O*rnithobacterium hominis* isolates from a South African birth cohort

Celine De Allende

Dissertation presented for the degree of

Master of Science

in the Department of Molecular and Cell Biology

University of Cape Town

2024

Supervisors: Dr Felix Dube

Co-supervisors: Dr Siobhan Brigg and Susannah Salter

Plagiarism Declaration

I know the meaning of plagiarism and declare that all of the work in the dissertation (or thesis), save for that which is properly acknowledged, is my own.

Signature: Signed by candidate

Date: 18/09/2024

Acknowledgments

I would like to express my sincerest gratitude to my supervisor, Dr Felix Dube, for his invaluable guidance, encouragement, expertise, and support throughout my research project and the writing of this dissertation.

I wish to thank Susannah Salter for her seemingly endless expertise and unwavering support and Dr Takudzwa Matuvhunye for the invaluable help provided throughout the dissertation writing. I am also thankful to my co-supervisor, Dr Siobhan Brigg, for her encouragement and assistance.

I would also like to express my gratitude to the members of the Dube Lab for their support and words of encouragement. Additionally, I extend my sincere thanks to the Department of Molecular and Cell Biology. A special thank you to the DAs and SOs for ensuring the smooth operation of our lab and department.

I am incredibly grateful for the funding provided by the Cambridge-Africa Alborada research fund, which made this work possible, and I am immensely appreciative of the financial support provided by the MCB departmental awards and the UCT Merit Award.

I would like to express my heartfelt thanks to my friends and family for their unwavering support. Thank you for always being there when I needed it the most. I am also deeply grateful to my partner, Jurgens, for their unwavering support, patience, and understanding throughout this journey, it has been a long road but I'm glad to have achieved this with you by my side. Finally, a special thanks to my parents, Olga and Mario, for always believing in my endeavours. This achievement would not have been possible without your support and encouragement. You have been a solid foundation for me throughout this journey, and for that, I am deeply grateful.

Abstract

Ornithobacterium hominis is a gram-negative bacterium that colonises the infant nasopharynx. *Ornithobacterium hominis* 16S rRNA gene sequences have been identified in South-East Asia, Australia, Gambia and Kenya; however, this species is notably absent from cohorts based in the northern hemisphere. Recent microbiome data from the Drakenstein Child Health Study (DCHS) has shown that *O. hominis* is present in South Africa; however, there is a lack of microbiological data regarding this species in Africa. Only a few strains have been isolated since its discovery, all of which are from Australian samples, and most data available for this species is generated from metagenomic data and short-read assemblies. *Ornithobacterium hominis* has been shown to colonizes the nasopharynx at high proportional abundances, especially in settings with disproportionately high pneumonia burden. However, due to the limited data, the implications of this species on disease progression are unclear; furthermore, genomic data and longitudinal carriage data suggest that *O. hominis* could impact the microbiome of the nasopharynx, possibly influencing disease progression. We describe the carriage and evolutionary dynamics of *O. hominis* in a South African birth cohort by screening available 16S rRNA data and employing standard microbiological and whole genome sequencing techniques.

Methods: Available 16S rRNA sequencing data from the DCHS was screened for *O. hominis* sequences. Candidate samples were selected for microbiological isolation of *O. hominis. Ornithobacterium hominis* was isolated on Columbia blood agar in a humid, high CO_2, microaerophilic incubator and identity was confirmed using a combination of standard PCR, 16S gene end-sequencing, and Gram staining. Colony morphologies and Gram-stain characteristics of the South African *O. hominis* isolates were determined and antibiotic susceptibility determined. For whole genome sequencing, DNA was extracted, assessed for quality, and sequenced using the Oxford Nanopore technologies platform. Genomes were assembled using an in-house bioinformatic pipeline. Genomes were analysed and compared using standard bioinformatic tools and a core genome pangenome analysis was performed.

Results: *Ornithobacterium hominis* was present in 32% of samples and abundance increased after six months of age in the cohort. *Ornithobacterium hominis* was isolated and characterised. The species exhibited four main colony morphologies: punctiform, uniformly sized, mixed and mucoid. Antibiotic susceptibility testing demonstrated that all isolates were susceptible to β-lactam antibiotics and resistant to polymyxin B. Six *O. hominis* genomes were assembled into 2 Mb-sized circular contigs and annotated. Virulence factors such as *vap*D, antimicrobial resistance genes such as *cfx*a, including several adhesins, and a putative plasmid were identified. In addition, three distinct lipopolysaccharide biosynthesis clusters were identified, and the rearrangement hotspot cluster structure was resolved. South African *O. hominis* strains were found to be more closely related to Australian strains than to Maela strains from Thailand. Finally, core genes were assigned functions and categorised.

Conclusion: We report some of the first *O. hominis* isolates and genomes assembled from cultured isolates in Africa. More strains will have to be isolated to gain a better understanding of species diversity. It is essential to investigate the interactions of this species with other respiratory tract organisms to obtain better insights into its function/role in nasopharyngeal microbiome and how it may influence disease risk. Further characterisation of the species, especially in sub-Saharan Africa, is warranted.

Contents

List of Tables

List of Figures

Figure 3.21. Multiple sequence alignment (MSA) of *O. hominis* rearrangement hotspot (RHS) C-terminal tips from full-length *rhs* genes.

Figure 3.22. Phylogenetic tree of *O. hominis* reconstructed from the core genome alignment of 13 genomes from Thailand (Maela cohort), Australia, and South Africa.

Figure 3.23. Summary of *O. hominis* core genes COG functional assignments.

List of abbreviations

® - Registered trademark

16S rRNA - 16S ribosomal RNA

AMR - Antimicrobial resistance

ASV - Amplicon sequence variants

BHI - Brain heart infusion

BLAST - Basic Local Alignment Search Tool

CARD - Comprehensive Antibiotic Resistance Database

CBA - Columbia agar plates supplemented with 5% sheep's blood

DCHS - Drakenstein Child Health Study

EUCAST- European Committee on Antimicrobial Susceptibility testing

FSD - *Fibrobacter succinogenes* major paralogous domain

gDNA - genomic DNA

gld - Gliding motility associated

HMW - High-molecular weight

IQR - Interquartile range

KEGG - Kyoto Encyclopedia of Genes and Genomes

LMIC - Low to middle-income countries

LPS - Lipopolysaccharide

LRTI - Lower respiratory tract infections

MAG - Metagenome-assembled genome

MIC - Minimum inhibitory concentration

NP - Nasopharyngeal

ONT - Oxford Nanopore Technologies

OriC - Origin of replication

OTU - Operational taxonomic units

PCI - Phenol chloroform isoamyl alcohol

PCR - Polymerase chain reaction

PCV - Pneumococcal conjugate vaccine

PMT - Pasteurella mitogenic toxin

POCP - Percentage of conserved proteins

Q20 - Quality score of 20

RARS - Replication-associated structural rearrangements

rhs - Rearrangement hotspot

SA OH - South African *Ornithobacterium hominis*

SCV - Small colony variants

SD - Standard deviation

STGG - Skim milk-tryptone-glucose-glycerol

T9SS - Type IX protein secretion system

TSA - Tryptic soy agar

™ - Trademark

vapD - Virulence associated protein D

WAZ - Weight-for-age Z-score

WGS - Whole Genome Sequencing

β - beta

Chapter 1: Literature Review

1.1 Introduction

Respiratory infections are one of the leading infectious causes of mortality globally, attributed to approximately 2.18 million deaths worldwide in 2021[1]. Young children and the elderly are particularly at risk and have higher mortality rates. Every year, approximately half a million deaths in children under the age of five are attributable to lower respiratory tract infections (LRTI)[2]. Furthermore, the incidence and mortality rate of LRTIs is higher in low to middle-income countries (LMICs), with disproportionately high rates of respiratory infections in Sub-Saharan Africa[2]. The burden of LRTI is further exacerbated by the global rise in antibiotic resistance.

Nasopharyngeal (NP) colonization typically precedes invasive disease progression[3]. The infant nasopharynx is an important microbiological niche, connecting the lower- and upper-respiratory tract, including the sinuses and middle ear[4]. Since colonization by pathogens is a pre-requisite for infection, the composition of the NP microbiome significantly influences disease risk[3]. Although a substantial body of research has established connections between microbial profiles and disease outcomes, these studies primarily rely on 16S rRNA amplicon sequencing, which provides limited insights into the biological processes contributing to disease progression[5-7].

Streptococcus pneumoniae, *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Staphylococcus aureus* are common colonisers of the nasopharynx[4]. Nevertheless, these species are also implicated in LRTIs due to their ability to switch from commensals to pathogens, also referred to as pathobionts. Factors contributing to the transition of organisms from asymptomatic colonizers to pathogens are not well studied and the exact mechanisms remain elusive. Previous studies suggests that the switching of commensals to pathogens is influenced by interactions with other bacterial species that colonize the nasopharynx, and research is increasingly focusing on the role of co-colonising organisms and genetic factors contributing to pathogenicity of commensals[8-11].

While the colonization dynamics of abundant bacterial species such as *S. aureus* and *Moraxella catarrhalis* are extensively studied, the colonization dynamics of most fastidious, lesser-common bacterial species in the NP microbiome are not well explored. Additionally, the investigation of the functional roles of less common or lower-abundance species often receives less attention in microbiota research, which hinders understanding of interactions between more characterised and lesser-known species and how these interactions influence respiratory health and disease. Investigating these low abundance colonizing species may reveal important insights into microbial interactions and mechanisms underlying disease risk and development.

Ornithobacterium hominis, a newly identified bacterial species of the *Weeksellaceae* family, was first identified in the nasopharyngeal microbiome of infants in the Maela refugee camp in Thailand[12]. Notably, *O. hominis was* found in high abundances and seems to be a persistent colonizer in the infant population of Maela, which experiences high rates of respiratory disease[12]. Although this species has also been detected and cultured in NP samples from Australia and identified through 16S rRNA gene data from The Gambia and Kenya, it has not been thoroughly characterized including by culturing methods in African samples. As seen in Table 1.1, most available genomic data is derived from metagenomic sources and short-read sequencing, with only one complete genome assembled from a cultured isolate, from Australia[13]. Furthermore, the isolates characterized to date are exclusively from Australian sources (Table 1.1).

Previous genome analyses have identified virulence features and competition factors in the *O. hominis* genome, suggesting that it interacts with both the microbiome and the host environment[14]. *Ornithobacterium hominis* colonization may have significant implications for respiratory illness outcomes. Therefore, it is essential to characterize *O. hominis* strains from diverse geographic regions, particularly from African contexts with high rates of respiratory infections to understand the potential role of *O. hominis* in respiratory health and disease.

