoculation, the average $C_T$ of 16S was not found to differ significantly between the collection materials (p=0.909) **(Table 3.4)**.

| | N | Average $C_T$ (cycles) | |
|---|---|---|---|
| | | ***16S*** | ***β-Actin*** |
| Sugi® Eyespear | 4 | 19.93 (17.71 - 22.81) | 23.26 (21.71 - 26.18) |
| Isohelix™ SK-2S | 6 | 18.26 (16.82 - 24.03) | 24.04 (22.53 - 29.76) |
| MW1041 Cotton | 6 | 20.61 (19.27 - 24.20) | 27.19 (23.51 - 28.45) |
| MW1021 Dryswab™ Rayon | 6 | 20.21 (16.05 - 24.42) | 27.95 (24.62 - 30.73) |
| MW100 Fine tip Dryswab™ Rayon | 6 | 20.76 (18.06 - 23.79) | 27.45 (21.51 - 32.02) |
| MW840 Hydraflock® Plastic | 4 | 20.68 (17.69 - 23.98) | 27.77 (26.76 - 29.76) |
| MW130 Hospiswab™ Albumin | 4 | 21.32 (17.88 - 23.99) | 28.43 (26.31 - 30.77) |
| MW1021D Dryswab™ Polyester | 6 | 20.16 (13.90 - 23.71) | 29.36 (24.74 - 31.65) |
| MW946 Sigma Swab® Purfoam | 4 | 21.82 (15.65 - 24.06) | 29.85 (29.00 - 30.49) |
| MW821 Dryswab™ Flock | 4 | 21.07 (17.68 - 24.16) | 34.65 (33.39 - 35.00) |
| BD Needle 21G | 4 | 23.13 (21.19 - 24.07) | 32.85 (28.92 - 35.00) |
| Biopore® PTFE (6mm) | 2 | 23.12 (23.00 - 23.23) | 31.37 (27.93 - 34.80) |
| Biopore® PTFE (4mm) | 4 | 22.50 (18.64 - 23.95) | 32.32 (26.61 - 34.30) |
| **p-Value** | | 0.909 | **0.002** |

*Table 3.4 Pre-inoculation average $C_T$ by collection material.*

*Data are reported as medians and ranges, with p-values from Kruskal-Wallis tests; bold p-values are significant at p<0.05.*

Post-inoculation, the DNA concentration was found to differ significantly across the collection materials (p<0.001) **(Table 3.5 & Figure 3.4)**. The highest concentrations were observed for the Sugi® Eyespear and Isohelix™ SK-2S, with medians of 82.2 and 96.6 ng/ul respectively, whilst the lowest concentrations were observed for the PTFE 4mm, at a median of 0.19 ng/ul. For the average $C_T$, no significant difference across the collection methods was detected for 16S (p=0.242). The 16S: β-Actin ratio was consistent across all of the collection materials (p=1.000).

| | N | Order* | DNA Concentration (ng/µl) | Average $C_T$ (cycles) *16S* | *β-Actin* | *16S: β-Actin Ratio* |
|---|---|---|---|---|---|---|
| Sugi® Eyespear | 4 | 1 | 82.20 (32.60 - 261.60) | 16.58 (15.95 - 17.94) | 23.07 (20.37 - 29.01) | 0.72 (0.62 - 0.78) |
| Isohelix™ SK-2S | 6 | 2 | 96.60 (22.80 - 308.00) | 14.96 (14.46 - 22.04) | 21.55 (20.18 - 24.01) | 0.73 (0.65 - 0.92) |
| MW1041 Cotton | 6 | 3 | 30.00 (9.82 - 106.00) | 16.14 (10.44 - 22.45) | 23.40 (22.44 - 26.49) | 0.71 (0.47 - 0.93) |
| MW1021 Dryswab™ Rayon | 6 | 4 | 22.40 (4.16 - 172.00) | 15.56 (13.02 - 22.16) | 23.67 (22.67 - 29.73) | 0.68 (0.45 - 0.91) |
| MW100 Fine tip Dryswab™ Rayon | 6 | 5 | 17.45 (7.30 - 31.20) | 17.71 (11.37 - 20.76) | 23.89 (23.04 - 29.82) | 0.77 (0.38 - 0.87) |
| MW840 Hydraflock® Plastic | 4 | 6 | 15.85 (5.00 - 21.60) | 18.83 (14.69 - 22.39) | 24.54 (23.88 - 26.02) | 0.75 (0.62 - 0.90) |
| MW130 Hospiswab™ Albumin | 4 | 7 | 11.25 (10.30 - 25.40) | 18.38 (11.93 - 22.15) | 24.79 (23.82 - 28.92) | 0.75 (0.41 - 0.91) |
| MW1021D Dryswab™ Polyester | 6 | 8 | 2.37 (1.32 - 142.40) | 16.48 (15.21 - 24.23) | 27.83 (21.44 - 31.12) | 0.76 (0.49 - 0.86) |
| MW946 Sigma Swab® Purfoam | 4 | 8 | 8.42 (2.14 - 17.50) | 18.34 (14.71 - 22.45) | 26.65 (23.60 - 31.05) | 0.72 (0.47 - 0.90) |
| MW821 Dryswab™ Flock | 4 | 10 | 9.33 (4.32 - 11.30) | 17.58 (12.18 - 23.14) | 27.21 (24.76 - 29.88) | 0.66 (0.41 - 0.92) |
| BD Needle 21G | 4 | 11 | 0.92 (0.12 - 7.76) | 21.46 (18.63 - 23.02) | 28.53 (21.59 - 35.00) | 0.73 (0.57 - 1.07) |
| Biopore® PTFE (6mm) | 2 | 12 | 0.95 (0.91 - 0.98) | 22.26 (21.79 - 22.72) | 28.34 (28.14 - 28.54) | 0.79 (0.76 - 0.81) |
| Biopore® PTFE (4mm) | 4 | 13 | 0.19 (0.11 - 3.18) | 23.75 (20.57 - 23.99) | 32.82 (29.93 - 35.00) | 0.73 (0.59 - 0.79) |
| p-Value | - | - | **<0.001** | 0.242 | **0.017** | 1.000 |

*Table 3.5 Post-inoculation average $C_T$ and DNA concentrations by collection material.*

*Data are reported as medians and ranges, with p-values from Kruskal-Wallis tests; bold p -values are significant at p<0.05. *The ordering of collection materials, sorted from largest to smallest on the mean rank of the DNA concentration.*

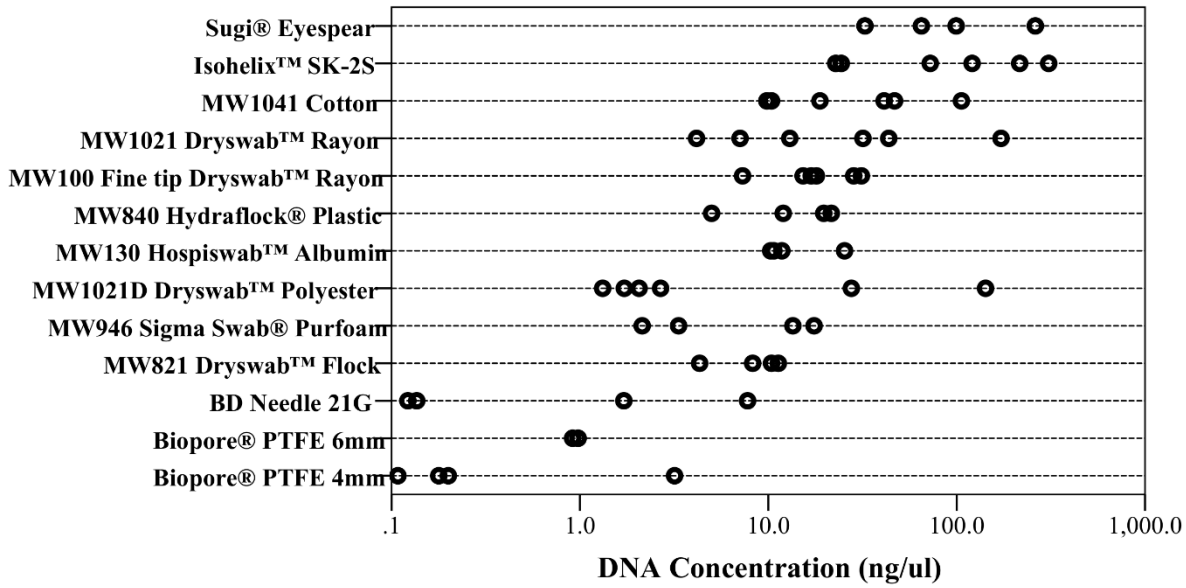



*Figure 3.4 Post-inoculation DNA concentration by collection material.*

*Collection materials are sorted in order of the mean rank of DNA concentration.*

Additional analyses were performed in order to quantify within-material variability, and compare this between collection devices. DNA concentration was found to follow a skewed distribution, which is why previous analyses had used a non-parametric approach, and used medians as summary statistics. Further assessment of the distribution found this to be approximately log-normal, hence values were log10-transformed to normalise the distribution, and allow parametric analysis. The variances of the log10-transformed data were found to differ significantly between collection devices (Levene's test: $p<0.001$), implying that the within-material variability differed significantly between collection devices.

In order to visualise this effect, the width of the 95% prediction interval (95% PI) was calculated for each material by anti-logging the standard deviation (SD) of the log10 DNA concentration, after multiplying by 1.96. As such, it would be expected that 95% of DNA concentrations collected from the current model by the stated material would be within the range [geometric mean/95% PI width] and [geometric mean*95% PI width]. For example, the

Sugi® Eyespear had a geometric mean DNA concentration of 86 and a 95% PI width of 5.5, implying that 95% of measurements would be expected to be between 16 (86/5.5) and 473 (86*5.5). The values for the other materials are reported in the table and figure below.

The lowest variability was observed for the Biopore® PTFE 6mm, with a 95% PI width of 1.1. However, this was only based on N=2 measurements, hence is likely to be unreliable. Of the other materials considered, the variability was lowest in the MW130 Hospiswab™ Albumin and highest in the BD Needle 21G, with 95% PI widths of 2.3 and 53.1, respectively.

| | N | $Log_{10}$ DNA Concentration | 95% PI Width |
|---|---|---|---|
| Biopore® PTFE 6mm | 2 | -0.02 ± 0.02 | 1.1 |
| MW130 Hospiswab™ Albumin | 4 | 1.13 ± 0.19 | 2.3 |
| MW821 Dryswab™ Flock | 4 | 0.91 ± 0.19 | 2.3 |
| MW100 Fine tip Dryswab™ Rayon | 6 | 1.25 ± 0.23 | 2.8 |
| MW840 Hydraflock® Plastic | 4 | 1.10 ± 0.29 | 3.7 |
| Sugi® Eyespear | 4 | 1.94 ± 0.38 | 5.5 |
| MW1041 Cotton | 6 | 1.43 ± 0.41 | 6.3 |
| MW946 Sigma Swab® Purfoam | 4 | 0.81 ± 0.45 | 7.6 |
| Isohelix™ SK-2S | 6 | 1.92 ± 0.47 | 8.5 |
| MW1021 Dryswab™ Rayon | 6 | 1.33 ± 0.59 | 14.2 |
| Biopore® PTFE 4mm | 4 | -0.48 ± 0.66 | 20.0 |
| MW1021D Dryswab™ Polyester | 6 | 0.78 ± 0.82 | 41.2 |
| BD Needle 21G | 4 | -0.16 ± 0.88 | 53.1 |

*Table 3.6 Variability in post-inoculation DNA concentrations by collection method*

*The $log_{10}$ DNA concentration is reported as mean ± SD. The 95% prediction interval (PI) width was then derived by multiplying the SD by 1.96, and anti-logging the result. As such, the resulting value represents the fold difference above or below the geometric mean that would include 95% of DNA concentrations measured by the stated material.*

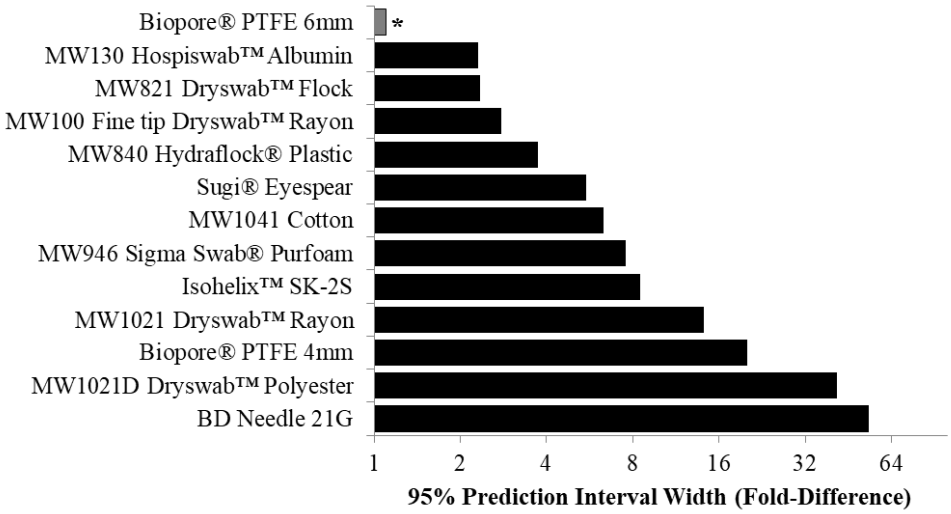



*Figure 3.5 Variability in post-inoculation DNA concentrations by collection material*

*The 95% prediction interval width is a quantification of variability, as previously described, and is plotted on scale with logarithmic spacing. *Variability for the Biopore® PTFE 6mm was estimated from only N=2 cases, hence may not be reliable.*

Six out of the twelve collection methods provided sufficient DNA yield for 16S Nanopore sequencing and 16S copy number quantification – Sugi® Eyespear, Isohelix™ SK-2S, MW1021D Dryswab™ Polyester, MW130 Hospiswab™ Albumin, MW1041 Cotton and MW1021 Dryswab™ Rayon **(Figure 3.5)**. All three species of the mock microbial community inoculated onto the ex-vivo porcine eyes were detectable on nanopore sequencing. The Sugi® Eyespear swab was chosen for patient sample collection as it provided the highest mean rank DNA concentration that was sufficient for nanopore sequencing.

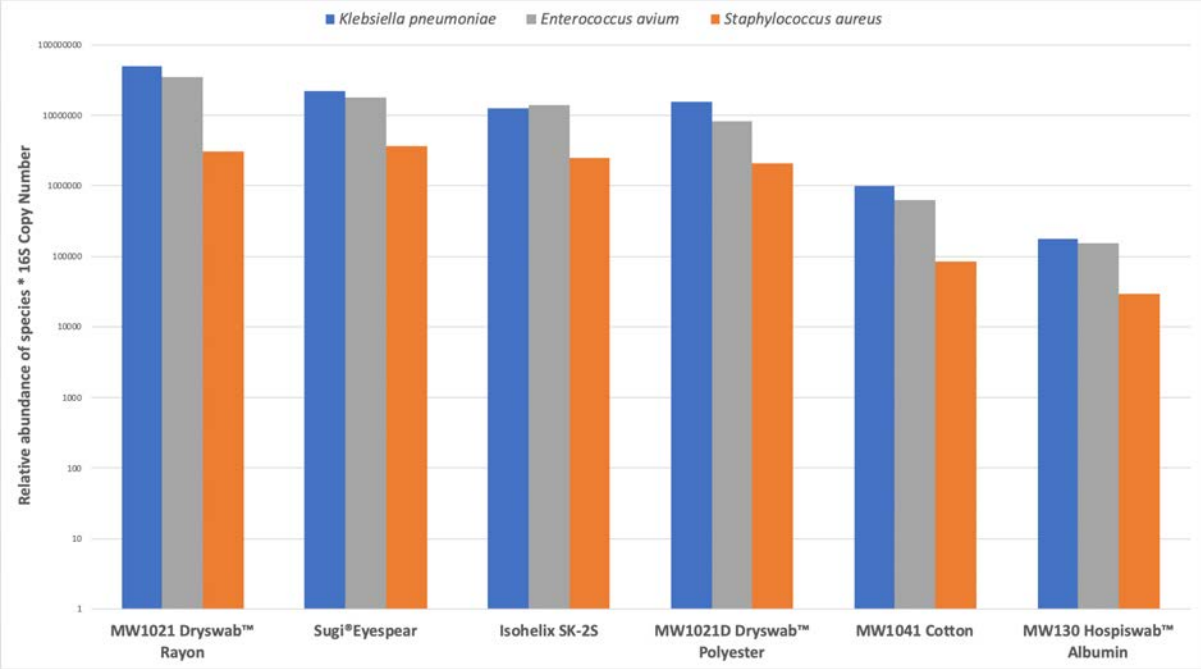



*Figure 3.6 Comparison of swab collection yield, defined by relative abundances of species normalised by 16S copy number.*

### 3.3.3 Proof-of-concept clinical study on microbial keratitis patient samples

The optimised sampling technique and bioinformatics pipeline were applied directly on two test patients with suspected microbial keratitis. Corneal, conjunctival and negative control swabs were collected from two consecutive patients presenting to the emergency department **(Table 3.6 & Supplementary Table 3.2).** Patient A had a majority of reads (497995 out of total 613482 reads, 81.2%) mapping to *Serratia marcescens,* which was in agreement with culture results. Patient B, who had been started on topical antibiotics prior to sampling, had no growth on culture after 5 days, but the majority of reads mapped to *Bacillus subtilis* (33188 out of a total of 92810 reads, 35.8%) on nanopore sequencing. Further Illumina Miseq 16S rRNA sequencing confirmed presence of *Bacillus* in the affected eye of Patient B. Control samples from the unaffected contralateral conjunctival swabs and the respective negative control 'air swabs' had fewer reads (<1%) compared to the number of reads generated from the affected eye.

| Sample | Total mapped reads of sample (Proportion of sample reads/total reads in sequencing run) | Bacterial taxa identified on nanopore sequencing (number of reads, relative abundance) | Culture results | Most abundant bacterial taxa identified on 16S V4 MiSeq (relative abundance) |
|---|---|---|---|---|
| **Patient A – Total reads of sequencing run: 1097318** | | | | |
| Affected eye (Right cornea) | 613,482 (55.9%) | *Serratia marcescens* (497995, 81.2%); *Serratia nematodiphila* (60000, 9.78%); *Klebsiella aerogenes* (9247, 1.51%); *Kluyvera ascorbate* (1981, 0.32%); *Cutibacterium acnes* (1942, 0.32%) | *Serratia marcescens* | - |
| Unaffected eye (Left conjunctiva) | 28 (0.0026%) | *Serratia marcescens* (6, 21.4%); *Streptococcus cristatus* (3, 10.7%); *Hungateiclostridium cellulolyticum* (2, 7.14%); *Streptococcus parauberis* (2, 7.14%); *Enterococcus avium* (2, 7.14%) | - | - |
| Negative control | 26 (0.0024%) | *Klebsiella pneumoniae* (6, 23.1%); *Enterococcus avium* (4, 15.4%); *Serratia marcescens* (4, 15.4%); *Staphylococcus aureus* (2, 7.69%); *Kluyvera ascorbate* (1, 3.85%) | - | - |
| **Patient B – Total reads of sequencing run: 273987** | | | | |
| Affected eye (Left cornea) | 92,810 (33.9%) | *Bacillus subtilis* (33188, 35.8%); *Staphylococcus caprae* (5875, 6.3%); *Staphylococcus saccharolyticus* (5206, 5.61%); *Aggregatibacter segnis* (4109, 4.43%); *Cutibacterium acnes* (2998, 3.23%) | No growth in culture after 5 days | *Bacillus* (13.9%); *Dialister* (8.6%); *Actinobacter* (6.9%); *Rubrobacter* (5.5%); *Staphylococcus* (4.3%) |
| Unaffected eye (Right conjunctiva) | 1,301 (0.45%) | *Snodgrassella alvi* (935, 71.9%); *Escherichia fergusonii* (48, 3.69%); *Anoxybacillus flavithermus* (35, 2.69%); *Eikenella corrodens* (24, 1.84%); *Thermicanus aegyptius* (20, 1.54%) | - | - |
| Negative control | 6 (0.0022%) | *Klebsiella pneumoniae* (3, 50%); *Staphylococcus aureus* (2, 33.3%); *Enterococcus avium* (1, 16.7%) | - | - |

*Table 3.7 Results of culture and 16S rRNA gene sequencing of patient samples aligned to NCBI 16S RefSeq Database at 100% similarity.*

## 3.4 Discussion

I have developed a proof-of-concept study optimising the sample collection method and downstream bioinformatics pipeline for full-length 16S rRNA gene identification by nanopore sequencing in the setting of microbial keratitis.

The use of different swabs and collection methods for microbial keratitis has a direct effect on the DNA yield – the Sugi® Eyespear and Isohelix™ SK-2S swabs provided the highest DNA concentration. However, the ratio of host to microbial DNA recovery is similar across all collection methods. Differences in the DNA yield could be explained by the absorption efficacy of the swabs[207]. The absorption capacity of the swab materials is dependent upon the swab tip dimensions and morphology of the sorbent material - how tightly wound the sorbent material is to the shaft[207]. The Sugi® Eyespear swab, primarily designed for use in ophthalmic theatres due to its high tensile strength in absorbing fluids, has been proven to be effective in recovering DNA from corneal tissue in our study. Previous studies have also demonstrated that DNA recovery is inversely proportional to the fiber density on the swabs[208,209]. Both Sugi® Eyespear and Isohelix™ SK-2S have swab tips made of cellulose fiber, which have high DNA-binding capacity[210]. Other limitations of the study include potential variations in the porcine eye inoculated model between sampling devices, and performance of a device may vary with different concentrations and types of bacteria. Currently, there is "no model system that is a perfect surrogate for the natural human infection" and this includes *in vitro, ex vivo* or *in vivo* animal systems[211]. Differences in the eyes (i.e. paired eyes from the same animal, contamination from the abattoir or market place), location, size and depth of the corneal wound along with stability of the inoculum in the wound, taking into account potential spillage, would be important considerations for standardising an eye model of corneal

infections[212]. Study done by Pinnock and associates found that injecting inoculum into the stroma resulted in more bacterial recovery compared to corneal wounding in both human and rabbit corneas[211]. Additionally, there were small variations in colony forming units following 24 hours of infection compared to 48 hours[211]. It would be difficult to control for the rate of infection in porcine eyes that have been obtained from the abattoir/food market due to high background contamination in those environments. Therefore, use of a synthetic sterile model with 3D topology of the human cornea would ideal to control for background contamination. The use of varying concentrations and types of microorganisms may also impact the use of corneal models for studying bacterial keratitis infections, such as co-infections with different combinations of bacterial species and fungi/viruses, along with the differences in the development of biofilm morphology and adherence to the corneal surface[212]. Further standardisation of a corneal infection eye model could be undertaken with considerations of the location, size and depth of wound, sampling pressure and time, along with different concentrations and type of microorganisms.

Factors such as adsorption capacity, DNA extraction and recovery efficiency of the type of swabs could be considered in future studies on optimising the sampling and collection devices, whereby known quantities of corneal epithelial cells (± intact bacterial cells) or pure DNA at varying concentrations are directly placed onto the collection devices for comparisons between the collection devices. Bruijns and associates have found that the adsorption capacity of a collection device is dependent upon the tip dimensions and morphology – swabs with larger tip dimensions and woven, more tightly wound fibers (as examined on electron microscopy) have better adsorption capacity than swabs with small tips or open, porous, sponge-like morphology[207]. The DNA extraction efficacy, defined as effectiveness of DNA

transfer from collection device to the DNA extraction solution, could be influenced by the type of swab material, sample type (i.e. epithelial cells vs. intact microorganisms vs. bacterial spores), swabbing method and angle, or even swabbing medium and volume[207]. Rose and associates reported high extraction efficacy in rayon swabs (90%) and polyester (80%) on direct inoculation with *Bacillus anthracis* spores, however Bruijns and associates reported less than 50% extraction efficacy on direct inoculation with pure DNA[207,213]. Similarly, Adamowicz and associates found that over 50% of the DNA from buccal cells were retained on the cotton swab, however, was able to increase the recovered DNA yield by two-folds with alterations in the incubation time and by re-suspending the swab tip in the extraction buffer[214].

Choice of primers affects the amplification efficiency of the 16S rRNA gene region, with primers that amplify the entire 16S rRNA gene spanning the V1-9 variable regions (27F and 1492R) providing better classification of reads compared to primers that only amplify a portion of the 16S rRNA region, as shown in studies by Winand et al and Nygaard et al[144,145]. Therefore, we have used the 27F and 1492R primer pairs provided in the commercially available ONT Rapid 16S barcoding kit. Other preliminary studies have also shown that sequencing the whole *rrn* operon (~4300bp), which includes the 16S rRNA gene, internal transcribed spacer (ITS) region, and the 23S rRNA gene, may provide better taxonomic resolution[154]. However, as there is a lack of updated and curated *rrn* operon databases, users will need to retrieve and compile whole ribosomal operon reference database for their specific usage[215], whereas, curated or updated 16S rRNA reference databases are more readily available[148,150].

Another factor influencing the quality of nanopore sequencing is the choice of basecalling software[134]. The Guppy 'flip-flop' high accuracy model was used for basecalling, based on Wick et al's study showing that the model, which utilises the recurrent neural network algorithm, performed better than the other basecalling programs (Albacore, Scrappie and Flappie)[216]. Using Minimap2[202], nanopore sequencing reads were aligned against the three different publicly available databases, NCBI 16S RefSeq, RDP and SILVA. Performance matrix comparison of thirteen different classification tools by Urban and colleagues revealed that Minimap2 provided robust alignments that were closely aligned to their mock community taxa[217]. However, similar to the challenges encountered by Urban et al[217], I have had issues of high memory usage on Minimap2, which necessitated a reduction in the number bases loaded into memory to process in the query batch (command -K 100M). By comparing against a defined mock community, differences in the taxonomic assignments were observed between the databases - with the NCBI 16S RefSeq database clustered at 100% providing the most accurate assignments, which could be attributed to the differences in the database size and sequence validation steps[218,146]. The NCBI 16S RefSeq database is manually curated and near full length 16S sequences are preferentially selected[148]. In the benchmarking steps, the CD-HIT program[219] was used, which employs heuristic greedy incremental clustering algorithm to cluster the reference sequences of the databases into 97%, 99% and 100% similarity to approximate taxonomic assignments. Schmidt et al assessed different clustering programs (hierarchical and heuristic algorithms) and observed that the CD-HIT program was robust, computationally efficient and provided reproducible clusterings[220]. In this study, the 100% identity threshold provided a more optimal OTU assignment compared to 97% or 99% identity thresholds, which is consistent with previously published studies[221,222].

Nanopore sequencing reads were in concordance with current gold standard clinical microbiology culture techniques in Patient A. In the case of Patient B where cultures were negative, in the setting of previous antibiotic use, the nanopore sequencing result was in agreement with Illumina short-read sequencing suggesting an identification of a putative organism in the context of a false negative culture. The human ocular surface is paucibacterial. The conjunctival microbiome predominantly consists of *Corynebacteria*, *Propionibacteria*, *Staphylococcus*, and *Streptococcus* – with "approximately 1 bacterium for every 20 human conjunctival epithelial cell collected on conjunctival swab"[36]. It is likely that the presence of *Serratia marcescens* and *Bacillus subtilis* at such high relative abundance in the patient samples, with correlating clinical signs and symptoms of infective keratitis, would constitute a positive test. A major challenge in applying high-throughput sequencing in clinical practice is distinguishing between true polymicrobial keratitis, ocular commensal or contaminant[82]. Hence, meticulous steps have been taken to account for any potential contamination in our study - swabs of the unaffected contralateral eye and negative control swabs at the same time point and clinical environment of the patient had been taken, processed and sequenced in the same manner as the swabs of the affected eye. These negative control swabs had significantly fewer reads, less than 1% of the reads from the affected eye.

Another challenge in using full-length 16S rRNA sequencing is the difficulty in differentiating species and subspecies strains in certain bacterial genera with high sequence homology and similarity, notably the *Bacillus subtilis* species complex[223]. This is illustrated in Patient B, whereby different taxonomic species within the *Bacillus subtilis* group such as *Bacillus amyloliquefacians*, *Bacillus licheniformis*, *Bacillus velenzensis*, and *Bacillus halotolerans* have

been assigned. Classical phenotypic tests by colonial appearance on culture, presence of ß-haemolysis, or biochemical tests to discriminate between these subspecies are unreliable. Other laboratory identification methods such as Matrix-Assisted Laser Desorption Ionisation – Time of Flight (MALDI-TOF) mass spectrometry to detect microbial protein composition is highly variable dependent upon endospore formation[224], whilst Multilocus Sequence Typing (MLST), which relies on PCR amplification and sequencing of six or seven well-conserved, housekeeping genes within the bacterial genome, is unable to provide distinct phylogenetic typing of *Bacillus* owing to the difficulty in designing primers for genetic sequences with high similarity. Although Pulse Field Gel Electrophoresis (PFGE), which utilises endonuclease restriction enzymes, subsequent separation of DNA fragments by gel electrophoresis and staining under ultraviolet light for bands, is highly discriminatory, it is time consuming (over 30 hours) and requires specialist laboratory equipment[223]. Hence, Public Health England advocates initial clustering and genus identification by 16S rRNA for *Bacillus,* which could then be followed by more in-depth whole genome sequencing for more accurate strain characterisation[223].

Nanopore sequencing technology has the potential to provide rapid, real-time diagnosis of causative pathogens in a healthcare setting, with relatively low cost. Although in this proof of concept study, 16S primers were utilised, which specifically amplify bacterial DNA, non-biased deep metagenomic detection of pathogens and its antimicrobial resistance genomes from cultured clinical isolates using nanopore sequencing have previously been demonstrated[132,147]. Direct RNA sequencing on the nanopore platform would also enable identification of 'live' pathogens and host gene profiling for transcriptome signatures related to the infection[117]. However, challenges still remain in terms of limiting background

contamination[225], reducing the error rates of sequencing and improving base-calling algorithms[216]. Performing molecular diagnostics on ocular samples is inherently more difficult compared to other types of clinical samples, such as blood, urine or cerebrospinal fluid, as ocular samples are many magnitudes smaller in volume[106] and are highly abundant in human cells[117].

This is the first study to evaluate the use of full-length 16S nanopore sequencing for detection of pathogens in microbial keratitis. I have optimised collection methods and demonstrated the bioinformatics pipeline for bacterial microbial keratitis. The study is limited by the small sample size of patient cohort and the use of 16S rRNA primers, which specifically amplifies bacterial genome. To resolve this, larger clinical sample studies involving unbiased metagenomic sequencing are required to determine the compare the clinical utility of 16S versus whole genome nanopore sequencing in ocular infections.

# Chapter 4

## Evaluation of full length 16S and whole genome nanopore sequencing for detection of pathogens in endophthalmitis

# Chapter 4 Evaluation of full length 16S and whole genome nanopore sequencing for detection of pathogens in endophthalmitis

## 4.1 Introduction

Endophthalmitis is a severe, sight-threatening intraocular infection and is one of the most feared complications for any ocular surgical procedures[90]. Conventional microbiology techniques in endophthalmitis rely heavily on culture-based assays with low sensitivities, ranging between 40-70%[91,92,107,226] as most causative pathogens are not easily cultured *in-vitro*, whilst PCR-based molecular tests require prior hypothesis and target only a few known, common pathogens[114,227]. Deep metagenomic, Illumina whole genome sequencing (WGS) techniques have allowed for discovery of novel pathogens and may even provide vital prognostic information for clinical outcomes[106,107,118]. However, samples have to be pooled for batch processing at select sequencing centres, due to the prohibitively high costs of Illumina sequencing machines, resulting in delayed turnaround time[157]. One of the latest break-throughs in the sequencing field is the introduction of a portable, pocket-sized, relatively low-cost nanopore sequencer (Oxford Nanopore Technologies' MinION sequencer)[157]. This technology measures the translocation of ionic currents as nucleotides pass through nanoscopic pores and generates real-time sequencing data for analysis[228]. In comparison to Illumina short-read (~500 base pairs) sequencing platforms, nanopore sequencing is capable of generating ultra-long reads (~1500 – 882k bases), allowing for better coverage of genomic elements, is portable, and provides real-time sequencing data.[229,230]

The sensitivity of WGS in detecting pathogens decreases with high background noise, either originating from host genome or from laboratory/environmental contamination[130,157,158]. In intraocular fluid biopsy samples of endophthalmitis, Kirstahler and associates reported over 99% of sequence reads mapping to the human genome, with similar proportions described by Thuy and associates[106,119]. Host depletion method using differential lysis of human cells with saponins and subsequent DNA degradation with DNase have been developed and optimised for clinical samples of larger volumes, such as the cerebrospinal fluid and nasopharyngeal aspirate[130,158]. Several techniques to selectively enrich for microbial reads have been developed, such as the use of antibodies targeting the methyl-CpG motifs that are present in human/vertebrate DNA but lacking in most microbial DNA and differential lysis of host cells using chaotropic buffers followed by degradation of released background DNA with DNase[158–160]. Charalampous and associates have optimised a saponin-based host DNA depletion protocol for nanopore sequencing on 41 clinical respiratory samples of patients with suspected lower respiratory tract infections and reported up to 10,000 fold depletion or 99.9% of human DNA[130]. Firstly, differential centrifugation was used to pellet the host and microbial cells. Then, saponin was used to permeabilise host cell membrane by solubilising cholesterol with subsequent addition of high concentrations of NaCl to deliver an osmotic shock to lyse the damaged host cells[130,161]. Finally, DNase was added to deplete the released host DNA[130,158]. The respiratory samples in study conducted by Charalampous and associates were derived from sputum, endotracheal secretions and bronchoalveolar lavage, and therefore, had larger starting volumes of 400µl[130]. However, these host depletion techniques have not been tested on ocular clinical samples that are of much smaller in volume.