Table 1.1. Studies investigating *O. hominis*

Reference	Region/Country	Study population	Methods	Major findings
Salter *et al.* (2019)[11]	Thailand	Infants born in the Maela refugee camp on the Thailand–Myanmar border (2007–2010). Investigating pneumococcal colonization and pneumonia epidemiology (Salter *et al*., 2017)[9]. Prevalence screen included 100 random-selected infants from Maela cohort (median age of 12 months old) and concurrent swab samples from mothers.	NP swab samples collected from infants (Salter *et al*., 2017)[10]. DNA was extracted from NP swabs and sequenced using the MiSeq™ 250 bp paired-end protocol (Illumina®). Manual curation led to two syntenic draft genomes, which were annotated using Prokka.	Seven *O. hominis* genomes were assembled from metagenomic data, two of which were ~~to~~ improved to draft standard, composed of nine and fifteen contigs each. Prevalence screen: Carriage prevalence of 42% of 12-month-old infants and 2% of mothers
Lawrence *et al.* (2019)[12]	Australia	A subset of children (<10 years of age) enrolled in either (i) an observational study of chronic suppurative lung disease in Darwin, Australia, between October 2011–October 2012 or (ii) an observational study of children undergoing bronchoscopy in Brisbane, Australia, between February 2011–November 2012 (Marsh *et al.*, 2016)[13].	NP swabs collected from four Australian children (age 1–2 years) with chronic suppurative lung disease from the Marsh *et al.* (2016) study[13]. NP swab media was inoculated onto various media and incubated under various conditions at 35°C for up to five days. *Ornithobacterium hominis* colonies were identified using PCR screening targeting 16S rRNA and *toxA* genes. Sequencing was performed using NextSeq 500 (Illumina®) and *de novo* genomes were assembled.	The first successful culture conditions for *O. hominis* were identified. First descriptions of *O. hominis* colony morphology.

Reference	Region/Country	Study population	Methods	Major findings
Ahmed *et al.* (2024)[14]	Egypt	NP swab samples collected from 10 infants (0–24 months of age) being investigated for pneumonia and 10 infants that were healthy, between November 2021 – April 2022 and from December 2022 – March 2023.	DNA extracted from NP swab samples. Shotgun metagenomic sequencing performed using MiSeq™ Reporter (Illumina®). Reads aligned to representative *O. hominis genome*. Draft genomes were assembled and annotated.	Two draft genome assemblies of *O. hominis* that were generated from metagenomic data. Identified associated virulence genes, antimicrobial genes, and distinct defence mechanisms.
Salter *et al.* (2023)[10]	Australia	A subset of children (<10 years of age) enrolled in either (i) an observational study of paediatric chronic suppurative lung disease at Royal Darwin Hospital, Darwin, Australia, between October 2011–October 2012 or (ii) an observational study of children undergoing bronchoscopy at the Royal Children's Hospital, Brisbane, Australia, between February 2011 and November 2012 (Marsh *et al.*, 2016)[13].	Isolate from NP swab sample collected from an infant based Marsh *et al.* (2016) study[13]. Isolate was cultured, DNA extracted from plate scrapes and sequenced using Oxford Nanopore Technologies. Preliminary assemblies generated using Flye then polished with Pilon using NextSeq 500 paired-end 150-bp reads (Illumina®).	The complete genome sequence of the *O. hominis* type strain MSHR-COH1 (ATCC TSD-185/NCTC 14317) isolated from~~orm~~ a NP sample. A circular genome of the type strain achieved using long-read data polished with short reads.

1.2 The nasopharyngeal bacteriome

The human NP bacteriome is a complex and dynamic microbiological niche bordering the lower respiratory tract, middle ear and sinuses which are generally considered microbially scarce[15]. The warm and oxygen-rich environment allows for colonization with a diverse bacteriome[15,16]. Mounting evidence indicates that the bacteriome plays a major role in overall respiratory health. The NP is home to commensal organisms that can serve a protective function against pathogens, thereby preventing subsequent disease[16]. Commensals can compete with pathogens by nutrient deprivation, secretion of antimicrobials or by modulating the local immune environment. Respiratory tract infections can be attributed to changes within the microbiome leading to the loss of commensals and subsequent colonization by opportunistic pathogens[4]. Dysbiosis of the NP microbiome is associated with various respiratory tract infections including pneumonia (lung infection) and otitis media (middle-ear infection)[17]. The NP pathogens can also invade the bloodstream, causing life-threating conditions such as sepsis and meningitis (Figure 1.1). Thus, the NP microbiome plays a major role in maintaining overall health.

1.3 The infant nasopharyngeal microbiome

The NP of infants exhibits high diversity, being predominated colonised by six genera, *Moraxella, Streptococcus, Corynebacterium, Staphylococcus, Haemophilus,* and *Alloiococcus/Dolosigranulum*[15]. The infant NP tends to be heavily colonized by organisms, such as *S. pneumoniae* and *M. catarrhalis* which poses a greater risk for the development of respiratory infections in infants[18]. Conversely, commensal species might have a protective influence. Therefore, the bacteria colonizing the infant NP play a major role in respiratory health, and investigations into the NP bacteriome are pertinent.

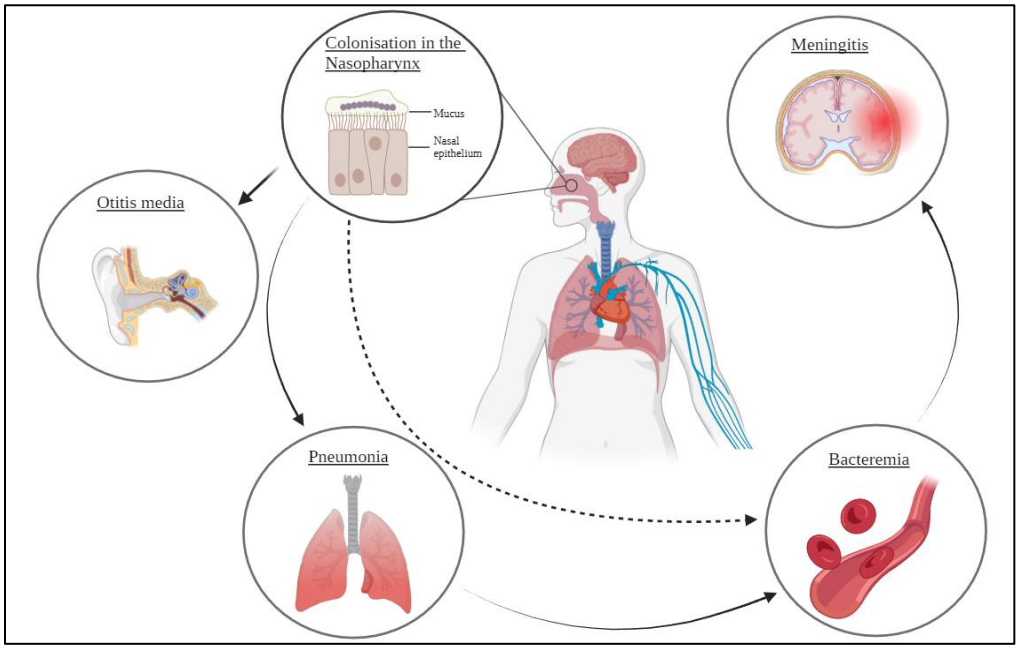

Figure 1.1. Nasopharyngeal pathogen colonisation is a pre-requisite for invasive disease. Image generated in Biorender.com

1.3.1. Development of the infant NP microbiome

The NP of infants is colonised rapidly after birth, and the bacterial profile of the NP microbiome shifts dramatically during the first two years of life[6,19]. Initially, *Streptococcus* and *Staphylococcus* tend to dominate the NP bacteriome during the first few months of life; however, these taxa usually decrease in abundance during the first year[19]. Then, by 24 months of age, *Staphylococcal* dominance is diminished. These taxa are initially replaced by *Dolosigranulum* and *Corynebacterium*, followed by *Moraxella* after 24 months of age[19]. Previous work conducted on the DCHS cohort sought to investigate the early life NP microbiota in healthy infants, without LRTIs[20]. Nasopharyngeal swabs were collected from infants at monthly intervals during the first year of life to monitor the longitudinal changes in the NP bacteriome and the bacterial profiles were characterised using 16S rRNA gene amplicon sequencing[20]. They found that, of the 20 most abundant bacterial genera, NP bacterial profiles were dominated by *Staphylococcus*, *Streptococcus* and *Corynebacterium* around the first month of life, which rapidly decreased in relative abundance at around 4 months of age. *Moraxella* increased in relative abundance at around 6 months of

age, which was followed by a plateau observed until 24 months of age[20]. In summary, the microbiome profiles typically shift to became more *Moraxella* and *Haemophilus* dominant during infancy. The dynamics of the bacteriome during infancy are important to investigate as these could impact respiratory health.

1.3.2. Impact of NP microbiota composition on respiratory illness risk

A study by Teo *et al*. (2015) investigated and compared the bacterial communities in the nasopharynx of Australian infants during the first year of life[5]. Nasopharyngeal swab samples were collected from children during episodes of respiratory illness and from children that were healthy, without respiratory illness[5]. The microbiome composition was determined using 16S rRNA gene deep sequencing. In healthy infants, *Staphylococcus* and *Corynebacterium* species dominated the nasopharyngeal profiles at 2 months of age; however, the relative abundance of these genera decreased over time, while *Alloiococcus* and *Moraxella* species increased in relative abundance with age[5]. In contrast, samples collected during illness episodes exhibited a slightly different pattern. While *Staphylococcus* and *Corynebacterium* declined in relative abundance with age, *Moraxella*, *Haemophilus,* and *Streptococcus* increased in relative abundance[5]. Notably, *Haemophilus* was found exclusively in samples collected during respiratory illness and was associated with chronic infections[5].

Additionally, a study by Xu *et al.* (2021) investigated the NP microbiome composition of otitis-prone and otitis-free children[7]. Microbiota composition was determined using 16S rRNA gene sequencing. The microbiota profiles of children that were otitis prone were less diverse than that of children who remain otitis-free, which is consistent with dysbiosis of the microbiome; however, the exact cause of this dysbiosis was not elucidated[7]. In addition, the microbiome composition differed between the two groups of children at 6 months of age; *Fusobacterium*, *Veillonella*, and *Moraxella* were observed at greater abundances in otitis-prone children than in otitis-free children[7]. However, a major limitation of this study was that 16S sequencing could not determine the bacterial identity to the species level. Overall, these studies show that the NP microbiome composition differs between children experiencing respiratory

illness episodes and children that are healthy; thus, the composition of the microbiome is implicated in respiratory health.

Children with a more stable NP microbiota have a lower risk of colonization by potential pathogens. In contrast, significant disruptions in the microbiota are linked to an increased likelihood of pathogenic bacterial colonization. A study by Biesbrook *et al*. (2014) investigated the NP microbiota profiles and microbiota stability of healthy children at 1.5, 6, 12 and 24 months of age and determine whether there was any link to respiratory health outcomes[6]. Nasopharyngeal samples were collected from infants enrolled in PCV-7 vaccine trail in the Netherlands and their bacterial profiles were characterised using 16S rRNA genes-based pyrosequencing. They found that that the presence of *Moraxella* or *Dolosigranulum*/*Corynebacterium* species in infancy was associated with more stable bacterial colonization patterns during the first two years of life and lower rates of parental-reported respiratory infections, while colonisation with *Streptococcus* and *Haemophilus* species is associated with less stable bacterial profiles over time[6].

A study by Bosch *et al*. (2017), investigating infant nasopharyngeal microbiota using 16S rRNA gene amplicon sequencing, found that less stable community patterns were associated with increased respiratory tract infection episodes during the first year of life, corroborating with findings by Biesbrook *et al*. (2014)[19]. Children with increased respiratory tract infections susceptibility in the first year of life also exhibited microbiota profiles characterised by a premature predominance of *Moraxella* species, less-prolonged colonization by *Corynebacterium* and *Dolosigranulum* species, and a prolonged colonisation by oral-niche bacteria such as *Prevotella* species. These studies show that the colonization dynamics of bacterial profiles over time greatly influence respiratory disease-risk outcomes and highlight the utility of longitudinal studies[19].

1.3.3. Impact of early life exposures on the infant NP microbiome composition

The development of the infant NP microbiome is heavily influenced by a range of environmental factors including mode of birth, infant feeding type, antibiotic exposure, vaccinations, social interactions, and exposure to pollutants[15,3]. These early life exposures may lead to perturbations in the composition of the NP microbiome that impact respiratory health promoting the colonisation of pathogenic or protective bacterial populations.

Infant feeding type has strong associations with infant microbial composition, especially during the first weeks of life. A study by Biesbrook *et al.* (2014), investigated the effect of feeding type on the NP microbiome of healthy infants from the Netherlands, and found that at 6 weeks of age, breastfed children showed increased abundance of *Dolosigranulum* and *Corynebacterium* species, while formula fed infants showed increased abundance of *Staphylococcus*, *Prevotella*, and *Veillonella* [21]. Thus, breastfeeding impacts NP bacterial colonization of infants and may have a protective effect against early childhood respiratory infections. Similarly, regarding mode of delivery, caesarean deliveries, compared to vaginal deliveries, are associated with reduced colonization by protective commensals such as *Corynebacterium* and *Dolosigranulum*[22].

Infants are exposed to numerous vaccines, and there is evidence to suggest that vaccine exposure results in perturbations of NP microbiome. The introduction of the pneumococcal conjugate vaccine (PCV) has led to the global decline of pneumococcal disease burden and severity[23]. PCVS have been used in national immunization programs for infants in high-income countries and are being introduced more progressively in LMICs over the last decade[24]. These vaccines have proven efficacy for reducing pneumococcal acquisition, transmission and disease severity. However, there is increasing concern over the impact of the vaccine on the bacterial community as evidence suggest that the vaccine serotype is replaced by other species, potentially pathogenic, in the bacterial community. One study by Biesbrook *et al.* (2014) investigated the effects of PCV7 of the NP microbiota of children in a randomized controlled trail in the Netherlands[25]. They found that vaccination with PCV7 lead to a shift in bacterial community structure

and composition characterized by an influx of oropharyngeal species such as *Prevotella* and *Rothia*. However, they were unable to link microbiota changes with respiratory disease risk as samples were not collected during respiratory tract infection episodes[25].

Although the consequences and potential mechanisms of the bacterial community shifts were not elucidated, it is known that dysbiosis of bacterial communities is associated with increased disease risk. Another study investigating the effects of the PCV on the NP microbiota in infants was conducted in Switzerland[26]. Here, a case-control study was conducted with samples from an infant cohort, and they investigated the impact of PCV7 and PCV13 on the infant microbiota. They found that samples from infants vaccinated with PCV7 were characterized by reduced diversity and richness compared to samples from infants vaccinated with PCV13, but increased stability in infants vaccinated with PCV13 compared to those vaccinated with PCV7; however, the prevalence of respiratory illness did not differ between the PCV-7 and PCV-13 vaccinated groups[26]. Overall, these studies show effects of vaccination on the bacterial communities should be investigated and not just the effect on the pneumococcus.

Antibiotic usage is another major factor influencing the NP microbiota profiles in infants. Respiratory tract infections are commonly treated with antibiotics, with otitis media being the most frequent diagnosis for antibiotic prescription in young children[7]. Teo *et al*. (2015) found that, compared to children that were not exposed to antibiotics, samples from healthy children that were administered antibiotics 4-weeks prior to sampling exhibited higher abundances of *Haemophilus*, *Streptococcus*, and *Moraxella*[5]. Conversely, *Alloiococcus* and *Corynebacterium* relative abundances were reduced with antibiotic exposure[5]. Thus, antibiotic exposure significantly disrupts the NP microbiome, facilitating the colonization of pathogenic bacteria. Furthermore, the overprescription of antibiotics has contributed to the alarming rise in antibiotic resistance among respiratory tract bacterial pathogens, leading to treatment failure and life-threatening infections[27]. The NP is a reservoir for many respiratory pathogens and may also serve as a source of antimicrobial-resistance gene transfer from non-pathogenic to pathogenic bacteria[27]. In many cases, horizontal gene transfer of mobile genetic elements results in antibiotic resistance, and in the case of

pneumococci, which are naturally competent, free DNA can be taken up from the surrounding environment[27]. Therefore, perturbations in the microbiome due to antibiotic exposure, coupled with horizontal gene transfer mechanisms, highlight the role of the NP as a reservoir for antibiotic-resistant pathogens.

1.3.4. Health implications and research limitations

Respiratory tract infections are highly prevalent among children, particularly infants during the first year of life. Research indicates that the composition of the microbiota plays a crucial role in influencing disease risk. While some species are associated with healthy microbiomes and have a protective effect, certain colonizers increase the risk of disease development and severity. Common colonisers, namely, *S. pneumoniae*, *S. aureus*, *H. influenzae*, and *M. catarrhalis* are major contributors to respiratory tract infections in children and often colonize the upper respiratory tract asymptomatically[15]. Importantly, colonisation is a pre-requisite for disease development. Studies show that the colonization by *S. pneumoniae* can lead to invasive, life-threatening infections such as pneumonia, meningitis, bacteraemia, and sepsis[15]. *Streptococcus pneumoniae* colonization can also lead to non-invasive disease in the respiratory tract such as sinusitis and otitis media and lower respiratory tract.

Respiratory infections are a leading cause for childhood mortality globally[28]. Consequently, it is important to investigate the development of the infant NP microbiome and early-life respiratory tract infections. Infections during this critical stage of development could have long-term health effects, may even lead to chronic lung disease into adulthood[29], and are associated with decreased lung function, asthma and chronic obstructive pulmonary disease (COPD)[30]. Further evidence from multiple cohort studies suggests that early-life lung health may impact lung health into adulthood[28].

Despite numerous studies linking microbial profiles to disease outcomes, most of these studies rely on 16S-amplicon sequencing to determine microbial composition, which has limited taxonomic resolution. These studies often sequence only two to four of the variable regions (V1–V9) of the 16S rRNA gene of

the 16S rRNA gene, allowing for bacterial identity to be resolved to the genus level[31]. Although species-level identification can be available in some taxa, the species of many taxa cannot be accurately determined as the variability in the sequenced 16S rRNA region may not be sufficient to distinguish species[32]. Furthermore, while 16S-amplicon sequencing is a widely utilised and cost-effective way to study microbiome composition and changes, it offers limited insight into the functional roles of species and the biological processes involved in disease progression[31]. As a result, the clinical impact of colonisation patterns remains unclear.

1.4 Inter-bacterial interactions play a major role in the microbiome composition

The interactions between bacteria are incredibly complex and can be species or strain specific[33]. These interactions in the NP could promote or inhibit the growth of the bacteria, influence carriage prevalence and may even affect virulence. This could have major implications for commensals with pathogenic potential such as *S. aureus* and *S. pneumoniae*. Thus, bacterial interactions between strains and other species may have clinically important outcomes. Previous research has focused on the positive association of *S. pneumoniae* with other colonizers such as *H. influenzae* and *M. catarrhalis*[34,35]. As well as negative associations between *S. pneumoniae* and other Streptococcal species and with *Corynebacterium and S. aureus*[8,34]. These associations can be attributed to various inter-species interactions[36,37].

A study by Xu *et al*. (2021) examined the NP microbiome of children aged 6 and 12 months, with or without *S. pneumonia*e colonization, using 16S rRNA gene sequencing and found that the number of detected *Corynebacterium* species was inversely correlated with *S. pneumoniae* colonization[38]. Two species of *Corynebacterium* were isolated, *C. pseudodiphtheriticum* and *C. propinquum*, and the effects of *Corynebacterium* on *S. pneumonia*e growth were investigated. *Corynebacterium* was found to inhibit the growth of *S. pneumoniae in vitro*. However, the exact mechanism of this inhibition could not be determined[38].

In has also been observed that non-diphtheriae *Corynebacterium* species are associated with a decreased risk of pneumococcal colonization during infancy[33]. In this study, the NP microbiomes of infants in Botswana were analysed using 16S rRNA gene sequencing and *S. pneumoniae* colonization was identified with a species-specific PCR assay. In addition, *in vitro* experiments showed that pneumococcal growth was inhibited by secreted factors (present in cell-free supernatants) from several *Corynebacterium* strains[33]. However, there are conflicting results suggesting that *Corynebacterium* does not appear to impact *S. pneumoniae* colonization, further highlighting the complex and intricate nature of bacterial interactions[12].

There is mounting evidence that indicates inter-species interactions during polymicrobial infections can influence the antibiotic susceptibility of the involved members. For example, a neighbouring bacterial species could confer a protective effect against antibiotics[39]. *Moraxella catarrhalis* is one such bacterial species that is positively associated with *S. pneumoniae* colonization. *M. catarrhalis* can express β-lactamases which could passively protect β-lactam susceptible *S. pneumoniae* in polymicrobial biofilms[40]. Additionally, β-lactamases released by *M. catarrhalis* in polymicrobial biofilms have been shown to confer a protective effect on *H. influenzae* against ampicillin[41].

1.5 Culturing and investigating fastidious bacteria

Bacterial interactions in the NP can alter the microbiome composition, which in turn, can alter the risk of respiratory infections[42]. However, the complex nature of these interactions in the NP remains underexplored and requires more extensive investigation.

Most research on bacterial interactions within the NP relies on correlation data derived from 16S rRNA sequencing, primarily focusing on well-characterized members of the bacteriome, such as *Staphylococcus aureus*, *Streptococcus pneumoniae*, and *Moraxella catarrhalis*. However, there is a significant gap in understanding how these well-studied bacteria interact with lesser studied, often uncultivated species. These poorly characterised bacteria may exist in lower abundances within their ecological niches or

exhibit fastidious growth requirements and have not been thoroughly studied due to the challenges of cultivation[43,44]. The limited research on poorly characterized or less abundant taxa limits the interpretations of microbiota composition data in the context of respiratory health and disease.

A substantial portion of the many bacterial communities remains uncultivated, this is often referred to as microbial "dark matter" [45]. Despite widespread detection through 16S rRNA gene sequencing methods, a large proportion of bacteria have proven to be difficult to culture under standard laboratory conditions due to their fastidious nature, meaning that they require specific growth conditions for cultivation, such as temperature, pH, nutrient sources, or oxygen availability[46]. This can hinder successful cultivation from primary or polymicrobial samples. Furthermore, some of these bacteria may depend on interactions with other microbial species for survival, making isolation difficult in a laboratory setting[47]. Importantly, uncultivated or difficult-to-cultivate bacteria are not necessarily present in low abundance; they may dominate their microbial niches and play pivotal roles in microbiome dynamics[48]. For instance, the oral bacterium *Porphyromonas pasteri*, is a highly abundant species in the oral cavity and saliva that is strongly associated with oral health[48]. This species represents an important yet understudied organism due to its fastidious growth characteristics.