The aim of this study was to evaluate the utility of nanopore sequencing for identification of potential causative pathogen in endophthalmitis. First, the effects of saponification on host and microbial DNA was tested on murine peripheral blood mononuclear cells spiked with serial dilutions of *E.coli* cells and quantified by qPCR. Then, the saponification protocol was applied to a small subset (n=5) of intraocular fluid biopsy samples from patients with clinically suspected endophthalmitis and the proportion of human and non-human reads were compared by whole genome nanopore sequencing. Lastly, intraocular fluid biopsy clinical samples (n=23) were processed for molecular sequencing and the culture results were compared against full-length 16S rRNA nanopore sequencing (16S nanopore) and whole genome nanopore sequencing (WGS nanopore) and Illumina whole genome sequencing (Illumina WGS).

## 4.2 Methods

The study was approved by the United Kingdom Health Research Authority Ethics Committee [Rapid Diagnosis of Ocular Infections (RADAR); Reference: 11/EM/0274], the Institutional Review Board from the University of Washington, Seattle, Washington, United States of America and conducted in accordance with the Declaration of Helsinki. All animal procedures were performed in accordance with the Association for Research in Vision and Ophthalmology (ARVO) Statement for the Use of Animals in Ophthalmic and Vision Research and approved by the Institutional Animal Care and Use Committee at the University of Washington, Seattle, Washington, United States of America.

### 4.2.1 Isolation of murine peripheral blood mononuclear cells

Peripheral blood mononuclear cells (PBMCs) were isolated using density centrifugation (Ficoll-Paque™). Fresh whole blood were collected from the tail vein of mice and placed into BD Vacutainer™ Heparin Plasma tube (Thermo Fisher Scientific, Waltham, MA, USA). In a 15mL conical tube, 1mL of blood was diluted with 3mL of Dulbecco's Phosphate Buffered Saline (PBS) (Gibco™, ThermoFisher Scientific, Waltham, Massachusetts, USA). The diluted blood suspension was then carefully layered over 3mL of Ficoll-Paque™ (Cytiva, ThermoFisher Scientific, Waltham, Massachusetts, USA) in a 15mL conical tube and centrifuged at 400 x *g* for 25 minutes at room temperature. The buffy coat 'interphase' mononuclear cell layer, containing PBMCs, was carefully transferred to a new 15mL conical tube filled with 4mL of PBS, mixed and centrifuged at 300 x *g* for 15 minutes at 19$^{o}$C. The supernatant was carefully removed, leaving a pellet containing the PBMCs. The pellet was then resuspended in PBS and the number of cells were quantified using a glass haemocytometer and microscope.

### 4.2.2 Serial dilutions of cultured *Escherichia coli*

The NEB® 5-alpha Competent *Escherichia coli* (New England Biolabs, Ipswich, MA, USA) was grown overnight at 37$^{o}$C in lysogeny broth (LB). Overnight culture was diluted into fresh medium to an optical density of 600nm (OD600) of 0.05 and incubated at 37$^{o}$C with aeration. The culture OD600 was measured every 30 minutes for 5 hours using spectrophotometer (Nanodrop™ 1000 Spectrophotometer, ThermoFisher Scientific, Waltham, Massachusetts, USA). The bacterial growth curve was plotted and the mid-exponential growth phases of the sample was taken. For enumeration of bacteria, the *E.coli* culture was plated onto LB agar plates and incubated overnight at 37$^{o}$C. The colony-forming units (CFUs) was enumerated the following day. Serial dilutions of cultured *E.coli* (10 – $10^5$ CFU/mL) was mixed with 2.4 x $10^5$ of

murine PBMC samples (this concentration was chosen because liquid-based cytology on vitreous samples from patients with endophthalmitis were reported to be greater than 10,000 cells per 200 microliters[231]) and used to determine the effects of saponification on the limit of detection by qPCR for bacteria and host DNA.

### 4.2.3 Host DNA depletion by saponification

The protocol for saponin-based host DNA depletion has been previously published[130]. The volume of the sample (mixture of *E.coli* with PBMC) was made up to 200µl with PBS (Gibco™, ThermoFisher Scientific, Waltham, Massachusetts, USA) and centrifuged at 8,000 x *g* for 5 minutes. The supernatant was carefully removed, leaving approximately 50µl of supernatant so as not to disturb the pellet. The pellet was then resuspended in 250µl pf PBS, and 200µl of 5% Saponin (Tokyo Chemical Industry, Tokyo, Japan) was added to a make a final concentration of 2.5% Saponin and mixed well. The mixture was incubated in room temperature for 10 minutes to promote host cell lysis. After incubation of 10 minutes, 350µl of water was added and incubation at room temperature was continued for a further 30 seconds. To deliver an osmotic shock to lyse the damaged host cell, 12µl of 5M NaCl (Invitrogen™, ThermoFisher Scientific, Waltham, Massachusetts, USA) was added to the mixture and centrifuged at 6,000 x *g* for 5 minutes. The supernatant was removed, leaving 50µl of supernatant and pellet resuspended with 50µl of PBS, giving a total volume of 100µl. Equal volume (100µl) of HL-SAN buffer, consisting of 5M NaCl and 100mM $MgCl_2$ (Invitrogen™, ThermoFisher Scientific, Waltham, Massachusetts, USA) in nuclease-free water, with 10µl (25,000 units) of HL-SAN DNase (Technologies ASA, Tromsø, Norway) was added to the mixture and incubated for 15 minutes at 37°C on a shaking plate at 800RPM to allow for host DNA digestion. The sample was then centrifuged at 6,000 x *g* for 3 minutes. Finally,

the host DNA depleted sample was washed twice with increasing volumes of PBS (800µl and then 1000µl of PBS). After each wash, the sample was centrifuged at 6,000 x *g* for 3 minutes. The supernatant was discarded and pellet resuspended in PBS.

### 4.2.4 Patient recruitment and sample collection

Patients presenting to the Birmingham and Midland Eye Centre, Birmingham, United Kingdom (n = 11) with suspected endophthalmitis were invited to participate in the study and written informed consent was obtained. Patients underwent either intraocular fluid biopsy or pars plana vitrectomy according to standard-of-care protocol for endophthalmitis at their respective clinical institutions. The vitreous tap was attempted and if unsuccessful, an aqueous tap was performed. The samples were sent for routine culture at the clinical microbiology laboratory, and remaining samples for research were frozen at -80$^{o}$C and sent frozen on dry ice to the University of Washington.

DNA samples from intraocular fluid biopsy (aqueous humour or vitreous "tap") of patients with suspected endophthalmitis from the previously published USA cohort[107] (n = 12) were also used in this study.

### 4.2.5 DNA extraction

All samples were stored at -80$^{o}$C until DNA extraction. DNA from the UK cohort was extracted with Quick-DNA/RNA™ Microprep Plus Kit (Zymo Research, Irvine, CA, USA) per manufacturer's protocol. Briefly, intraocular biopsy samples were placed directly in 800µl DNA/RNA Shield™ (1X) (Zymo Research, Irvine, CA, USA) in ZR BashingBead™ Lysis Tubes® and mechanically homogenised at 2500 – 4000 RPM for 45 seconds. Samples were then

centrifuged at 16,000 x *g* for 30 seconds. The supernatant (~600μl) was then transferred into an RNase-free tube containing equal volume (~600μl) of DNA/RNA Lysis buffer and mixed well. Sample was transferred into a Zymo-Spin™ IC-XM Column in a Collection Tube and centrifuged at 16,000 x *g* for 30 seconds. The DNA is bound to column. The Zymo-Spin™ IC-XM Column was transferred into a new collection tube. DNA/RNA Prep Buffer (400μl) was added to the column and centrifuged at 16,000 x *g* for 30 seconds. The flow-through was discarded, and 700μl of DNA/RNA Wash Buffer was added to the sample and centrifuged at 16,000 x *g* for 30 seconds. The flow-through was discarded. DNA/RNA Wash Buffer (400μl) was added to column and centrifuged at 16,000 x *g* for 2 minutes to ensure complete removal of the wash buffer. The column was carefully transferred into a clean microcentrifuge tube. DNase/RNase-Free Water (25μl) was added directly to column matrix and incubated at room temperature for 5 minutes. The sample was then centrifuged at 16,000 x *g* for 30 seconds to elute the DNA from the column. DNA concentration was determined fluorometrically using a Qubit® dsDNA high-sensitivity assay (Thermo Fisher Scientific, Waltham, MA, USA) on the Qubit®2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) . The eluted DNA was then stored at -80$^{o}$C.

DNA from the USA cohort had been previously extracted with Qiagen Blood and Tissue kit according to manufacturer's protocol and published in the American Journal of Ophthalmology[107].

### 4.2.6 Quantitative Polymerase Chain Reaction (PCR)

The quantitative PCR (qPCR) assay to detect β-mouse actin and universal pan-bacterial 16S ribosomal DNA was performed on the Applied Biosystems 7500 Fast Real-Time PCR System

(Foster City, CA, USA). The primer sequence pairs for β-mouse actin were 5'-GCTGCCTCAACACCTCAAC-3' and 5'-AGGTGACAGCATTGCTTCTG-3'. The primer sequence pairs for 16S ribosomal DNA universal primers (Integrated DNA Technologies, San Diego, CA, USA) were 5'-GAGGAAGGTGGGGATGACGT-3' and 5'-AGGCCCGGGAACGTATTCAC-3'. The final PCR mix consisted of 0.8μl each of forward and reverse primers, 10μl of PowerUp™ SYBR™ Green Master Mix (Applied Biosystems, ThermoFisher Scientific, Waltham, Massachusetts, USA), 6.4μl of UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™, ThermoFisher Scientific, Waltham, Massachusetts, USA) and 2μl of unamplified genomic DNA. The final reaction volume was 20μl. All samples were performed in duplicate. The run consisted of initial holding stages at 50°C for 2 minutes and at 95°C for 2 minutes, followed by a cycling stage (40 cycles) at 95°C for 15 seconds and 60°C for 1 minute.

Fold changes in the relative quantity of host (β-mouse actin) and bacterial (16S) amplicons were calculated based on the $C_T$ values, using the following equation: "Fold change $= 2^{\Delta C_T}$, where $\Delta C_T = (C_{T\ un\text{-}saponified} - C_{T\ saponified})$"[158]

4.2.7 Full-length 16S rRNA nanopore sequencing and bioinformatics

The full-length 16S rRNA genes (16S sequencing primers: 27F-AGAGTTTGATCMTGGCTCAG; 1492R-CGGTTACCTTGTTACGACTT) were amplified using the 16S Barcoding Kit (Oxford Nanopore Technologies, Oxford, UK). Genomic DNA (10ng) was transferred into a 1.5mL DNA LoBind tube (Eppendorf, Hamburg, Germany) and volume was adjusted to 10μl with UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™, ThermoFisher Scientific, Waltham, Massachusetts, USA). The mixture was then mixed thoroughly by flicking the tube and spun down briefly in a microcentrifuge before transferring into a 0.2mL thin-walled PCR tube. For samples with less than 1ng/μl of genomic DNA, 10μl of genomic DNA was

transferred directly into a 0.2mL thin-walled PCR tube. The final PCR mix consisted of 10μl of input DNA, 14μl of UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™, ThermoFisher Scientific, Waltham, Massachusetts, USA), 1μl of 16S Barcode at 10μM, and 25μl of LongAmp Taq 2X master mix (New England BioLabs, Ipswich, Massachusetts, USA). The PCR amplification was performed on the Eppendorf Mastercycler® using the following cycling conditions: initial denaturation of 1 min at 95°C (1 cycle), followed by 35 cycles of denaturation for 20 seconds at 95°C, annealing for 30 seconds at 55°C, extension for 2 mins at 65°C , and final extension for 5 minutes at 65°C (1 cycle). The PCR amplicon was then transferred to a 1.5mL Eppendorf DNA LoBind tube containing 30μl of AMPure XP beads (Beckman Coulter Life Sciences, Indianapolis, Indiana, USA), and incubated on a rotator mixer for 5 minutes at room temperature. The sample was then spun down and pelleted on a magnet. The supernatant was removed and beads were washed with 200μl of freshly prepared 70% ethanol (3 washes in total). Any residual ethanol was removed and the pellet was allowed to dry for 30 seconds. The tube was removed from the magnetic rack, resuspended in 10μl of 10mM Tris-HCl (Invitrogen™, ThermoFisher Scientific, Waltham, Massachusetts, USA) at pH 8.0 with 50mM NaCl (Invitrogen™, ThermoFisher Scientific, Waltham, Massachusetts, USA), and incubated for 2 minutes at room temperature. The beads were pelleted on a magnet until the elute was clear and colourless. The eluate (10μl) was removed and retained in a clean 1.5mL Eppendorf DNA LoBind tube. The DNA concentration was determined fluorometrically using a Qubit® dsDNA high-sensitivity assay on the Qubit®2.0 Fluorometer. The barcoded samples were then pooled to a total of 150 – 200 fmoles in 10μl of 10mM Tris-HCl pH 8.0 with 50mM NaCl. The RAP solution (1μl) was added to the barcoded DNA and incubated for 5 minutes at room temperature. The library was prepared and loaded onto the primed flow cell with the following reagents: 34μl Sequencing

Buffer, 25.5μl of Loading Buffer, 4.5μl of UltraPure™ DNase/RNase-Free Distilled Water and 11μl of DNA library. The barcoded 16S libraries were then sequenced on MinION (Oxford Nanopore Technologies, Oxford, UK) using the R9 flow cell.

Raw sequence reads were basecalled using ONT's MinKNOW software Guppy v.3.2.2. Raw reads were demultiplexed using qcat version 1.0.1. 16S rRNA database were obtained from National Center for Biotechnology Information (NCBI) 16S RefSeq (Nucleotide search details: 33175 [BioProject] or 33317 [BioProject]). Reference sequences from the 16S rRNA database was then clustered into 100% similarity threshold using heuristic clustering method (greedy incremental clustering algorithm) on CD-HIT version 4.8.1 (commands: -c 0.97, 0.99, 1.0, -M 62900, -d 250). Minimap2 version 2.12-r87 (commands: -K 100M, -ax map-ont) was used to align the demultiplexed reads to the 16S database and the resulting files were processed using the Sequence Alignment/Map (SAM)tools version 1.9 [commands: samtools view -b -F 2308 (to remove unmapped, non-primary and supplementary reads), samtools sort, samtools index, samtools idxstats].

4.2.8 Nanopore Whole Genome Sequencing (Nanopore WGS) and bioinformatics

The ONT Rapid Barcoding Sequencing kit was used to make DNA libraries for nanopore whole genome sequencing. Genomic DNA (5ng) was transferred into a 1.5mL DNA LoBind tube and volume adjusted to 3μl with UltraPure™ DNase/RNase-Free Distilled Water. The mixture was then mixed thoroughly by flicking the tube and spun down briefly in a microcentrifuge before transferring into a 0.2mL thin-walled PCR tube containing 1μl of FRM solution. For samples with less than 1ng/μl of genomic DNA, 3μl of genomic DNA was transferred directly into a 0.2mL thin-walled PCR tube containing 1μl of FRM solution. The mixture was incubated at 30$^{o}$C for 1 minute and then at 80$^{o}$C for 1 minute in a thermocycler for DNA tagmentation and

cooled down in ice briefly. The final PCR mixture contained: 20μl of UltraPure™ DNase/RNase-Free Distilled Water, 4μl of tagmented DNA, 1μl of RLB barcodes at 10μM and 25μl of LongAmp Taq 2X master mix. The PCR amplification was performed on the Eppendorf Mastercycler® using the following cycling conditions: initial denaturation of 3 minutes at 95°C (1 cycle), followed by 35 cycles of denaturation for 15 seconds at 95°C, annealing for 15 seconds at 55°C, extension for 6 minutes at 65°C , and final extension for 6 minutes at 65°C (1 cycle). The PCR product was then transferred to a 1.5mL Eppendorf DNA LoBind tube containing 30μl of AMPure XP beads, and incubated on a rotator mixer for 5 minutes at room temperature. The sample was then spun down and pelleted on a magnet. The supernatant was removed and beads were washed with 200μl of freshly prepared 70% ethanol (3 washes in total). Any residual ethanol was removed and the pellet was allowed to dry for 30 seconds. The tube was removed from the magnetic rack, resuspended in 10μl of 10mM Tris-HCl at pH 8.0 with 50mM NaCl, and incubated for 2 minutes at room temperature. The beads were pelleted on a magnet until the elute was clear and colourless. The eluate (10μl) was removed and retained in a clean 1.5mL Eppendorf DNA LoBind tube. The DNA concentration was determined fluorometrically using a Qubit® dsDNA high-sensitivity assay on the Qubit®2.0 Fluorometer. The barcoded samples were then pooled to a total of 150 – 200 fmoles in 10μl of 10mM Tris-HCl pH 8.0 with 50mM NaCl. The RAP solution (1μl) was added to the barcoded DNA and incubated for 5 minutes at room temperature. The library was prepared and loaded onto the primed flow cell with the following reagents: 34μl Sequencing Buffer, 25.5μl of Loading Buffer, 4.5μl of UltraPure™ DNase/RNase-Free Distilled Water and 11μl of DNA library. The barcoded nanopore whole genome libraries were then sequenced on MinION (Oxford Nanopore Technologies, Oxford, UK) using the R9 flow cell. Raw sequence reads were basecalled using ONT's MinKNOW software Guppy v.3.2.2 and analysed using Epi2ME What's

In My Pot (WIMP) pipeline (v.3.2.1) for metagenomic analysis. For genome alignment to predominant microorganism, the basecalled reads were demultiplexed using qcat version 1.0.1, aligned to the human Genome Reference Consortium Human Build 38 patch release 13 (GRCh18.p13) using minimap2 and unassigned (i.e. non-human reads) were exported to bam file using SAMtools (v.1.9) with the flag (-f 4) and converted back to fasta file with the command (fasta). Further downstream analysis for genome coverage to the predominant microorganism was performed using minimap2 with the parameters for ONT long-read data (-ax map -ont); SAMtools [commands: samtools view -b -F 2308 -q 30 (to remove unmapped, non-primary, supplementary, low-quality reads), samtools sort, samtools index, samtools idxstats]; and bedtools (v.2.28.0) with the following parameter: genomecov -ibam. Bioinformatics scripts are available in the Supplementary material 4.1 – 4.3.

### 4.2.9 Illumina Whole Genome Sequencing (Illumina WGS) and bioinformatics

The libraries for Illumina WGS was prepared using Illumina Nextera™ XT DNA Flex Library Prep (Illumina Inc, San Diego, California, USA). Genomic DNA (30ng) was transferred to a 96-well PCR plate and volume adjusted to a total of 30μl with UltraPure™ DNase/RNase-Free Distilled Water. For samples with less than 1ng/μl of genomic DNA, 30μl of genomic DNA was transferred to a 96-well PCR plate. The tagmentation master mix, consisting of 10μl of BLT and 10μl of TB1 per reaction, was mixed well with each sample. The plate was incubated at 55$^{o}$C for 15 minutes followed by a 10$^{o}$C hold. For the post-tagmentation clean up, 10μl of TSB was added to the tagmentation reaction and further incubated at 37$^{o}$C for 15 minutes on the Eppendorf Mastercycler®. The plate was then placed on a magnet, supernatant removed and washed with 100μl of TWB (for a total of two washes). The tagmented DNA is then amplified with the PCR master mix consisting of 20μl pf EPM and 20μl of UltraPure™ DNase/RNase-Free

Distilled Water per reaction. Index adapters, 5μl each of i5 and i7 adapters, from the Nextera™ DNA CD Indexes (Illumina Inc, San Diego, California, USA) were added to each sample, and ran in the Eppendorf Mastercycler® with the following programme: 68°C for 3 minutes, 98°C for 3 minutes, 14 cycles of 98 °C for 45 seconds, 62 °C for 30 seconds, 68 °C for 2 minutes, and then 68 °C for 1 minute and held at 10 °C. After the PCR cycling, the plate was centrifuged at 280 x *g* for 1 minute and placed onto the magnet until the supernatant was clear. The PCR supernatant (45μl) was transferred to fresh midi plate and 85μl of diluted SPB solution was added to each PCR product. The mixture was incubated at room temperature for 5 minutes and placed onto the magnet. The supernatant (125μl) was transferred to new midi plate containing 17μl of SPB and incubated at room temperature for 5 minutes. The midi plate was placed on the magnet until clear, supernatant discarded, washed with 300μl of 80% ethanol and wash repeated for a total of two washes. The beads were left to airdry on the magnet for 5 minutes, and 32μl of RSB were added to resuspend the beads. The mixture was incubated at room temperature for 2 minutes, and placed back onto the magnet until clear. The supernatant (30μl) was transferred to a new plate. The DNA concentration of each individual library was determined fluorometrically using a Qubit® dsDNA high-sensitivity assay on the Qubit®2.0 Fluorometer. The libraries were pooled with RSB to a concentration of 4 nM. Then, 5μl of the pooled library (at 4nM) was added to 5μl of 0.2N NaOH and incubated at room temperature for 5 minutes. After the incubation, HTI buffer (990μl) was added and 900μl of the mixture was mixed with 100μl of PhiX control (20pM). The final loading concentration of the pooled library was 4pM. The pooled library was sequence on the Illumina MiSeq platform. The bioinformatics analysis of Illumina MiSeq FASTQ files were performed using the Scalable Metagenomic Analysis Research Tool (SMART) pipeline[232] at the University of Washington. Bioinformatics script available in the Supplementary material 4.4.

### 4.2.10 Negative controls

Negative control DNA extraction and sequencing was performed on reagents without a DNA template and processed in the same manner as the clinical intraocular fluid biopsy samples to account for any potential contamination.

### 4.2.11 Definitions

The microorganism was considered to be a potential pathogen or credible 'hit' on sequencing taking into account the background contamination (i.e. sequences present in the negative control samples) and if

1. "the organism was known to be potentially pathogenic given the clinical context of the particular patient"[106]
2. the organism represented the highest bacterial load (for WGS) or most abundant reads (for 16S) in the sample[105,106]

Bacterial load is defined as "the median number of the presumed pathogen per recovered human genome by whole genome sequencing"[107], calculated as:

$$\text{Bacterial load} = \frac{\text{Total number of pathogen reads}}{\text{Total number of human reads}} \times \frac{\text{Size of human genome}}{\text{Size of pathogen genome}}$$

Alignment breadth is calculated as[107]:

$$\text{Alignment breadth} = \frac{\text{Number of bases with} \geq 1 \text{ aligned base}}{\text{Size of reference pathogen genome}}$$

Alignment depth is calculated as[107]:

$$\text{Alignment depth} = \frac{\text{Sum of all aligned bases of the query sequence}}{\text{Length of regions with } \geq 1 \text{ aligned base of the reference sequence}}$$

Coverage is calculated as[107]:

$$\text{Coverage} = \text{Alignment depth} \times \text{Alignment breadth}$$

### 4.2.12 Statistical analyses

Comparisons of the proportion of human and non-human reads on saponified and un-saponified samples were performed using Mann-Whitney tests. Data were summarised using medians and ranges. Statistical analyses were performed using GraphPad Prism version 8.3.1 (GraphPad Software, San Diego, California, USA), with $p<0.05$ deemed to be indicative of statistical significance throughout.

## 4.3 Results

4.3.1 Effects of differential centrifugation and saponification on host DNA depletion and bacterial DNA recovery

To examine the effects of saponification on host DNA depletion by qPCR, murine peripheral blood mononuclear cells (PBMCs) were spiked with varying concentrations of *E.coli*. There was minimal bacterial DNA loss between saponified and un-saponified samples, with difference in quantification cycles ($\Delta C_T$) of less than 2 [**Figure 4.1**]. A median of 1573-fold (range in $\Delta C_{T\ mouse\ actin}$ : 9.01 – 11.67) depletion of host DNA (median $C_T$ of mouse actin in saponified samples: 36.54 vs. non-saponified samples: 25.92, p<0.0001) was observed [**Figure 4.2**].

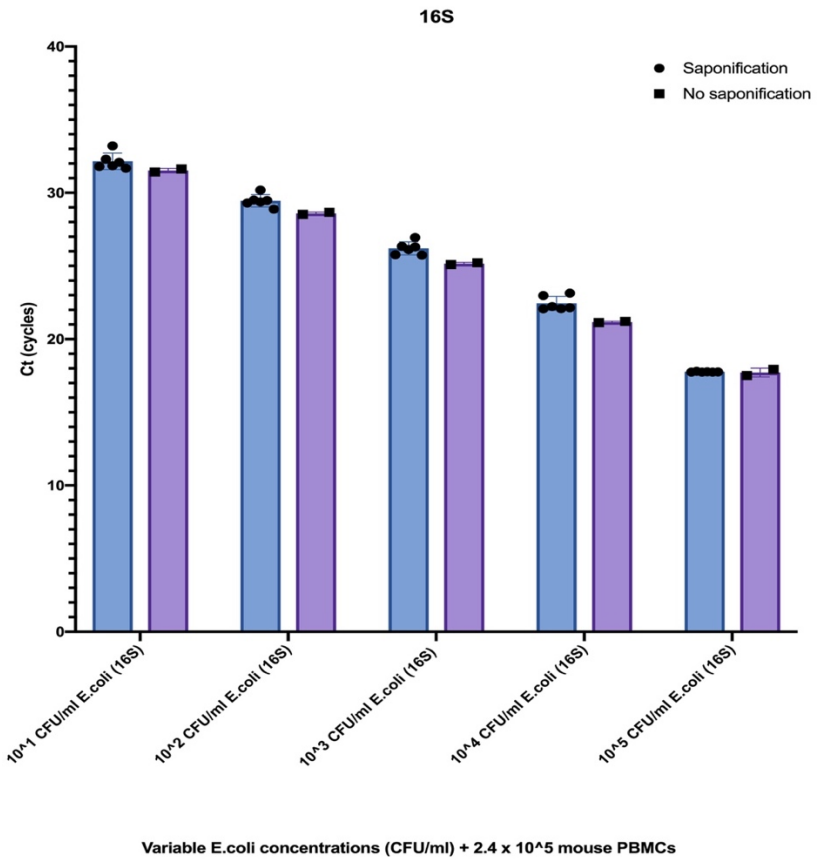



*Figure 4.1 Bar graph comparing the effects of saponification on 16S bacterial recovery by PCR in murine peripheral blood mononuclear cells (PBMC) samples spiked with varying E.coli concentrations, ranging from $10^1$ to $10^5$ colony-forming units (CFU)/ml.*

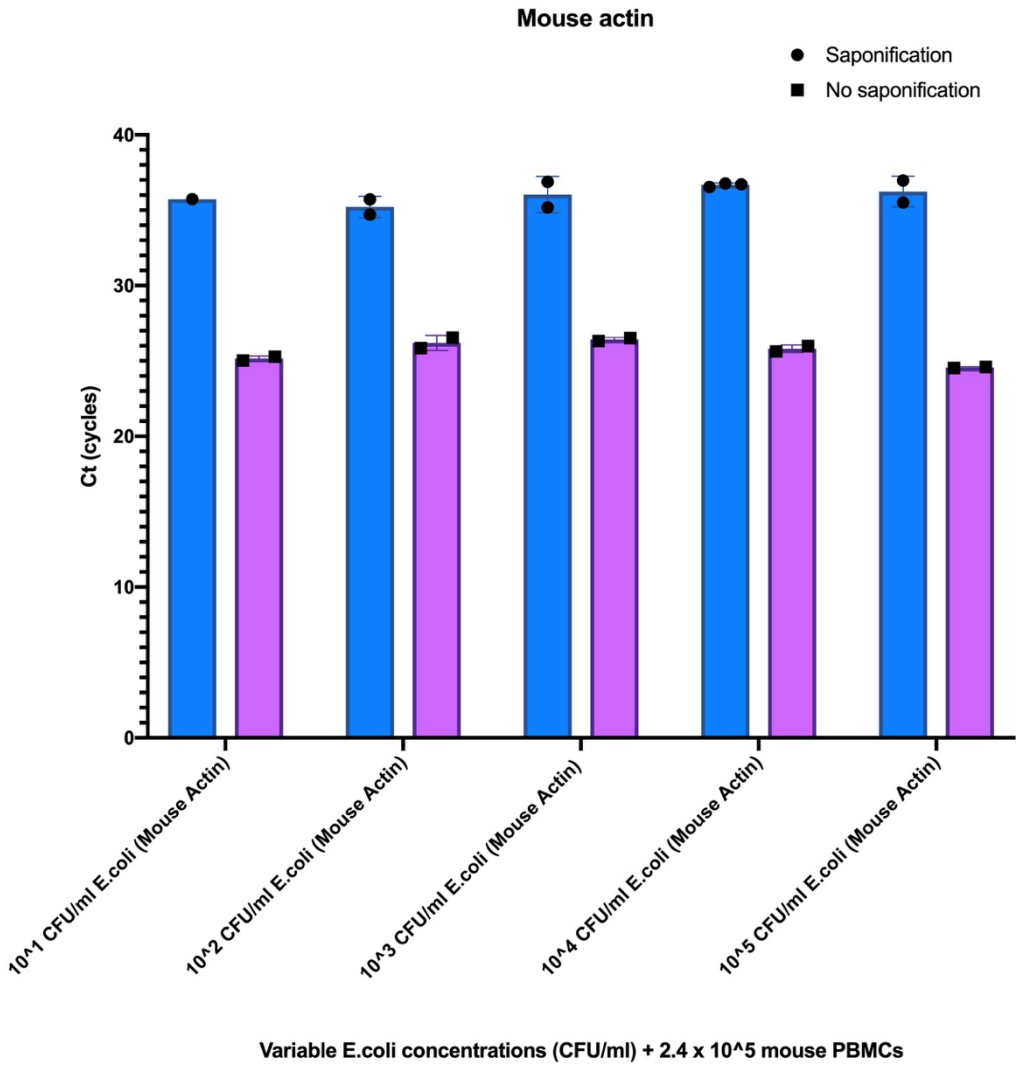



*Figure 4.2 Bar graph comparing the effects of saponification on murine actin recovery in murine peripheral blood mononuclear cells (PBMC) samples spiked with varying E.coli concentrations, ranging from $10^1$ to $10^5$ colony-forming units (CFU)/ml.*

To compare the effects of saponification on host DNA depletion by whole genome nanopore sequencing, vitreous samples from a subset of patients with endophthalmitis (n=5) were split into two aliquots: one aliquot of vitreous subjected to saponification and another aliquot had no saponification. The proportion of human reads in the saponified clinical samples were lower compared to un-saponified samples [median %human reads in saponified samples was 71.35% (range: 28.9% - 97.95%) vs 99.41% (range: 97.7% - 99.63% in un-saponified samples (p=0.016)]. Saponified samples had higher proportion of bacterial reads (calculated by sequencing) compared to un-saponified samples [median of bacterial load in saponified samples: 92.5 (range: 9.7 – 6019.1) vs 2.9 (range: 0.04 – 9.49) in un-saponified samples (p=0.016)] **Table 4.1**.

The predominant microbial organism detected by whole genome nanopore sequencing agreed with culture results in all saponified samples (4/4 culture positive saponified samples) and in 3 of 4 un-saponified samples. The predominant genus identified in a saponified vitreous sample (NSEQ005) was *Pseudomonas*, which agreed with the culture results, however, the un-saponified vitreous from the same sample (NSEQ005) did not detect any significant taxon by nanopore sequencing.

As saponification was shown to reduce the proportion of host DNA without inadvertent loss in bacterial DNA, the process of saponification was applied to the intraocular fluid biopsy samples from the UK. The samples from the USA had already undergone DNA extraction, therefore, the process of saponification could not be applied to these samples.

| Sample ID | Saponification | Total Reads | Total human reads | %human reads | Total non-human reads | %non-human reads | Pathogen detected (Genus) | Bacterial load |
|---|---|---|---|---|---|---|---|---|
| NSEQ004 | Saponified | 89829 | 64089 | 71.35 | 25740 | 28.65 | Pseudomonas | 104.74 |
| NSEQ004 | No saponin | 1091540 | 1087541 | 99.63 | 3999 | 0.37 | Pseudomonas | 0.04 |
| NSEQ005 | Saponified | 615898 | 603270 | 97.95 | 12628 | 2.05 | Pseudomonas | 9.76 |
| NSEQ005 | No saponin | 1066266 | 1059948 | 99.41 | 6,318 | 0.60 | No significant taxon | N/A |
| NSEQ006 | Saponified | 58172 | 54851 | 94.29 | 3321 | 5.71 | No significant taxon | N/A |
| NSEQ006 | No saponin | 631793 | 628238 | 99.44 | 3555 | 0.56 | No significant taxon | N/A |
| NSEQ007 | Saponified | 596285 | 172340 | 28.90 | 423945 | 71.10 | Staphylococcus | 6019.91 |
| NSEQ007 | No saponin | 848392 | 843164 | 99.38 | 5228 | 0.62 | Staphylococcus | 2.92 |
| NSEQ011 | Saponified | 335668 | 171558 | 51.11 | 164110 | 48.89 | Escherichia | 80.36 |
| NSEQ011 | No saponin | 652640 | 637609 | 97.70 | 15031 | 2.30 | Escherichia | 9.49 |

*Table 4.1 Comparisons of the effects of saponification on proportion of human, non-human reads and bacterial load*

### 4.3.2 Sequences detected in negative controls

Sequences detected in the negative controls (extraction blanks as no template control) for full-length 16S rRNA nanopore, whole genome nanopore and Illumina sequencing runs are summarized in **Supplementary Table 4.1 – 4.3**. Microorganisms commonly found in plants such as *Mesorhizobium loti* and *Sinorhizobium meliloti* were detected in negative control samples that have been undergone saponification. Contaminating sequences that match water- and soil-associated genera were also detected including *Strenotrophomonas*, *Acinetobacter, Pseudomonas* and *Burkholderia.* Other contaminating sequences associated with skin commensal bacteria, such as *Streptococcus* spp. and *Cutibacterium acnes* were detected in the negative controls. The composition of the contaminating sequences varied greatly between the sequencing platforms and runs.