The advent of metagenomic sequencing, which involves the direct sequencing of unamplified environmental DNA, has provided a valuable tool for studying microbial communities without the need for cultivation[31]. This approach can be applied in a targeted manner, such as through 16S rRNA gene sequencing, to analyse microbial composition[31]. However, the targeted sequencing and analysis of one gene restricts the amount of functional information that can be inferred. In contrast, shotgun metagenomics, which involves the untargeted sequencing of all microbial DNA/nucleic acids in a sample, offers the potential to profile taxonomic composition, functional capabilities, and genomic diversity[31]. Shotgun metagenomic data can reveal novel genes and functions, providing deeper insights into microbial communities.

However, despite these advances, bacterial cultivation remains indispensable. Culturing bacteria allows for the simultaneous characterization of their genomes, metabolic roles in the environment, pathogenicity, and antibiotic susceptibilities[49]. Although metagenomics, particularly whole-genome sequencing, can facilitate the prediction of genotypic characteristics and sometimes phenotypic traits, cultivation is required for confirming these predictions, evaluating phenotypic traits, and conducting functional studies[45]. Culturing organisms is the also important for studying their interactions with other species[50].

Therefore, the cultivation of fastidious bacterial species, remains crucial even in the era of metagenomics. Culturing fastidious bacteria that are poorly characterised is important for investigating their functions and roles in health and disease. Expanding research to include theses bacteria can enhance our understanding of the NP microbiome and its impact on respiratory health.

1.6 The discovery of *O. hominis*

A longitudinal study by Salter *et al*. (2017) identified an unknown bacterial species prevalent in the children under the age of two years in Maela, a refugee camp in Thailand[12]. This study was undertaken between 2007–2010, investigated the NP microbiome of 955 infants living in the Maela refugee camp situated on the Thailand-Myanmar border. NP swabs were collected monthly from birth up to 24 months of age to investigate the associations of infant age with changes in microbiome composition[12].

Bacterial 16S rRNA gene sequencing was performed on the swab samples and data analysed to investigate microbiome changes. The researchers found that *Staphylococcus* and *Corynebacterium* species proportional abundances were higher in the first three months but decreased from 3–24 months[12]. Conversely, *Moraxella* increased in abundance from three to twelve months followed by a plateau, while an uncharacterised *Flavobaceriaceae* taxon, with 93% nucleotide identity to the *Ornithobacterium rhinotracheale*, increased in proportional abundance over the first year as seen in Figure 1.2.

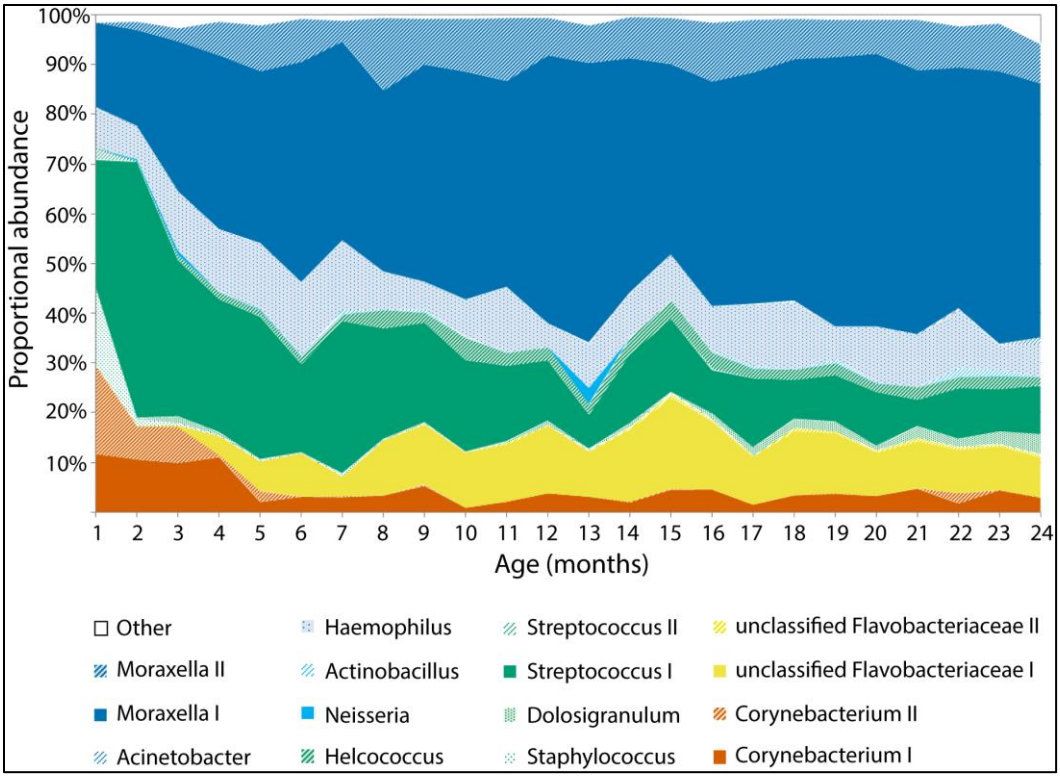

Figure 1.2. Aggregate of all routine swabs from all infants, proportional abundance of the 15 operational taxonomic units (OTUs) that account for >98% of the cohort microbiota, by age in months (1–24). OTUs are colour grouped by phylum: Actinobacteria (orange), Bacteroidetes (yellow), Firmicutes (green), Proteobacteria (blue). Adapted from Salter *et al.* (2017)[51]. CC BY 4.0.

This uncharacterized *Flavobacteriaceae* was found in 62.5% of samples from a sub-set of infants from the Maela cohort and appeared to be a persistent colonizer of the infant nasopharynx at high proportional abundances (up to 70% relative abundance in some samples), which warranted further investigation of this species[12]. Additionally, 16S rRNA sequences were found to be nearly identical (>99% identity) to sequences of the unidentified species deposited from cohort studies based in Gambian[52] and Kenyan[53], and in Australian cohorts[54], indicating the global distribution of this species.

In 2019, the first draft genomes of the unidentified species were assembled using shotgun metagenomic sequencing methods as culturing methods were unsuccessful at the time[14]. Candidate swab samples with high proportional abundance of the bacterium, based on 16S rRNA gene sequencing data, were selected from the Maela study. The DNA was extracted and subsequently amplified using multiple displacement amplification. Sequencing was performed on 300 bp paired reads using the Illumina® MiSeq™ platform and draft genomes were assembled. Additionally, fluorescent probes were designed using genomic data to

identify putative cells in mixed culture samples[55]. Two nearly complete draft genomes were assembled (i) draft genome OH-22767 (acc. no. GCF_900538225.1) consisted of 15 contigs, with a total size of ~1.9 Mb and the largest fragment of 611 kb and (ii) OH-22803 (acc. no. GCF_900538215.1) consisted of 15 contigs, with a size ~1.87 Mb, and the largest fragment of 684 kb. The presumptive species was proposed as "Candidatus *Ornithobacterium hominis*".

Ornithobacterium hominis was placed within the *Flavobacteriaceae* family, based on the 16S rRNA gene sequence data. Comparison of the draft genomes to the *O. rhinotracheale* reference strain (UMN-88) revealed that *O. hominis* appeared to be very distantly related to *O. rhinotracheale*, which is an avian respiratory pathogen[55]. The percentage of conserved proteins (POCP) was used to determine the relatedness of the species at the genus level. This concept states that two members of the same genus should share at least half of their proteins. The POCP between UMN-88 and the *O. hominis* was approximately 58%; therefore, based on this measure, they may be considered members of the same genus.

Since culturing was unsuccessful, a fluorescent probe targeting the V5 region of the *O. hominis* 16S rRNA was used to directly visualize the cells directly in NP swab samples[14]. The results revealed that *O. hominis* is a gram-negative bacillus, often observed in pairs or longer chains, which is a similar cell morphology to *O. rhinotracheale*[55].

1.7 *Ornithobacterium hominis* genomic insights

Analysis of the *O. hominis* draft genomes revealed several genomic features of interest. At the time of publication only 50% of the draft genome was identified as "core" genome due to the poor quality of the assemblies. The core genome refers to a sub-set of genes present in all members of the species; these tend to be more conserved and may be considered essential genes. The *O. hominis* core genomes include three copies of the rRNA operon, the large (3.8 kb) toxin gene (*toxA*), and 14 gliding motility-associated genes (*gld*)[55]. The accessory genome is a pool of genes that are less conserved, highly variable, and may be used

to distinguish strains[56]. The *O. hominis* accessory genomes include evidence of mobile elements, drug-resistance genes, rearrangement (*rhs*) genes and six distinct lipopolysaccharide (LPS) production clusters[55]. Such features are of interest as they indicate potential for host interaction, inter-species competition and may influence the growth and virulence of other respiratory tract pathogens[55].

1.7.1. The core genome

The core genome included the *toxA* gene. The *toxA* gene is predicted to encode a toxin similar to the *Pasteurella* mitogenic toxin (PMT)[55]. The PMT is a 146 kDa protein expressed by toxigenic strains of the zoonotic bacterium *Pasteurella multocida.* Toxigenic strains are more commonly identified in pigs and other farm animals, whereas the majority of *P. multocida* infections in humans are non-toxigenic and transmitted though bites or scratches from household pets. *P. multocida* infections can lead to atrophic rhinitis in pigs and rabbits, and dermonecrosis in human bite or scratch wounds from infected animals[57].

The PMT is structured into an N-terminus receptor-binding domain (involved in receptor binding and translocation into the cell) and a C-terminus catalytic site. The C-terminus is sub-divided into domains C1 (membrane-binding domain), C2 (unknown function) and C3 (containing the active site)[57]. Once inside the cell, the PMT targets the alpha-subunit of specific heterotrimeric G proteins and constitutively activates them through the deamination of a glutamide residue to a glutamate[57,58]. This activates the G-signalling pathway, resulting in a myriad of cellular responses, and modulates several signalling pathways, such as calcium signalling, cytoskeletal signalling, adipogenic signalling and osteogenic signalling[58]. PMT may also be potentially carcinogenic as it can activate several mitogenic pathways such as Map kinase and STAT signalling[58]. The overall amino acid identity shared between the predicted *O. hominis* toxin and PMT is 35%; however, protein modelling indicates a greater conservation in the predicted active site suggesting that this *toxA* gene might have similar functions (Figure 1.3)[55]. Preliminary assays with the dermonecrotic toxin confirm that it deamidates and binds to human targets [pers. comm. Professor Klaus Aktories].

A full complement of *gld* genes was also present in the core genome[55]. Most bacteria rely on external features such as flagella or pilli for movement while some taxa, specifically some species in the *Bacteroidetes* phylum, rely on gliding motility[59]. Movement in the environment is advantageous for nutrient scavenging, biofilm formation and for contact with other bacterial cells or host cells[55,59]. *Flavobacterium johnsoniae* is a well-studied gliding motility model organism, which uses cell surface-anchored adhesins to move on surfaces[64]. The major motility adhesins required for gliding motility are sprB and RemA are the major motility, while SprB is the main adhesin in *F. johnsoniae* that interacts with the substratum to propel the cell[55,60]. The *gld* genes identified in *F. johnsoniae* were initially believed to be essentially for gliding motility, however it was later discovered that many of the *gld* genes are actually involved in a transport system, the type IX protein secretion system (T9SS), that transports the adhesin. *Flavobacterium johnsoniae* utilises the T9SS for surface localisation of these adhesins[59,60].

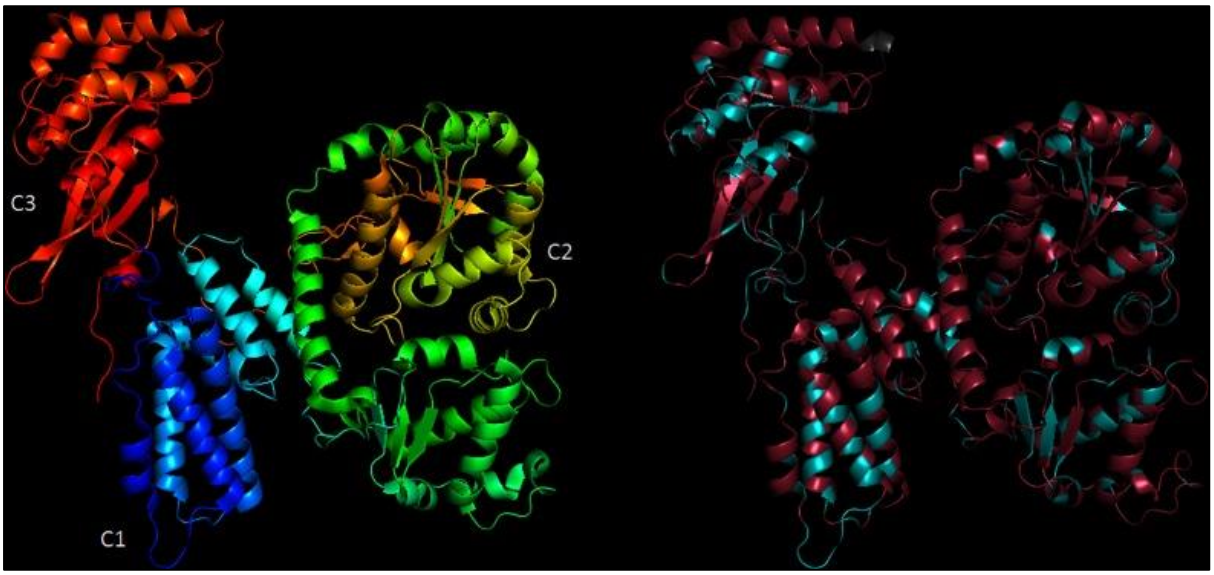

Figure 1.3. Structure of PMT C-terminal domain, and a model of the equivalent region of 'Candidatus O. hominis' ToxA. The PMT structure (left) is coloured on a rainbow spectrum to indicate position. The 'Candidatus O. hominis' model (right) is coloured according to amino acid identity to PMT with identical residues in blue, non-identical in purple and insertion in grey. In PMT, the C1 region is responsible for plasma membrane localization, the C3 region forms an active pocket and is responsible for mitogenic activity in mammalian cells. Adapted from Salter *et al*. (2019)[61]. CC BY 4.0.

The *gld* genes present in the *O. hominis* genomes are homologues of *F. johnsoniae* gliding motility genes[12]. Although *O. hominis* possesses all 14 *gld* genes and the additional sprAET (which is required for

the secretion of the SprB in *F. johnsoniae*), sprB homologues have not been identified in *O. hominis*[55]. Furthermore, *O. hominis* is a non-motile bacterium; thus, these *gld* genes are likely not involved in gliding motility, but rather encode for the components of the T9SS; this protein translocation system is not only involved in the secretion of gliding motility proteins in motile species but is also essential for the export of large proteins in non-motile relatives of the *Bacteriodetes* phylum[62]. These exported proteins include extracellular adhesins, virulence factors, and factors involved in biofilm formation[63].

1.7.2. The accessory genome

Approximately half the accessory genome are hypothetical genes, some of which encode for bacteriophage-associated genes, mobilisation genes, drug-resistance genes, and proteins containing the *Fibrobacter succinogenes* major paralogous domain (FSD). At the time of publication of the draft genomes, Salter *et al*. (2019), reported that up to 17 of these genes containing the FSD were present in the genomes, but this has since been update to approximately 20 genes[13]. Furthermore, approximately a quarter of these FSD-containing proteins were predicted to contain an immunoglobulin-like fold[12]. The accessory genomes also include evidence of *rhs* genes and LPS biosynthesis clusters.

The accessory genome contains evidence of *rhs* genes which encode for polymorphic toxins[14]. Polymorphic toxins are composed of an N-terminal region that specifies their mode of secretion into the extracellular or into the target cell, and a C-terminal domain that has toxin activity[64]. The hypervariable C-terminal domain of RHS proteins can encode a range of toxin activities, including endonuclease, peptidase, and deaminase functions[65]. RHS toxins achieve their effects by targeting conserved structures, including the cell wall, cell membrane, and nucleic acids[65]. These RHS proteins may function either as antibacterial agents, serving as a competitive mechanism, or as factors targeting the host[64].

Another feature of the *O. hominis* accessory genome are the genes of the LPS biosynthesis cluster, which is an important contributor to virulence[14]. The bacterial outer membrane, in particular, the surface exposed LPS, is essential for the survival of the bacterium in the host environment, it is implicated in

multi-drug resistance and the overall protection of the bacterial cell[66]. In addition, LPS molecules are the immunodominant antigens of gram-negative bacteria and play an essential role in host-microbial interactions. The LPS is a glycoconjugate with 3 domains: 1) lipid A domain, 2) the core oligosaccharide, and 3) the O polysaccharide[66]. The O-antigen is the immunodominant portion and is recognised by the innate immune system[66]. The genes involved in the biosynthesis of the O-antigen are found in clusters[66]. In the case of *O. hominis*, the presence of six LPS clusters, determined from Australian genomes and Thai genomes from the Maela cohort, suggests that there are at least six serotypes in this species (Figure 1.4)[55].

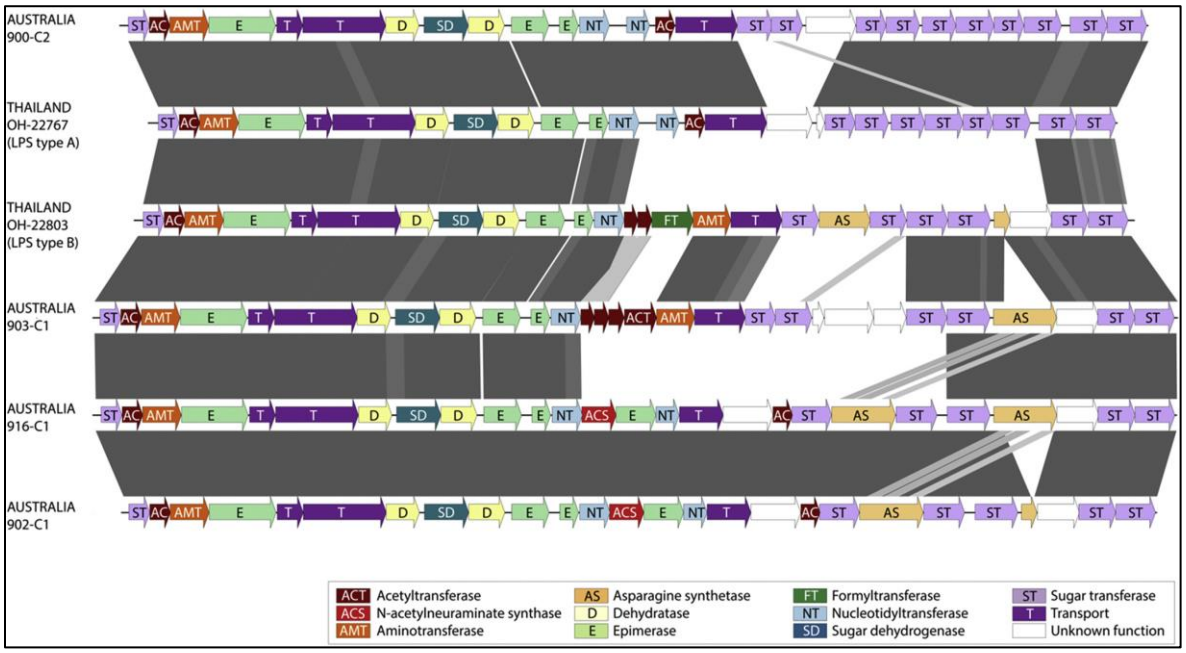

Figure 1.4. Comparison of Ca. *Ornithobacterium hominis* lipopolysaccharide biosynthesis loci. A tblastx alignment of the lipopolysaccharide biosynthesis clusters in the four Australian OH isolates compared to draft genomes OH-22767 and OH-22803 derived from Thailand (2). Adapted from Lawrence *et al*. (2019)[67]. STM signatory.

1.8 Insights from culturing *O. hominis*

At the time the draft genomes of *O. hominis* were published, culturing efforts were unsuccessful; however, a method for culturing this bacterial species was recently developed, allowing for further studies and investigations into the clinical significance of this species[68]. NP swabs were collected from Australian children (1–2 years old) and stored in skim milk-tryptone-glucose-glycerol broth (STGG) were confirmed to be *O. hominis*-positive by 16S rRNA gene sequencing[68]. Samples with a 5–55% relative abundance

were regarded as *O. hominis* positive. The bacterium was cultured on Tryptic Soy Agar (TSA) and incubated in a microaerophilic atmosphere at 35 °C for up to five days[68]. These culturing and incubation conditions were optimal for primary isolation of *O. hominis* bacterium. Colonies also formed on horse-blood agar, chocolate agar and brain-heart infusion agar, but isolates were inconsistently recovered. All isolates were pleomorphic and confirmed to be gram-negative bacilli[68].

Analysis of draft genomes from the Maela cohort in Thailand by Salter *et al*. (2019) predicted that the *O. hominis* isolates were oxidase-positive, catalase-negative, and produced β-lactamase[14]. These characteristics were confirmed in the cultured isolates, demonstrating the predictive value of genomic analyses[68]. In addition to these characteristics, the isolates also hydrolysed tributyrin[68]. Notably, some pure isolates produced two colony morphologies, which is suggestive of small-cell variants[68]. The isolate genomes were also analysed and exhibited an average nucleotide identity of 97.86–98.23% with draft genomes from the Maela cohort in Thailand, OH-22803 and OH-22767[68]. B-lactamase production was also found to be associated with multiple mobile genetic elements, suggesting that β-lactam resistance was acquired over multiple independent events[68]. Finally, consistent with previous genomic work, LPS clusters were heterogenous among the isolates, further suggesting multiple capsular types[55,68].