### 4.3.3 Clinical samples

A total of 23 samples were used in the study, 22 of which were vitreous samples and 1 aqueous humour. Samples were derived from patients with a median age of 68.5 years (range: 47 – 88 years) and 14 (61%) were male. The most common preceding clinical history was post-intravitreal injection (43%) followed by post-cataract surgery (39%). The majority of the patients presented with visual acuity worse than 6/60 (87%) [**Table 4.2**].

| Sample ID | Gender | Age (years) | Laterality | Clinical history | VA at presentation |
|---|---|---|---|---|---|
| **NSEQ004** | M | 84 | Right | 2 days post cataract surgery & IOL insertion | HM |
| **NSEQ005** | F | 88 | Left | 10 days post-left IOL exchange (Artisan lens) & anterior vitrectomy for subluxed IOL (previous bilateral cataract surgery in 2008) | PL |
| **NSEQ006** | F | 54 | Right | 2 days post-IVT injection | HM |
| **NSEQ007** | M | 83 | Left | 3 days post-IVT injection | HM |
| **NSEQ011** | F | 47 | Left | Endogenous endophthalmitis secondary to *E.coli* septicaemia treated with antibiotics. PMH of renal transplant on immunosuppression | 6/18 |
| **NSEQ003** | M | 59 | Left | 14 days post-corneal graft surgery | HM |
| **NSEQ014** | F | 75 | Right | 5 days post cataract surgery & IOL insertion | HM |
| **NSEQ015** | F | 87 | Left | 11 days post-IVT injection | NPL |
| **NSEQ008** | M | 56 | Left | 3 days post-traumatic (wire from grinding wheel) penetrating eye injury with full-thickness corneal laceration | HM |
| **NSEQ009** | M | 66 | Left | 3 days h/o microbial keratitis and hypopyon treated with topical antibiotics. PMH Rheumatoid arthritis on Methotrexate and Hydroxychloroquine | PL |
| **NSEQ012** | M | 62 | Left | 7 days post-IVT injection | 6/60 |
| **WE164D** | M | 76 | Left | 3 days post-IVT injection | HM |
| **WE25D** | M | 73 | Left | 8 days post cataract surgery & IOL insertion | HM |
| **WE6D** | F | 65 | Right | 18 days post cataract surgery & IOL insertion | HM |
| **WE13D** | F | 85 | Right | 5 days post cataract surgery & IOL insertion | LP |
| **WE26D** | F | 76 | Right | 4 days post cataract surgery & IOL insertion | CF |
| **WE141D** | F | 61 | Left | 2 days post cataract surgery & IOL insertion | HM |
| **WE152D** | M | 54 | Left | 1 day post-IVT injection | CF |
| **WE27D** | M | 70 | Right | 9 days post-IVT injection | 6/12 |
| **WE30D** | M | 61 | Right | 4 days post-IVT injection | 6/7.5 |
| **WE23D** | M | 73 | Left | 5 days post cataract surgery & IOL insertion | HM |
| **WE155D** | F | 51 | Right | 5 days post-IVT injection | HM |
| **WE164D** | M | 76 | Left | 3 days post-IVT injection | HM |
| **WE135D** | M | 67 | Left | 6 days post cataract surgery & IOL insertion | CF |

*Table 4.2 Demographics of patient participants.*

*Abbreviations: history of, h/o; intraocular lens, IOL; intravitreal injection, IVT; past medical history, PMH; hand movement, HM; perception of light, PL; counting fingers, CF.*

4.3.4 Agreement between microbial culture and full-length 16S rRNA nanopore sequencing

A total of 18 samples were culture-positive and 5 were culture negative. The most commonly cultured organism was *Staphylococcus epidermidis*, followed by other *Staphylococcus* spp, *Streptococcus* spp and *Pseudomonas* spp (**Table 4.3**).

Twenty samples had sufficient volume for full-length 16S rRNA nanopore sequencing. The predominant organism detected by full-length 16S rRNA nanopore sequencing agreed with culture result in 15 cases (15/20, 75%) (**Table 4.4**).

Two culture-positive cases did not yield any significant organism on 16S rRNA nanopore sequencing - NSEQ004 grew *Pseudomonas aeruginosa*, whilst NSEQ003 grew *Propionibacterium acnes*. In two culture-negative cases, 16S rRNA nanopore sequencing detected potential pathogens including *Staphylococcus epidermidis* in sample WE25D and polymicrobial *Massillia oculi* and *Cutibacterium acnes* in sample WE27D. These 2 culture-negative samples were tested by Illumina WGS and the results agreed with 16S rRNA nanopore sequencing results. *Staphylococcus epidermidis* was detected by Illumina WGS in sample WE25D. In sample WE27D, *Propionibacterium acnes* (previous name for *Cutibacterium acnes*) was detected by Illumina WGS in sample WE27D, but *Massillia oculi* was not detected. One sample (NSEQ011) grew *Escherichia coli* in culture while the predominant organism detected by 16S nanopore sequencing was *Staphylococcus sanguinis.* WGS Nanopore for NSEQ011 detected polymicrobial *Escherichia coli* and *Streptococcus mitis*, however no significant taxon was detected by WGS Illumina.

| Sample ID | Organism cultured by conventional microbiology | Organism detected from 16S rRNA nanopore sequencing | Organism detected from whole genome nanopore sequencing | Organism detected by Illumina Miseq |
|---|---|---|---|---|
| NSEQ004 | *Pseudomonas aeruginosa* | No significant taxon | *Pseudomonas sp. LPH1* | Pseudomonas flourescences |
| NSEQ005 | *Pseudomonas aeruginosa* | *Pseudomonas aeruginosa* | *Pseudomonas pseudoalcaligenes* | *Pseudomonas aeruginosa* |
| NSEQ006 | No growth | No significant taxon | No significant taxon | No significant taxon |
| NSEQ007 | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* |
| NSEQ011 | *Escherichia coli* | *Streptococcus sanguinis* | *Escherichia coli*<br>*Streptococcus mitis* | No significant taxon |
| NSEQ003 | *Propionibacterium acnes* | No significant taxon | *Cutibacterium acnes* | No significant taxon |
| NSEQ014 | *Staphylococcus epidermidis* | *Staphylococcus saccharolyticus*<br>*Staphylococcus epidermidis* | *Staphylococcus aureus* | No significant taxon |
| NSEQ015 | *Staphylococcus aureus* | *Staphylococcus aureus* | *Staphylococcus aureus*<br>Phietavirus<br>(Staphylococcus bacteriophage) | *Staphylococcus aureus* |
| NSEQ008 | No growth | No significant taxon | No significant taxon | No significant taxon |
| NSEQ009 | No growth | No significant taxon | No significant taxon | No significant taxon |
| NSEQ012 | *Staphylococcus aureus* | *Staphylococcus aureus* | *Staphylococcus aureus* | *Staphylococcus aureus* |
| WE13D | Alpha-haemolytic streptococcus | *Streptococcus sanguinis* | *Streptococcus sanguinis* | *Streptococcus sanguinis* |
| WE141D | *Streptococcus mitis* | *Streptococcus sanguinis* | *Streptococcus sanguinis* | *Streptococcus sanguinis* |
| WE152D | *Enterococcus faecalis* | *Enterococcus faecalis* | *Enterococcus faecalis* | *Enterococcus faecalis* |
| WE164D | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* |
| WE23D | *Staphylococcus epidermidis* | N/A | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* |
| WE155D | *Staphylococcus epidermidis* | N/A | *Staphylococcus epidermidis*<br>Phietavirus<br>(Staphylococcus bacteriophage) | *Staphylococcus epidermidis* |
| WE135D | Coagulase-negative Staphylococcus | N/A | *Cutibacterium acnes*<br>*Staphylococcus haemolyticus* | *Staphylococcus haemolyticus* |
| WE25D | No growth | *Staphylococcus epidermidis* | N/A | *Staphylococcus epidermidis* |
| WE6D | *Staphylococcus lugdunesis* | *Staphylococcus lugdunesis* | N/A | *Staphylococcus lugdunesis* |
| WE26D | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | N/A | *Staphylococcus epidermidis* |
| WE27D | No growth | *Massillia oculi*<br>*Cutibacterium acnes* | N/A | *Propionibacterium acnes* |
| WE30D | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | N/A | *Staphylococcus epidermidis* |

*Table 4.3 Summary of predominant microorganism detected by culture, 16S rRNA nanopore sequencing, whole genome nanopore and Illumina sequencing.*

*N/A denotes insufficient sample to be processed for sequencing.*

| | 16S Nanopore | | | Nanopore WGS | | | Illumina WGS | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***Positive*** | ***Negative*** | ***Total*** | ***Positive*** | ***Negative*** | ***Total*** | ***Positive*** | ***Negative*** | ***Total*** |
| **Culture Positive** | 13 | 2 | 15 | 15 | 0 | 15 | 15 | 3 | 18 |
| **Culture Negative** | 2 | 3 | 5 | 0 | 3 | 3 | 2 | 3 | 5 |
| **Total** | 15 | 5 | 20 | 15 | 3 | 18 | 17 | 6 | 23 |

*Table 4.4 Comparison of microbial culture, full-length 16S nanopore sequencing, nanopore and Illumina whole genome sequencing (WGS) results for endophthalmitis.*

4.3.5 Agreement between microbial culture and nanopore whole genome sequencing

A total of 18 samples had sufficient volume for whole genome nanopore sequencing (WGS Nanopore). The predominant genus detected by Nanopore WGS agreed with culture results in all of the cases (18/18). Speciation by Nanopore WGS agreed with culture results in 16 out of the 18 cases.

Nanopore sequencing reads of culture-positive cases were aligned to the corresponding reference genome of the predominant organism detected by Nanopore WGS (**Table 4.5**). Low levels of alignment breadth were observed in samples from the United Kingdom (NSEQ003 – NSEQ015), ranging from 0.17% to 12.89%, and could be due to DNA degradation as the frozen samples thawed to room temperature during the shipment process. These UK samples had relatively high bacterial load (by sequencing reads), with a median of 608.4 (range: 5.62 – 418445.89). The samples from the USA cohort (WE13D – WE164D) had higher levels of alignment breadth, ranging from 6.96% to 88.77%, compared to the UK samples.

| Sample ID | Organism detected by culture | Organism detected by WGS Nanopore | Size of reference genome | Bacterial load | Alignment breadth | Alignment depth | Coverage |
|---|---|---|---|---|---|---|---|
| NSEQ003 | Propionibacterium acnes | *Cutibacterium acnes* | 2560265 | 1311.80 | 2.54% | 1824 | 46.36 |
| NSEQ004 | *Pseudomonas aeruginosa* | *Pseudomonas sp LPH* | 5229495 | 104.74 | 0.17% | 950 | 1.64 |
| NSEQ005 | *Pseudomonas aeruginosa* | *Pseudomomnas oleovarans*[a] | 5005894 | 5.62 | 0.09% | 94 | 0.08 |
| NSEQ007 | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | 2564615 | 4386.11 | 7.58% | 1817 | 137.83 |
| NSEQ011 | *Escherichia coli* | *Escherichia coli* | 4641652 | 79.39 | 0.37% | 1445 | 5.4 |
| NSEQ012 | *Staphylococcus aureus* | *Staphylococcus aureus* | 2821361 | 919.11 | 0.87% | 463 | 4.05 |
| NSEQ014 | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | 2564615 | 297.70 | 0.91% | 1 | 0.01 |
| NSEQ015 | *Staphylococcus aureus* | *Staphylococcus aureus* | 2821361 | 418445.89 | 12.89% | 9906 | 1277.15 |
| WE13D | Alpha-hemolytic *streptococcus* | *Streptococcus sanguinis* | 2388435 | 0.75 | 6.96% | 1 | 0.07 |
| WE23 | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | 2564615 | 4.94 | 71.84% | 2 | 1.55 |
| WE135D | Coagulase-negative *Staphylococcus* | *Staphylococcus haemolyticus* | 2560146 | 0.09 | 0.32% | 8 | 0.02 |
| WE141D | *Streptococcus mitis* | *Streptococcus sanguinis* | 2388435 | 199.32 | 88.77% | 82 | 72.96 |
| WE152D | *Enterococcus faecalis* | *Enterococcus faecalis* | 3359974 | 753.79 | 75.89% | 154 | 116.64 |
| WE155D | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | 2564615 | 234.21 | 15.60% | 167 | 26.17 |
| WE164D | *Staphylococcus epidermidis* | *Staphylococcus epidermidis* | 2564615 | 5.53 | 8.59% | 15 | 1.27 |

*Table 4.5 Comparisons of bacterial load, mean alignment breadth, depth and genomic coverage.*

4.3.6 Agreement between culture and Illumina whole genome sequencing (Illumina WGS)

All 23 samples were processed for Illumina WGS. The predominant organism detected by Illumina WGS agreed with culture results in 18 cases (78%). Illumina WGS detected the cultured pathogen as the predominant genus in 15 of 18 (83%) culture-positive samples. The remaining 3 culture-positive samples grew *Escherichia coli* (NSEQ011), *Propionibacterium acnes* (NSEQ003), and *Staphylococcus epidermidis* (NSEQ014) respectively and were detected but did not constitute the most abundant read by WGS Illumina as the predominant sequences were similar to the sequences in the negative controls.

## 4.4 Discussion

This is the first study to evaluate the utility of full length 16S and whole genome nanopore sequencing for the identification of pathogen in patients with endophthalmitis. In this pilot study, saponification was able to reduce the proportion of host DNA, without inadvertent loss in bacterial DNA, thereby improving the sensitivity of identifying the pathogen. Incorporation of the negative 'no-template' control into each saponification, DNA extraction and sequencing run was useful in the limiting over-interpretation of the sequencing results. In two cases of culture-negative endophthalmitis, full-length 16S rRNA nanopore sequencing was able to detect the potential causative pathogen and the results were confirmed by Illumina sequencing. Finally, whole genome nanopore sequencing performed better than full-length 16S rRNA nanopore sequencing in detecting the predominant organism in endophthalmitis. Taken together, these results suggest that WGS nanopore sequencing may provide useful information on the pathogens associated with endophthalmitis.

Charalampous and associates have optimised a saponin-based host DNA depletion protocol for nanopore sequencing on 41 clinical respiratory samples of patients with suspected lower respiratory tract infections and reported up to 10,000 fold depletion or 99.9% of human DNA[130]. The respiratory samples in study conducted by Charalampous and associates were derived from sputum, endotracheal secretions and bronchoalveolar lavage, and therefore, had larger starting volumes of 400μl. In contrast, the starting volumes from the intraocular fluid biopsy samples in this study were much smaller, of around 30μl. The performance of the saponin-based host depletion in this study was modest, with up to 70% reduction in human DNA, and in one case (NSEQ005), the pathogen was detectable in the saponified sample but not in the un-saponified sample. However, there was an increase in the background contamination with *Mesorhizobium* spp. and *Sinorhizobium* spp. (bacteria commonly found in soil and aquatic legume roots[233]) in negative 'no-template' control and clinical samples that have undergone saponification. Saponins are foaming agents, commonly derived from plant extracts[234]. This suggests that the use of additional reagents in the saponin-based host DNA depletion may have paradoxically increased the exogenous background contamination in those samples. Furthermore, saponin-based host DNA depletion may limit the sensitivity for detecting nucleic acid from viruses, microorganisms without cell walls (e.g. parasites or *Mycoplasma* spp.), or free nucleic acid (i.e. cell-free DNA) from dead organisms that have either been cleared by the host immune system or through antimicrobial treatments[157]. A potential solution to improve the detection of viral and other cell-free nucleic acid in ocular clinical samples could be by retaining the supernatant fraction from the initial differential centrifugation step[130]. Another feasible hybrid option could be combining both targeted and untargeted sequencing, by spiking in variably sized primers to

enrich for a specific organism of interest whilst simultaneously sequencing other untargeted organisms to retain broad metagenomic sensitivity[157,235].

Other methods of host DNA enrichment/depletion include cell-free DNA extraction by differential centrifugation at 16,000g, selective host DNA degradation and lysis by enzymes such as benzonase (QIAamp DNA Microbiome kit), MolDNaseB (Molzym Ultra-Deep Microbiome Prep) or Proteinase K (HostZERO Microbial DNA kit), immunoprecipitation and magnetic bead capture of methylated eukaryotic host DNA (NEBNext Microbiome DNA Enrichment) or osmotic lysis of host cells and subsequent use of propidium monoazide for covalent attachment to extracellular DNA[158,236,237]. However, there are varying efficiencies across these host DNA removal methods dependent upon sample type[237]. For example, selective enzymatic degradation and osmotic lysis was shown to be effective for clinical samples from nasal aspirates, but not for milk substrates due to high fat and protein content[130,237,238]. Therefore, it is imperative to tailor the host depletion method to the desired clinical sample.

In dealing with samples of low biomass, which are inherent in ocular samples, the presence of contaminating DNA, either from the time of sample collection, handling, storage, laboratory reagents and kits, can generate misleading false positive results[239]. In this study, the composition of the contaminating sequences varied between the batches of DNA extraction and sequencing runs. Consistent with previous studies, the contaminating sequences were mainly derived from water- and soil-associated bacteria, such as *Acinetobacter*, *Pseudomonas*, *Stenotrophomonas*, *Bradyrhizobium*, *Burkholderia* and *Ralstonia*, and human-skin associated organisms, such as *Corynebacterium*, and

*Propionibacterium* [239,240]. Salter and associates performed 16S rRNA gene sequencing and shotgun metagenomic sequencing on a dilutional series of *Salmonella bongori* (acting as positive control) along with concomitant blank 'no template' negative control extracted with four commercially available DNA extraction kits at three different sites. Their study demonstrated that the profile of contaminating sequences varied greatly between different sites, extraction kits and even kit batches. Samples with low biomass, of less than $10^4$ *Salmonella bongori* cells, were critically impacted, as the predominant sequencing reads were mainly derived from the background contamination[239]. As the negative controls were processed concomitantly with the clinical samples, the sequencing results or "hits" were compared against the profile of the contaminating sequences to ensure that the sequencing results were not driven by over-amplified background DNA contamination.

The concordance of whole genome nanopore sequencing results with cultured organism was excellent, with 100% agreement (15/15). There was good agreement between culture results and 16S nanopore (13/15, 87%) and Illumina whole genome sequencing (15/18, 83%). Identification of organisms using targeted 16S sequencing is restricted by the primer sequence used and quantification can be challenging given the highly variable per-genome copy number of the 16S rRNA gene[144,241]. Whole genome sequencing provides better bacterial taxonomic resolution to the level of species- or strain compared to 16S sequencing, which is usually limited to genus level classifications[154]. Furthermore, by sequencing the entire community of genomes, the sequences of all organisms including viruses and fungi, could be captured, along with their functional genomic capacity and phylogeny. To illustrate, in this study, the bacteriophage *Staphylococcus phietavirus* was identified in two cases of *Staphylococcus* endophthalmitis through whole genome sequencing but not on 16S amplicon

sequencing. In terms of taxa classification, Pearman and associates reported that longer reads improved the accuracy in taxa classification compared to short reads, albeit being having higher sequencing error rates[242]. This might explain the higher agreement rates between culture and WGS nanopore compared to WGS Illumina seen in this study.

Two of the five culture-negative samples in this study revealed potential organisms on molecular sequencing. The 16S nanopore sequencing results were in agreement with Illumina short-read sequencing suggesting an identification of a putative organism in the context of a false negative culture. The cause of endophthalmitis in the three culture- and sequencing-negative cases is unknown and could either be due to clearance of the organism by the host immune system prior to biopsy, extra-ocular source of infection not present in the aqueous or vitreous humour, or pathogen undetected due to potential bias in the study laboratory or bioinformatics workflow such as RNA virus or parasite[107,130].

Molecular sequencing technology has the potential in advancing diagnostics for ocular infections[106,107,117–119]. Using metagenomic deep sequencing, Thuy and associates were able to detect the presence of RV virus (*Rubivirus* genus) in the intraocular fluid sample of a patient with a 16-year history of idiopathic uveitis. By performing phylogenetic analysis of the genomic sequence in comparison to other RV strains deposited in public repositories (GenBank), along with estimates of the nucleotide substitution rate, they were able to approximate the time and place of when the patient might have been exposed to the virus[106]. Seminal work by Lee and associates, on 50 patients (24 culture-positive and 26 culture-negative) enrolled prospectively, demonstrated that detection of pathogens and their bacterial load by Illumina whole genome sequencing and directed PCR had prognostic

significance for clinical outcomes in post-procedural endophthalmitis[107]. In comparison to the large and expensive Illumina sequencing platforms, the nanopore MINion sequencer is pocket-sized, weighing less than 450g, costs less than £1000 and is able to provide rapid sequencing results[228]. In the present study, Nanopore WGS was able to reliably detect the pathogen and its bacterial load directly from intraocular fluid biopsy samples. This suggests that with further optimisation, the future use of the portable, real-time nanopore sequencing technology as a point-of-care testing for pathogen identification and quantification may be beneficial in the initial management of endophthalmitis[107].

There were several limitations to this study. First, there was limited sample size and small amounts of ocular samples available for molecular sequencing. The present study included ocular samples from the USA cohort that have previously been processed for Illumina WGS and other molecular testing, therefore in certain cases there was insufficient DNA leftover to perform both 16S and WGS nanopore sequencing. Additionally, there could be potential impact from the shipping process as frozen samples from the UK cohort thawed during shipment, and impact from the different DNA extraction kits used for both cohorts. However, the nanopore and Illumina sequencing protocol were completed in a single centre to minimise any potential sequencing bias.

In summary, this study demonstrated the utility of nanopore sequencing in identifying potential pathogens and its bacterial load in endophthalmitis. Further optimisation through the use of internal spike-in controls to determine the limit of detection, single-use flow-cell (Nanopore Flongle) to reduce cost and minimise risk of contamination between samples, miniaturisation of laboratory steps through microfluidic devices, and automation of

bioinformatic workflow may facilitate the adoption of nanopore sequencing as a point-of-care testing in a clinical environment[157]. The combination of transcriptomic analyses of host immune responses and pathogenic virulence may elucidate the pathophysiologic mechanisms underpinning poor clinical outcomes and provide personalised, targeted treatment in endophthalmitis.

# Chapter 5

# Conclusions

dissect the host-microbiome interact traditionally rely on animal models, typically involving germ-free or specific pathogen free mice, however these models have limited human translational or reproducible capabilities as the microbiota of laboratory mice are divergent from 'wild' mice and humans, the 'normal' mucosal immunity requires priming and stimulation by the gut microbiome and is further compounded by wide inter- and intra-individual gut microbial heterogeneity[18,187,243]. Therefore, metagenomic temporal and spatial surveys of the effects of disrupting 'take-away' and reconstituting 'add-back' the immune system would be highly informative, as exemplified in a seminal study conducted by Schluter and associates, whereby tracking day-by-day changes in the peripheral immune cell counts after auto-faecal microbiota transplant following allogeneic haematopoietic cell transplant, showed higher relative abundances of *Akkermansia*, *Ruminococcus*, and *Faecalibacterium* were associated with increased neutrophil rates[244]. There is also emerging evidence suggesting that the gut microbiome modulates the host response to immunotherapies and mediates immunotherapy-related toxicity, specifically immune checkpoint inhibitors in cancer treatments[185,189,245]. Manipulation of the gut microbiome could be a pathway to improve the efficacy, prolong remission and reduce the toxicity of immunosuppressive treatments in patients with inflammatory eye disease.

In addition to gut microbiome, the ocular surface microbiome is also thought to modulate the immune responses in ocular inflammatory diseases and could potentially serve as reservoir for opportunistic and recurrent infections in microbial keratitis[5,37,246]. However, approximately 40% of clinically suspected microbial keratitis are culture-negative either due to poor sampling techniques, difficulty in culturing fastidious microorganisms, inactivation or clearance of microorganisms by the host immune system or prior use of antibiotics[82,85,107].

Therefore, culture-independent techniques using molecular sequencing of the microbial DNA might be useful in the rapid identification of pathogens in microbial keratitis[82]. Sequencing of ocular samples are technically challenging, owing to the high proportion of background host DNA compared to microbial genome. The overcome this background 'noise', the sample would need to be sequenced deeply, which could be cost prohibitive[105]. Therefore, an alternative cost-effective method would be to perform targeted sequencing of the bacterial 16S rRNA genome[247]. In the second results chapter, sample collection method of the ocular surface was optimised on an *ex vivo* porcine model of microbial keratitis. Using a defined mock microbial community, a full-length nanopore 16S rRNA sequencing bioinformatics pipeline was evaluated and benchmarked. Applying the sampling technique that provided the highest DNA recovery and optimised bioinformatics pipeline directly to samples from patients with suspected microbial keratitis, the bacterial pathogens were detectable by full-length nanopore 16S rRNA sequencing.

Metagenomic sequencing using a untargeted approach allows for the discovery of novel pathogens. In the third results chapter, intraocular fluid biopsy samples of endophthalmitis were optimised (host depletion by saponification) for whole genome and 16S nanopore sequencing. Results showed that metagenomic nanopore sequencing was able to detect potential pathogens directly from clinical ocular samples and performed better than culture. Removal of host cell contamination by saponification could also be utilised for metagenomic sequencing workflow in microbial keratitis. Corneal and conjunctival samples are thought to compose of larger proportion of host cells compared to microbial cells compared to vitreous or aqueous samples, hence steps to deplete host cells would be of utmost importance to improve the sensitivity of sequencing in detecting causative pathogens in these infections[82].

Another approach would be utilising the host sequencing data to establish host-specific responses to different microbial infections through differential gene expressions[117]. With optimisation of the sequencing and bioinformatics workflow, future large prospective studies comparing whole genome nanopore and Illumina sequencing with culture and 16S sequencing results in microbial keratitis could be undertaken.

Gu and associates have shown that cell-free metagenomic DNA nanopore sequencing coupled with real-time computational analyses enabled pathogen identification in body fluids (pleural, peritoneal, abscess, cerebrospinal fluid, joint aspirate, bronchial alveolar lavage and urine) within 6 hours[196]. The portability of the nanopore sequencing platform, MinION, has been demonstrated through in-field studies of Ebola virus in Guinea, Zika virus in Brazil, and have been utilised for antimicrobial resistance gene surveillance in Malawi and Catalan[126,128,248,249]. Recent advances in the nanopore platform has enabled native RNA sequencing, without the need for prior cDNA generation, and has revealed novel insights to viral pathogen transcriptomes that were not previously detectable by recoded short-read RNA (Illumina) sequencing alone[250]. Incorporating both the host and microbial gene expression transcriptomic and proteomic data would be useful in identifying pathogen-specific host responses and discriminating between non-infectious vs infectious ocular inflammation[117,251]. Taken together, nanopore sequencing offers a rapid, accurate and portable technology for identifying microbial species and their antimicrobial resistance genes in ocular infectious and inflammatory diseases.

At present, there are still limitations of interpreting sequencing results from clinical samples, including difficulty in distinguishing contamination from true or polymicrobial infections, quality control across multi-sequencing sites and standardisation of bioinformatics analysis.

Overcoming these challenges will be essential before sequencing is adopted in routine clinical practice.

## 5.1 Future directions

Integrating clinical metagenomics into routine clinical care will revolutionise the delivery of personalised medicine. Figure 5.1 summarises the potential implementation of clinical metagenomic workflow in the management of an unwell patient.

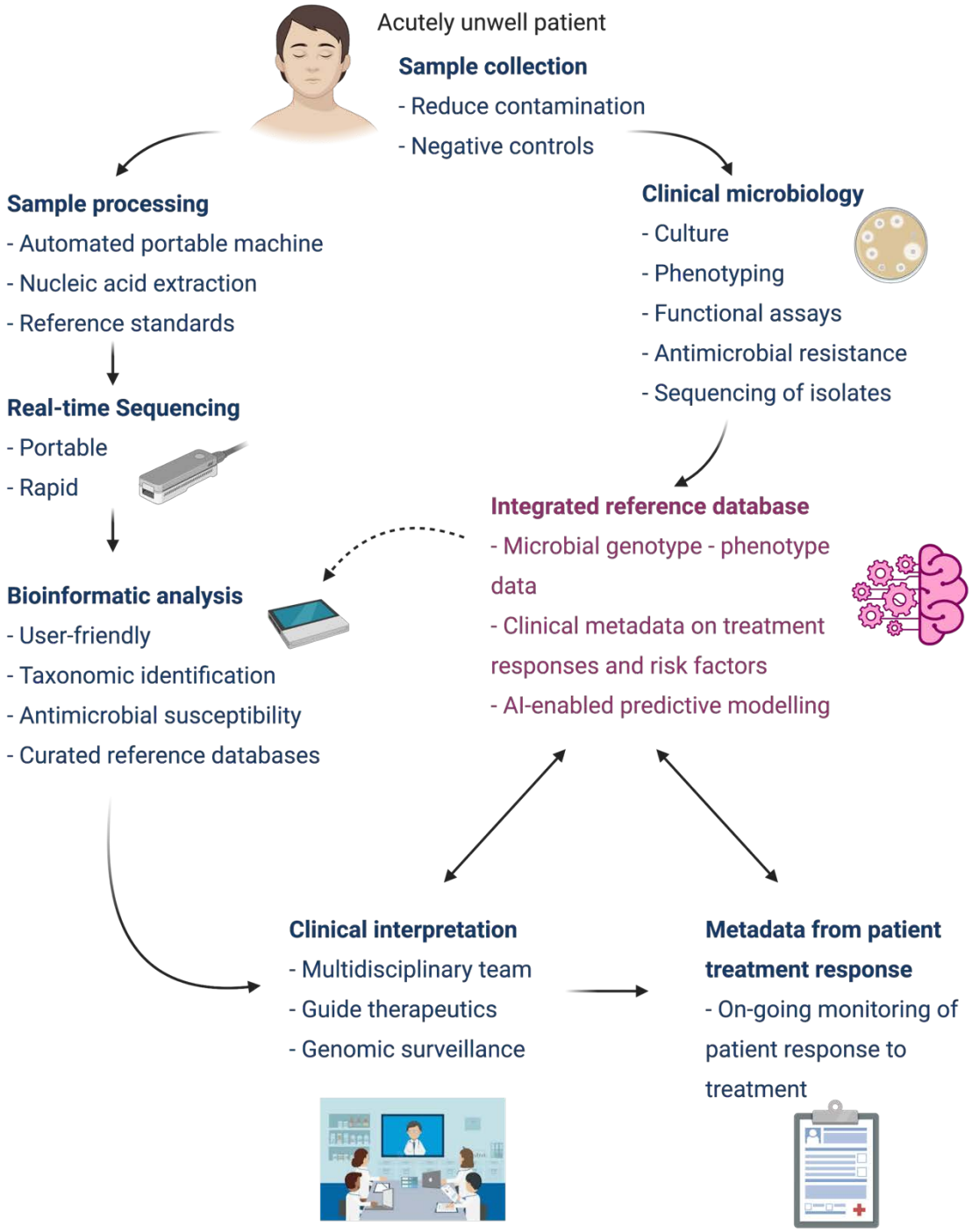



*Figure 5.1 Potential metagenomic workflow in the management of an unwell patient in the clinical setting.*

When a patient presents with an acute illness, clinical samples will be collected, with stringent measures to reduce contamination. Samples from the surrounding environment or adjacent body site, such as the skin around the eye for a case of potential endophthalmitis, will be

collected and processed as negative controls. An aliquot of the clinical sample will sent for parallel routine microbiology culture to test for phenotypic – genotypic correlations.

With the remaining aliquot, nucleic acid will be extracted and library preparation performed using standardised protocol on automated microfluidic devices that are portable and requires the least hands-on time. Reference standards for positive and negative controls will be incorporated, along with spike-in material to quantify microbial load on a standard curve, and to control for any assay inhibitors.

Sequencing on a portable nanopore sequencer using smaller, single-use flow-cells (i.e. one flow cell per patient) will enable the generation of real-time sequencing data that is cost-effective without the need to batch samples from multiple patients for sequencing. Although still under development, native RNA nanopore sequencing of both microbial and host transcriptomic data may enable detection of potential 'live' microbial nucleic acid along with host responses. For samples with low bio-mass additional microbial enrichment, host depletion, and targeted sequencing steps may be necessary.

Real-time bioinformatic analyses conducted in a user-friendly graphical interface software on a curated and updated reference database, supported by secure computational infrastructure either local or remote 'cloud' network, will enable actionable metagenomic data (microbial taxonomic identification and predicted antimicrobial genetic susceptibility) to be reported quickly and directly to the clinical team. A multidisciplinary team discussion involving the treating physician, clinical microbiologist, and bioinformatician will then interpret the sequencing results to guide therapeutic management in the context of the individual patient

case. Anonymised metadata on the patient's risk factors, and response to treatment will then feedback to a centralised database.

As for the clinical sample aliquot that was sent for culture and microbiology testing, improvements in culture techniques will enable isolation of the causative pathogen and further functional and antimicrobial susceptibility tests to be performed in vitro. Nucleic acid from the isolated pathogen will then be extracted and sequenced in dedicated sequencing hubs. Genomic data correlated with information on the microbial phenotype, will be deposited in curated microbial reference databases.

Longitudinal host-microbial genomic studies complemented with clinical data on patient response to therapy is currently lacking. The true power of clinical metagenomics will come from integrating curated microbial genomic-phenotypic data with high-quality metadata on patient clinical outcomes, similar to the Cancer Genome Atlas initiative[252], and provide the necessary training dataset for future AI predictive modelling to prognosticate and guide therapeutic interventions. Additionally, real-time sequencing and clinical data will facilitate genomic surveillance in the event of an outbreak within a geographical area, similar to national initiative developed by the COVID-19 Genomics UK consortium[253].

Advances in user-friendly hand-held sequencing technologies complemented with portable communications technologies, such as the smartphone, drive the emergence of mobile healthcare, whereby self-testing at the convenience of the patient's own home would be a possibility in the future[254]. To illustrate, a patient obtains a swab, performs nucleic acid extraction, library preparation and sequencing on a portable, automated, hand-held

microfluidic and sequencing device in their own home. Data uploads directly onto a remote computational network for real-time analysis and informs the clinical team who will run a parallel virtual clinical consultation with the patient. Based on the predictive modelling and clinical indication, clinicians can make an informed clinical judgement on whether the patient should admitted to the hospital or could be treated at home with the appropriate antimicrobial therapies. On-going monitoring continues for the patient until resolution of infection. This integration of multi-omic data along with real-time clinical data provides vital information for longitudinal genomic surveillance and for discovery of biomarkers indicative of early disease, whereby subclinical inflammation and infection can be detected before onset of symptoms. Adoption of these new technologies into routine clinical care has the potential to accelerate precision, personalised medicine for all[255].