1.9 *Ornithobacterium hominis* in high respiratory illness populations

Ornithobacterium hominis has been detected in populations experiencing high rates of respiratory illness. In the Maela refugee camp, where childhood pneumonia is prevalent (0.73 pneumonia episodes per child per year), *O. hominis* was detected in children at an estimated prevalence of 42% of 12-month-old infants, reaching up to 70% relative abundance in NP samples[12]. Additionally, in an Australian study investigating the microbiota of children aged 2–3 years with chronic lung disease, including protracted bacterial bronchitis and chronic suppurative lung disease, *O. hominis* was detected at a maximum relative abundance of 54% in NP samples[69]. Thus, *O. hominis* has been detected at high relative abundances in children with chronic lung infections. Another study based in Australia explored the relationship between the bacteriome and ear and nose health in Indigenous Australian children aged 2–7 years old[70]. In this

cohort, the presence of an *Ornithobacterium* species, which the authors believed to be *O. hominis*, was associated with otitis media. The presumed *O. hominis* was notably absent or present in low relative abundances in children with no history of otitis media[70]. This suggests that *O. hominis* may represent a novel otopathogen. Thus, *O. hominis* colonization may be linked to poor ear health[70]. Furthermore, potential relationships between *O. hominis*, *Helcococcus*, and *Dichelobacter* species were demonstrated by network analyses, and these associations might be clinically important. Although the role of *O. hominis* in respiratory or ear disease is still unclear, these findings suggest that it may play a role in respiratory health outcomes.

1.10 *Ornithobacterium hominis* in Africa

Despite recent advances in culturing *O. hominis*, there is little information available about the prevalence of this species from other countries. Although *O. hominis* has been studied from samples from the Maela cohort in Thailand and Australian samples, there is even less information available for this species in Africa. *Ornithobacterium hominis* has been detected in nasopharyngeal 16S rRNA datasets from The Gambia, Kenya, and most recently Egypt[71]. A study by Ahmed *et al*. (2024) reported draft metagenome-assembled genomes (MAGs) of *O. hominis* from NP samples of infants in Egypt[71]. DNA was extracted from NP swab samples and subjected to shotgun metagenomic sequencing using the Illumina® platform. The assembled genome was comprised of 16 contigs and was a total length of 1.84 Mb[71]. The Egyptian draft genome (acc. no. GCA_030149645.1) included antibiotic resistance genes such as *murA*, which confers resistance to Fosfomycin, as well as other genes associated with resistance to aminoglycosides, isoniazid, and β-lactam antibiotics[71]. Genes associated with virulence were also identified and include the *vapD* gene and a gene encoding for a parE toxin, which possibly targets DNA gyrase[71]. Although genotypic information indicates potential resistance mechanisms, the contribution of these genes to antibiotic resistance and virulence must be validated phenotypically; however, the limited number of cultured isolates hinders this further work.

Recently, *O hominis* 16S rRNA sequences were identified in a South African birth cohort nested within the DCHS[72]. The DCHS population experience high rates of HIV infections and respiratory illness; however, the mother-to-child transmission is very low in this birth cohort due to effective mother-to-child HIV prevention programs[73]. Furthermore, there is high vaccination coverage with vaccines contained in the South African national immunisation programme which includes PCVs, specifically PCV13[73]. Despite effective primary care, pneumonia is still major cause of illness and hospitalization in this cohort, particularly amongst children less than 1 year of age, with an incidence rate of 0.27 episodes of LRTIs per child every year[73]. Nevertheless, mortality rates due to severe pneumonia are low due to effective health-care programs, which encompass access to hospitals and antibiotic treatment[73]. Thus, despite effective primary care, high vaccination coverage, and access to hospitals; this population is burdened by respiratory illness.

1.11 Rationale, aims, and objectives

The role of *O. hominis* in respiratory health and disease risk remains unclear, despite its identification as a persistent colonizer at high proportional abundances in several studies. Genomic evidence suggests that *O. hominis* may influence the NP environment through interactions with the host and potentially pathogenic species, thereby affecting disease outcomes. However, the specific impact of *O. hominis* colonization on respiratory disease development is not well studied. Although *O. hominis* has been detected in 16S datasets from The Gambia, Kenya, Egypt, and South Africa, it remains poorly characterized from African samples, with no cultured isolates or complete genomes available, apart from Australian isolates and the type-strain genome. Given that *O. hominis* was detected in our study population, which experiences high rates of respiratory illness despite extensive vaccine coverage, characterizing *O. hominis* in the South African context is crucial. This would provide a foundation for future research on this species in Southern Africa, where respiratory infections are disproportionately prevalent.

Aim and objectives: The aim of this project was to characterise *O. hominis* isolates from NP samples from a South African birth cohort. The first objective was to (i) isolate the *O. hominis* from archived samples from the DCHS using standard microbiological techniques and then characterize standard morphological aspects such as colony morphology and antibiotic susceptibility. The second objective was to (ii) obtain whole genome sequences of these isolates using the Oxford Nanopore Platform (ONT). Finally, (iii) genomic characteristics were determined using various bioinformatic tools and genomes analysed and compared using a pangenome analysis. This research could provide insights into the genomic characteristics of *O. hominis* and pave the way for future research studies into this species.

Chapter 2: Methods and Materials

2.1 Materials

Archived nasopharyngeal samples: Archived nasopharyngeal samples were obtained from the DCHS. Culture media and reagents: CampyGen 2.5 L sachets, Columbia agar, Todd-Hewitt's broth, and brain-heart infusion broth (BHI) was purchased from Oxoid. Tryptic soy agar (TSA) was purchased from Merck. Sheep's blood was obtained from Media Mage. PCR reagents: OneTaq® Quick-Load® 2× MasterMix with standard buffer (M04865) was purchased from New England Biolabs. Primers were synthesised and purchased from Inqaba Biotech. Antibiotic susceptibility disks were obtained from Oxoid and Nitrocefin disks were obtained from ThermoFisher Scientific. Genomic DNA extraction kit: Wizard HMW (High molecular weight) DNA extraction kit from Promega. RNase and Proteinase K were also from Promega. Sequencing platform: Oxford Nanopore Technologies (ONT)

2.2 Study design and participants

Samples utilised for this study are based on a case-control study of infants enrolled in the DCHS[72]. This study took place in the Drakenstein municipality, a peri-urban region outside of Cape Town, South Africa. Ethical approval was received from the Human Research Ethics Committee of the University of Cape Town, South Africa (401/2009 and 585/2015). Pregnant mothers were enrolled in their 2nd trimester of pregnancy and 1,137 mother-child pairs were followed until the children were five years of age. All births and hospital care occurred at a central public hospital, Paarl Hospital. Nasopharyngeal samples were collected every two weeks from birth until six months of age, then monthly until 12 months of age and six monthly thereafter (Figure 2.1)[72]. Demographic and clinical data were recorded antenatally at birth, and postnatally during scheduled study visits. Episodes of LRTIs during the first year of life were identified using active surveillance at local clinics and hospitals, following the World Health Organisation criteria[74]. Mothers were educated about respiratory symptoms and advised to contact study staff whenever their infant developed respiratory illness symptoms[74]. Infants were monitored during hospitalization or ambulatory illness. LRTI cases and non-LRTI controls were matched 1:1 by birth date and study site.

2.3 Specimen collection and matching

Paired NP flocked swabs (Copan Diagnostics, CA, USA) were collected fortnightly from infants across the first year of life (0–365, +14 days). One of these swabs was placed in PrimeStore® Molecular Transport medium (Longhorn Vaccines & Diagnostics, MD, USA) for 16S rRNA sequencing and the other swab was placed in STGG for microbiological culture. A window of +/-14 days was allowed between LRTI episode date and specimen collection date for inclusion as a LRTI-case specimen. Samples were assigned unique identification number to ensure confidentiality. Archived NP samples were stored at -80 °C.

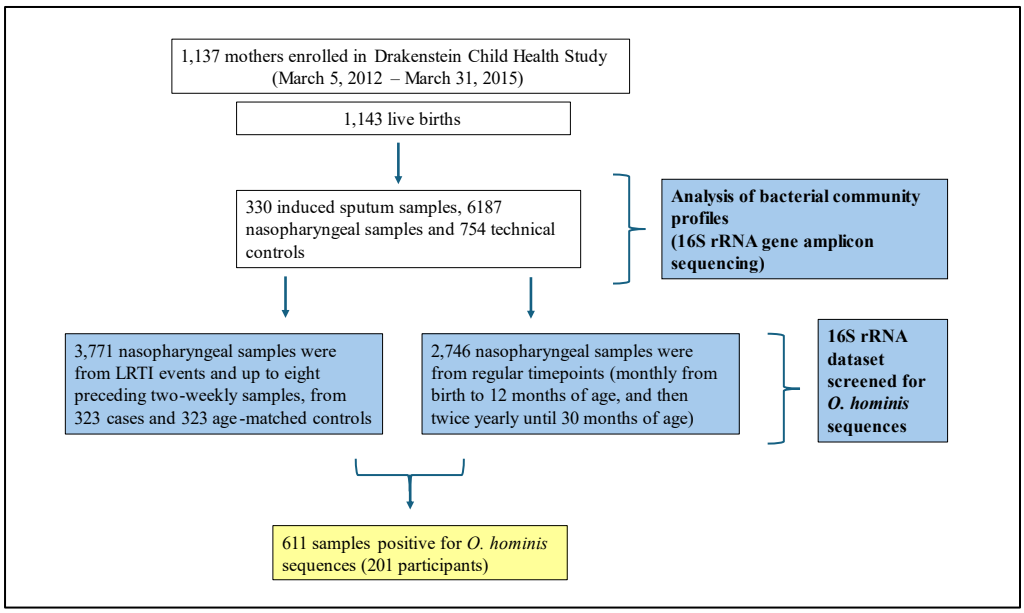

Figure 2.1. Outline of study design, participants, and sample matching. LRTI, Lower respiratory tract infection.

2.4 16S rRNA gene sequencing

16S rRNA gene amplicon sequencing was performed on total nucleic acids that were extracted from 330 induced sputum samples, 6,187 nasopharyngeal specimens, and 754 technical controls, as described elsewhere[20]. Each sequencing run included sequencing controls. Total 16S rRNA gene copy numbers were measured via qPCR. 16S rRNA gene amplicons were generated using a two-step PCR protocol targeting the V4 hypervariable region of the 16S rRNA gene. Pooled libraries were sequenced on the Illumina® MiSeq™ platform using V3 chemistry with 2 x 301 cycles [MiSeq™ Reagent Kit v3 (600-cycle) Reagent Cartridge (Illumina, CA, USA)]. The DADA2 pipeline was used to filter reads and infer amplicon sequence

variants (ASVs). Taxonomy was assigned using the RDP classifier. Then, Rv4.1.2 and RStudio v2021.09.2 were used to remove Eukaryota and unclassified Kingdom-levels ASVs.

2.5 *Ornithobacterium hominis* screening

To identify the presence of *O. hominis* in the samples, the amplicon dataset was screened for perfect matches to three sequences unique to this species in the 16S V4 region:

i. GAGCGTTATCCGGATTCATTGGGTTTAAAGGGTCYGTAGGCGGGCTRATAAGTCAGTG GTGAAATCTCAC
ii. GAGTGAGTTTGATGTTGCTGGAATGTGTAGTGTAGCGGTG
iii. ATGCGAAGGCAGGTAACAAAGACTTAACTG.

2.6 Selection of NP samples for *O. hominis* isolation

Based on 16S rRNA gene abundance data, the samples with the highest abundances of *O. hominis* reads (> 1%) were selected for isolation attempts maximize chances of recovery from primary culture (Figure 2.2).

2.7 Isolation of *O. hominis*

Frozen samples were thawed at room temperature and aliquots of 20 µl were streaked onto 2% Columbia agar plates supplemented with 5% sheep's blood (CBA). Plates were incubated at 30 °C and 37 °C for at least 5 days under microaerophilic conditions. CampyGen sachets were placed in a secondary containment with the plates to generate a microaerophilic environment. Plates were incubated at 30 °C. Bacterial colonies resembling *O. hominis* were selected and sub-cultured on CBA and incubated at 30 °C for at least 48 hours, as described by Lawrence *et al*. (2019)[68]. Colonies from primary culture are reportedly greyish and pinprick-sized, and only apparent after 5 days on incubation[68]. Colonies suggestive of *O. hominis* were sub-cultured until pure cultures were obtained.

Potential *O. hominis* colonies were gram-stained to confirm cell morphology (Figure 2.2). Colonies were subjected to PCR targeting the conserved *O. hominis* t*oxA* gene, as described previously[14]. PCR was performed on single colonies. Briefly, a single colony was picked with a sterile loop and transferred to

200 µl of AVE buffer (Qiagen, Germany) and heated at 95 °C for 5 minutes to lyse the cells. Lysate was diluted 1:10 in nuclease free water and PCRs were performed with 1 µl of diluted lysate in 25 µl reactions using OneTaq® Quick-Load® 2× MM and OH_TOXIN primers (Inqaba biotech, South Africa) targeting the *toxA* gene.

Primer sequences:

i. Forward primer (OH_TOXIN-F_ii): 5′-GATGTATTGATAGATACTCCCGCCATTACG-3′.
ii. Reverse primer (OH_TOXIN-R_ii): 5′-CTATATTTGGGAAAGGCGCATGAATACC-3′.

The TOXIN PCR cycling conditions were as follows: Initial denaturation at 94 °C for 30 sec; 30 cycles of 94 °C for 30 sec; 55 °C for 30 sec; 68 °C for 1.20 min; Final extension at 68 °C for 5 min; Hold: 4–10 °C. Product visualisation was done as follows: PCR products were stored at 4 °C and visualized on an 1.6% agarose gel stained with SYBER safe dye. Cultures that were positive for the t*oxA* gene (~1,7 kb) were identified to be *O. hominis* isolates. Isolates were stored in BHI with 50% glycerol solutions (1 mL) and stored at -80 °C for further analysis.

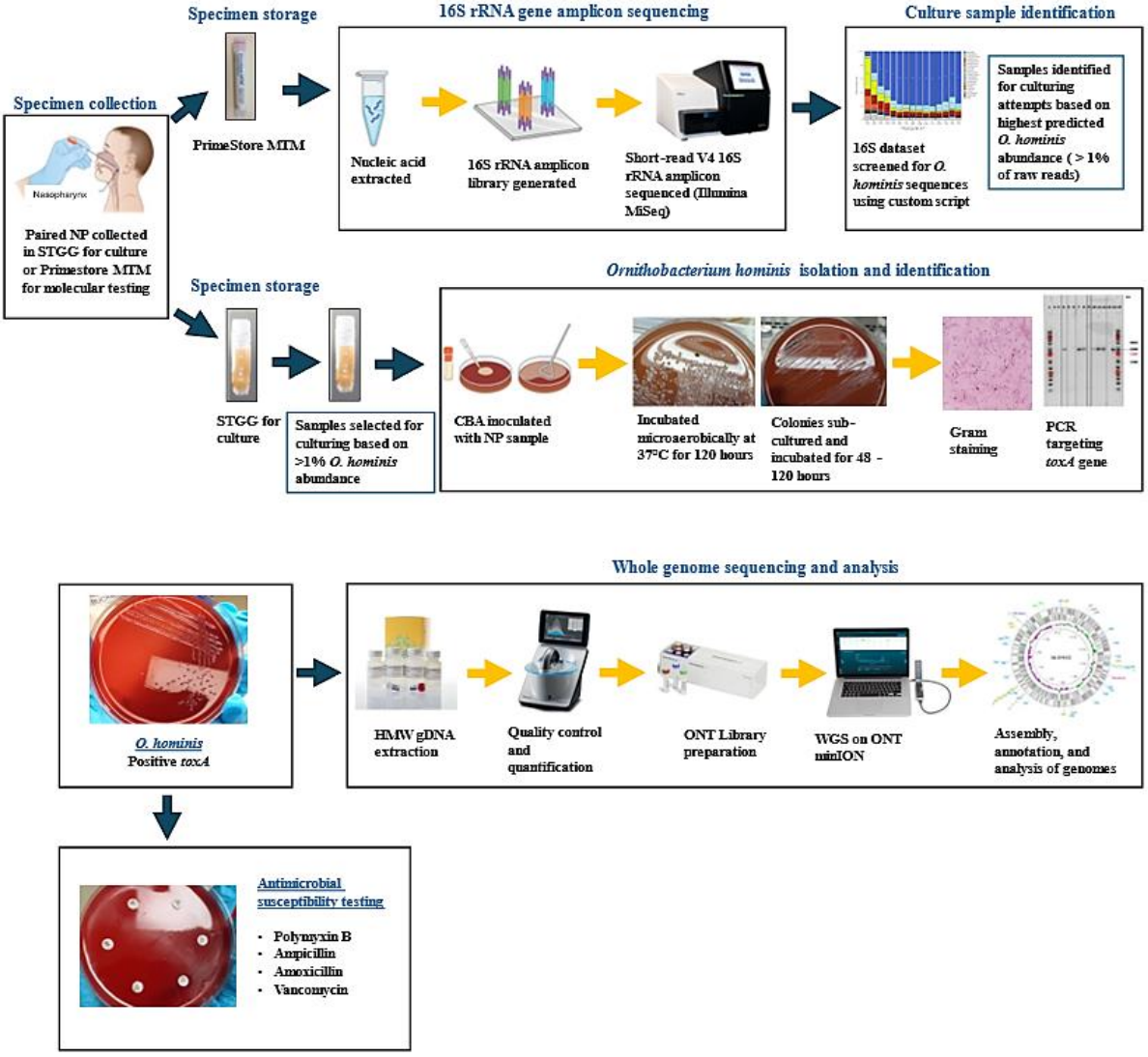

Figure 2.2. Study workflow. NP, Nasopharynx; STGG, skim milk, tryptone, glucose, and glycerine; CBA, Columbia Blood agar; HMW, High-molecular weight; gDNA, genomic DNA; ONT, Oxford Nanopore Technologies; WGS, Whole Genome Sequencing.

2.8 Microbiological Characterization

2.8.1 Optimizing *O. hominis* growth and culturing conditions

Colony growth and morphology of *O. hominis* were evaluated under different cultivation conditions to optimize growth and appearance. Pure cultures of *O. hominis* were inoculated onto CBA and incubated under microaerophilic conditions. Cultures were incubated at 35 °C and 37 °C, and colony morphology

was assessed after five days of incubation. Colony size, texture, shape and growth density were evaluated and compared across multiple isolates.

2.8.2. Gram Staining

Gram staining was performed by following standard procedures to determine the bacterial cell wall characteristics and cell shape. Gram staining was performed as described by the American Society for Microbiology[75].

2.8.3. Antibiotic Susceptibility Testing and β-lactamase production testing

The antibiotic susceptibility tests were determined using a modified version of the disk-diffusion method[76]. Frozen stocks were thawed and inoculated onto CBA plates, followed by incubation at 30 °C for 72 hours. A loopful of colony was inoculated into 3 mL (~5 mm depth) of Todd-Hewitts broth in a flat-bottomed vial, lined with approximately 1 mL of TSA. Liquid cultures were incubated static for 48 hours at 37 °C, and thereafter cultures were resuspended and assessed visually for turbidity. A 1:1 dilution of the culture was made in Todd-Hewitts and further incubated static for 3–4 hours or until an increase in turbidity was apparent using McFarland standards; this allowed the *O. hominis* growth to recover. A volume of 100µl of the diluted broth was inoculated onto CBA plates using the spread-plate method. Antibiotic disks of amoxicillin (25 µg), ampicillin (10 µg), vancomycin (5 µg) and polymyxin B (300 µg) were placed onto the agar surface using sterile forceps. Plates were incubated for 48 hours at 30 °C, growth was evaluated, and zone sizes were measured using callipers. Isolates were tested in triplicate. According to the European Committee on Antimicrobial Susceptibility testing (EUCAST), categorical reporting of "resistant" or "susceptible" is not recommended for species without established breakpoints. Since there are no established EUCAST breakpoints for *O. hominis*, only zone diameters were reported. Colour-changing nitrocefin disks were used according to the manufacturer's instructions, briefly, a sterile disk was placed onto a petri dish lid and moistened with a drop of sterile, distilled water. A loopful of *O. hominis* from each plate was spread onto the disk and incubated for 5 minutes at room temperature,

aerobically. A colour change from white to pink indicated a positive result for β-lactamase production, while no change in colour indicated a negative result. Isolates were tested in triplicate.

2.9 Genome sequencing and assembly

2.9.1. DNA extraction

2.9.1.1 Phenol chloroform isoamyl (PCI) alcohol method

Total whole cell genomic DNA was initially extracted using the phenol-chloroform-isoamyl (PCI) method. Briefly, the colonies were inoculated into biphasic media (TSA and BHI) and grown overnight. The liquid portion of the overnight cultures was centrifuged at 16, 000 × *g* for 5 minutes to pellet the cells. Cell pellets were stored at -80 °C until ready for DNA extraction. All centrifugation steps occurred at room temperature unless otherwise stated. Pellets were thawed and 200 µl of Lysis buffer with 8 µL of Proteinase K was added. The sample was heated for 2 hours at 56 °C to lyse the cells, followed by adding 200 µl of PCI, and centrifugation was performed at 16, 000 × *g* for 5 minutes. The upper aqueous layer containing the nucleic acids was gently pipetted off the organic layer, taking care not to disturb the interface, and dispensed into a sterile tube. Nucleic acids were precipitated using ethanol precipitation. 200 µl of absolute ethanol was added to the sample and incubated at -20 °C overnight. After overnight precipitation, the samples were centrifuged for 30 minutes at 16 000 × *g* to pellet the nucleic acids. Supernatant was decanted and the pellet was washed with 150 µl of 70 % ethanol, followed by centrifugation for 10 min at 16 000 × *g*. Ethanol was decanted and the DNA pellet was left to air dry for 10 minutes. The dried pellets were resuspended in 50 µl of TE buffer and stored at 4 °C until ready for further analysis.