# Bibliography


1. Kels BD, Grzybowski A, Grant-Kels JM. Human ocular anatomy. *Clin Dermatol*. 2015;33(2):140-146. doi:10.1016/j.clindermatol.2014.10.006
2. Purves D, Augustine GJ, Fitzpatrick D, Katz LC, LaMantia A-S, McNamara JO, Williams SM. Anatomy of the Eye. Published online 2001. Accessed December 13, 2020. https://www.ncbi.nlm.nih.gov/books/NBK11120/
3. Anatomy of the Eye - Neuroscience - NCBI Bookshelf. Accessed December 13, 2020. https://www.ncbi.nlm.nih.gov/books/NBK11120/
4. Gipson IK. The ocular surface: The challenge to enable and protect vision. The Friedenwald lecture. In: *Investigative Ophthalmology and Visual Science*. Vol 48. NIH Public Access; 2007:4391-4398. doi:10.1167/iovs.07-0770
5. St. Leger AJ, Desai J V., Drummond RA, Kugadas A, Almaghrabi F, Silver P, Raychaudhuri K, Gadjeva M, Iwakura Y, Lionakis MS, Caspi RR. An Ocular Commensal Protects against Corneal Infection by Driving an Interleukin-17 Response from Mucosal γδ T Cells. *Immunity*. 2017;47(1):148-158.e5. doi:10.1016/j.immuni.2017.06.014
6. Akpek EK, Gottsch JD. Immune of defense at the ocular surface. *Eye*. Published online 2003. doi:10.1038/sj.eye.6700617
7. Forrester J V., Kuffova L, Dick AD. Autoimmunity, Autoinflammation, and Infection in Uveitis. *Am J Ophthalmol*. 2018;189:77-85. doi:10.1016/j.ajo.2018.02.019
8. Wilbanks GA, Streilein JW. Characterization of suppressor cells in anterior chamber-associated immune deviation (ACAID) induced by soluble antigen. Evidence of two functionally and phenotypically distinct T-suppressor cell populations. *Immunology*. 1990;71(3):383-389. Accessed December 31, 2020. /pmc/articles/PMC1384437/?report=abstract
9. Spadoni I, Fornasa G, Rescigno M. Organ-specific protection mediated by cooperation between vascular and epithelial barriers. *Nat Rev Immunol*. 2017;17(12):761-773. doi:10.1038/nri.2017.100
10. Niederkorn JY, Stein-Streilein J. History and physiology of immune privilege. *Ocul Immunol Inflamm*. 2010;18(1):19-23. doi:10.3109/09273940903564766
11. Forrester J V., Xu H. Good news-bad news: The Yin and Yang of immune privilege in the eye. *Front Immunol*. 2012;3(NOV):338. doi:10.3389/fimmu.2012.00338
12. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. *Nature*. Published online 2007. doi:10.1038/nature06244
13. Belkaid Y, Harrison OJ. Homeostatic Immunity and the Microbiota. *Immunity*. 2017;46(4):562-576. doi:10.1016/j.immuni.2017.04.008
14. Zelante T, Iannitti RG, Cunha C, DeLuca A, Giovannini G, Pieraccini G, Zecchi R, D'Angelo C, Massi-Benedetti C, Fallarino F, Carvalho A, Puccetti P, Romani L. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity*. Published online 2013. doi:10.1016/j.immuni.2013.08.003
15. Chang MC, Tsai YL, Chen YW, Chan CP, Huang CF, Lan WC, Lin CC, Lan WH, Jeng JH. Butyrate induces reactive oxygen species production and affects cell cycle progression in human gingival fibroblasts. *J Periodontal Res*. Published online 2013.

doi:10.1111/j.1600-0765.2012.01504.x
16. Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee CA, Lynch S V., Tanoue T, Imaoka A, Itoh K, Takeda K, Umesaki Y, Honda K, Littman DR. Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria. *Cell*. Published online 2009. doi:10.1016/j.cell.2009.09.033
17. Mazmanian SK, Cui HL, Tzianabos AO, Kasper DL. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*. Published online 2005. doi:10.1016/j.cell.2005.05.007
18. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. *Cell Res*. Published online 2020. doi:10.1038/s41422-020-0332-7
19. Caruso R, Warner N, Inohara N, Nú N ˜ Ez G. Immunity Review NOD1 and NOD2: Signaling, Host Defense, and Inflammatory Disease. Published online 2014. doi:10.1016/j.immuni.2014.12.010
20. Yang J, Yan H. TLR5: Beyond the recognition of flagellin. *Cell Mol Immunol*. 2017;14(12):1017-1019. doi:10.1038/cmi.2017.122
21. Ghimire L, Paudel S, Jin L, Jeyaseelan S. The NLRP6 inflammasome in health and disease. *Mucosal Immunol*. 2020;13(3):388-398. doi:10.1038/s41385-020-0256-z
22. Levy M, Kolodziejczyk AA, Thaiss CA, Elinav E. Dysbiosis and the immune system. *Nat Rev Immunol*. Published online 2017. doi:10.1038/nri.2017.7
23. Schroeder BO, Bäckhed F. Signals from the gut microbiota to distant organs in physiology and disease. *Nat Med*. Published online 2016. doi:10.1038/nm.4185
24. Horai R, Caspi RR. Microbiome and autoimmune uveitis. *Front Immunol*. 2019;10(FEB):232. doi:10.3389/fimmu.2019.00232
25. St. Leger AJ, Caspi RR. Visions of Eye Commensals: The Known and the Unknown About How the Microbiome Affects Eye Disease. *BioEssays*. 2018;40(11):e1800046. doi:10.1002/bies.201800046
26. Horai R, Zárate-Bladés CR, Dillenburg-Pilla P, Chen J, Kielczewski JL, Silver PB, Jittayasothorn Y, Chan CC, Yamane H, Honda K, Caspi RR. Microbiota-Dependent Activation of an Autoreactive T Cell Receptor Provokes Autoimmunity in an Immunologically Privileged Site. *Immunity*. 2015;43(2):343-353. doi:10.1016/j.immuni.2015.07.014
27. Sen H, Chaigne-Delalande B, Li Z, Lagishetty V, Jacobs J, Braun J. Gut Microbiome in Uveitis . *Investig Ophthalmol Vis Sci*. 2017;58(8):846. Accessed December 20, 2020. https://iovs.arvojournals.org/article.aspx?articleid=2639043
28. Ye Z, Wu C, Zhang N, Du L, Cao Q, Huang X, Tang J, Wang Q, Li F, Zhou C, Xu Q, Xiong X, Kijlstra A, Qin N, Yang P. Altered gut microbiome composition in patients with Vogt-Koyanagi-Harada disease. *Gut Microbes*. Published online 2020. doi:10.1080/19490976.2019.1700754
29. Andriessen EM, Wilson AM, Mawambo G, Dejda A, Miloudi K, Sennlaub F, Sapieha P. Gut microbiota influences pathological angiogenesis in obesity-driven choroidal neovascularization. *EMBO Mol Med*. Published online 2016. doi:10.15252/emmm.201606531
30. Zinkernagel MS, Zysset-Burri DC, Keller I, Berger LE, Leichtle AB, Largiadèr CR, Fiedler GM, Wolf S. Association of the intestinal microbiome with the development of neovascular age-related macular degeneration. *Sci Rep*. Published online 2017. doi:10.1038/srep40826
31. Astafurov K, Elhawy E, Ren L, Dong CQ, Igboin C, Hyman L, Griffen A, Mittag T, Danias

J. Oral microbiome link to neurodegeneration in glaucoma. *PLoS One*. 2014;9(9). doi:10.1371/journal.pone.0104416
32. Kugadas A, Wright Q, Geddes-McAlister J, Gadjeva M. Role of microbiota in strengthening ocular mucosal barrier function through secretory IgA. *Investig Ophthalmol Vis Sci*. 2017;58(11):4593-4600. doi:10.1167/iovs.17-22119
33. Zegans ME, Van Gelder RN. Considerations in understanding the ocular surface microbiome. *Am J Ophthalmol*. Published online 2014. doi:10.1016/j.ajo.2014.06.014
34. Perkins RE, Kundsin RB, Pratt M V. Bacteriology of normal and infected conjunctiva. *J Clin Microbiol*. 1975;1(2):147-149. doi:10.1128/jcm.1.2.147-149.1975
35. Ozkan J, Nielsen S, Diez-Vives C, Coroneo M, Thomas T, Willcox M. Temporal Stability and Composition of the Ocular Surface Microbiome. *Sci Rep*. Published online 2017. doi:10.1038/s41598-017-10494-9
36. Doan T, Akileswaran L, Andersen D, Johnson B, Ko N, Shrestha A, Shestopalov V, Lee CS, Lee AY, Van Gelder RN. Paucibacterial microbiome and resident DNA virome of the healthy conjunctiva. *Investig Ophthalmol Vis Sci*. 2016;57(13):5116-5126. doi:10.1167/iovs.16-19803
37. De Paiva CS, Jones DB, Stern ME, Bian F, Moore QL, Corbiere S, Streckfus CF, Hutchinson DS, Ajami NJ, Petrosino JF, Pflugfelder SC. Altered Mucosal Microbiome Diversity and Disease Severity in Sjögren Syndrome. *Sci Rep*. 2016;6. doi:10.1038/srep23561
38. Kugadas A, Christiansen SH, Sankaranarayanan S, Surana NK, Gauguet S, Kunz R, Fichorova R, Vorup-Jensen T, Gadjeva M. Impact of Microbiota on Resistance to Ocular Pseudomonas aeruginosa-Induced Keratitis. *PLoS Pathog*. 2016;12(9). doi:10.1371/journal.ppat.1005855
39. Zaidi T, Zaidi T, Cywes-Bentley C, Lu R, Priebe GP, Pier GB. Microbiota-driven immune cellular maturation is essential for Antibody-mediated adaptive immunity to Staphylococcus aureus Infection in the eye. *Infect Immun*. 2014;82(8):3483-3491. doi:10.1128/IAI.01951-14
40. Graham JE, Moore JEJE, Jiru X, Moore JEJE, Goodall EA, Dooley JSG, Hayes VEA, Dartt DA, Downes CS, Moore TCB. Ocular pathogen or commensal: A PCR-based study of surface bacterial flora in normal and dry eyes. *Investig Ophthalmol Vis Sci*. 2007;48(12):5616-5623. doi:10.1167/iovs.07-0588
41. Shimizu E, Ogawa Y, Saijo Y, Yamane M, Uchino M, Kamoi M, Fukui M, Yang F, He J, Mukai S, Tsubota K. Commensal microflora in human conjunctiva; characteristics of microflora in the patients with chronic ocular graft-versus-host disease. *Ocul Surf*. 2019;17(2):265-271. doi:10.1016/j.jtos.2019.02.001
42. Dong Q, Brulc JM, Iovieno A, Bates B, Garoutte A, Miller D, Revanna K V., Gao X, Antonopoulos DA, Slepak VZ, Shestopalov VI. Diversity of bacteria at healthy human conjunctiva. *Investig Ophthalmol Vis Sci*. 2011;52(8):5408-5413. doi:10.1167/iovs.10-6939
43. Lee SH, Oh DH, Jung JY, Kim JC, Jeon CO. Comparative ocular microbial communities in humans with and without blepharitis. *Investig Ophthalmol Vis Sci*. 2012;53(9):5585-5593. doi:10.1167/iovs.12-9922
44. Wen X, Miao L, Deng Y, Bible PW, Hu X, Zou Y, Liu Y, Guo S, Liang J, Chen T, Peng GH, Chen W, Liang L, Wei L. The influence of age and sex on ocular surface microbiota in healthy adults. *Investig Ophthalmol Vis Sci*. 2017;58(14):6030-6037. doi:10.1167/iovs.17-22957

45. Williams GP, Radford C, Nightingale P, Dart JKG, Rauz S. Evaluation of early and late presentation of patients with ocular mucous membrane pemphigoid to two major tertiary referral hospitals in the United Kingdom. *Eye*. 2011;25(9):1207-1218. doi:10.1038/eye.2011.175
46. Dart JK. The 2016 Bowman Lecture Conjunctival curses: Scarring conjunctivitis 30 years on. *Eye*. 2017;31(2):301-332. doi:10.1038/eye.2016.284
47. Chan LS, Razzaque Ahmed A, Anhalt GJ, Bernauer W, Cooper KD, Elder MJ, Fine JD, Stephen Foster C, Ghohestani R, Hashimoto T, Hoang-Xuan T, Kirtschig G, Korman NJ, Lightman S, Lozada-Nur F, Peter Marinkovich M, Mondino BJ, Prost-Squarcioni C, Rogers RS, Setterfield JF, West DP, Wojnarowska F, Woodley DT, Yancey KB, Zillikens D, Zone JJ. The first international consensus on mucous membrane pemphigoid: Definition, diagnostic criteria, pathogenic factors, medical treatment and prognostic indicators. *Arch Dermatol*. 2002;138(3):370-379. doi:10.1001/archderm.138.3.370
48. Radford CF, Rauz S, Williams GP, Saw VPJ, Dart JKG. Incidence, presenting features, and diagnosis of cicatrising conjunctivitis in the United Kingdom. *Eye*. 2012;26(9):1199-1208. doi:10.1038/eye.2012.119
49. Xu HH, Werth VP, Parisi E, Sollecito TP. Mucous Membrane Pemphigoid. *Dent Clin North Am*. 2013;57(4):611-630. doi:10.1016/j.cden.2013.07.003
50. Ong HS, Minassian D, Rauz S, Mehta JS, Dart JK. Validation of a clinical assessment tool for cicatrising conjunctivitis. *Ocul Surf*. 2020;18(1):121-129. doi:10.1016/j.jtos.2019.10.010
51. Black APB, Seneviratne SL, Jones L, King AS, Winsey S, Arsecularatne G, Wojnarowska F, Ogg GS. Rapid effector function of circulating NC16A-specific T cells in individuals with mucous membrane pemphigoid. *Br J Dermatol*. 2004;151(6):1160-1164. doi:10.1111/j.1365-2133.2004.06219.x
52. Oyama N, Setterfield JF, Powell AM, Sakuma-Oyama Y, Albert S, Bhogal BS, Vaughan RW, Kaneko F, Challacombe SJ, Black MM. Bullous pemphigoid antigen II (BP180) and its soluble extracellular domains are major autoantigens in mucous membrane pemphigoid: The pathogenic relevance to HLA class II alleles and disease severity. *Br J Dermatol*. Published online 2006. doi:10.1111/j.1365-2133.2005.06998.x
53. Domloge-Hultsch N, Gammon WR, Briggaman RA, Gil SG, Carter WG, Yancey KB. Epiligrin, the major human keratinocyte integrin ligand, is a target in both an acquired autoimmune and an inherited subepidermal blistering skin disease. *J Clin Invest*. Published online 1992. doi:10.1172/JCI116033
54. Lazarova Z, Salato VK, Lanschuetzer CM, Janson M, Fairley JA, Yancey KB. IgG anti-laminin-332 autoantibodies are present in a subset of patients with mucous membrane, but not bullous, pemphigoid. *J Am Acad Dermatol*. Published online 2008. doi:10.1016/j.jaad.2008.02.035
55. Rashid KA, Gürcan HM, Ahmed AR. Antigen specificity in subsets of mucous membrane pemphigoid. *J Invest Dermatol*. Published online 2006. doi:10.1038/sj.jid.5700465
56. Bhol KC, Goss L, Kumari S, Colon JE, Ahmed AR. Autoantibodies to human α6 integrin in patients with oral pemphigoid. *J Dent Res*. 2001;80(8):1711-1715. doi:10.1177/00220345010800080601
57. Cozzani E, di Zenzo G, Calabresi V, Carrozzo M, Burlando M, Longanesi L, Cerri A, Caproni M, Sera F, Antiga E, Quaglino P, Marzano A V., Parodi A. Autoantibody profile of a cohort of 78 Italian patients with mucous membrane pemphigoid: Correlation

between reactivity profile and clinical involvement. *Acta Derm Venereol*. 2016;96(6):768-773. doi:10.2340/00015555-2311
58. Schmidt E, Zillikens D. Pemphigoid diseases. *Lancet*. 2013;381(9863):320-332. doi:10.1016/S0140-6736(12)61140-4
59. Delgado JC, Turbay D, Yunis EJ, Yunis JJ, Morton ED, Bhol K, Norman R, Alper CA, Good RA, Ahmed R. A common major histocompatibility complex class II allele HLA-DQB1*0301 is present in clinical variants of pemphigoid. *Proc Natl Acad Sci U S A*. 1996;93(16):8569-8571. doi:10.1073/pnas.93.16.8569
60. Elhusseiny AM, Fakhari H, Bishehsari F, Kang K, Djalilian AR. A gut connection in mucous membrane pemphigoid: Insights into the role of the microbiome. *Ocul Surf*. 2019;17(4):615-616. doi:10.1016/j.jtos.2019.09.005
61. Bernauer W, Wright P, Dart JK, Leonard JN, Lightman S. The Conjunctiva in Acute and Chronic Mucous Membrane Pemphigoid: An Immunohistochemical Analysis. *Ophthalmology*. 1993;100(3):339-346. doi:10.1016/S0161-6420(93)31644-1
62. Williams GP, Nightingale P, Southworth S, Denniston AKO, Tomlins PJ, Turner S, Hamburger J, Bowman SJ, John Curnow S, Rauz S. Conjunctival neutrophils predict progressive scarring in ocular mucous membrane pemphigoid. *Investig Ophthalmol Vis Sci*. 2016;57(13):5457-5469. doi:10.1167/iovs.16-19247
63. Tesavibul N, Dorfman D, Sangwan VS, Christen W, Panayotis Z, Rojas B, Foster CS. Costimulatory molecules in ocular cicatricial pemphigoid. *Investig Ophthalmol Vis Sci*. Published online 1998.
64. Saw VPJ, Offiah I, Dart RJ, Galatowicz G, Dart JKG, Daniels JT, Calder VL. Conjunctival interleukin-13 expression in mucous membrane pemphigoid and functional effects of interleukin-13 on conjunctival fibroblasts in vitro. *Am J Pathol*. 2009;175(6):2406-2415. doi:10.2353/ajpath.2009.090579
65. Saw VPJ, Dart RJC, Galatowicz G, Daniels JT, Dart JKG, Calder VL. Tumor necrosis factor-α in ocular mucous membrane pemphigoid and its effect on conjunctival fibroblasts. *Investig Ophthalmol Vis Sci*. Published online 2009. doi:10.1167/iovs.08-3345
66. Razzaque MS, Ahmed BS, Foster CS, Ahmed AR. Effects of IL-4 on conjunctival fibroblasts: Possible role in ocular cicatricial pemphigoid. *Investig Ophthalmol Vis Sci*. 2003;44(8):3417-3423. doi:10.1167/iovs.02-1084
67. Razzaque MS, Foster CS, Ahmed AR. Role of collagen-binding heat shock protein 47 and transforming growth factor-β1 in conjunctival scarring in ocular cicatricial pemphigoid. *Investig Ophthalmol Vis Sci*. 2003;44(4):1616-1621. doi:10.1167/iovs.02-0644
68. Ahadome SD, Abraham DJ, Rayapureddi S, Saw VP, Saban DR, Calder VL, Norman JT, Ponticos M, Daniels JT, Dart JK. Aldehyde dehydrogenase inhibition blocks mucosal fibrosis in human and mouse ocular scarring. *JCI Insight*. Published online 2016. doi:10.1172/jci.insight.87001
69. Ahadome SD, Mathew R, Reyes NJ, Mettu PS, Cousins SW, Calder VL, Saban DR. Classical dendritic cells mediate fibrosis directly via the retinoic acid pathway in severe eye allergy. *JCI Insight*. Published online 2016. doi:10.1172/jci.insight.87012
70. Zaltas MM, Ahmed R, Foster CS. Association of HLA-DR4 with ocular cicatricial pemphigoid. *Curr Eye Res*. Published online 1989. doi:10.3109/02713688908995191
71. Setterfield J, Theron J, Vaughan RW, Welsh KI, Mallon E, Wojnarowska F, Challacombe SJ, Black MM. Mucous membrane pemphigoid: HLA-DQB1*0301 is

associated with all clinical sites of involvement and may be linked to antibasement membrane IgG production. *Br J Dermatol*. Published online 2001. doi:10.1046/j.1365-2133.2001.04380.x
72. Kamada N, Seo SU, Chen GY, Núñez G. Role of the gut microbiota in immunity and inflammatory disease. *Nat Rev Immunol*. 2013;13(5):321-335. doi:10.1038/nri3430
73. Russell JT, Roesch LFW, Ördberg M, Ilonen J, Atkinson MA, Schatz DA, Triplett EW, Ludvigsson J. Genetic risk for autoimmunity is associated with distinct changes in the human gut microbiome. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-11460-x
74. Ong HS, Setterfield JF, Minassian DC, Dart JK, Booth D, Reid E, Carnt N, Gugliemetti S, Shanmuganathan V, Watson M, Saw V, Wilkins M, McCudden V, Ahmad S, Calder V, Waseem N, Scott B, Bunce C, Gage W, Gleeson M, Lund V, Sandhu G. Mucous Membrane Pemphigoid with Ocular Involvement: The Clinical Phenotype and Its Relationship to Direct Immunofluorescence Findings. *Ophthalmology*. 2018;125(4):496-504. doi:10.1016/j.ophtha.2017.10.004
75. Di Caprio R, Lembo S, Di Costanzo L, Balato A, Monfrecola G. Anti-inflammatory properties of low and high doxycycline doses: An in vitro study. *Mediators Inflamm*. 2015;2015. doi:10.1155/2015/329418
76. Williams HC, Wojnarowska F, Kirtschig G, Mason J, Godec TR, Schmidt E, Chalmers JR, Childs M, Walton S, Harman K, Chapman A, Whitham D, Nunn AJ, Adams J, Akhras V, Anstey A, Barnard C, Bell H, Blackford S, Bröcker E, Carmichael A, Coelho RR, Craig F, Davies K, Ellis R, English J, Gläser R, Groves R, Günthert C, Hampton PJ, Hepburn N, Hügel R, Hussain K, Ingram J, Layton AM, Levell NJ, Lewis V, Malhomme H, Omerod A, Patel G, Rallan R, Ravenscroft J, Santander H, Steinbrink K, Sticherling M, Thomas C, Vatve M, van Beek N, Venning V, Veysey E, Wachsmuth R, Wahie S, Walker B, Walsh M, Wee J, Westmoreland M, Wong G, Ferguson A, Verpetinske I, Duarte-Williamson E, Antony F, Bower C, Gawkrodger D, Taghipour K, Dunnill MGS, Waters A, Bottomley W, Wright A, Sterling J, Azam A, Gibbs S, Luger T, Salvary I, Lovell C, Ilchyshyn A, Gibbon K, Nik M, Charles-Holmes R, Lavery AL. Doxycycline versus prednisolone as an initial treatment strategy for bullous pemphigoid: a pragmatic, non-inferiority, randomised controlled trial. *Lancet*. 2017;389(10079):1630-1638. doi:10.1016/S0140-6736(17)30560-3
77. Campolongo MG, Val M, Lupatelli M, Gandolfo S, Pentenero M. Oral Doxycycline as an alternative initial treatment for desquamative gingivitis. *Front Physiol*. 2019;10. doi:10.3389/conf.fphys.2019.27.00036
78. Rauz S, Maddison PG, Dart JKG. Evaluation of mucous membrane pemphigoid with ocular involvement in young patients. *Ophthalmology*. 2005;112(7):1268-1274. doi:10.1016/j.ophtha.2005.01.039
79. Whitcher JP, Srinivasan M, Upadhyay MP. Corneal blindness: A global perspective. *Bull World Health Organ*. Published online 2001. doi:10.1590/S0042-96862001000300009
80. Tuft S, Burton M. *Focus: Microbial Keratitis*.; 2013. Accessed December 17, 2020. https://www.rcophth.ac.uk/wp-content/uploads/2014/08/Focus-Autumn-2013.pdf
81. Moussa G, Hodson J, Gooch N, Virdee J, Penaloza C, Kigozi J, Rauz S. Calculating the economic burden of presumed microbial keratitis admissions at a tertiary referral centre in the UK. *Eye*. Published online December 7, 2020. doi:10.1038/s41433-020-01333-9
82. Ung L, Bispo PJM, Doan T, Van Gelder RN, Gilmore MS, Lietman T, Margolis TP,

Zegans ME, Lee CS, Chodosh J. Clinical metagenomics for infectious corneal ulcers: Rags to riches? *Ocul Surf*. 2020;18(1):1-12. doi:10.1016/j.jtos.2019.10.007
83. Al-Mujaini A, Al-Kharusi N, Thakral A, Wali UK. Bacterial keratitis: Perspective on epidemiology, Clinico-Pathogenesis, diagnosis and treatment. *Sultan Qaboos Univ Med J*. Published online 2009.
84. Ibrahim YW, Boase DL, Cree IA. Epidemiological characteristics, predisposing factors and microbiological profiles of infectious corneal ulcers: The Portsmouth corneal ulcer study. *Br J Ophthalmol*. Published online 2009. doi:10.1136/bjo.2008.151167
85. Ung L, Bispo PJM, Shanbhag SS, Gilmore MS, Chodosh J. The persistent dilemma of microbial keratitis: Global burden, diagnosis, and antimicrobial resistance. *Surv Ophthalmol*. 2019;64(3):255-271. doi:10.1016/j.survophthal.2018.12.003
86. Callegan MC, Engelbert M, Parke DW, Jett BD, Gilmore MS. Bacterial endophthalmitis: Epidemiology, therapeutics, and bacterium-host interactions. *Clin Microbiol Rev*. 2002;15(1):111-124. doi:10.1128/CMR.15.1.111-124.2002
87. Taube MA, Del Mar Cendra M, Elsahn A, Christodoulides M, Hossain P. Pattern recognition receptors in microbial keratitis. *Eye*. 2015;29(11):1399-1415. doi:10.1038/eye.2015.118
88. Kaye S, Choudhary A. Herpes simplex keratitis. *Prog Retin Eye Res*. 2006;25(4):355-380. doi:10.1016/j.preteyeres.2006.05.001
89. Austin A, Lietman T, Rose-Nussbaumer J. Update on the Management of Infectious Keratitis. Published online 2017. doi:10.1016/j.ophtha.2017.05.012
90. Relhan N, Forster RK, Flynn HW. Endophthalmitis: Then and Now. *Am J Ophthalmol*. 2018;187:xx-xxvii. doi:10.1016/j.ajo.2017.11.021
91. Moussa G, Bhatt H, Reekie I, Butt G, Ng A, Blanch R, Fusi-Rubiano W, Al-Ibrahim J, Quinlan M, Virdee J, Dawson S, Patel A, Low L, Mushtaq F, Nambiar S, Routledge C, Yang Y, Barry RJ, Burdon M, Barua A, De Silva I, Panthagani J, Qureshi M, Pherwani A, Sigona M, Morphis G, Rauz S. Using the West Midlands CONCERT to characterise regional incidence of acute-onset post cataract surgery endophthalmitis. *Eye*. Published online September 1, 2020:1-11. doi:10.1038/s41433-020-01158-6
92. Barry P, Gettinby G, Lees F, Peterson M, Revie C, Seal D, D'Ardis M, Bradley B, Grabner G, Egger SF, Ruckhofer J, Blanckaert J, Budo C, Galand A, Crommen J, Rakic J, Xhauflaire G, Tassignon MJ, Verbraeken H, de Donker R, Schmickler S, Bellucci R, Morselli S, Soldati S, Vigasio F, Bertelli M, Bottoli A, Cassamali M, Danieli F, Khuri S, Rosa L, Szaflik J, Izdebska J, Szaflik JP, Lobo C, Cunha-Vaz J, Mira J, Abreu A, Aguilar J, Arteaga V, Cordovés L, Díaz-Aleman VT, Abreu P, Alvarez-Marin J, Gil MA, Alió J, Teus M, Alvarez MT, Román JM, Kaynak S, Saatcý O, Durak Ý, Günenç Ü, Berk T, Söylev M, Öner H, Koçak N, Ergin M, Allen D, Phelan P, Steel D, Wood C, Cunningham C, Miller M, Moorthy R, Zafar A, Ionides A, Lake D, Thompson G, Jacob J, Byles D, Gibbon C, Kleinschmidt A, Ling R, Quinn A, Simcock P, Sturrock G, Kiel W, Mabey D, Spalton D, Kumar A, Rosen P, Patel CK, Salmon J, Callens E, Mathysen D, Molemans F, de Rassenfosse G, Schenkeveld C, Eckelmann A, Haselhoff A, Franczuk R, Carvalho L, Catarina A, Bataille L, Jiménez E, de Nicolás AM, Kuheylan H, Bovill S, Tate I, Cabral S, Withers T, Cox D, Yeates T, Dash J, Lindsell L, Turner R, Ferrer C, Koerner R, Reischl U, Behr A, Barrett S, Gardner S, Kennedy S, Ludgate J, Montan P, Wilhelmus K, Barrett S, Gardner S, Kennedy S, Ludgate J, Montan P, Wilhelmus K. Prophylaxis of postoperative endophthalmitis following cataract surgery: Results of the ESCRS multicenter study and identification of risk factors. *J Cataract Refract Surg*. Published

online 2007. doi:10.1016/j.jcrs.2007.02.032
93. Coleman AL. How Big Data Informs Us about Cataract Surgery: The LXXII Edward Jackson Memorial Lecture. *Am J Ophthalmol*. Published online 2015. doi:10.1016/j.ajo.2015.09.028
94. Diago T, McCannel CA, Bakri SJ, Pulido JS, Edwards AO, Pach JM. Infectious endophthalmitis after intravitreal injection of antiangiogenic agents. *Retina*. Published online 2009. doi:10.1097/IAE.0b013e31819d2591
95. Klein KS, Walsh MK, Hassan TS, Halperin LS, Castellarin AA, Roth D, Driscoll S, Prenner JL. Endophthalmitis After Anti-VEGF Injections. *Ophthalmology*. Published online 2009. doi:10.1016/j.ophtha.2009.02.031
96. Yannuzzi NA, Si N, Relhan N, Kuriyan AE, Albini TA, Berrocal AM, Davis JL, Smiddy WE, Townsend J, Miller D, Flynn HW. Endophthalmitis After Clear Corneal Cataract Surgery: Outcomes Over Two Decades. *Am J Ophthalmol*. 2017;174:155-159. doi:10.1016/j.ajo.2016.11.006
97. Kernt M, Kampik A. Endophthalmitis: Pathogenesis, clinical presentation, management, and perspectives. *Clin Ophthalmol*. 2010;4(1):121-135. doi:10.2147/opth.s6461
98. Moshfeghi AA, Rosenfeld PJ, Flynn HW, Schwartz SG, Davis JL, Murray TG, Smiddy WE, Berrocal AM, Dubovy SR, Lee WH, Albini TA, Lalwani GA, Kovach JL, Puliafito CA. Endophthalmitis after intravitreal anti-vascular endothelial growth factor antagonists: A six-year experience at a university referral center. *Retina*. 2011;31(4):662-668. doi:10.1097/IAE.0b013e31821067c4
99. Fileta JB, Scott IU, Flynn HW. Meta-analysis of infectious endophthalmitis after intravitreal injection of anti-vascular endothelial growth factor agents. *Ophthalmic Surg Lasers Imaging Retin*. Published online 2014. doi:10.3928/23258160-20140306-08
100. Cao H, Zhang L, Li L, Lo SK. Risk Factors for Acute Endophthalmitis following Cataract Surgery: A Systematic Review and Meta-Analysis. *PLoS One*. 2013;8(8). doi:10.1371/journal.pone.0071731
101. Stringham JD, Relhan N, Miller D, Flynn HW. Trends in fluoroquinolone nonsusceptibility among coagulase-negative Staphylococcus Isolates causing endophthalmitis, 1995-2016. *JAMA Ophthalmol*. 2017;135(7):814-815. doi:10.1001/jamaophthalmol.2017.1826
102. Dave VP, Pathengay A, Nishant K, Pappuru RR, Sharma S, Sharma P, Narayanan R, Jalali S, Mathai A, Das T. Clinical presentations, risk factors and outcomes of ceftazidime-resistant Gram-negative endophthalmitis. *Clin Exp Ophthalmol*. 2017;45(3):254-260. doi:10.1111/ceo.12833
103. Miller FC, Coburn PS, Huzzatul MM, LaGrow AL, Livingston E, Callegan MC. Targets of immunomodulation in bacterial endophthalmitis. *Prog Retin Eye Res*. 2019;73:100763. doi:10.1016/j.preteyeres.2019.05.004
104. Otri AM, Fares U, Al-Aqaba MA, Miri A, Faraj LA, Said DG, Maharajan S, Dua HS. Profile of sight-threatening infectious keratitis: A prospective study. *Acta Ophthalmol*. Published online 2013. doi:10.1111/j.1755-3768.2012.02489.x
105. Seitzman GD, Hinterwirth A, Zhong L, Cummings S, Chen C, Driver TH, Lee MD, Doan T. Metagenomic Deep Sequencing for the Diagnosis of Corneal and External Disease Infections. *Ophthalmology*. 2019;126(12):1724-1726. doi:10.1016/j.ophtha.2019.06.013

106. Doan T, Wilson MR, Crawford ED, Chow ED, Khan LM, Knopp KA, O'Donovan BD, Xia D, Hacker JK, Stewart JM, Gonzales JA, Acharya NR, DeRisi JL. Illuminating uveitis: Metagenomic deep sequencing identifies common and rare pathogens. *Genome Med*. Published online 2016. doi:10.1186/s13073-016-0344-6
107. Lee CS, Hong B, Kasi SK, Aderman C, Talcott KE, Adam MK, Yue B, Akileswaran L, Nakamichi K, Wu Y, Rezaei KA, Olmos de Koo LC, Chee YE, Lee AY, Garg SJ, Van Gelder RN. Prognostic Utility of Whole-Genome Sequencing and Polymerase Chain Reaction Tests of Ocular Fluids in Postprocedural Endophthalmitis. *Am J Ophthalmol*. Published online 2020. doi:10.1016/j.ajo.2020.03.008
108. Kaye S. Microbial keratitis and the selection of topical antimicrobials. *BMJ Open Ophthalmol*. Published online 2017. doi:10.1136/bmjophth-2017-000086
109. Asbell PA, Colby KA, Deng S, McDonnell P, Meisler DM, Raizman MB, Sheppard JD, Sahm DF. Ocular TRUST: Nationwide Antimicrobial Susceptibility Patterns in Ocular Isolates. *Am J Ophthalmol*. 2008;145(6):951-958.e1. doi:10.1016/j.ajo.2008.01.025
110. Asbell PA, Sanfilippo CM, Sahm DF, Decory HH. Trends in Antibiotic Resistance among Ocular Microorganisms in the United States from 2009 to 2018. *JAMA Ophthalmol*. 2020;138(5):439-450. doi:10.1001/jamaophthalmol.2020.0155
111. Kaye S, Tuft S, Neal T, Tole D, Leeming J, Figueiredo F, Armstrong M, McDonnell P, Tullo A, Parry C. Bacterial susceptibility to topical antimicrobials and clinical outcome in bacterial keratitis. *Investig Ophthalmol Vis Sci*. 2010;51(1):362-368. doi:10.1167/iovs.09-3933
112. Hilliam Y, Kaye S, Winstanley C. Pseudomonas aeruginosa and microbial keratitis. *J Med Microbiol*. Published online 2020. doi:10.1099/jmm.0.001110
113. Sugita S, Ogawa M, Shimizu N, Morio T, Ohguro N, Nakai K, Maruyama K, Nagata K, Takeda A, Usui Y, Sonoda KH, Takeuchi M, Mochizuki M. Use of a comprehensive polymerase chain reaction system for diagnosis of ocular infectious diseases. *Ophthalmology*. Published online 2013. doi:10.1016/j.ophtha.2013.02.020
114. Kim E, Chidambaram JD, Srinivasan M, Lalitha P, Wee D, Lietman TM, Whitcher JP, Van Gelder RN. Prospective Comparison of Microbial Culture and Polymerase Chain Reaction in the Diagnosis of Corneal Ulcer. *Am J Ophthalmol*. 2008;146(5):714-723.e1. doi:10.1016/j.ajo.2008.06.009
115. Dworkin LL, Gibler TM, Van Gelder RN. Real-time quantitative polymerase chain reaction diagnosis of infectious posterior uveitis. *Arch Ophthalmol*. 2002;120(11):1534-1539. doi:10.1001/archopht.120.11.1534
116. Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol*. 2015;13(12):787-794. doi:10.1038/nrmicro3565
117. Lalitha P, Seitzman GD, Kotecha R, Hinterwirth A, Chen C, Zhong L, Cummings S, Lebas E, Sahoo MK, Pinsky BA, Lietman TM, Doan T. Unbiased Pathogen Detection and Host Gene Profiling for Conjunctivitis. *Ophthalmology*. Published online 2019. doi:10.1016/j.ophtha.2019.03.039
118. Lee AY, Akileswaran L, Tibbetts MD, Garg SJ, Van Gelder RN. Identification of torque teno virus in culture-negative endophthalmitis by representational deep DNA sequencing. *Ophthalmology*. 2015;122(3):524-530. doi:10.1016/j.ophtha.2014.09.001
119. Kirstahler P, Bjerrum SS, Friis-Møller A, La Cour M, Aarestrup FM, Westh H, Pamp SJ. Genomics-Based Identification of Microorganisms in Human Ocular Body Fluid. *Sci Rep*. 2018;8(1):1-14. doi:10.1038/s41598-018-22416-4
120. Parekh M, Romano V, Franch A, Leon P, Birattari F, Borroni D, Kaye SB, Ponzin D,

Ahmad S, Ferrari S. Shotgun sequencing to determine corneal infection. *Am J Ophthalmol Case Reports*. 2020;19:100737. doi:10.1016/j.ajoc.2020.100737
121. De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND, Swann J, Wick R, Oun MA, Stubberfield E, Hoosdally SJ, Crook DW, Peto TEA, Sheppard AE, Bailey MJ, Read DS, Anjum MF, Sarah Walker A, Stoesser N. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genomics*. 2019;5(9). doi:10.1099/mgen.0.000294
122. 16S sequencing and analysis. Accessed July 21, 2020. https://nanoporetech.com/analyse/16s
123. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, Wain J, O'Grady J. *MinION Nanopore Sequencing Identifies the Position and Structure of a Bacterial Antibiotic Resistance Island*. Vol 33. Nature Publishing Group; 2015:296-300. doi:10.1038/nbt.3103
124. PacBio Sequel Systems - PacBio. Accessed September 8, 2020. https://www.pacb.com/products-and-services/sequel-system/
125. Workflow - PacBio. Accessed September 8, 2020. https://www.pacb.com/products-and-services/sequel-system/workflow/
126. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, Ouédraogo N, Afrough B, Bah A, Baum JHJ, Becker-Ziaja B, Boettcher JP, Cabeza-Cabrerizo M, Camino-Sánchez Á, Carter LL, Doerrbecker J, Enkirch T, Garciá-Dorival I, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui M, Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J, Ngabo D, Nitzsche K, Pallasch E, Patrono LV, Portmann J, Repits JG, Rickett NY, Sachse A, Singethan K, Vitoriano I, Yemanaberhan RL, Zekeng EG, Racine T, Bello A, Sall AA, Faye OO, Faye OO, Magassouba N, Williams CJC V., Amburgey V, Winona L, Davis E, Gerlach J, Washington F, Monteil V, Jourdain M, Bererd M, Camara AA, Somlare H, Camara AA, Gerard M, Bado G, Baillet B, Delaune D, Nebie KY, Diarra A, Savane Y, Pallawo RB, Gutierrez GJ, Milhano N, Roger I, Williams CJC V., Yattara F, Lewandowski K, Taylor J, Rachwal P, Turner DJ, Pollakis G, Hiscox JA, Matthews DA, O'Shea MK, Johnston AMD, Wilson D, Hutley E, Smit E, Di Caro A, Wolfel R, Stoecker K, Fleischmann E, Gabriel M, Weller SA, Koivogui L, Diallo B, Keita S, Rambaut A, Formenty P, Gunther S, Carroll MW. Real-time, portable genome sequencing for Ebola surveillance. *Nature*. 2016;530(7589):228-232. doi:10.1038/nature16996
127. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, De Jesus JG, Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC, Sabino EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017;12(6):1261-1266. doi:10.1038/nprot.2017.066
128. Faria NR, Quick J, Claro IM, Thézé J, De Jesus JG, Giovanetti M, Kraemer MUG, Hill SC, Black A, Da Costa AC, Franco LC, Silva SP, Wu CH, Raghwani J, Cauchemez S, Du Plessis L, Verotti MP, De Oliveira WK, Carmo EH, Coelho GE, Santelli ACFS, Vinhal LC, Henriques CM, Simpson JT, Loose M, Andersen KG, Grubaugh ND, Somasekar S, Chiu CY, Muñoz-Medina JE, Gonzalez-Bonilla CR, Arias CF, Lewis-Ximenez LL, Baylis SA, Chieppe AO, Aguiar SF, Fernandes CA, Lemos PS, Nascimento BLS, Monteiro HAO, Siqueira IC, De Queiroz MG, De Souza TR, Bezerra JF, Lemos MR, Pereira GF, Loudal D,

Moura LC, Dhalia R, França RF, Magalhães T, Marques ET, Jaenisch T, Wallau GL, De Lima MMC, Nascimento V, De Cerqueira EM, De Lima MMC, Mascarenhas DL, Neto JPM, Levin AS, Tozetto-Mendoza TR, Fonseca SN, Mendes-Correa MC, Milagres FP, Segurado A, Holmes EC, Rambaut A, Bedford T, Nunes MRT, Sabino EC, Alcantara LCJ, Loman NJ, Pybus OG. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature*. 2017;546(7658):406-410. doi:10.1038/nature22401

129. Leggett RM, Alcon-Giner C, Heavens D, Caim S, Brook TC, Kujawska M, Martin S, Peel N, Acford-Palmer H, Hoyles L, Clarke P, Hall LJ, Clark MD. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. *Nat Microbiol*. Published online 2019. doi:10.1038/s41564-019-0626-z
130. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, Rae D, Grundy S, Turner DJ, Wain J, Leggett RM, Livermore DM, O'Grady J. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat Biotechnol*. 2019;37(7):783-792. doi:10.1038/s41587-019-0156-5
131. Yang L, Haidar G, Zia H, Nettles R, Qin S, Wang X, Shah F, Rapport SF, Charalampous T, Methé B, Fitch A, Morris A, McVerry BJ, O'Grady J, Kitsios GD. Metagenomic identification of severe pneumonia pathogens in mechanically-ventilated patients: A feasibility and clinical validity study. *Respir Res*. 2019;20(1). doi:10.1186/s12931-019-1218-4
132. Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, M. Khan A, Woodford N, Saunders NJ, Wain J, O'Grady J, Livermore DM. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother*. Published online 2017. doi:10.1093/jac/dkw397
133. Sanderson ND, Street TL, Foster D, Swann J, Atkins BL, Brent AJ, McNally MA, Oakley S, Taylor A, Peto TEA, Crook DW, Eyre DW. Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. *BMC Genomics*. 2018;19(1):714. doi:10.1186/s12864-018-5094-y
134. Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol*. 2018;19(1):90. doi:10.1186/s13059-018-1462-9
135. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet*. Published online 2020. doi:10.1038/s41576-020-0236-x
136. Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev*. 2004;17(4):840-862. doi:10.1128/CMR.17.4.840-862.2004
137. Raina V, Nayak T, Ray L, Kumari K, Suar M. A Polyphasic Taxonomic Approach for Designation and Description of Novel Microbial Species. In: *Microbial Diversity in the Genomic Era*. Elsevier; 2019:137-152. doi:10.1016/b978-0-12-814849-5.00009-5
138. Akram A, Maley M, Gosbell I, Nguyen T, Chavada R. Utility of 16S rRNA PCR performed on clinical specimens in patient management. *Int J Infect Dis*. 2017;57:144-149. doi:10.1016/j.ijid.2017.02.006
139. Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol*. 2008;6(6):431-440. doi:10.1038/nrmicro1872
140. Yang B, Wang Y, Qian PY. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*. 2016;17(1):135. doi:10.1186/s12859-016-0992-y

141. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M, Sodergren E, Weinstock GM. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*. 2019;10(1):1-11. doi:10.1038/s41467-019-13036-1
142. Holm JB, Humphrys MS, Robinson CK, Settles ML, Ott S, Fu L, Yang H, Gajer P, He X, McComb E, Gravitt PE, Ghanem KG, Brotman RM, Ravel J. Ultrahigh-Throughput Multiplexing and Sequencing of >500-Base-Pair Amplicon Regions on the Illumina HiSeq 2500 Platform. *mSystems*. 2019;4(1). doi:10.1128/msystems.00029-19
143. Santos A, van Aerle R, Barrientos L, Martinez-Urtaza J. Computational methods for 16S metabarcoding studies using Nanopore sequencing data. *Comput Struct Biotechnol J*. 2020;18:296-305. doi:10.1016/j.csbj.2020.01.005
144. Winand R, Bogaerts B, Hoffman S, Lefevre L, Delvoye M, Van Braekel J, Fu Q, Roosens NHC, De Keersmaecker SCJ, Vanneste K. Targeting the 16s rRNA gene for bacterial identification in complex mixed samples: Comparative evaluation of second (illumina) and third (oxford nanopore technologies) generation sequencing technologies. *Int J Mol Sci*. Published online 2020. doi:10.3390/ijms21010298
145. Nygaard AB, Tunsjø HS, Meisal R, Charnock C. A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes. *Sci Rep*. 2020;10(1):3209. doi:10.1038/s41598-020-59771-0
146. Park S-C, Won S. Evaluation of 16S rRNA Databases for Taxonomic Assignments Using a Mock Community. *Genomics Inform*. 2018;16(4):e24. doi:10.5808/gi.2018.16.4.e24
147. Szabó M, Nagy T, Wilk T, Farkas T, Hegyi A, Olasz F, Kiss J. Characterization of Two Multidrug-Resistant IncA/C Plasmids from the 1960s by Using the MinION Sequencer Device. *Antimicrob Agents Chemother*. Published online 2016. doi:10.1128/aac.01121-16
148. 16S RefSeq records processing and curation. Accessed February 17, 2020. https://www.ncbi.nlm.nih.gov/refseq/targetedloci/16S_process/
149. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014;42(D1):D633. doi:10.1093/nar/gkt1244
150. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(D1):D590. doi:10.1093/nar/gks1219
151. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems*. 2018;3(6). doi:10.1128/msystems.00187-18
152. Corless CE, Guiver M, Borrow R, Edwards-Jones V, Kaczmarski EB, Fox AJ. Contamination and sensitivity issues with a real-time universal 16s rRNA PCR. *J Clin Microbiol*. 2000;38(5):1747-1752. doi:10.1128/jcm.38.5.1747-1752.2000
153. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, Kooistra-Smid AMD, Raangs EC, Rosema S, Veloo ACM, Zhou K, Friedrich AW, Rossen JWA. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol*. 2017;243:16-24. doi:10.1016/j.jbiotec.2016.12.022
154. Cuscó A, Catozzi C, Viñes J, Sanchez A, Francino O. Microbiota profiling with long amplicons using Nanopore sequencing: full-length 16S rRNA gene and the 16S-ITS-23S

of the rrn operon. *F1000Research*. 2019;7:1755. doi:10.12688/f1000research.16817.2
155. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J Clin Microbiol*. 2007;45(9):2761-2764. doi:10.1128/JCM.01228-07
156. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol*. 2017;35(9):833-844. doi:10.1038/nbt.3935
157. Chiu CY, Miller SA. *Clinical Metagenomics*. Vol 20. Nature Publishing Group; 2019:341-355. doi:10.1038/s41576-019-0113-7
158. Hasan MR, Rawat A, Tang P, Jithesh P V., Thomas E, Tan R, Tilley P. Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *J Clin Microbiol*. 2016;54(4):919-927. doi:10.1128/JCM.03050-15
159. Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, Dimalanta ET, Amaral-Zettler LA, Davis T, Quail MA, Pradhan S. A Method for Selectively Enriching Microbial DNA from Contaminating Vertebrate Host DNA. *PLoS One*. 2013;8(10):76096. doi:10.1371/journal.pone.0076096
160. Horz HP, Scheer S, Vianna ME, Conrads G. New methods for selective isolation of bacterial DNA from human clinical specimens. *Anaerobe*. 2010;16(1):47-53. doi:10.1016/j.anaerobe.2009.04.009
161. Jacob MC, Favre M, Bensa J-C. Membrane cell permeabilisation with saponin and multiparametric analysis by flow cytometry. *Cytometry*. 1991;12(6):550-558. doi:10.1002/cyto.990120612
162. Willis AD. Rarefaction, alpha diversity, and statistics. *Front Microbiol*. 2019;10(OCT):2407. doi:10.3389/fmicb.2019.02407
163. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J*. 2017;11(12):2639-2643. doi:10.1038/ismej.2017.119
164. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537-7541. doi:10.1128/AEM.01541-09
165. Wagner BD, Grunwald GK, Zerbe GO, Mikulich-Gilbertson SK, Robertson CE, Zemanick ET, Harris JK. On the use of diversity measures in longitudinal sequencing studies of microbial communities. *Front Microbiol*. 2018;9(MAY):1037. doi:10.3389/fmicb.2018.01037
166. Schloss PD. Evaluating different approaches that test whether microbial communities have the same structure. *ISME J*. 2008;2(3):265-275. doi:10.1038/ismej.2008.5
167. Khan IJ, Barry RJ, Amissah-Arthur KN, Carruthers D, Elamanchi SR, Situnayake D, Murray PI, Denniston AK, Rauz S. Ten-year experience of pulsed intravenous cyclophosphamide and methylprednisolone protocol (PICM protocol) in severe ocular inflammatory disease. *Br J Ophthalmol*. Published online 2013. doi:10.1136/bjophthalmol-2012-302130
168. Saw VPJ, Dart JKG, Rauz S, Ramsay A, Bunce C, Xing W, Maddison PG, Phillips M. Immunosuppressive Therapy for Ocular Mucous Membrane Pemphigoid. Strategies and Outcomes. *Ophthalmology*. 2008;115(2). doi:10.1016/j.ophtha.2007.04.027
169. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a Dual-

Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl Environ Microbiol*. 2013;79(17):5112-5120. doi:10.1128/aem.01043-13
170. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6):R60. doi:10.1186/gb-2011-12-6-r60
171. R: The R Project for Statistical Computing. Accessed December 6, 2020. https://www.r-project.org/
172. Boer CG, Radjabzadeh D, Medina-Gomez C, Garmaeva S, Schiphof D, Arp P, Koet T, Kurilshikov A, Fu J, Ikram MA, Bierma-Zeinstra S, Uitterlinden AG, Kraaij R, Zhernakova A, van Meurs JBJ. Intestinal microbiome composition and its relation to joint pain and inflammation. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-12873-4
173. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, Wang J, Imhann F, Brandsma E, Jankipersadsing SA, Joossens M, Cenit MC, Deelen P, Swertz MA, Weersma RK, Feskens EJM, Netea MG, Gevers D, Jonkers D, Franke L, Aulchenko YS, Huttenhower C, Raes J, Hofker MH, Xavier RJ, Wijmenga C, Fu J. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science (80- )*. 2016;352(6285):565-569. doi:10.1126/science.aad3369
174. Deschasaux M, Bouter KE, Prodan A, Levin E, Groen AK, Herrema H, Tremaroli V, Bakker GJ, Attaye I, Pinto-Sietsma SJ, van Raalte DH, Snijder MB, Nicolaou M, Peters R, Zwinderman AH, Bäckhed F, Nieuwdorp M. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nat Med*. 2018;24(10):1526-1531. doi:10.1038/s41591-018-0160-1
175. Kalyana Chakravarthy S, Jayasudha R, Sai Prashanthi G, Ali MH, Sharma S, Tyagi M, Shivaji S. Dysbiosis in the Gut Bacterial Microbiome of Patients with Uveitis, an Inflammatory Disease of the Eye. *Indian J Microbiol*. 2018;58(4):457-469. doi:10.1007/s12088-018-0746-9
176. Consolandi C, Turroni S, Emmi G, Severgnini M, Fiori J, Peano C, Biagi E, Grassi A, Rampelli S, Silvestri E, Centanni M, Cianchi F, Gotti R, Emmi L, Brigidi P, Bizzaro N, De Bellis G, Prisco D, Candela M, D'Elios MM. Behçet's syndrome patients exhibit specific microbiome signature. *Autoimmun Rev*. 2015;14(4):269-276. doi:10.1016/j.autrev.2014.11.009
177. Furusawa Y, Obata Y, Fukuda S, Endo TA, Nakato G, Takahashi D, Nakanishi Y, Uetake C, Kato K, Kato T, Takahashi M, Fukuda NN, Murakami S, Miyauchi E, Hino S, Atarashi K, Onawa S, Fujimura Y, Lockett T, Clarke JM, Topping DL, Tomita M, Hori S, Ohara O, Morita T, Koseki H, Kikuchi J, Honda K, Hase K, Ohno H. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature*. 2013;504(7480):446-450. doi:10.1038/nature12721
178. Hernandez H, de Souza R, Yu Z, Britton RA, de Paiva CS. Anti-inflammatory properties of butyrate on the ocular surface epithelium. *Investig Ophthalmol Vis Sci*. 2019;60(9):2818.
179. Moon J, Choi SH, Yoon CH, Kim MK. Gut dysbiosis is prevailing in Sjögren's syndrome and is related to dry eye severity. Appel S, ed. *PLoS One*. 2020;15(2):e0229029. doi:10.1371/journal.pone.0229029
180. BNF: British National Formulary - NICE.

181. Maier L, Goemans C V, Pruteanu M, Wirbel J, Kuhn M, Cacace E, Banerjee T, Erin Anderson E, Milanese A, Forslund SK, Raosaheb Patil K, Zeller G, Bork P, Typas A. Dissecting the collateral damage of antibiotics on gut microbes 1. doi:10.1101/2020.01.09.893560
182. Doan T, Worden L, Hinterwirth A, Arzika AM, Maliki R, Abdou A, Zhong L, Chen C, Cook C, Lebas E, O'Brien KS, Oldenburg CE, Chow ED, Porco TC, Lipsitch M, Keenan JD, Lietman TM. Macrolide and Nonmacrolide Resistance with Mass Azithromycin Distribution. *N Engl J Med*. 2020;383(20):1941-1950. doi:10.1056/nejmoa2002606
183. Bhat M, Pasini E, Copeland J, Angeli M, Husain S, Kumar D, Renner E, Teterina A, Allard J, Guttman DS, Humar & A. Impact of Immunosuppression on the Metagenomic Composition of the Intestinal Microbiome: a Systems Biology Approach to Post-Transplant Diabetes OPEN. doi:10.1038/s41598-017-10471-2
184. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H, Patil KR, Bork P, Typas A. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*. Published online 2018. doi:10.1038/nature25979
185. Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre ML, Luke JJ, Gajewski TF. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science (80- )*. Published online 2018. doi:10.1126/science.aao3290
186. Gopalakrishnan V, Helmink BA, Spencer CN, Reuben A, Wargo JA. The Influence of the Gut Microbiome on Cancer, Immunity, and Cancer Immunotherapy. *Cancer Cell*. 2018;33(4):570-580. doi:10.1016/J.CCELL.2018.03.015
187. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. *Nature*. Published online 2020. doi:10.1038/s41586-020-2881-9
188. Moon J, Yoon CH, Choi SH, Kim MK. Can gut microbiota affect dry eye syndrome? *Int J Mol Sci*. 2020;21(22):1-28. doi:10.3390/ijms21228443
189. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets T V., Prieto PA, Vicente D, Hoffman K, Wei SC, Cogdill AP, Zhao L, Hudgens CW, Hutchinson DS, Manzo T, Petaccia De Macedo M, Cotechini T, Kumar T, Chen WS, Reddy SM, Szczepaniak Sloane R, Galloway-Pena J, Jiang H, Chen PL, Shpall EJ, Rezvani K, Alousi AM, Chemaly RF, Shelburne S, Vence LM, Okhuysen PC, Jensen VB, Swennes AG, McAllister F, Marcelo Riquelme Sanchez E, Zhang Y, Le Chatelier E, Zitvogel L, Pons N, Austin-Breneman JL, Haydu LE, Burton EM, Gardner JM, Sirmans E, Hu J, Lazar AJ, Tsujikawa T, Diab A, Tawbi H, Glitza IC, Hwu WJ, Patel SP, Woodman SE, Amaria RN, Davies MA, Gershenwald JE, Hwu P, Lee JE, Zhang J, Coussens LM, Cooper ZA, Futreal PA, Daniel CR, Ajami NJ, Petrosino JF, Tetzlaff MT, Sharma P, Allison JP, Jenq RR, Wargo JA. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science (80- )*. Published online 2018. doi:10.1126/science.aan4236
190. Centers for Disease Control and Prevention. *Estimated Burden of Keratitis — United States, 2010*.; 2014. Accessed June 3, 2019. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6345a3.htm
191. Ting DSJ, Settle C, Morgan SJ, Baylis O, Ghosh S. A 10-year analysis of microbiological profiles of microbial keratitis: the North East England Study. *Eye*. 2018;32(8):1416-1417. doi:10.1038/s41433-018-0085-4
192. Tan SZ, Walkden A, Au L, Fullwood C, Hamilton A, Qamruddin A, Armstrong M,

Brahma AK, Carley F. Twelve-year analysis of microbial keratitis trends at a UK tertiary hospital. *Eye*. Published online 2017. doi:10.1038/eye.2017.55
193. Lichtinger A, Yeung SN, Kim P, Amiran MD, Iovieno A, Elbaz U, Ku JYF, Wolff R, Rootman DS, Slomovic AR. Shifting trends in bacterial keratitis in Toronto: An 11-year review. *Ophthalmology*. 2012;119(9):1785-1790. doi:10.1016/j.ophtha.2012.03.031
194. Bispo PJM, Davoudi S, Sahm ML, Ren A, Miller J, Romano J, Sobrin L, Gilmore MS. Rapid detection and identification of uveitis pathogens by qualitative multiplex real-time PCR. *Investig Ophthalmol Vis Sci*. 2018;59(1):582-589. doi:10.1167/iovs.17-22597
195. Kaye SB, Rao PG, Smith G, Scott JA, Hoyles S, Morton CE, Willoughby C, Batterbury M, Harvey G. Simplifying collection of corneal specimens in cases of suspected bacterial keratitis. *J Clin Microbiol*. 2003;41(7):3192-3197. doi:10.1128/JCM.41.7.3192-3197.2003
196. Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu Rev Pathol*. 2019;14:319-338. doi:10.1146/annurev-pathmechdis-012418-012751
197. Douglas CA, Ivey KL, Papanicolas LE, Best KP, Muhlhausler BS, Rogers GB. DNA extraction approaches substantially influence the assessment of the human breast milk microbiome. *Sci Rep*. 2020;10(1):1-10. doi:10.1038/s41598-019-55568-y
198. Sui H, Weil AA, Nuwagira E, Qadri F, Ryan ET, Mezzari MP, Phipatanakul W, Lai PS. Impact of DNA Extraction Method on Variation in Human and Built Environment Microbial Community and Functional Profiles Assessed by Shotgun Metagenomics Sequencing. *Front Microbiol*. 2020;11:953. doi:10.3389/fmicb.2020.00953
199. de Boer R, Peters R, Gierveld S, Schuurman T, Kooistra-Smid M, Savelkoul P. Improved detection of microbial DNA after bead-beating before DNA isolation. *J Microbiol Methods*. 2010;80(2):209-211. doi:10.1016/j.mimet.2009.11.009
200. Ojo-Okunola A, Claassen-Weitz S, Mwaikono KS, Gardner-Lubbe S, Zar HJ, Nicol MP, du Toit E. The Influence of DNA Extraction and Lipid Removal on Human Milk Bacterial Profiles. *Methods Protoc*. 2020;3(2):39. doi:10.3390/mps3020039
201. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28(23):3150-3152. doi:10.1093/bioinformatics/bts565
202. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100. doi:10.1093/bioinformatics/bty191
203. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352
204. Lin A, Rhee MK, Akpek EK, Amescua G, Farid M, Garcia-Ferrer FJ, Varu DM, Musch DC, Dunn SP, Mah FS. Bacterial Keratitis Preferred Practice Pattern®. *Ophthalmology*. 2019;126(1):P1-P55. doi:10.1016/j.ophtha.2018.10.018
205. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. Published online 2010. doi:10.1038/nmeth.f.303
206. Li W, Jaroszewski L, Godzik A. Sequence clustering strategies improve remote

homology recognitions while reducing search times. *Protein Eng*. 2002;15(8):643-649. doi:10.1093/protein/15.8.643
207. Bruijns BB, Tiggelaar RM, Gardeniers H. The Extraction and Recovery Efficiency of Pure DNA for Different Types of Swabs. *J Forensic Sci*. Published online 2018. doi:10.1111/1556-4029.13837
208. Brownlow RJ, Dagnall KE, Ames CE. A Comparison of DNA Collection and Retrieval from Two Swab Types (Cotton and Nylon Flocked Swab) when Processed Using Three QIAGEN Extraction Methods. *J Forensic Sci*. Published online 2012. doi:10.1111/j.1556-4029.2011.02022.x
209. Verdon TJ, Mitchell RJ, van Oorschot RAH. Swabs as DNA collection devices for sampling different biological materials from different substrates. *J Forensic Sci*. Published online 2014. doi:10.1111/1556-4029.12427
210. Su X, Comeau AM. Cellulose as a matrix for nucleic acid purification. *Anal Biochem*. 1999;267(2):415-418. doi:10.1006/abio.1998.2987
211. Pinnock A, Shivshetty N, Roy S, Rimmer S, Douglas I, MacNeil S, Garg P. Ex vivo rabbit and human corneas as models for bacterial and fungal keratitis. *Graefe's Arch Clin Exp Ophthalmol*. 2017;255(2):333-342. doi:10.1007/s00417-016-3546-0
212. Urwin L, Okurowska K, Crowther G, Roy S, Garg P, Karunakaran E, MacNeil S, Partridge LJ, Green LR, Monk PN. Corneal Infection Models: Tools to Investigate the Role of Biofilms in Bacterial Keratitis. *Cells*. 2020;9(11). doi:10.3390/cells9112450
213. Rose L, Jensen B, Peterson A, Banerjee SN, Arduino MJ. Swab materials and Bacillus anthracis spore recovery from nonporous surfaces. *Emerg Infect Dis*. 2004;10(6):1023-1029. doi:10.3201/eid1006.030716
214. Adamowicz MS, Stasulli DM, Sobestanovich EM, Bille TW. Evaluation of methods to improve the extraction and recovery of DNA from cotton swabs for forensic analysis. *PLoS One*. 2014;9(12). doi:10.1371/journal.pone.0116351
215. Benítez-Páez A, Sanz Y. Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinIONTM portable nanopore sequencer. *Gigascience*. 2017;6(7):1. doi:10.1093/gigascience/gix043
216. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol*. 2019;20(1):129. doi:10.1186/s13059-019-1727-y
217. Urban L, Holzer A, Baronas JJ, Hall M, Braeuninger-Weimer P, Scherm MJ, Kunz DJ, Perera SN, Martin-Herranz DE, Tipper ET, Salter SJ, Stammnitz MR. Freshwater monitoring by nanopore sequencing. *bioRxiv*. Published online February 2020:2020.02.06.936302. doi:10.1101/2020.02.06.936302
218. Balvočiūtė M, Huson DH. SILVA, RDP, Greengenes, NCBI and OTT — how do these taxonomies compare? *BMC Genomics*. 2017;18(S2):114. doi:10.1186/s12864-017-3501-4
219. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658-1659. doi:10.1093/bioinformatics/btl158
220. Schmidt TSBB, Matias Rodrigues JF, von Mering C. Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environ Microbiol*. 2015;17(5):1689-1706. doi:10.1111/1462-2920.12610
221. Edgar RC. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Valencia A, ed. *Bioinformatics*. 2018;34(14):2371-2375. doi:10.1093/bioinformatics/bty113

222. Mysara M, Vandamme P, Props R, Kerckhof F-MM, Leys N, Boon N, Raes J, Monsieurs P. Reconciliation between operational taxonomic units and species boundaries. *FEMS Microbiol Ecol*. 2017;93(4). doi:10.1093/femsec/fix029
223. Public Health England. *UK Standards for Microbiology Investigations Identification of Bacillus Species*.; 2018.
224. Shu LJ, Yang YL. Bacillus Classification Based on Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry - Effects of Culture Conditions. *Sci Rep*. 2017;7(1). doi:10.1038/s41598-017-15808-5
225. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog*. Published online 2016. doi:10.1186/s13099-016-0103-7
226. Johnson MW, Doft BH, Kelsey SF, Barza M, Wilson LA, Barr CC, Wisniewski SR. The endophthalmitis vitrectomy study: Relationship between clinical presentation and microbiologic spectrum. *Ophthalmology*. 1997;104(2):261-272. doi:10.1016/S0161-6420(97)30326-1
227. Taravati P, Lam D, Van Gelder RN. Role of Molecular Diagnostics in Ocular Microbiology. *Curr Ophthalmol Rep*. 2013;1(4):181-189. doi:10.1007/s40135-013-0025-1
228. MinION | Oxford Nanopore Technologies. Accessed October 30, 2020. https://nanoporetech.com/products/minion
229. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, Malla S, Marriott H, Nieto T, O'Grady J, Olsen HE, Pedersen BS, Rhie A, Richardson H, Quinlan AR, Snutch TP, Tee L, Paten B, Phillippy AM, Simpson JT, Loman NJ, Loose M. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*. 2018;36(4):338-345. doi:10.1038/nbt.4060
230. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. 2020;21(1):1-16. doi:10.1186/s13059-020-1935-5
231. Narumi M, Nishitsuka K, Yamakawa M, Yamashita H. A survey of vitreous cell components performed using liquid-based cytology. *Acta Ophthalmol*. 2015;93(5):e386-e390. doi:10.1111/aos.12623
232. Lee AY, Lee CS, Van Gelder RN. Scalable metagenomics alignment research tool (SMART): A scalable, rapid, and complete search heuristic for the classification of metagenomic sequences from complex sequence populations. *BMC Bioinformatics*. 2016;17(1):1-12. doi:10.1186/s12859-016-1159-6
233. Delgado MAJ, Casella S, Bedmar EJ. Denitrification in Rhizobia-Legume Symbiosis. In: *Biology of the Nitrogen Cycle*. Elsevier; 2007:83-91. doi:10.1016/B978-044452857-5.50007-2
234. Mugford ST, Osbourn A. Saponin synthesis and function. In: *Isoprenoid Synthesis in Plants and Microorganisms: New Concepts and Experimental Approaches*. Springer New York; 2013:405-424. doi:10.1007/978-1-4614-4063-5_28
235. Thézé J, Li T, du Plessis L, Bouquet J, Kraemer MUG, Somasekar S, Yu G, de Cesare M, Balmaseda A, Kuan G, Harris E, Wu C hsi, Ansari MA, Bowden R, Faria NR, Yagi S, Messenger S, Brooks T, Stone M, Bloch EM, Busch M, Muñoz-Medina JE, González-Bonilla CR, Wolinsky S, López S, Arias CF, Bonsall D, Chiu CY, Pybus OG. Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central

America and Mexico. *Cell Host Microbe*. 2018;23(6):855-864.e7. doi:10.1016/j.chom.2018.04.017
236. Gu W, Deng X, Lee M, Sucu YD, Arevalo S, Stryke D, Federman S, Gopez A, Reyes K, Zorn K, Sample H, Yu G, Ishpuniani G, Briggs B, Chow ED, Berger A, Wilson MR, Wang C, Hsu E, Miller S, DeRisi JL, Chiu CY. Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. *Nat Med*. 2021;27(1):115-124. doi:10.1038/s41591-020-1105-z
237. Ganda E, Beck KL, Haiminen N, Kawas B, Cronk B, Anderson RR, Goodman LB, Wiedmann M. DNA extraction and host depletion methods significantly impact and potentially bias bacterial detection in a biological fluid. *bioRxiv*. Published online August 23, 2020:2020.08.21.262337. doi:10.1101/2020.08.21.262337
238. Nelson MT, Pope CE, Marsh RL, Wolter DJ, Weiss EJ, Hager KR, Vo AT, Brittnacher MJ, Radey MC, Hayden HS, Eng A, Miller SI, Borenstein E, Hoffman LR. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. *Cell Rep*. 2019;26(8):2227-2240.e5. doi:10.1016/j.celrep.2019.01.091
239. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12(1). doi:10.1186/s12915-014-0087-z
240. Laurence M, Hatzis C, Brash DE. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One*. 2014;9(5):97876. doi:10.1371/journal.pone.0097876
241. Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res*. 2015;43(D1):D593-D598. doi:10.1093/nar/gku1201
242. Pearman WS, Freed NE, Silander OK. Testing the advantages and disadvantages of short- And long- read eukaryotic metagenomics using simulated reads. *BMC Bioinformatics*. 2020;21(1):220. doi:10.1186/s12859-020-3528-4
243. McCoy KD, Burkhard R, Geuking MB. The microbiome and immune memory formation. *Immunol Cell Biol*. 2019;97(7):625-635. doi:10.1111/imcb.12273
244. Schluter J, Peled JU, Taylor BP, Markey KA, Smith M, Taur Y, Niehus R, Staffas A, Dai A, Fontana E, Amoretti LA, Wright RJ, Morjaria S, Fenelus M, Pessin MS, Chao NJ, Lew M, Bohannon L, Bush A, Sung AD, Hohl TM, Perales MA, van den Brink MRM, Xavier JB. The gut microbiota is associated with immune cell dynamics in humans. *Nature*. 2020;588(7837). doi:10.1038/s41586-020-2971-8
245. Berman D, Parker SM, Siegel J, Chasalow SD, Weber J, Galbraith S, Targan SR, Wang HL. Blockade of cytotoxic T-lymphocyte antigen-4 by ipilimumab results in dysregulation of gastrointestinal immunity in patients with advanced melanoma. *Cancer Immun*. Published online 2010.
246. Somerville TF, Shankar J, Aldwinckle S, Sueke H, Neal T, Horsburgh MJ, Kaye SB. Recurrent microbial keratitis and endogenous site Staphylococcus aureus colonisation. *Sci Rep*. 2020;10(1):18559. doi:10.1038/s41598-020-75821-z
247. Salipante SJ, Sengupta DJ, Rosenthal C, Costa G, Spangler J, Sims EH, Jacobs MA, Miller SI, Hoogestraat DR, Cookson BT, McCoy C, Matsen FA, Shendure J, Lee CC, Harkins TT, Hoffman NG. Rapid 16S rRNA Next-Generation Sequencing of

Polymicrobial Clinical Samples for Diagnosis of Complex Bacterial Infections. *PLoS One*. 2013;8(5). doi:10.1371/journal.pone.0065226

248. Tegha G, Ciccone EJ, Krysiak R, Kaphatika J, Chikaonda T, Ndhlovu I, van Duin D, Hoffman I, Juliano JJ, Wang J. Genomic epidemiology of Escherichia coli isolates from a tertiary referral center in Lilongwe, Malawi. *Microb Genomics*. Published online December 9, 2020. doi:10.1099/mgen.0.000490
249. Marí-Almirall M, Cosgaya C, Pitart C, Viñes J, Muñoz L, Campo I, Cuscó A, Rodríguez-Serna L, Santana G, Del Río A, Francino O, Ciruela P, Pujol I, Ballester F, Marco F, Martínez JA, Soriano Á, Vila J, Roca I, Jove PP, Padilla E, Ballestero-Téllez M, Zboromyrska Y, Ángel Benítez M, Clivillé R, González S, Calvet I, Gallés C, Sauca G, Martí-Sala C, Angeles Pulido M, Vilamala A, González-Cuevas A, Gené A, Trujillo G, Lopez Madueño J, Raga X, Gómez F, Picó E, Sarvisé C, Font X. Dissemination of NDM-producing Klebsiella pneumoniae and Escherichia coli high-risk clones in Catalan healthcare institutions . *J Antimicrob Chemother*. Published online November 17, 2020. doi:10.1093/jac/dkaa459
250. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, Mohr I, Wilson AC. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-08734-9
251. Langelier C, Kalantar KL, Moazed F, Wilson MR, Crawford ED, Deiss T, Belzer A, Bolourchi S, Caldera S, Fung M, Jauregui A, Malcolm K, Lyden A, Khan L, Vessel K, Quan J, Zinter M, Chiu CY, Chow ED, Wilson J, Miller S, Matthay MA, Pollard KS, Christenson S, Calfee CS, DeRisi JL, Dickson RP, Fodor AA, Kau A. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci U S A*. 2018;115(52):E12353-E12362. doi:10.1073/pnas.1809700115
252. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Sander C, Stuart JM, Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, Ally A, Balasundaram M, Birol I, Butterfield YSN, Chu A, Chuah E, Chun HJE, Dhalla N, Guin R, Hirst M, Hirst C, Holt RA, Jones SJM, Lee D, Li HI, Marra MA, Mayo M, Moore RA, Mungall AJ, Robertson AG, Schein JE, Sipahimalani P, Tam A, Thiessen N, Varhol RJ, Beroukhim R, Bhatt AS, Brooks AN, Cherniack AD, Freeman SS, Gabriel SB, Helman E, Jung J, Meyerson M, Ojesina AI, Pedamallu CS, Saksena G, Schumacher SE, Tabak B, Zack T, Lander ES, Bristow CA, Hadjipanayis A, Haseley P, Kucherlapati R, Lee S, Lee E, Luquette LJ, Mahadeshwar HS, Pantazi A, Parfenov M, Park PJ, Protopopov A, Ren X, Santoso N, Seidman J, Seth S, Song X, Tang J, Xi R, Xu AW, Yang L, Zeng D, Auman JT, Balu S, Buda E, Fan C, Hoadley KA, Jones CD, Meng S, Mieczkowski PA, Parker JS, Perou CM, Roach J, Shi Y, Silva GO, Tan D, Veluvolu U, Waring S, Wilkerson MD, Wu J, Zhao W, Bodenheimer T, Hayes DN, Hoyle AP, Jeffreys SR, Mose LE, Simons J V., Soloway MG, Baylin SB, Berman BP, Bootwalla MS, Danilova L, Herman JG, Hinoue T, Laird PW, Rhie SK, Shen H, Triche T, Weisenberger DJ, Carter SL, Cibulskis K, Chin L, Zhang J, Sougnez C, Wang M, Getz G, Dinh H, Doddapaneni HV, Gibbs R, Gunaratne P, Han Y, Kalra D, Kovar C, Lewis L, Morgan M, Morton D, Muzny D, Reid J, Xi L, Cho J, Dicara D, Frazer S, Gehlenborg N, Heiman DI, Kim J, Lawrence MS, Lin P, Liu Y, Noble MS, Stojanov P, Voet D, Zhang H, Zou L, Stewart C, Bernard B, Bressler R, Eakin A, Iype L, Knijnenburg T, Kramer R, Kreisberg R, Leinonen K, Lin J, Liu Y, Miller M, Reynolds SM, Rovira H, Shmulevich I, Thorsson V, Yang D, Zhang W, Amin S, Wu CJ, Wu CC,

Akbani R, Aldape K, Baggerly KA, Broom B, Casasent TD, Cleland J, Dodda D, Edgerton M, Han L, Herbrich SM, Ju Z, Kim H, Lerner S, Li J, Liang H, Liu W, Lorenzi PL, Lu Y, Melott J, Nguyen L, Su X, Verhaak R, Wang W, Wong A, Yang Y, Yao J, Yao R, Yoshihara K, Yuan Y, Yung AK, Zhang N, Zheng S, Ryan M, Kane DW, Aksoy BA, Ciriello G, Dresdner G, Gao J, Gross B, Jacobsen A, Kahles A, Ladanyi M, Lee W, Lehmann K Van, Miller ML, Ramirez R, Rätsch G, Reva B, Schultz N, Senbabaoglu Y, Shen R, Sinha R, Sumer SO, Sun Y, Taylor BS, Weinhold N, Fei S, Spellman P, Benz C, Carlin D, Cline M, Craft B, Goldman M, Haussler D, Ma S, Ng S, Paull E, Radenbaugh A, Salama S, Sokolov A, Swatloski T, Uzunangelov V, Waltman P, Yau C, Zhu J, Hamilton SR, Abbott S, Abbott R, Dees ND, Delehaunty K, Ding L, Dooling DJ, Eldred JM, Fronick CC, Fulton R, Fulton LL, Kalicki-Veizer J, Kanchi KL, Kandoth C, Koboldt DC, Larson DE, Ley TJ, Lin L, Lu C, Magrini VJ, Mardis ER, McLellan MD, McMichael JF, Miller CA, O'Laughlin M, Pohl C, Schmidt H, Smith SM, Walker J, Wallis JW, Wendl MC, Wilson RK, Wylie T, Zhang Q, Burton R, Jensen MA, Kahn A, Pihl T, Pot D, Wan Y, Levine DA, Black AD, Bowen J, Frick J, Gastier-Foster JM, Harper HA, Helsel C, Leraas KM, Lichtenberg TM, McAllister C, Ramirez NC, Sharpe S, Wise L, Zmuda E, Chanock SJ, Davidsen T, Demchok JA, Eley G, Felau I, Sheth M, Sofia H, Staudt L, Tarnuzzer R, Wang Z, Yang L, Zhang J, Omberg L, Margolin A, Raphael BJ, Vandin F, Wu HT, Leiserson MDM, Benz SC, Vaske CJ, Noushmehr H, Wolf D, Veer LVT, Anastassiou D, Yang THO, Lopez-Bigas N, Gonzalez-Perez A, Tamborero D, Xia Z, Li W, Cho DY, Przytycka T, Hamilton M, McGuire S, Nelander S, Johansson P, Jörnsten R, Kling T. The cancer genome atlas pan-cancer analysis project. *Nat Genet*. 2013;45(10):1113-1120. doi:10.1038/ng.2764

253. The COVID-19 Genomics UK Consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. Published online 2020. doi:10.1016/S2666-5247(20)30054-9
254. Wood CS, Thomas MR, Budd J, Mashamba-Thompson TP, Herbst K, Pillay D, Peeling RW, Johnson AM, McKendry RA, Stevens MM. Taking connected mobile-health diagnostics of infectious diseases to the field. *Nature*. Published online 2019. doi:10.1038/s41586-019-0956-2
255. *Optimising the Research Environment for a Healthier, Fairer Future*.; 2016.

# Appendices

## Supplementary for Chapter 2

*Supplementary material 2.1 R script for generating principle coordination analysis plots.*

*Pcoa_script.R written by Liying Low.*

```
library(tidyverse)
library(readxl)
library(ggplot2)
pcoa <- read_tsv(file="raw_data/inputfileMMPHC.opti_mcc.thetayc.0.03.lt.ave.pcoa.axes")
metadata <- read_excel(path = "raw_data/MMP_HC_metadata.xlsx")
metadata_pcoa <- inner_join(metadata, pcoa, by=c('16SID' = 'group'))

ggplot(metadata_pcoa, aes(x=axis1, y=axis2, color=Cohort)) +
  geom_point(shape=19, size=2) +
  scale_color_manual(name=NULL,
                     values=c("#CC79A7", "#0072B2"),
                     breaks=c("Ocular Mucous Membrane Pemphigoid", "Healthy volunteer"),
                     labels=c("Ocular Mucous Membrane Pemphigoid (n=49)", "Healthy volunteer
(n=38)")) +
  coord_fixed() +
  labs(title="PCoA of ThetaYC Distances between Faecal Samples",
       x = "PCo Axis 1 (12.3%)",
       y = "PCo Axis 2 (9.3%)") +
  theme_classic() +
  theme(legend.position = "top")

ggsave(file  = "MMP_HC_ordination.png", dpi = 300)

ggplot(metadata_pcoa, aes(x=axis1, y=axis2, color=Tetracycline)) +
  geom_point(shape=19, size=2, alpha = 0.6) +
  scale_color_manual(name=NULL,
                     values=c("blue", "red", "purple", "grey"),
                     breaks=c("Never used", "Previously used medication, but now stopped",
"Currently using tetracycline", "Healthy volunteer"),
                     labels=c("Never used (n=24)", "Previously used medication, but now stopped
(n=7)", "Currently using tetracycline (n=18)", "Healthy volunteer (n=38)")) +
  coord_fixed() +
  labs(title="PCoA of ThetaYC Distances between Faecal Samples",
       x = "PCo Axis 1",
       y = "PCo Axis 2") +
  theme_classic()
```

```
ggsave("tetracycline_ordination_with_HC.pdf")

ggplot(metadata_pcoa, aes(x=axis1, y=axis2, color=Cohort, shape=Sex)) +
  geom_point(size=2, alpha = 0.7) +
  scale_color_manual(name=NULL,
                values=c("#CC79A7", "#0072B2"),
                breaks=c("Ocular Mucous Membrane Pemphigoid", "Healthy volunteer"),
                labels=c("Ocular Mucous Membrane Pemphigoid (n=49)", "Healthy volunteer
(n=38)")) +
  scale_shape_manual(name = NULL,
                values =c(19,17),
                breaks = c("Female", "Male")) +
  coord_fixed() +
  labs(title="PCoA of ThetaYC Distances between Faecal Samples",
      x = "PCo Axis 1",
      y = "PCo Axis 2") +
  theme_classic()

ggsave("gender_ordination.pdf")

summary(metadata)
count(metadata_pcoa, Biopsy)

ggplot(metadata_pcoa, aes(x=axis1, y=axis2, color=Biopsy)) +
  geom_point(shape=19, size=2, alpha = 0.6) +
  scale_color_manual(name=NULL,
                values=c("blue", "red"),
                breaks=c("Negative", "Positive"),
                labels=c("Negative (n=15)", "Positive (n=34)")) +
  coord_fixed() +
  labs(title="PCoA of ThetaYC Distances between Faecal Samples",
      x = "PCo Axis 1",
      y = "PCo Axis 2") +
  theme_classic()

ggsave("biopsy_ordination_with_HC.pdf")

summary(metadata_pcoa,OSDI)
count(metadata_pcoa, Current_Systemic_Immunotherapy)
count(metadata_pcoa, Tetracycline)
count(metadata_pcoa, Systemic _Immunosuppression_Ever)
count(metadata_pcoa, Systemic_Steroids)
count(metadata_pcoa, DMARDS)
count(metadata_pcoa, Cohort)
```

*Supplementary material 2.2 R script for processing 16S taxonomic data.*

*Processing_taxonomy_data.R written by Liying Low*

```
library(tidyverse)
library(readxl)
library (ggplot2)

metadata <- read_excel(path="raw_data/MMP_HC_metadata.xlsx")

taxonomy <- read_tsv(file="raw_data/inputfileMMPHC.taxonomy_v2.txt") %>%
  rename_all(tolower) %>%
  mutate(taxonomy=str_replace_all(string=taxonomy, pattern ="\\(\\d*\\)", replacement =
"")) %>%
  mutate(taxonomy=str_replace_all(string=taxonomy, pattern =";$", replacement="")) %>%
  separate(taxonomy, into=c("kingdom", "phylum", "class", "order", "family", "genus"),
sep=";")

otu_data <- read_tsv("raw_data/inputfileMMPHC.opti_mcc.0.03.subsample.shared",
col_types = cols(Group=col_character())) %>%
  select(-label, -numOtus) %>%
  rename(sample=Group) %>%
  pivot_longer(cols=-sample, names_to="otu", values_to="count") %>%
  mutate(rel_abund=count/25000)

otu_data %>% group_by(sample) %>% summarize(n=sum(count)) %>% summary()

agg_phylum_data <- inner_join(otu_data, taxonomy) %>%
  group_by(sample, phylum) %>%
  summarize(agg_rel_abund=sum(rel_abund)) %>%
  inner_join(., metadata, by = "sample") %>%
  ungroup()

top_phyla <- agg_phylum_data %>%
  group_by(phylum) %>%
  summarize(median = median(agg_rel_abund)) %>%
  arrange((desc(median))) %>%
  top_n(5, median) %>%
  pull(phylum)

agg_phylum_HC <- agg_phylum_data %>%
  group_by(phylum) %>%
  filter(Cohort == "Healthy volunteer") %>%
  summarize(enframe(quantile(agg_rel_abund*100, c(0.25, 0.5, 0.75)), "quantile",
"agg_rel_abund"))

agg_phylum_MMP <- agg_phylum_data %>%
```

```
  group_by(phylum) %>%
  filter(Cohort == "Ocular Mucous Membrane Pemphigoid") %>%
  summarize(enframe(quantile(agg_rel_abund*100, c(0.25, 0.5, 0.75)), "quantile",
"agg_rel_abund"))


agg_phylum_data %>%
  filter(phylum %in% top_phyla) %>%
  mutate(phylum=factor(phylum, levels = top_phyla)) %>%
  mutate(agg_rel_abund=agg_rel_abund + 1/21000) %>%
  ggplot(aes(x=phylum, y=agg_rel_abund, color=Cohort)) +
  geom_hline(yintercept=1/10530, color="gray") +
  geom_boxplot() +
  scale_color_manual(name=NULL,
                     values=c("#0072B2", "#CC79A7"),
                     breaks=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid"),
                     labels=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid")) +
  labs(title="Phylum level",
       x=NULL,
       y="Relative abundance (%)") +
  scale_y_log10(breaks=c(1e-4, 1e-3, 1e-2, 1e-1, 1), labels=c(1e-2, 1e-1, 1, 10, 100)) +
  theme_classic() +
  theme(axis.text.x = element_text(angle=45, hjust=1))

ggsave(file = "phlya.png", dpi=300)


library(broom)
library(purrr)

phylum_tests <- agg_phylum_data %>%
  nest(sample_data = c(-phylum)) %>%
  mutate(test = map(sample_data, ~tidy(kruskal.test(agg_rel_abund~Cohort, data=.)))) %>%
  unnest(test) %>%
  mutate(p.value.adj=p.adjust(p.value, method="BH")) %>%
  arrange(p.value.adj)

sig_phyla <- phylum_tests %>%
  filter(p.value.adj <= 0.05) %>%
  pull(phylum)

####class

agg_class_data <- inner_join(otu_data, taxonomy) %>%
  group_by(sample, class) %>%
  summarize(agg_rel_abund=sum(rel_abund)) %>%
  inner_join(., metadata, by = "sample") %>%
```

```
  ungroup()

top_class <- agg_class_data %>%
  group_by(class) %>%
  summarize(median = median(agg_rel_abund)) %>%
  arrange((desc(median))) %>%
  top_n(10, median) %>%
  pull(class)

view(top_class)

agg_class_data %>%
  filter(class %in% top_class) %>%
  mutate(class=factor(class, levels = top_class)) %>%
  mutate(agg_rel_abund=agg_rel_abund + 1/21000) %>%
  ggplot(aes(x=class, y=agg_rel_abund, color=Cohort)) +
  geom_hline(yintercept=1/10530, color="gray") +
  geom_boxplot() +
  scale_color_manual(name=NULL,
                     values=c("#0072B2", "#CC79A7"),
                     breaks=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid"),
                     labels=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid")) +
  labs(title="Relative abundance in class-level between OcMMP and HC",
       x=NULL,
       y="Relative abundance (%)") +
  scale_y_log10(breaks=c(1e-4, 1e-3, 1e-2, 1e-1, 1), labels=c(1e-2, 1e-1, 1, 10, 100)) +
  theme_classic() +
  theme(axis.text.x = element_text(angle=45, hjust=1))

ggsave("class.png")

library(broom)
library(purrr)

class_tests <- agg_class_data %>%
  nest(sample_data = c(-class)) %>%
  mutate(test = map(sample_data, ~tidy(kruskal.test(agg_rel_abund~Cohort, data=.)))) %>%
  unnest(test) %>%
  mutate(p.value.adj=p.adjust(p.value, method="BH")) %>%
  arrange(p.value.adj)

sig_class <- class_tests %>%
  filter(p.value.adj <= 0.05) %>%
  pull(class)

#### deepest
```

```
deep_taxonomy <- read_tsv(file="raw_data/inputfileMMPHC.taxonomy_v2.txt") %>%
  rename_all(tolower) %>%
  mutate(taxonomy=str_replace_all(string=taxonomy, pattern ="\\(\\d*\\)", replacement =
"")) %>%
  mutate(taxonomy=str_replace_all(string=taxonomy, pattern = "_unclassified*",
replacement = "")) %>%
  mutate(taxonomy=str_replace_all(string=taxonomy, pattern =";$", replacement="")) %>%
  mutate(taxonomy=str_replace_all(string=taxonomy, pattern=".*;", replacement=""))

agg_deep_data <- inner_join(otu_data, deep_taxonomy) %>%
  group_by(sample, taxonomy) %>%
  summarize(agg_rel_abund=sum(rel_abund)) %>%
  inner_join(., metadata, by = "sample") %>%
  ungroup()

agg_deep_HC <- agg_deep_data %>%
  group_by(taxonomy) %>%
  filter(Cohort == "Healthy volunteer") %>%
  summarize(enframe(quantile(agg_rel_abund*100, c(0.25, 0.5, 0.75)), "quantile",
"agg_rel_abund"))

agg_deep_MMP <- agg_deep_data %>%
  group_by(taxonomy) %>%
  filter(Cohort == "Ocular Mucous Membrane Pemphigoid") %>%
  summarize(enframe(quantile(agg_rel_abund*100, c(0.25, 0.5, 0.75)), "quantile",
"agg_rel_abund"))


deep_tests <- agg_deep_data %>%
  nest(sample_data = c(-taxonomy)) %>%
  mutate(test=map(sample_data, ~tidy(kruskal.test(agg_rel_abund~Cohort, data=.)))) %>%
  unnest(test) %>%
  mutate(p.value.adj=p.adjust(p.value, method="BH")) %>%
  arrange(p.value.adj)

sig_deep <- deep_tests %>%
  filter(p.value.adj <= 0.05) %>%
  pull(taxonomy)

view(sig_deep)

agg_deep_data %>%
  filter(taxonomy %in% sig_deep) %>%
  mutate(taxonomy = factor (taxonomy, levels=sig_deep)) %>%
  mutate(agg_rel_abund = agg_rel_abund + 1/25000) %>%
  ggplot(aes(x=taxonomy, y=agg_rel_abund, color=Cohort)) +
  geom_boxplot() +
```

```
  scale_color_manual (name=NULL,
                      values = c("#0072B2", "#CC79A7"),
                      breaks = c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid"),
                      labels = c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid" )) +
  labs(title = "Populations associated with OcMMP",
       x=NULL,
       y="Relative abundance (%)") +
  scale_y_log10(breaks=c(1e-4, 1e-3, 1e-2, 1e-1, 1), labels=c(1e-2, 1e-1, 1, 10, 100)) +
  coord_flip() +
  theme_classic()

######

sig_deep_lefse <- c("Eggerthella", "Lachnoclostridium", "Streptococcus")

view(sig_deep_lefse)

agg_deep_data %>%
  filter(taxonomy %in% sig_deep_lefse) %>%
  mutate(taxonomy = factor (taxonomy, levels=sig_deep_lefse)) %>%
  mutate(agg_rel_abund = agg_rel_abund + 1/25000) %>%
  ggplot(aes(x=taxonomy, y=agg_rel_abund, color=Cohort)) +
  geom_boxplot() +
  scale_color_manual (name=NULL,
                      values = c("#0072B2", "#CC79A7"),
                      breaks = c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid"),
                      labels = c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid" )) +
  labs(title = "Taxa with significantly higher relative abundances in OcMMP",
       x=NULL,
       y="Relative abundance (%)") +
  scale_y_log10(breaks=c(1e-4, 1e-3, 1e-2, 1e-1, 1), labels=c(1e-2, 1e-1, 1, 10, 100)) +
  coord_flip() +
  theme_classic()

ggsave("relative_abundance_higher_MMP.png")

####

sig_deep_lefse_lower <- c("Ruminococcaceae_ge", "UCG-005", "Roseburia",
"NK4A214_group", "Clostridia_UCG-014_ge", "Coprococcus", "Lachnospiraceae",
"Oscillospiraceae", "UCG-003", "Christensenellaceae_R-7_group", "UCG-002", "Oxalobacter"
)

agg_deep_data %>%
  filter(taxonomy %in% sig_deep_lefse_lower) %>%
  mutate(taxonomy = factor (taxonomy, levels=sig_deep_lefse_lower)) %>%
  mutate(agg_rel_abund = agg_rel_abund + 1/25000) %>%
```

```
 ggplot(aes(x=taxonomy, y=agg_rel_abund, color=Cohort)) +
 geom_boxplot() +
 scale_color_manual (name=NULL,
                values = c("#0072B2", "#CC79A7"),
                breaks = c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid"),
                labels = c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid" )) +
 labs(title = "Taxa with significantly lower relative abundances in OcMMP",
    x=NULL,
    y="Relative abundance (%)") +
 scale_y_log10(breaks=c(1e-4, 1e-3, 1e-2, 1e-1, 1), labels=c(1e-2, 1e-1, 1, 10, 100)) +
 coord_flip() +
 theme_classic()

ggsave("relative_abundance_lower_MMP.png")

agg_deep_data %>%
 filter(taxonomy %in% sig_deep_lefse) %>%
 mutate(taxonomy = factor (taxonomy, levels=sig_deep_lefse)) %>%
 nest(data = -Cohort) %>%
 mutate(summary_data=map(data, ~summary(.x$taxonomy) %>% tidy)) %>%
 unnest(cols=summary_data)

###
rel_abund_data <- read_excel(path="raw_data/inputfileMMPHC.opti_mcc.0.03.subsample-
1.relabund_v2.xlsx") %>%
 select(Group, Otu0021, Otu0055, Otu0062, Otu0063, Otu0071, Otu0072, Otu0086,
Otu0099, Otu0100, Otu0108, Otu0113, Otu0115, Otu0129, Otu0134, Otu0158, Otu0164,
Otu0168, Otu0194, Otu0215, Otu0237, Otu0277, Otu0357, Otu0359, Otu0441)

rel_abund_metadata <- inner_join(rel_abund_data, metadata, by = c("Group" = "sample"))

summary(rel_abund_metadata)

colnames(rel_abund_metadata)
glimpse(rel_abund_metadata)

rel_abund_metadata <- col_types=c(Group = "text", Otu0021 = "numeric", Otu0055 =
"numeric",
                        Otu0062 = "numeric", Otu0063 = "numeric", Otu0071 = "numeric",
                        Otu0072 = "numeric", Otu0086 = "numeric", Otu0099 = "numeric",
                        Otu0100 = "numeric", Otu0108 = "numeric", Otu0113 = "numeric",
                        Otu0115 = "numeric", Otu0129 = "numeric", Otu0134 = "numeric",
                        Otu0158 = "numeric", Otu0164 = "numeric", Otu0168 = "numeric",
                        Otu0194 = "numeric", Otu0215 = "numeric", Otu0237 = "numeric",
                        Otu0277 = "numeric", Otu0357 = "numeric", Otu0359 = "numeric",
                        Otu0441 = "numeric", ID = "text", `StudyID(STUDIOUS)` = "text",
                        `StudyID(Biopsy)` = "text", Age = "numeric", Sex = "text",
```

```
                    Cohort = "text", BMI = "numeric", Ethnicity = "text",
                    PPI_use = "text", Antacid = "text", Diet = "text",
                    `Duration of dx` = "numeric", OSDI = "numeric",
                    OSS_staining = "numeric", Biopsy = "text",
Current_Systemic_Immunotherapy = "text",
                    `Systemic _Immunosuppression_Ever` = "text", Systemic_Steroids =
"text",
                    DMARDS = "text", Biologics = "text", Tetracycline = "text")


library("xlsx")

write.xlsx(rel_abund_metadata, file = "rel_abund_metadata.xlsx")

lefse <- read_excel(path="raw_data/lefse.xlsx")

glimpse(lefse)

head(lefse)



ggplot(lefse, aes(x= reorder(Genus, LDA) , y=LDA, fill=Class))  +
 scale_fill_manual(name=NULL,
              values = c("#0072B2", "#CC79A7"),
              breaks = c("HC", "MMP"),
              labels = c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid" )) +
 geom_bar(stat="identity") +
 labs(y = expression(Log[10]~LDA~score),
    x = NULL)+
 coord_flip()+
 theme_classic() +
 theme(legend.position = "top")

ggsave("lefse.png")

#### order
agg_order_data <- inner_join(otu_data, taxonomy) %>%
 group_by(sample, order) %>%
 summarize(agg_rel_abund=sum(rel_abund)) %>%
 inner_join(., metadata, by = "sample") %>%
 ungroup()

top_order <- agg_order_data %>%
 group_by(order) %>%
 summarize(median = median(agg_rel_abund)) %>%
```

```
  arrange((desc(median))) %>%
  top_n(5, median) %>%
  pull(order)

view(top_order)

agg_order_data %>%
  filter(order %in% sig_order) %>%
  mutate(order=factor(order, levels = sig_order)) %>%
  mutate(agg_rel_abund=agg_rel_abund + 1/21000) %>%
  ggplot(aes(x=order, y=agg_rel_abund, color=Cohort)) +
  geom_hline(yintercept=1/10530, color="gray") +
  geom_boxplot() +
  scale_color_manual(name=NULL,
                     values=c("#0072B2", "#CC79A7"),
                     breaks=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid"),
                     labels=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid")) +
  labs(title="Order level",
       x=NULL,
       y="Relative abundance (%)") +
  scale_y_log10(breaks=c(1e-4, 1e-3, 1e-2, 1e-1, 1), labels=c(1e-2, 1e-1, 1, 10, 100)) +
  theme_classic() +
  coord_flip()

ggsave(file = "order.png", dpi=300)

library(broom)
library(purrr)

order_tests <- agg_order_data %>%
  nest(sample_data = c(-order)) %>%
  mutate(test = map(sample_data, ~tidy(kruskal.test(agg_rel_abund~Cohort, data=.)))) %>%
  unnest(test) %>%
  mutate(p.value.adj=p.adjust(p.value, method="BH")) %>%
  arrange(p.value.adj)

sig_order <- order_tests %>%
  filter(p.value.adj <= 0.05) %>%
  pull(order)

sig_order

#### family
agg_family_data <- inner_join(otu_data, taxonomy) %>%
  group_by(sample, family) %>%
  summarize(agg_rel_abund=sum(rel_abund)) %>%
  inner_join(., metadata, by = "sample") %>%
```

```
  ungroup()

top_family <- agg_family_data %>%
  group_by(family) %>%
  summarize(median = median(agg_rel_abund)) %>%
  arrange((desc(median))) %>%
  top_n(10, median) %>%
  pull(family)

view(top_family)

agg_family_data %>%
  filter(family %in% sig_family) %>%
  mutate(family=factor(family, levels = sig_family)) %>%
  mutate(agg_rel_abund=agg_rel_abund + 1/21000) %>%
  ggplot(aes(x=family, y=agg_rel_abund, color=Cohort)) +
  geom_hline(yintercept=1/10530, color="gray") +
  geom_boxplot() +
  scale_color_manual(name=NULL,
                     values=c("#0072B2", "#CC79A7"),
                     breaks=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid"),
                     labels=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid")) +
  labs(title="Family-level",
       x=NULL,
       y="Relative abundance (%)") +
  scale_y_log10(breaks=c(1e-4, 1e-3, 1e-2, 1e-1, 1), labels=c(1e-2, 1e-1, 1, 10, 100)) +
  theme_classic() +
  coord_flip()

ggsave("family.png")

library(broom)
library(purrr)

family_tests <- agg_family_data %>%
  nest(sample_data = c(-family)) %>%
  mutate(test = map(sample_data, ~tidy(kruskal.test(agg_rel_abund~Cohort, data=.)))) %>%
  unnest(test) %>%
  mutate(p.value.adj=p.adjust(p.value, method="BH")) %>%
  arrange(p.value.adj)

sig_family <- family_tests %>%
  filter(p.value.adj <= 0.05) %>%
  pull(family)

#### genus
agg_genus_data <- inner_join(otu_data, taxonomy) %>%
```

```
  group_by(sample, genus) %>%
  summarize(agg_rel_abund=sum(rel_abund)) %>%
  inner_join(., metadata, by = "sample") %>%
  ungroup()

top_genus <- agg_genus_data %>%
  group_by(genus) %>%
  summarize(median = median(agg_rel_abund)) %>%
  arrange((desc(median))) %>%
  top_n(10, median) %>%
  pull(genus)

view(top_genus)

agg_genus_data %>%
  filter(genus %in% sig_genus) %>%
  mutate(genus=factor(genus, levels = sig_genus)) %>%
  mutate(agg_rel_abund=agg_rel_abund + 1/21000) %>%
  ggplot(aes(x=genus, y=agg_rel_abund, color=Cohort)) +
  geom_hline(yintercept=1/10530, color="gray") +
  geom_boxplot() +
  scale_color_manual(name=NULL,
                     values=c("#0072B2", "#CC79A7"),
                     breaks=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid"),
                     labels=c("Healthy volunteer", "Ocular Mucous Membrane Pemphigoid")) +
  labs(title="Genus-level",
       x=NULL,
       y="Relative abundance (%)") +
  scale_y_log10(breaks=c(1e-4, 1e-3, 1e-2, 1e-1, 1), labels=c(1e-2, 1e-1, 1, 10, 100)) +
  theme_classic() +
  coord_flip()

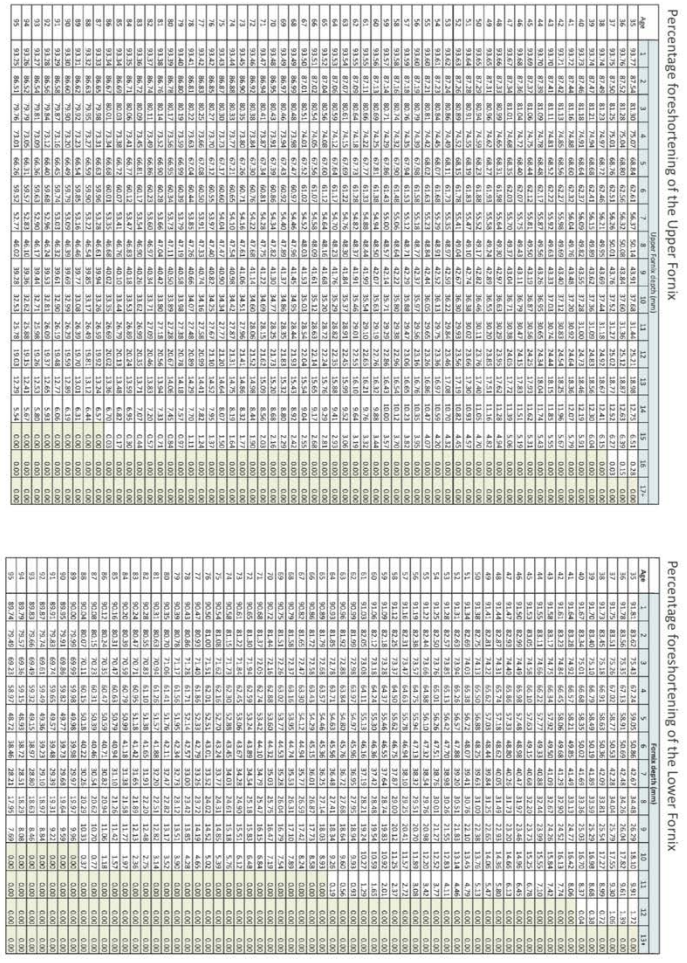
ggsave("genus.png")

library(broom)
library(purrr)

genus_tests <- agg_genus_data %>%
  nest(sample_data = c(-genus)) %>%
  mutate(test = map(sample_data, ~tidy(kruskal.test(agg_rel_abund~Cohort, data=.)))) %>%
  unnest(test) %>%
  mutate(p.value.adj=p.adjust(p.value, method="BH")) %>%
  arrange(p.value.adj)

sig_genus <- genus_tests %>%
  filter(p.value.adj <= 0.05) %>%
  pull(genus)
```

*Supplementary material 2.3 Cicatrising Conjunctivitis Assessment Tool reproduced from Ong et al*[50].

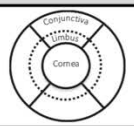
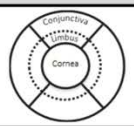
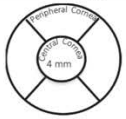
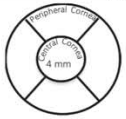
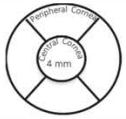
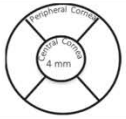


Percentage foreshortening of the Upper Fornix

| Age | Upper Fornix depth (mm) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 15 | 93.76 | 87.54 | 81.32 | 75.10 | 68.88 | 62.66 | 56.44 | 50.22 | 44.00 | 37.78 | 31.56 | 25.34 | 19.12 | 12.90 | 6.68 | 0.46 |
| 16 | 93.76 | 87.54 | 81.32 | 75.10 | 68.88 | 62.66 | 56.44 | 50.22 | 43.84 | 37.60 | 31.36 | 25.12 | 18.87 | 12.63 | 6.38 | 0.13 |
| 17 | 93.76 | 87.54 | 81.32 | 75.10 | 68.88 | 62.66 | 56.44 | 50.22 | 43.84 | 37.60 | 31.36 | 25.12 | 18.87 | 12.63 | 6.38 | 0.13 |
| 18 | 93.75 | 87.54 | 81.32 | 75.10 | 68.88 | 62.66 | 56.44 | 50.22 | 43.84 | 37.60 | 31.36 | 25.12 | 18.87 | 12.63 | 6.38 | 0.13 |
| 19 | 93.75 | 87.54 | 81.32 | 75.10 | 68.88 | 62.66 | 56.44 | 50.22 | 43.84 | 37.60 | 31.36 | 25.12 | 18.87 | 12.63 | 6.38 | 0.13 |
| 20 | 93.74 | 87.45 | 81.17 | 74.89 | 68.60 | 62.32 | 56.04 | 49.76 | 43.48 | 37.20 | 30.92 | 24.64 | 18.36 | 12.07 | 5.79 | 0.00 |
| 21 | 93.72 | 87.44 | 81.16 | 74.88 | 68.60 | 62.32 | 56.04 | 49.76 | 43.48 | 37.20 | 30.92 | 24.64 | 18.36 | 12.07 | 5.79 | 0.00 |
| 22 | 93.70 | 87.39 | 81.09 | 74.78 | 68.48 | 62.17 | 55.87 | 49.56 | 43.26 | 36.95 | 30.65 | 24.34 | 18.04 | 11.74 | 5.43 | 0.00 |
| 23 | 93.70 | 87.39 | 81.09 | 74.78 | 68.48 | 62.17 | 55.87 | 49.56 | 43.26 | 36.95 | 30.65 | 24.34 | 18.04 | 11.74 | 5.43 | 0.00 |
| 24 | 93.70 | 87.39 | 81.09 | 74.78 | 68.48 | 62.17 | 55.87 | 49.56 | 43.26 | 36.95 | 30.65 | 24.34 | 18.04 | 11.74 | 5.43 | 0.00 |
| 25 | 93.69 | 87.37 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 26 | 93.69 | 87.37 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 27 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 28 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 29 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 30 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 31 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 32 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 33 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 34 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 35 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 36 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 37 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 38 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 39 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 40 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 41 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 42 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 43 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 44 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 45 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 46 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 47 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 48 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 49 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 50 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 51 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 52 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 53 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 54 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 55 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 56 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 57 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 58 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 59 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 60 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 61 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 62 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 63 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 64 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 65 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 66 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 67 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 68 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 69 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 70 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 71 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 72 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 73 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 74 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 75 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 76 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 77 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 78 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 79 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 80 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 81 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 82 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 83 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 84 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 85 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 86 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 87 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 88 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 89 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 90 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 91 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 92 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 93 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 94 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |
| 95 | 93.68 | 87.36 | 81.04 | 74.71 | 68.39 | 62.06 | 55.73 | 49.40 | 43.08 | 36.75 | 30.42 | 24.08 | 17.75 | 11.42 | 5.08 | 0.00 |

Percentage foreshortening of the Lower Fornix

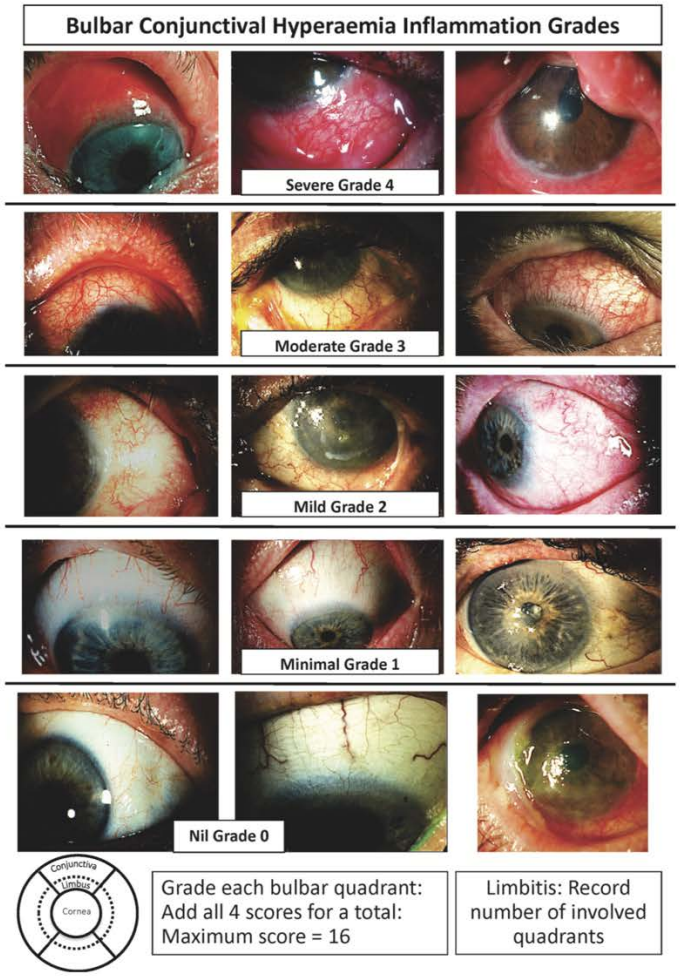
| Age | Fornix depth (mm) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 15 | 91.78 | 83.56 | 75.33 | 67.11 | 58.89 | 50.67 | 42.44 | 34.22 | 26.00 | 17.78 | 9.55 | 1.33 | 0.00 | 0.00 | 0.00 | 0.00 |
| 16 | 91.78 | 83.56 | 75.33 | 67.11 | 58.89 | 50.67 | 42.44 | 34.22 | 26.00 | 17.78 | 9.55 | 1.33 | 0.00 | 0.00 | 0.00 | 0.00 |
| 17 | 91.78 | 83.56 | 75.33 | 67.11 | 58.89 | 50.67 | 42.44 | 34.22 | 26.00 | 17.78 | 9.55 | 1.33 | 0.00 | 0.00 | 0.00 | 0.00 |
| 18 | 91.75 | 83.51 | 75.26 | 67.02 | 58.77 | 50.53 | 42.28 | 34.04 | 25.79 | 17.55 | 9.30 | 1.06 | 0.00 | 0.00 | 0.00 | 0.00 |
| 19 | 91.75 | 83.51 | 75.26 | 67.02 | 58.77 | 50.53 | 42.28 | 34.04 | 25.79 | 17.55 | 9.30 | 1.06 | 0.00 | 0.00 | 0.00 | 0.00 |
| 20 | 91.67 | 83.34 | 75.01 | 66.68 | 58.35 | 50.02 | 41.68 | 33.35 | 25.01 | 16.70 | 8.37 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 |
| 21 | 91.64 | 83.28 | 74.92 | 66.57 | 58.21 | 49.85 | 41.49 | 33.13 | 24.77 | 16.42 | 8.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 22 | 91.55 | 83.13 | 74.72 | 66.32 | 57.91 | 49.51 | 41.10 | 32.70 | 24.29 | 15.88 | 7.47 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 23 | 91.55 | 83.13 | 74.72 | 66.32 | 57.91 | 49.51 | 41.10 | 32.70 | 24.29 | 15.88 | 7.47 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 24 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 25 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 26 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 27 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 28 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 29 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 30 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 31 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 32 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 33 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 34 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 35 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 36 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 37 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 38 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 39 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 40 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 41 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 42 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 43 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 44 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 45 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 46 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 47 | 91.53 | 83.10 | 74.68 | 66.26 | 57.84 | 49.42 | 40.99 | 32.57 | 24.15 | 15.73 | 7.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 48 | 91.44 | 83.07 | 74.51 | 65.74 | 57.18 | 48.62 | 40.05 | 31.49 | 23.03 | 14.58 | 6.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 49 | 91.41 | 82.91 | 74.22 | 65.62 | 57.03 | 48.44 | 39.84 | 31.25 | 22.65 | 14.06 | 5.47 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 50 | 91.34 | 82.82 | 74.11 | 65.51 | 56.91 | 48.31 | 39.70 | 31.09 | 22.48 | 13.86 | 4.96 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 51 | 91.34 | 82.82 | 74.11 | 65.51 | 56.91 | 48.31 | 39.70 | 31.09 | 22.48 | 13.86 | 4.96 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 52 | 91.31 | 82.63 | 73.94 | 65.25 | 56.57 | 47.88 | 39.20 | 30.51 | 21.83 | 13.14 | 4.46 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 53 | 91.31 | 82.63 | 73.94 | 65.25 | 56.57 | 47.88 | 39.20 | 30.51 | 21.83 | 13.14 | 4.46 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 54 | 91.28 | 82.57 | 73.85 | 65.13 | 56.42 | 47.70 | 38.98 | 30.26 | 21.55 | 12.83 | 4.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 55 | 91.28 | 82.57 | 73.85 | 65.13 | 56.42 | 47.70 | 38.98 | 30.26 | 21.55 | 12.83 | 4.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 56 | 91.19 | 82.38 | 73.57 | 64.75 | 55.94 | 47.13 | 38.33 | 29.53 | 20.73 | 11.93 | 3.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 57 | 91.14 | 82.31 | 73.47 | 64.63 | 55.79 | 46.94 | 38.10 | 29.25 | 20.41 | 11.57 | 2.72 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 58 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 59 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 60 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 61 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 62 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 63 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 64 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 65 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 66 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 67 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 68 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 69 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 70 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 71 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 72 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 73 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 74 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 75 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 76 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 77 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 78 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 79 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 80 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 81 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 82 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 83 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 84 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 85 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 86 | 91.06 | 82.17 | 73.28 | 64.39 | 55.50 | 46.61 | 37.73 | 28.84 | 19.95 | 11.06 | 2.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.

| Right Eye | | Left Eye |
|---|---|---|
| **Inflammation*** | | |
| None (0) Minimal (1) Mild (2) Moderate (3) Severe (4)   | **Conjunctival Inflammation (score each quadrant using photos)† 'C' is the cornea and not graded.** | None (0) Minimal (1) Mild (2) Moderate (3) Severe (4)   |
| None ☐ One ☐ Two ☐ Three ☐ Four ☐ | **Limbitis – Number of quadrant(s)†** | None ☐ One ☐ Two ☐ Three ☐ Four ☐ |
| **Scarring*** | | |
| No ☐ Yes ☐ | **Sub-conjunctival fibrosis** | No ☐ Yes ☐ |
| Absent (0) ☐ <50% (1) ☐ 50-100% (2) ☐ | **Lower fornix symblepharon** | Absent (0) ☐ <50% (1) ☐ 50-100% (2) ☐ |
| Absent (0) ☐ <50% (1) ☐ 50-100% (2) ☐ | **Upper fornix symblepharon** | Absent (0) ☐ <50% (1) ☐ 50-100% (2) ☐ |
| Lower: _____ mm Upper: _____ mm | **Central Fornix Depth‡** | Lower: _____ mm Upper: _____ mm |
| **Morbidity*** | | |
| No ☐ Yes ☐ | **Distichiasis§** | No ☐ Yes ☐ |
| No ☐ Yes ☐ | **Keratinisation¤** Bulbar conjunctiva | No ☐ Yes ☐ |
| No ☐ Yes ☐ | Lid-margin/tarsal conj. | No ☐ Yes ☐ |
| No ☐ Yes ☐ | Cornea | No ☐ Yes ☐ |
| (0) ☐ (1) ☐ (2) ☐ (3) ☐ (4) ☐   | **Peripheral vascularisation‖ (number of quadrants**)** | (0) ☐ (1) ☐ (2) ☐ (3) ☐ (4) ☐   |
| Absent (0) ☐ Present (1) ☐ | **Central 4 mm vascularisation** | Absent (0) ☐ Present (1) ☐ |
| (0) ☐ (1) ☐ (2) ☐ (3) ☐ (4) ☐   | **Peripheral opacity¤ (number of quadrants**)** | (0) ☐ (1) ☐ (2) ☐ (3) ☐ (4) ☐   |
| Absent (0) ☐ Present (5) ☐ | **Central 4 mm opacity** | Absent (0) ☐ Present (5) ☐ |

**Definitions**

* Any one category of inflammation, scarring and morbidity can be used on its own.
† Refer to bulbar conjunctival hyperaemia grading panel of photographs.
‡ Using the fornix depth measurement ruler; record measurements to the nearest mm: round UP to the nearest mm if lid margin is over halfway and round DOWN if lid margin is under halfway between two marks.
§ Lashes arising from meibomian gland orifices / posterior lamellae of the lid.
¤ White, rough, shiny non-wetting deposits often with foam on the surface.
‖ Central cornea refers to central 4 mm of the cornea; Peripheral cornea refers to area outside central 4mm zone; use narrow slit lamp beam to delineate central 4 mm zone.
** Assess each quadrant as per diagram; quadrant scored as positive with any involvement.

©JKGD/SR/DM/HSO/2019 




©JKGD/SR/DM/HSO/2019 

## Supplementary for Chapter 3

*Supplementary Table 3.1 Pilot data comparing DNA yield of dry vs. pre-moistened swab at different time points*

| Inoculum | Condition | Duration | DNA concentration (ng/µl) |
|---|---|---|---|
| No inoculum (negative control) | Dry swab | 30 mins | 1.27 |
| With inoculum | Dry swab | 30 mins | 10.9 |
| With inoculum | Pre-moistened swab | 30 mins | 1.97 |
| No inoculum (negative control) | Dry swab | 12 hours | 65.6 |
| With inoculum | Dry swab | 12 hours | 96.4 |
| With inoculum | Pre-moistened swab | 12 hours | 22.2 |

Freshly enucleated porcine eyes (*Sus scrofa domestica*) were obtained as a by-product of the meat industry and transported to the laboratory under storage at 4$^{o}$C. Each eye was disinfected with Povidone-iodine 10% w/w for 1 minute, followed by two rinses of sterile 0.9% Sodium Chloride for 1 minute, and placed in an individual chamber of sterile 6-well culture plate (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany). Using stereoscopic surgical loupes, a 4mm trephine punch (Acu-Punch®, Acuderm, Fort Lauderdale, USA) was used to create a single central anterior stromal corneal lesion (with debridement of the central 4mm). Each eye was then inoculated with 20 µL of 1x10$^{5}$ CFU/ml each of the mock community (*Enterococcus avium*, *Staphylococcus aureus*, and *Klebsiella pneumoniae*). Negative control eyes were not inoculated with the mock community. The area was re-sampled with the respective swab conditions (dry swab vs. pre-moistened swab with sterile 0.9% Sodium Chloride) at two time points (30 minutes vs. 12 hours) using Purflock® Ultra Standard (MWE Medical Wire, Corsham, UK) at room temperature, and placed immediately into a ZR BashingBead™ Lysis Tube containing 750 µl of DNA Shield™ (Zymo Research, Irvine, CA, USA) and stored at -80$^{o}$C until DNA extraction. DNA was extracted using ZymoBIOMICS DNA Miniprep kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. DNA concentration was determined fluorometrically using a Qubit dsDNA high-sensitivity assay (Thermo Fisher Scientific, Waltham, MA, USA).

*Supplementary material 3.1 Bioinformatics script written in python for clustering of reference databases using the CD-HIT software.*

*cd-hit_clustering.py written by Liying Low*

```
# coding: UTF-8
#! /usr/bin/python
import sys
import re
import time
import os

def main():
   script = sys.argv[0]
   for filename in sys.argv[1:]:
      print(filename)
      x = filename.rsplit ('.', 1)[0]
      #the .rsplit function splits the filename at the '.' and returns the first chunk before the '.', because I have specified it to
be [0]
      print(x)
      command = "cd-hit-est -i " + str(filename) + " -o " + str(x) + "_cluster_100.fasta -c 1.0 -T 8 -M 62900 -d 250" + " ; "
      os.system(command)
      #print(command)
      command = "cd-hit-est -i " + str(filename) + " -o " + str(x) + "_cluster_99.fasta -c 0.99 -T 8 -M 62900 -d 250" + " ; "
      os.system(command)
      #print(command)
      command = "cd-hit-est -i " + str(filename) + " -o " + str(x) + "_cluster_97.fasta -c 0.97 -T 8 -M 62900 -d 250" + " ; "
      os.system(command)
      #print(command)
main()
```

*Supplementary material 3.2 Bioinformatics script written on python for read alignment using minimap2.*

*Nanopore_minimap2_submitjob_database.py written by Liying Low*

```
# coding: utf-8
#! /usr/bin/python
import os
import re
import time
import sys

def main():
   script = sys.argv[0]
   for database in sys.argv[1:]:
      print(database)
      db = database.rsplit('.',1)[0]
      print(db)
      filepath = "/data/gridion/swabcomparison/qcat_output_folder/"
      filelist = os.listdir(filepath)
      print(filelist)
      filelist = []
      count = 0
      filelist = os.listdir(filepath)
      print (filelist)
      for fname in filelist:
         print(fname)
         file = (fname)
         x = file.strip (".fastq")
         print(x)
         command = "minimap2 -K 100M -t 8 -ax map-ont " + str(database) + " " + str(filepath) + str(fname) + " > " +
"alignments/" + str(x) + "_aligned_" + str(db) + ".sam" + " ; "
         os.system(command)
         #print(command)
         print("Finished")

main()
```

*Supplementary material 3.3 Bioinformatics script written in python for read filtering, sorting, and indexing using SAMtools.*

*samtools_submit_job_divided_filtered_not_q30_v2.py written by Liying Low*

```
 # coding: utf-8
#! /usr/bin/python
import os
import re
import time

alignmentpath = "./alignments/"
print (alignmentpath)

bamoutputpath = "./bam_output_filtered_not_q30_ncbi/"
print (bamoutputpath)

command = "mkdir " + str(bamoutputpath)
os.system(command)
#print(command)

txtoutputpath = "./txt_output_filtered_not_q30_ncbi/"

command = "mkdir " + str(txtoutputpath)
os.system(command)
#print(command)


txtoutputpath_divided = "./txt_output_filtered_divided_not_q30/"
command = "mkdir " + str(txtoutputpath_divided)
os.system(command)
#print(command)

alignmentlist = os.listdir(alignmentpath)
print (alignmentlist)


for aname in alignmentlist:
   print(aname)
   alignment = (aname)
   y = re.sub(".sam", "", aname)
   print(y)
   command = "samtools view --threads 8 -b -F 2308 " + str(alignmentpath) + str (aname) + " -o " + str(bamoutputpath) +
str(y) + "_combinedflags_notq30.bam" + " ; "
   os.system(command)
#    print(command)

#SAM flag 2308: read unmapped (0X4), not primary alignment (0x100) and supplementary alignment (0x800)

   command = "samtools sort " + str(bamoutputpath) + str(y) + "_combinedflags_notq30.bam" + " -o " +
str(bamoutputpath) + str(y) + "_combinedflags_notq30_sorted.bam" + " ; "
```

```
    os.system(command)
#    print(command)

    command = "samtools index " + str(bamoutputpath) + str(y) + "_combinedflags_notq30_sorted.bam" + " ; "
    os.system(command)
#    print(command)

    command = "samtools idxstats " + str(bamoutputpath) + str(y) + "_combinedflags_notq30_sorted.bam" + " > " +
str(txtoutputpath) + str(y) + "_combinedflags_notq30_sorted.txt" + " ; "
    os.system(command)
#    print(command)

#    keyword = 'RDP'
#    if keyword in aname:
#        command = "python3 divide_by_underscore_RDP.py " + str(txtoutputpath) + str(y) +
"_combinedflags_notq30_sorted.txt" + " > " + str(txtoutputpath_divided) + str(y) +
"_combinedflags_notq30_sorted_dividedbyunderscores.txt" + " ; "
#        os.system(command)
#        print(command)

#    keyword_1 = 'SILVA'
#    if keyword_1 in aname:
#        command = "sed '$d' " + str(txtoutputpath) + str(y) + "_combinedflags_notq30_sorted.txt"  + " > " +
str(txtoutputpath) + str(y) + "_combinedflags_notq30_sorted_last_line_removed.txt ; "
#        os.system(command)
#        print(command)
#        command = "rm " + str(txtoutputpath) + str(y) + " ; "
#        os.system(command)
#        print(command)

#        command = "python3 divide_by_underscore_SILVA.py " + str(txtoutputpath) + str(y) +
"_combinedflags_notq30_sorted_last_line_removed.txt" + " > " + str(txtoutputpath_divided) + str(y) +
"_combinedflags_notq30_sorted_last_line_removed_dividedbyunderscores.txt" + " ; "
#        os.system(command)
#        print(command)

    keyword_2 = 'ncbi'
    if keyword_2 in aname:
        command = "python3 divide_by_underscore.py " + str(txtoutputpath) + str(y) +  "_combinedflags_notq30_sorted.txt"
+ " > " + str(txtoutputpath_divided) + str(y) + "_combinedflags_notq30_sorted_dividedbyunderscores.txt" + " ; "
        os.system(command)
#        print(command)
        print("Finished")
```

*Supplementary material 3.4 Bioinformatics script written in python for removing the last redundant line on SILVA reference database.*

*SILVA_remove_last_line.py written by Liying Low*

```
# coding: UTF-8
#! /usr/bin/python

import os
import re
import sys



script = sys.argv[0]
inputpath = sys.argv[1]
print(inputpath)
outputpath = sys.argv[2]
print(outputpath)
command = "mkdir " + str(outputpath)
os.system(command)
#print(command)

filelist = os.listdir(inputpath)
#print(filelist)

for filename in filelist:
   print(filename)
   x = re.sub(".txt", "", filename)
   print(x)

   command = "sed '$d' " + str(inputpath) + "/" + str(filename) + " > " + str(outputpath) + "/" + str(x) +
"_last_line_removed.txt ;"
   os.system(command)
   #print(command)
```

*Supplementary material 3.5 Python script for dividing the taxonomic assignments into genus and species groupings.*

*divide_by_underscore.py script obtained from Prof Nicholas Loman.*

```
import sys
import csv

for rec in csv.reader(open(sys.argv[1]), dialect='excel-tab'):
   #print (rec)

   header_line = rec[0]

   columns = header_line.split("_")
   if len(columns) > 2:
      genus = columns[2]
      species = columns[2] + " " + columns[3]

      print ("\t".join([genus, species, rec[1], rec[2], rec[3]]))
```

*Supplementary material 3.6 Python script for calculating relative abundances of taxa from 16S nanopore output.*

*nanopore_MK_genus_species_assignment_by_directory.py written by Liying Low*

```
import pandas as pd
import os
import re

inputpath = "/Users/lowl/Documents/Research_2020/Nanopore_Endophthalmitis/Zymo_mock/txt_output_filtered_divided_not_q3
genus_outputpath = "/Users/lowl/Documents/Research_2020/Nanopore_Endophthalmitis/Zymo_mock/Genus/"

species_outputpath = "/Users/lowl/Documents/Research_2020/Nanopore_Endophthalmitis/Zymo_mock/Species/"

filelist = os.listdir(inputpath)

for filename in filelist:
#    print(inputpath+filename)
    x = re.sub(".txt", "", filename)
#    print(x)
    y = str(inputpath) + str(filename)
    print(y)
    g = str(genus_outputpath) + str(filename)
    s = str(species_outputpath) + str(filename)

    data = pd.read_csv(y, sep='\t', names=["Genus", "Species","Seq_Length", "#Mapped", "#Unmapped"])
    #making data frame from csv file, add header namesdat)
    data.sort_values("#Mapped", axis=0, ascending = False, inplace = True, na_position='last')
    del data['Seq_Length']
    del data['#Unmapped']
    data.sort_values("#Mapped", axis=0, ascending = False, inplace = True, na_position='last')
    Total_mapped = data['#Mapped'].sum()
    print(Total_mapped)

    data = data.assign(Relative_abundance = lambda x: (x['#Mapped']/Total_mapped)*100)

    summary_species = data.groupby(['Species']).sum()

    summary_species.sort_values("#Mapped", axis=0, ascending = False, inplace = True, na_position='last')

    summary_genus = data.groupby(['Genus']).sum()

    summary_genus.sort_values("#Mapped", axis=0, ascending = False, inplace = True, na_position='last')

    summary_species.to_csv(s, encoding='utf-8')

    summary_genus.to_csv(g, encoding='utf-8')
```

*Supplementary Table 3.2 Comparing the effects of using NCBI 16S, RDP and SILVA databases at different levels of clustering (97%, 99% and 100% similarity) on patient samples*

| **Patient A Affected Eye - Culture: Serratia marcescens** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **NCBI 16S** | | | | **RDP** | | | | **SILVA** | | | |
| **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** |
| 612600 | **97% similarity** | | | 612792 | **97% similarity** | | | 610279 | **97% similarity** | | |
| | *Enterobacter soli* | 425586 | 69.47 | | *Pectobacterium aroidearum* | 291442 | 47.56 | | *Serratia marcescens* | 561823 | 92.06 |
| | *Serratia rubidaea* | 60511 | 9.88 | | *Enterobacter soli* | 240393 | 39.23 | | *Lelliottia amnigena* | 6084 | 1.00 |
| | *Serratia liquefaciens* | 39320 | 6.42 | | *Serratia vespertilionis* | 18516 | 3.02 | | *Klebsiella pneumoniae* | 4185 | 0.69 |
| | *Gibbsiella dentisursi* | 27820 | 4.54 | | *Serratia rubidaea* | 17540 | 2.86 | | *Kluyvera ascorbata* | 4006 | 0.66 |
| | *Serratia odorifera* | 20522 | 3.35 | | *Gibbsiella dentisursi* | 14993 | 2.45 | | *Staphylococcus aureus* | 2345 | 0.38 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 613214 | **99% similarity** | | | 613244 | **99% similarity** | | | 610876 | **99% similarity** | | |
| | *Serratia marcescens* | 550188 | 89.72 | | *Serratia marcescens* | 555615 | 90.60 | | *Serratia marcescens* | 408294 | 66.84 |
| | *Klebsiella aerogenes* | 12622 | 2.06 | | *Kluyvera ascorbata* | 6236 | 1.02 | | *Serratia nematodiphila* | 150295 | 24.60 |
| | *Citrobacter gillenii* | 3770 | 0.61 | | *Citrobacter gillenii* | 4522 | 0.74 | | *Klebsiella pneumoniae* | 4160 | 0.68 |
| | *Raoultella ornithinolytica* | 3126 | 0.51 | | *Enterobacter soli* | 3404 | 0.56 | | *Kluyvera ascorbata* | 3463 | 0.57 |
| | *Kluyvera ascorbata* | 2573 | 0.42 | | *Citrobacter murliniae* | 3075 | 0.50 | | *Raoultella ornithinolytica* | 2250 | 0.37 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 613482 | **100% similarity** | | | 613385 | **100% similarity** | | | 610637 | **100% similarity** | | |
| | *Serratia marcescens* | 497995 | 81.18 | | *Serratia marcescens* | 499687 | 81.46 | | *Serratia marcescens* | 510738 | 83.64 |
| | *Serratia nematodiphila* | 60000 | 9.78 | | *Serratia nematodiphila* | 64090 | 10.45 | | *Escherichia coli* | 41634 | 6.82 |
| | *Klebsiella aerogenes* | 9247 | 1.51 | | *Kluyvera ascorbata* | 5039 | 0.82 | | *Klebsiella pneumoniae* | 5054 | 0.83 |
| | *Kluyvera ascorbata* | 1981 | 0.32 | | *Citrobacter gillenii* | 2624 | 0.43 | | *Serratia ureilytica* | 3399 | 0.56 |
| | *Cutibacterium acnes* | 1942 | 0.32 | | *Propionibacterium acnes* | 1942 | 0.32 | | *Serratia nematodiphila* | 2787 | 0.46 |

*Supplementary Table 3.2 (cont)*

**Patient A - Unaffected Eye**

| NCBI 16S | | | | RDP | | | | SILVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** |
| 28 | **97% similarity** | | | 28 | **97% similarity** | | | 29 | **97% similarity** | | |
| | *Enterobacter soli* | 7 | 25.00 | | *Enterobacter soli* | 4 | 14.29 | | *Serratia marcescens* | 7 | 24.14 |
| | *Streptococcus cristatus* | 3 | 10.71 | | *Pectobacterium aroidearum* | 4 | 14.29 | | *Clostridiales bacterium* | 2 | 6.90 |
| | *Enterococcus hirae* | 2 | 7.14 | | *Clostridium aldrichii* | 2 | 7.14 | | *Streptococcus cristatus* | 2 | 6.90 |
| | *Peptoniphilus duerdenii* | 2 | 7.14 | | *Peptoniphilus duerdenii* | 2 | 7.14 | | *Peptoniphilus duerdenii* | 2 | 6.90 |
| | *Streptococcus iniae* | 2 | 7.14 | | *Enterococcus phoeniculicola* | 2 | 7.14 | | *Streptococcus parauberis* | 1 | 3.45 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 28 | **99% similarity** | | | 28 | **99% similarity** | | | 29 | **99% similarity** | | |
| | *Serratia marcescens* | 8 | 28.57 | | *Serratia marcescens* | 8 | 28.57 | | *Serratia marcescens* | 5 | 17.24 |
| | *Streptococcus cristatus* | 3 | 10.71 | | *Streptococcus cristatus* | 3 | 10.71 | | *Streptococcus cristatus* | 3 | 10.34 |
| | *Hungateiclostridium cellulolyticum* | 2 | 7.14 | | *Streptococcus parauberis* | 2 | 7.14 | | *Enterococcus devriesei* | 2 | 6.90 |
| | *Enterococcus avium* | 2 | 7.14 | | *Enterococcus avium* | 2 | 7.14 | | *Clostridiales bacterium* | 2 | 6.90 |
| | *Streptococcus parauberis* | 2 | 7.14 | | *Peptoniphilus duerdenii* | 2 | 7.14 | | *Peptoniphilus duerdenii* | 2 | 6.90 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 28 | **100% similarity** | | | 28 | **100% similarity** | | | 29 | **100% similarity** | | |
| | *Serratia marcescens* | 6 | 21.43 | | *Serratia marcescens* | 7 | 25.00 | | *Serratia marcescens* | 7 | 24.14 |
| | *Streptococcus cristatus* | 3 | 10.71 | | *Streptococcus cristatus* | 3 | 10.71 | | *Streptococcus cristatus* | 3 | 10.34 |
| | *Hungateiclostridium cellulolyticum* | 2 | 7.14 | | *Enterococcus avium* | 2 | 7.14 | | *Enterococcus avium* | 3 | 10.34 |
| | *Streptococcus parauberis* | 2 | 7.14 | | *Acetivibrio cellulolyticus* | 2 | 7.14 | | *Peptoniphilus duerdenii* | 2 | 6.90 |
| | *Enterococcus avium* | 2 | 7.14 | | *Streptococcus parauberis* | 2 | 7.14 | | *Streptococcus parauberis* | 2 | 6.90 |

*Supplementary Table 3.2 (cont)*

| Patient A - Negative control swab | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **NCBI 16S** | | | | **RDP** | | | | **SILVA** | | | |
| **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** |
| 26 | **97% similarity** | | | 26 | **97% similarity** | | | 27 | **97% similarity** | | |
| | *Enterobacter soli* | 13 | 50.00 | | *Enterobacter soli* | 12 | 46.15 | | *Serratia marcescens* | 6 | 22.22 |
| | *Enterococcus hirae* | 4 | 15.38 | | *Staphylococcus aureus* | 3 | 11.54 | | *Klebsiella quasipneumoniae* | 4 | 14.81 |
| | *Staphylococcus aureus* | 3 | 11.54 | | *Melissococcus plutonius* | 1 | 3.85 | | *Enterococcus faecium* | 3 | 11.11 |
| | *Moraxella osloensis* | 1 | 3.85 | | *Corynebacterium mastitidis* | 1 | 3.85 | | *Staphylococcus aureus* | 3 | 11.11 |
| | *Methylorubrum thiocyanatum* | 1 | 3.85 | | *Methylobacterium thiocyanatum* | 1 | 3.85 | | *Klebsiella pneumoniae* | 2 | 7.41 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 26 | **99% similarity** | | | 26 | **99% similarity** | | | 26 | **99% similarity** | | |
| | *Klebsiella pneumoniae* | 6 | 23.08 | | *Klebsiella pneumoniae* | 7 | 26.92 | | *Klebsiella pneumoniae* | 6 | 23.08 |
| | *Serratia marcescens* | 5 | 19.23 | | *Serratia marcescens* | 5 | 19.23 | | *Serratia marcescens* | 5 | 19.23 |
| | *Enterococcus avium* | 4 | 15.38 | | *Enterococcus avium* | 4 | 15.38 | | *Enterococcus avium* | 3 | 11.54 |
| | *Staphylococcus aureus* | 2 | 7.69 | | *Staphylococcus aureus* | 2 | 7.69 | | *Staphylococcus aureus* | 2 | 7.69 |
| | *Staphylococcus saccharolyticus* | 1 | 3.85 | | *Corynebacterium mastitidis* | 1 | 3.85 | | *synthetic construct* | 1 | 3.85 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 26 | **100% similarity** | | | 26 | **100% similarity** | | | 27 | **100% similarity** | | |
| | *Klebsiella pneumoniae* | 6 | 23.08 | | *Serratia marcescens* | 5 | 19.23 | | *Klebsiella pneumoniae* | 6 | 22.22 |
| | *Enterococcus avium* | 4 | 15.38 | | *Enterococcus avium* | 4 | 15.38 | | *Serratia marcescens* | 4 | 14.81 |
| | *Serratia marcescens* | 4 | 15.38 | | *Klebsiella quasipneumoniae* | 3 | 11.54 | | *Staphylococcus aureus* | 3 | 11.11 |
| | *Staphylococcus aureus* | 2 | 7.69 | | *Klebsiella pneumoniae* | 3 | 11.54 | | *Enterococcus avium* | 3 | 11.11 |
| | *Kluyvera ascorbata* | 1 | 3.85 | | *Staphylococcus argenteus* | 2 | 7.69 | | *bacterium NTL235* | 1 | 3.70 |

*Supplementary Table 3.2 (cont)*

| Patient B Affected Eye - Culture: Negative | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **NCBI 16S** | | | | **RDP** | | | | **SILVA** | | | |
| **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** |
| 92702 | **97% similarity** | | | 92725 | **97% similarity** | | | 89250 | **97% similarity** | | |
| | *Bacillus velezensis* | 35498 | 38.29 | | *Bacillus licheniformis* | 33919 | 36.58 | | *Bacillus subtilis* | 32176 | 36.58 |
| | *Staphylococcus aureus* | 12430 | 13.41 | | *Staphylococcus aureus* | 12497 | 13.48 | | *Staphylococcus caprae* | 10072 | 13.48 |
| | *Aggregatibacter segnis* | 4238 | 4.57 | | *Aggregatibacter segnis* | 4558 | 4.92 | | *Aggregatibacter segnis* | 5061 | 4.92 |
| | *Cutibacterium acnes* | 3018 | 3.26 | | *Propionibacterium acnes* | 3023 | 3.26 | | *Telluria mixta* | 2996 | 3.26 |
| | *Moraxella osloensis* | 2658 | 2.87 | | *Erwinia billingiae* | 2666 | 2.88 | | *Propionibacterium acnes* | 2985 | 2.88 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 92786 | **99% similarity** | | | 92813 | **99% similarity** | | | 84071 | **99% similarity** | | |
| | *Bacillus halotolerans* | 29589 | 31.89 | | *Bacillus amyloliquefaciens* | 31441 | 33.88 | | *Bacillus subtilis* | 26535 | 33.88 |
| | *Staphylococcus saccharolyticus* | 12063 | 13.00 | | *Staphylococcus caprae* | 11801 | 12.71 | | *Staphylococcus epidermidis* | 9352 | 12.71 |
| | *Bacillus velezensis* | 5794 | 6.24 | | *Aggregatibacter segnis* | 4505 | 4.85 | | *Aggregatibacter segnis* | 5092 | 4.85 |
| | *Aggregatibacter segnis* | 4189 | 4.51 | | *Bacillus subtilis;* | 3995 | 4.30 | | *Propionibacterium acnes* | 2858 | 4.30 |
| | *Cutibacterium acnes* | 3002 | 3.24 | | *Propionibacterium acnes* | 3017 | 3.25 | | *Erwinia billingiae* | 2308 | 3.25 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 92810 | **100% similarity** | | | 92844 | **100% similarity** | | | 86792 | **100% similarity** | | |
| | *Bacillus subtilis* | 33188 | 35.76 | | *Bacillus subtilis* | 25842 | 11.22 | | *Bacillus subtilis* | 27889 | 11.22 |
| | *Staphylococcus caprae* | 5875 | 6.33 | | *Staphylococcus caprae* | 10413 | 9.72 | | *Staphylococcus capitis* | 8881 | 9.72 |
| | *Staphylococcus saccharolyticus* | 5206 | 5.61 | | *Bacillus amyloliquefaciens* | 9021 | 4.94 | | *Aggregatibacter segnis* | 5061 | 4.94 |
| | *Aggregatibacter segnis* | 4109 | 4.43 | | *Aggregatibacter segnis* | 4584 | 3.25 | | *Staphylococcus epidermidis* | 3054 | 3.25 |
| | *Cutibacterium acnes* | 2998 | 3.23 | | *Propionibacterium acnes* | 3018 | 0.00 | | *Massilia putida* | 2180 | 0.00 |

*Supplementary Table 3.2 (cont)*

**Patient B - Unaffected Eye**

| NCBI 16S | | | | RDP | | | | SILVA | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of reads | Cluster | #Mapped | Relative abundance | Number of reads | Cluster | #Mapped | Relative abundance | Number of reads | Cluster | #Mapped | Relative abundance |
| 1302 | **97% similarity** | | | 1301 | **97% similarity** | | | 1270 | **97% similarity** | | |
| | *Snodgrassella alvi* | 970 | 74.5 | | *Snodgrassella alvi* | 972 | 74.71 | | *Snodgrassella alvi* | 931 | 73.31 |
| | *Escherichia fergusonii* | 64 | 4.92 | | *Shigella sonnei* | 65 | 5.00 | | *Escherichia coli* | 57 | 4.49 |
| | *Anoxybacillus kamchatkensis* | 50 | 3.84 | | *Anoxybacillus kamchatkensis* | 50 | 3.84 | | *Anoxybacillus gonensis* | 47 | 3.70 |
| | *Thermicanus aegyptius* | 20 | 1.54 | | *Eikenella corrodens* | 22 | 1.69 | | *Vitreoscilla stercoraria* | 31 | 2.44 |
| | *Bacillus toyonensis* | 18 | 1.38 | | *Thermicanus aegyptius* | 20 | 1.54 | | *Thermicanus aegyptius* | 20 | 1.57 |
| 1300 | **99% similarity** | | | 1301 | **99% similarity** | | | 1236 | **99% similarity** | | |
| | *Snodgrassella alvi* | 948 | 72.92 | | *Snodgrassella alvi* | 958 | 73.64 | | *Snodgrassella alvi* | 900 | 72.82 |
| | *Escherichia fergusonii* | 55 | 4.23 | | *Shigella sonnei* | 60 | 4.61 | | *Escherichia coli* | 59 | 4.77 |
| | *Anoxybacillus flavithermus* | 45 | 3.46 | | *Anoxybacillus flavithermus* | 45 | 3.46 | | *Anoxybacillus pushchinoensis* | 35 | 2.83 |
| | *Thermicanus aegyptius* | 20 | 1.54 | | *Thermicanus aegyptius* | 20 | 1.54 | | *Thermicanus aegyptius* | 20 | 1.62 |
| | *Eikenella corrodens* | 20 | 1.54 | | *Eikenella corrodens* | 19 | 1.46 | | *Eikenella corrodens* | 16 | 1.29 |
| 1301 | **100% similarity** | | | 1302 | **100% similarity** | | | 1230 | **100% similarity** | | |
| | *Snodgrassella alvi* | 935 | 71.87 | | *Snodgrassella alvi* | 950 | 72.96 | | *Snodgrassella alvi* | 861 | 70.00 |
| | *Escherichia fergusonii* | 48 | 3.69 | | *Escherichia/Shigella flexneri* | 49 | 3.76 | | *Anoxybacillus flavithermus* | 51 | 4.15 |
| | *Anoxybacillus flavithermus* | 35 | 2.69 | | *Anoxybacillus flavithermus* | 45 | 3.46 | | *Escherichia coli* | 51 | 4.15 |
| | *Eikenella corrodens* | 24 | 1.84 | | *Thermicanus aegyptius* | 20 | 1.54 | | *Eikenella corrodens* | 37 | 3.01 |
| | *Thermicanus aegyptius* | 20 | 1.54 | | *Eikenella corrodens* | 18 | 1.38 | | *Thermicanus aegyptius* | 20 | 1.63 |

*Supplementary Table 3.2 (cont)*

| Patient B - Negative control swab | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **NCBI 16S** | | | | **RDP** | | | | **SILVA** | | | |
| **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** | **Number of reads** | **Cluster** | **#Mapped** | **Relative abundance** |
| 6 | **97% similarity** | | | 6 | **97% similarity** | | | 6 | **97% similarity** | | |
| | *Enterobacter soli* | 2 | 33.33 | | *Enterobacter soli* | 3 | 50.00 | | *Klebsiella quasipneumoniae* | 2 | 33.33 |
| | *Staphylococcus aureus* | 2 | 33.33 | | *Staphylococcus aureus* | 2 | 33.33 | | *Staphylococcus aureus* | 2 | 33.33 |
| | *Salmonella enterica* | 1 | 16.67 | | *Enterococcus faecium* | 1 | 16.67 | | *Klebsiella pneumoniae* | 1 | 16.67 |
| | *Enterococcus hirae* | 1 | 16.67 | | | | | | *Enterococcus faecium* | 1 | 16.67 |
| 6 | **99% similarity** | | | 6 | **99% similarity** | | | 6 | **99% similarity** | | |
| | *Klebsiella pneumoniae* | 3 | 50.00 | | *Klebsiella pneumoniae* | 3 | 50.00 | | *Klebsiella pneumoniae* | 3 | 50.00 |
| | *Staphylococcus aureus* | 2 | 33.33 | | *Staphylococcus aureus* | 2 | 33.33 | | *Staphylococcus aureus* | 2 | 33.33 |
| | *Enterococcus avium* | 1 | 16.67 | | *Enterococcus avium* | 1 | 16.67 | | *Enterococcus avium* | 1 | 16.67 |
| 6 | **100% similarity** | | | 6 | **100% similarity** | | | 6 | **100% similarity** | | |
| | *Klebsiella pneumoniae* | 3 | 50.00 | | *Klebsiella pneumoniae* | 3 | 50.00 | | *Klebsiella pneumoniae* | 3 | 50.00 |
| | *Staphylococcus aureus* | 2 | 33.33 | | *Staphylococcus aureus* | 2 | 33.33 | | *Staphylococcus aureus* | 2 | 33.33 |
| | *Enterococcus avium* | 1 | 16.67 | | *Enterococcus avium* | 1 | 16.67 | | *Enterococcus avium* | 1 | 16.67 |

## Supplementary for Chapter 4

*Supplementary Table 4.1 Sequences detected in negative controls for full-length 16S rRNA nanopore sequencing runs.*

| (Mapped reads (%total reads in run) | NC - Saponin (30/7/19) | Run 1 | | | (Mapped reads (%total reads in run) | NC - 8/9/19 | Run 2 | | | (Mapped reads (%total reads in run) | NC -20190610 Dneasy Qiagen | Run 3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **16S06** | **Species** | **#Mapped** | **%Relative abundance** | | **16S04** | **Species** | **#Mapped** | **%Relative abundance** | | **16S01** | **Species** | **#Mapped** | **%Relative abundance** |
| 94504 (0.99%) | *Aliterella antarctica* | 29631 | 31.35 | | 7528 (0.546%) | *Haemophilus parainfluenzae* | 1975 | 26.24 | | 67 (0.001%) | *Lactobacillus fermentum* | 24 | 35.82 |
| | *Sphingomonas metalli* | 4500 | 4.76 | | | *Streptococcus parasanguinis* | 594 | 7.89 | | | *Burkholderia contaminans* | 23 | 34.33 |
| | *Sphingomonas roseiflava* | 3328 | 3.52 | | | *Streptococcus mitis* | 593 | 7.88 | | | *Burkholderia metallica* | 4 | 5.97 |
| | *Truepera radiovictrix* | 3286 | 3.48 | | | *Streptococcus oralis* | 477 | 6.34 | | | *Burkholderia lata* | 3 | 4.48 |
| | *Rubrobacter radiotolerans* | 2360 | 2.50 | | | *Neisseria perflava* | 347 | 4.61 | | | *Burkholderia arboris* | 2 | 2.99 |
| | *Rhodocytophaga aerolata* | 2251 | 2.38 | | | *Streptococcus salivarius* | 338 | 4.49 | | | *Escherichia fergusonii* | 2 | 2.99 |
| | *Potamolinea magna* | 1823 | 1.93 | | | *Veillonella parvula* | 336 | 4.46 | | | *Streptococcus sanguinis* | 2 | 2.99 |
| | *Cylindrospermum stagnale* | 1789 | 1.89 | | | *Veillonella dispar* | 236 | 3.13 | | | *Bacillus subtilis* | 2 | 2.99 |
| | *Kallotenue papyrolyticum* | 1735 | 1.84 | | | *Neisseria sicca* | 227 | 3.02 | | | *Enterococcus faecalis* | 1 | 1.49 |
| | *Sphingomonas mali* | 1534 | 1.62 | | | *Veillonella infantium* | 159 | 2.11 | | | *Moraxella osloensis* | 1 | 1.49 |
| | | | | | | | | | | | | | |
| | **Genus** | **#Mapped** | **%Relative abundance** | | | **Genus** | **#Mapped** | **%Relative abundance** | | | **Genus** | **#Mapped** | **%Relative abundance** |
| | *Aliterella* | 29631 | 31.35 | | | *Streptococcus* | 2500 | 33.21 | | | *Burkholderia* | 33 | 49.25 |
| | *Sphingomonas* | 17839 | 18.88 | | | *Haemophilus* | 2045 | 27.17 | | | *Lactobacillus* | 24 | 35.82 |
| | *Truepera* | 3286 | 3.48 | | | *Veillonella* | 853 | 11.33 | | | *Bacillus* | 3 | 4.48 |
| | *Brasilonema* | 3051 | 3.23 | | | *Neisseria* | 764 | 10.15 | | | *Escherichia* | 2 | 2.99 |
| | *Potamolinea* | 2528 | 2.68 | | | *Campylobacter* | 211 | 2.80 | | | *Streptococcus* | 2 | 2.99 |
| | *Rubrobacter* | 2401 | 2.54 | | | *Granulicatella* | 177 | 2.35 | | | *Salmonella* | 1 | 1.49 |
| | *Rhodocytophaga* | 2251 | 2.38 | | | *Sphingomonas* | 171 | 2.27 | | | *Moraxella* | 1 | 1.49 |
| | *Cylindrospermum* | 1796 | 1.90 | | | *Gemella* | 80 | 1.06 | | | *Enterococcus* | 1 | 1.49 |
| | *Kallotenue* | 1735 | 1.84 | | | *Aggregatibacter* | 63 | 0.84 | | | | | |
| | *Nitratireductor* | 1456 | 1.54 | | | *Capnocytophaga* | 61 | 0.81 | | | | | |

*Supplementary Table 4.2 Sequences detected in negative controls for whole genome nanopore sequencing runs.*

| Sample ID | NC - saponified - 30/7/2019 | | Sample ID | NC - No saponin 01/08/2019 | | Sample ID | NC - 8/8/19 | |
|---|---|---|---|---|---|---|---|---|
| Run | Nanopore WGS Run 1 | | Run | Nanopore WGS Run 1 | | Run | Nanopore WGS Run 2 | |
| Barcode | RLB-BC06 | | Barcode | RLB-BC12 | | Barcode | RLB-BC06 | |
| | | | | | | | | |
| Total Reads | 381789 | | Total Reads | 197 | | Total Reads | 180068 | |
| Total human reads | 130153 | | Total human reads | 169 | | Total human reads | 21495 | |
| %human reads | 34.09 | | %human reads | 85.7868 | | %human reads | 11.94 | |
| Total non-human reads | 251636 | | Total non-human reads | 28 | | Total non-human reads | 158573 | |
| %non-human reads | 65.91 | | %non-human reads | 14.2132 | | %non-human reads | 88.06 | |
| | | | | | | | | |
| **Genus** | **# Genus Reads** | **%Relative abundance** | **Genus** | **# Genus Reads** | **%Relative abundance** | **Genus** | **# Genus Reads** | **%Relative abundance** |
| *Acinetobacter* | 42264 | 16.80 | *Cutibacterium* | 6 | 21.43 | *Mesorhizobium* | 20273 | 12.78 |
| *Escherichia* | 11684 | 4.64 | *Staphylococcus* | 2 | 7.14 | *Rhizobium* | 8543 | 5.39 |
| *Glutamicibacter* | 5954 | 2.37 | *Cyanothece* | 1 | 3.57 | *Phyllobacterium* | 4591 | 2.90 |
| *Colwellia* | 4573 | 1.82 | *Singulisphaera* | 1 | 3.57 | *Agrobacterium* | 4005 | 2.53 |
| *Oscillatoria* | 3118 | 1.24 | *Planctomyces* | 1 | 3.57 | *Chelativorans* | 3534 | 2.23 |
| *Phialophora* | 2225 | 0.88 | | | | *Sinorhizobium* | 3401 | 2.14 |
| *Microcystis* | 2163 | 0.86 | | | | *Bacillus* | 2951 | 1.86 |
| *Chroococcidiopsis* | 1707 | 0.68 | | | | *Cutibacterium* | 2726 | 1.72 |
| *Thermosynechococcus* | 1338 | 0.53 | | | | *Aminobacter* | 2486 | 1.57 |
| *Exophiala* | 1217 | 0.48 | | | | *Kocuria* | 2437 | 1.54 |
| **Species** | **# Species Reads** | **%Relative abudance** | **Species** | **# Species Reads** | **%Relative abudance** | **Species** | **# Species Reads** | **%Relative abudance** |
| *Escherichia coli* | 11679 | 4.64 | *Staphylococcus epidermidis* | 1 | 3.57 | *Mesorhizobium loti* | 9186 | 5.79 |
| *Acinetobacter schindleri* | 10683 | 4.25 | *Staphylococcus aureus* | 1 | 3.57 | *Phyllobacterium sp. Tri-48* | 4591 | 2.90 |
| *Glutamicibacter arilaitensis* | 5954 | 2.37 | | | | *Mesorhizobium australicum* | 3804 | 2.40 |
| *Acinetobacter baumannii* | 4805 | 1.91 | | | | *Chelativorans sp. BNC1* | 3534 | 2.23 |
| *Colwellia sp. MT41* | 4461 | 1.77 | | | | *Bacillus subtilis* | 2833 | 1.79 |
| *Acinetobacter junii* | 4278 | 1.70 | | | | *Cutibacterium acnes* | 2725 | 1.72 |
| *Acinetobacter pittii* | 3365 | 1.34 | | | | *Aminobacter aminovorans* | 2486 | 1.57 |
| *Oscillatoria nigro-viridis* | 3118 | 1.24 | | | | *Kocuria rhizophila* | 2423 | 1.53 |
| *Acinetobacter sp. TGL-Y2* | 2781 | 1.11 | | | | *Mesorhizobium ciceri* | 2300 | 1.45 |
| *Phialophora attae* | 2225 | 0.88 | | | | *Rhizobium leguminosarum* | 2224 | 1.40 |

*Supplementary Table 4.2 (cont)*

| Sample ID | NC - 8/9/19 | | | Sample ID | NC - 02/07/2019 | | | Sample ID | NC - saponin 14/08/19 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Run | Nanopore WGS Run 2 | | | Run | Nanopore WGS Run 3 | | | Run | Nanopore WGS Run 3 | |
| Barcode | RLB-BC10 | | | Barcode | RLB-BC01 | | | Barcode | RLB-BC12 | |
| | | | | | | | | | | |
| Total Reads | 239920 | | | Total Reads | 9 | | | Total Reads | 34 | |
| Total human reads | 140863 | | | Total human reads | 8 | | | Total human reads | 30 | |
| %human reads | 58.71 | | | %human reads | 88.89 | | | %human reads | 88.235294 | |
| Total non-human reads | 99057 | | | Total non-human re | 1 | | | Total non-human reads | 4 | |
| %non-human reads | 41.29 | | | %non-human reads | 11.11 | | | %non-human reads | 11.764706 | |
| | | | | | | | | | | |
| **Genus** | **# Genus Reads** | **%Relative abundance** | | **Genus** | **# Genus Reads** | **%Relative abundance** | | **Genus** | **# Genus Reads** | **%Relative abundance** |
| Cutibacterium | 23760 | 23.99 | | Staphylococcus | 1 | 100 | | Staphylococcus | 3 | 75 |
| Acinetobacter | 5299 | 5.35 | | | | | | Phietavirus | 1 | 25 |
| Pseudomonas | 2643 | 2.67 | | | | | | | | |
| Salmonella | 1166 | 1.18 | | | | | | | | |
| Brevibacterium | 1160 | 1.17 | | | | | | | | |
| Novosphingobium | 843 | 0.85 | | | | | | | | |
| Sphingomonas | 659 | 0.67 | | | | | | | | |
| Planococcus | 639 | 0.65 | | | | | | | | |
| Burkholderia | 576 | 0.58 | | | | | | | | |
| Arthrobacter | 549 | 0.55 | | | | | | | | |
| **Species** | **# Species Reads** | **%Relative abudance** | | **Species** | **# Species Reads** | **%Relative abudance** | | **Species** | **# Species Reads** | **%Relative abudance** |
| Cutibacterium acnes | 23755 | 23.98 | | Staphylococcus epi | 1 | 100 | | Staphylococcus epidermidis | 2 | 50 |
| Escherichia coli | 403 | 0.41 | | | | | | Staphylococcus aureus | 1 | 25 |
| Arthrobacter sp. LS16 | 258 | 0.26 | | | | | | Staphylococcus virus PH15 | 1 | 25 |
| Arthrobacter sp. YC-RL1 | 245 | 0.25 | | | | | | | | |
| Sphingomonas wittichii | 199 | 0.20 | | | | | | | | |
| Novosphingobium sp. PP1Y | 101 | 0.10 | | | | | | | | |
| Corynebacterium falsenii | 67 | 0.07 | | | | | | | | |
| Microsporum canis | 57 | 0.06 | | | | | | | | |
| Eutypa lata | 40 | 0.04 | | | | | | | | |
| Kocuria rhizophila | 39 | 0.04 | | | | | | | | |

*Supplementary Table 4.3 Sequences detected in negative controls for whole genome Illumina sequencing runs.*

| **Sample ID** | NC-80919 | | | **Sample ID** | NC-RUN1 | | | **Sample ID** | NC-RUN2 | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Run** | Illumina WGS Run 3 | | | **Run** | Illumina WGS Run 1 | | | **Run** | Illumina WGS Run 2 | |
| **Total non-mammalian reads** | 1264931 | | | **Total non-mammalian reads** | 152590 | | | **Total non-mammalian reads** | 1903374 | |
| | | | | | | | | | | |
| **Species** | **#Reads** | **%Relative abundance** | | **Species** | **#Reads** | **%Relative abundance** | | **Species** | **#Reads** | **%Relative abundance** |
| *Mesorhizobium loti* | 174950 | 13.83 | | *Stenotrophomonas maltophilia* | 9091 | 5.96 | | *Elaeophora elaphi* | 147817 | 7.77 |
| *Sinorhizobium meliloti* | 139193 | 11.00 | | *Delftia sp. Cs1-4* | 7059 | 4.63 | | *Stenotrophomonas maltophilia* | 129088 | 6.78 |
| *Delftia sp. Cs1-4* | 137585 | 10.88 | | *Delftia acidovorans* | 6878 | 4.51 | | *Delftia sp. Cs1-4* | 101048 | 5.31 |
| *Methylobacterium oryzae* | 124577 | 9.85 | | *Pseudomonas fluorescens* | 6603 | 4.33 | | *Delftia acidovorans* | 100261 | 5.27 |
| *Burkholderia lata* | 59309 | 4.69 | | *marine metagenome* | 4870 | 3.19 | | *Pseudomonas fluorescens* | 91119 | 4.79 |
| *Delftia acidovorans* | 46079 | 3.64 | | *Actinoplanes utahensis* | 4362 | 2.86 | | *Mesorhizobium loti* | 71456 | 3.75 |
| *Rhizobium leguminosarum* | 36113 | 2.85 | | *Methylobacterium populi* | 3123 | 2.05 | | *Sinorhizobium meliloti* | 49069 | 2.58 |
| *marine metagenome* | 27585 | 2.18 | | *Methylobacterium extorquens* | 3084 | 2.02 | | *Methylobacterium populi* | 45154 | 2.37 |
| *Burkholderia ambifaria* | 21596 | 1.71 | | *Actinoplanes friuliensis* | 3024 | 1.98 | | *Methylobacterium extorquens* | 39715 | 2.09 |
| *Ralstonia mannitolilytica* | 20896 | 1.65 | | *Pseudomonas aeruginosa* | 2701 | 1.77 | | *Mesorhizobium huakuii* | 37906 | 1.99 |

*Supplementary material 4.1 Bioinformatics script written in python for alignment and subsequent removal of human reads using minimap2 and SAMtools.*

*nanopore_minimap2_submitjob_endophalmitis_wgs_GRCh38v1.py written by Liying Low*

```
# coding: utf-8
#! /usr/bin/python
import os
import re
import time
import sys

filepath =
"/data/endophthalmitis_wgs/20190805_WGS_Nanopore_UK_Endo_Run1/qcat_output_folder"
print(filepath)
filelist = os.listdir(filepath)
print(filelist)
filelist = []
count = 0
filelist = os.listdir(filepath)
print (filelist)
for fname in filelist:
    print(fname)
    file = (fname)
    x = file.strip (".fastq")
    print(x)
    command = "minimap2 -K 100M -t 8 -ax map-ont
/data/endophthalmitis_wgs/GRCh38.mmi " + str(filepath) + str(fname) + " > " +
"alignments/" + str(x) + "_aligned_GRCh38" + ".sam" + " ; "
    #os.system(command)
    print(command)
    print("Finished")
#
```

*Supplementary material 4.2 Bioinformatics script written in python for alignment of reads to reference genomes using minimap2 and SAMtools.*

*nanopore_minimap2_submitjob_endophalmitis_wgs_GRCh38v1.py written by Liying Low*

```
# coding: utf-8
#! /usr/bin/python
import os
import re
import time

alignmentpath = "./alignedto_referencegenome_nseq004_005/"
print (alignmentpath)

bamoutputpath = "./bam_output_filtered_nseq004_005/"
print (bamoutputpath)

command = "mkdir " + str(bamoutputpath)
os.system(command)
#print(command)

txtoutputpath = "./txt_output_filtered_nseq004_005/"

command = "mkdir " + str(txtoutputpath)
os.system(command)
#print(command)

bedtools_outputpath = "./bedtools_output_nseq004_005/"
command = "mkdir " + str(bedtools_outputpath)
os.system(command)
#print(command)

alignmentlist = os.listdir(alignmentpath)
print (alignmentlist)


for aname in alignmentlist:
    print(aname)
    alignment = (aname)
    y = re.sub(".sam", "", aname)
    print(y)
    command = "samtools view --threads 8 -b -F 2308 -q 30 " + str(alignmentpath) +
str (aname) + " -o " + str(bamoutputpath) +   str(y) + "_combinedflags_q30.bam" + "
; "
    os.system(command)
#    print(command)

#SAM flag 2308: read unmapped (0X4), not primary alignment (0x100) and
supplementary alignment (0x800)
```

```
    command = "samtools sort " + str(bamoutputpath) + str(y) +
"_combinedflags_q30.bam" + " -o " + str(bamoutputpath) + str(y) +
"_combinedflags_q30_sorted.bam" + " ; "
    os.system(command)
#    print(command)

    command = "samtools index " + str(bamoutputpath) + str(y) +
"_combinedflags_q30_sorted.bam" + " ; "
    os.system(command)
#    print(command)

    command = "samtools idxstats " + str(bamoutputpath) + str(y) +
"_combinedflags_q30_sorted.bam" + " > " + str(txtoutputpath) + str(y) +
"_combinedflags_q30_sorted.txt" + " ; "
    os.system(command)
#    print(command)

    command = "bedtools genomecov -ibam " + str(bamoutputpath) + str(y) +
"_combinedflags_q30_sorted.bam -d -split > " + str(bedtools_outputpath) + str(y) +
"_combinedflags_q30_sorted_split"
    os.system(command)
#    print(command)

    print("Finished")
```

*Supplementary material 4.3 Python script for calculating breadth of coverage, mean alignment depth and coverage.*

*Alignment_to_histogram_trial_by_directory.py written by Liying Low*

```
import pandas as pd
import numpy as np
import matplotlib
import seaborn as sns
import os
import re
import matplotlib.pyplot as plt
%matplotlib inline

inputpath = "/Users/lowl/Documents/Research_2020/Nanopore_Endophthalmitis/bedtools_output_nseq004_005/"
outputpath = "/Users/lowl/Documents/Research_2020/Nanopore_Endophthalmitis/WGS_coverage/"
outputpath_image = "/Users/lowl/Documents/Research_2020/Nanopore_Endophthalmitis/WGS_coverage_plots/"

filelist = os.listdir(inputpath)

for filename in filelist:
#    print(inputpath+filename)
    x = re.sub("_combinedflags_q30_sorted_split", "", filename)
    print(x)
    y = str(inputpath) + str(filename)
    print(y)
    p = str(outputpath_image) + str(filename) + ".png"
    print(p)
    stats = str(outputpath) + str(filename) + ".txt"
    print(stats)

    df = pd.read_csv(y, sep='\t', names =["Reference", "Index", "Count"])
    df["Index"],df["Count"]
    %matplotlib notebook
    ax1 = df.plot.line(x='Index', y='Count')
    plt.savefig(p, dpi = 300)
    with open(stats, 'a') as f:
        print("Sum of all aligned bases", file = f)
        print(sum(df["Count"]), file = f)
        tmpcount = 0
        for i in range(0,len(df["Count"])):
            if df["Count"][i] != 0:
                tmpcount = tmpcount + 1
        print("Number of bases with depth >= 1", file = f)
        print(tmpcount, file = f)
        reference_length = len(df["Count"])
        print("Breadth of coverage "+ str((tmpcount/reference_length)*100)+" %", file = f)
        print("Mean alignment depth " + str(sum(df["Count"])/tmpcount) + " x", file = f)
        #depth is sum of all aligned bases divided by number of bases with depth >=1
        print("Length of reference genome " + str(reference_length)+ " bases", file = f)
        breadth = (tmpcount/reference_length)
        depth = sum(df["Count"])/tmpcount
        coverage = breadth*depth
        print("Coverage " + str(coverage) + " x", file = f)
        print("Finished")
```

*Supplementary material 4.4 Python script for calculating relative abundances of taxa from Illumina WGS output.*

*Illumina_WGS_genus_by_directory.py written by Liying Low*

```
#Received output from SMART (UW) for the Illumina WGS Endophthalmitis - Excel format table of all the genus and #reads
#To aim here is to summarise the reads in to top 10 with the %relative abundances

import pandas as pd
import os
import re

df = pd.read_csv('/Users/lowl/Documents/Research_2020/Nanopore_Endophthalmitis/Illumina_SMART_genus.csv', sep=',')

x = "NC-RUN2"

#Replace NSEQ001 with the name of the sample - have to correspond to the name on the headings

df_x = df[['genus', str(x)]]
df_x.head()
sort_by_x = df_x.sort_values(by=[str(x)], ascending = False)
sort_by_x.head(10)
Total_mapped = df_x[str(x)].sum()
print("Total mapped " + str(x) + " = " + str(Total_mapped))
df_x = df_x.assign(Relative_abundance = lambda y: (y[str(x)]/Total_mapped)*100)
sort_by_x = df_x.sort_values(by=[str(x)], ascending = False)
sort_by_x.head(10)
output_path = "/Users/lowl/Documents/Research_2020/Nanopore_Endophthalmitis/Illumina_Miseq_SMART_genus/"
z = str(output_path) + str(x)+ "_genus.xlsx"
sort_by_x.to_excel(z, encoding='utf-8')
print("Finished")
```

## Publication arising from this thesis

Low L, *et al*. Evaluation of Full-Length Nanopore 16S Sequencing for Detection of Pathogens in Microbial Keratitis. PeerJ. (in Press)