2.9.1.2. Promega's Wizard® HMW DNA extraction kit method

Genomic DNA was extracted from pure bacterial cultures using the Wizard HMW DNA extraction kit (Promega, USA). The protocol was performed according to the manufacturer's instructions with minor modifications[77]. Cell material was obtained from plate scrapes, instead of overnight liquid cultures, and pelleted in PBS by centrifugation at 16, 000 × *g* for 2 minutes. Supernatant was removed and the cell

pellet was resuspended in PBS. Resuspensions were incubated at 85 °C for 5 minutes for cell lysis. Then, 500 µL of lysis buffer was added along with 6 µL of RNase A and incubated at 37 °C for 30 minutes. Then, 20 µl of Proteinase K was added and the samples were incubated at 56 °C for 15 minutes. Protein was precipitated using the protein precipitation solution and the samples were left on ice for 5 minutes. Thereafter, the samples were centrifuged at 16, 000 × *g* for 10 minutes at room temperature or until protein pellet was visible and supernatant appeared clear/debris free. The supernatant was carefully transferred, without disturbing the protein pellet, to a tube containing room temperature isopropanol for DNA precipitation. The samples were gently inverted a few times until white/translucent threads of DNA were visible. Once threads were visible, the samples were centrifuged at 16, 000 × *g* for 2 minutes to pellet the DNA. An ethanol wash followed after the supernatant was discarded, without disturbing the DNA pellet, and 70 % ethanol was added. The samples were inverted to wash the sides of the tube and centrifuged at 16, 000 × *g* for 2 minutes. Then, ethanol was discarded, and the DNA pellet was air dried at room temperature for 10 minutes. To achieve a higher concentration, the DNA was left to resuspend in 50 µL of DNA resuspension buffer, instead of 100 µL, overnight at room temperature. DNA was stored at 4 °C until ready for further analysis.

2.9.2. DNA quality assessment

DNA quality was determined using a Nanodrop (Nanodrop One, ThermoFisher Scientific, USA) and quantified using a Qubit Fluorometer (Qubit 4, ThermoFisher, USA). DNA was extracted from strains SA (South Africa)-OH (*O. hominis*) -C4 and SA-OH-C6 to compare extraction methods. For each strain, DNA was extracted using both methods: three isolates were processed with the kit-based method and three isolates with the PCI method. Nanodrop and Qubit measurements were recorded in triplicate for each sample. DNA integrity was assessed by DNA gel electrophoresis on a 1.6% agarose gel stained with Syber Safe DNA gel stain (ThermoFisher Scientific, USA) to visualise nucleic acids.

2.9.3. Library preparation and sequencing

Genomic libraries were prepared following ONT barcoding kit. The sequencing library was prepared with the Native Barcoding Kit 24 v12 (SQK-NBD112.24; Oxford Nanopore Technologies, UK) according to the manufacturer's guidelines. Briefly, DNA was FFPE repaired and end-prepped/dA-tailed, then a unique dT-tailed barcode adapter was ligated on the dA-tailed template. Barcoded samples are pooled together, and then sequencing adapters are ligated. The library was loaded onto a primed R10.4 flow cell on a MinION device for a 72-hour run (FLO-MIN114; Oxford Nanopore Technologies, UK). Real-time basecalling was conducted with the MinKNOW software.

2.9.4 Genome assembly: Pipelines

Post-hoc base calling was undertaken using the Super Accurate model and barcode removal settings in Guppy v6.5.7 (GPU) within MinKNOW. Raw reads were subjected to two pipelines, separately, to assemble genomes *de novo*.

Pipeline 1

Pipeline 1 was conducted according to Salter *et al*. (2023), but with minor alterations in the pipeline tools used (Figure 2.3)[13]. Reads were trimmed using Prowler to a Q20 average in a 5,000-nucleotide window, discarding reads of <5 kb[78]. Alternatively, reads were trimmed to a 1 kb length cut-off for assemblies with low coverage. Preliminary assemblies were generated with Flye v2.9.2 using default settings[79]. Consensus genomes were generated with Medaka v1.11.3 and polished using trimmed reads. The genomes were annotated with the web-based version of Bakta v1.8.1[80].

Pipeline 2

Pipeline 2 is based on an unpublished pipeline, assembleBAC-ONT v1.1, developed by Andries van Tonder (Figure 2.3)[81]. This is an automated pipeline using Nextflow as a workflow manager[82]. Briefly, Artic guppyplex aggregates pre-demultiplexed reads from MinKNOW/Guppy. Long reads are filtered using Filtlong, discarding reads of < 5 kb. The assemblies were generated using Flye v 2.9.3 and polished with trimmed reads using Medaka. Genome assemblies were annotated with Bakta.

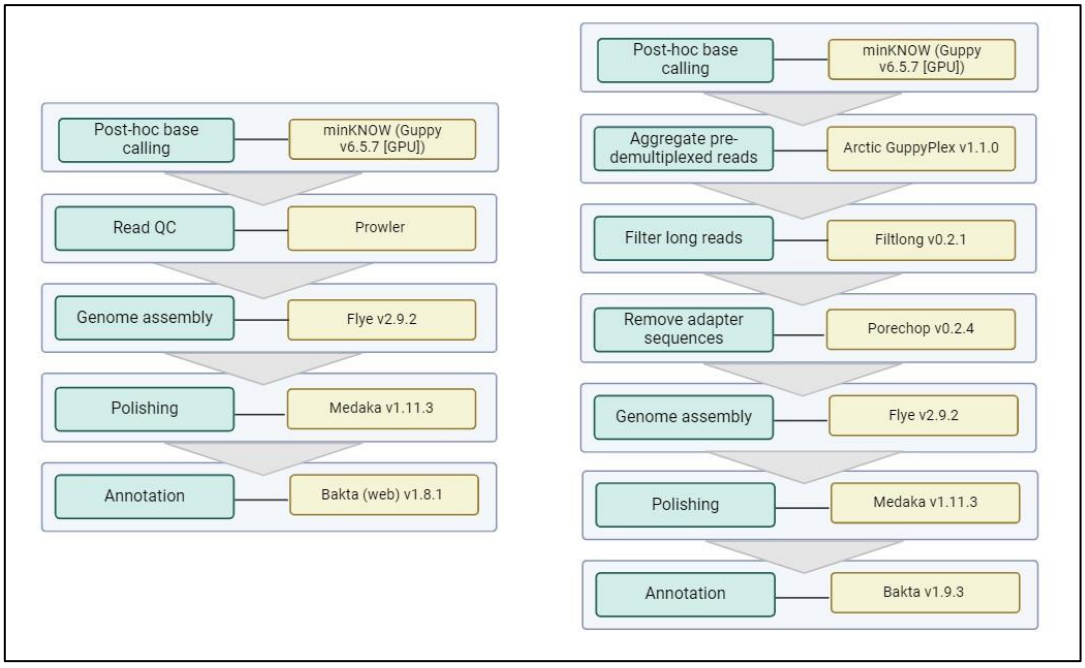

Figure 2.3. Workflow for Pipeline 1 (left) and Pipeline 2 (right).

2.10 Genomic analysis

After assembly, genomes were assessed for coverage, size, number of contigs, and gene annotation. Genomes were rotated to the origin of replication as determined by the position of the *mnmG* gene and GC skew. Assemblies of more than one contig were subjected to nucleotide Basic Local Alignment Search Tool (BLASTn) alignment to account for possible contamination by other species. Mapping was performed using Minimap2 v2.24 to map reads to contigs to improve select assemblies[83]. Circular genomes plots were generated using Proksee with integrated GC skew plot tool v1.0.2[84]. Antimicrobial

resistance (AMR) genes were identified using the NCBI's AMRFinderPlus tool v3.12.8 and the Proksee-integrated Comprehensive Antibiotic Resistance Database (CARD) Resistance Gene Identifier (RGI) tool v1.2.1.

2.11 Pangenome analysis, core genome alignment, and functional prediction

Pangenome analysis was performed using the Panaroo v1.4.3 pipeline[85]. Annotated South African and Australian *O. hominis* genomes as well as the MAGs from the Maela cohort in Thailand were included in pangenome analysis. Panaroo was performed using default settings for the strictest mode and generated a core genome alignment (all genes present in 98% of isolates) using the ClustalW multiple sequence aligner. The core genome alignment tree was calculated in MEGA11 and visualised in the iTOL tree visualizer[86,87]. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model. The bootstrap consensus tree inferred from 10 replicates is taken to represent the evolutionary history of the taxa analysed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10 replicates) are shown next to the branches. Functional prediction of core genes performed using eggnog-mapper v2.0.1[88].

2.12 Statistical analyses

All statistical analyses were conducted using GraphPad Prism version 10.2.0 (GraphPad Software, USA). The comparison of 260/230 and 260/280 ratios between PCI- and Kit-extracted DNA was performed using unpaired t-tests to determine if there were significant differences in nucleic acid purity between the two methods. A one-way analysis of variance was conducted, followed by Bonferroni's multiple comparisons test to determine if there was a significant difference between the Qubit and Nanodrop measurements for PCI- and Kit-extracted DNA. A p-value of < 0.05 was considered statistically significant.

Chapter 3: Results

3.1 Nasopharyngeal specimens and participant characteristics

Screening of the 16S rRNA sequencing dataset from prior DCHS research identified multiple candidate samples containing *O. hominis* sequences, which were selected for culture attempts. Based on this initial screening, a more extensive analysis was conducted on *O. hominis*-positive samples. Of the 6,188 samples collected from 646 infants, *O. hominis* was detected in 32% of the children (201 read-positive) during the first 2.5 years of life. Notably, most samples from these 201 children exhibited a low relative abundance of *O. hominis* (< 0.1%), and in most cases, *O. hominis* was not carried longitudinally (i.e., present in multiple samples over time). Therefore, we narrowed the selection to participants with multiple *O. hominis*-positive samples and proportional abundances > 0.1%. This approach increases the confidence that these children, who display longitudinal *O. hominis* carriage at higher proportional abundance, are *O. hominis* carriers rather than cases of contamination. Furthermore, this criterion ensured that our study focused on individuals with sustained and possibly significant *O. hominis* colonization.

Of the 201 *O. hominis*-positive children, 46 children showed longitudinal carriage of *O. hominis* with at least 0.1% proportional abundance (Table 3.1). Of the 46 participants, most were female (58.69%, 27/46), delivered vaginally (84.8%, 39/46) and full-term (78.3%, 36/46). Less than one quarter of infants were HIV-exposed (17%, 8/46). Maternal smoking occurred in 26% (12/46) of participants, while maternal alcohol consumption occurred in 13% (6/46) of participants. A higher proportion of births occurred in the spring season (34.8%, 16/46) and fewer in autumn (15.2%, 7/46). The majority of infants in this cohort exhibit a weight-for-age z-score (WAZ) at birth within the standard range, with a median WAZ of 0, indicating that there is no substantial occurrence of underweight or overweight status at birth. In terms of sociodemographic factors, the average household size was 5 individuals (IQR 3–7). Most participants came from households where the parents were unemployed (29/46, 63%) and most households reported incomes of less than R5000, reflective of poor socio-economic status. In addition, limited maternal education is observed with 71.7% (33/46) of mothers not completing secondary school. Furthermore,

there was high immunization coverage (100%, 46/46), including for PCVs (13-valent pneumococcal conjugate vaccine).

Table 3.1. Characteristics of children with high relative abundance of *O. hominis*

Participant Characteristics	**16S rRNA gene amplicon data**
	N = 46
	n (%)
Environmental Exposures	
Prenatal smoking (self-report):	
Yes	12 (26.1)
Prenatal alcohol consumption (self-report):	
Yes	6 (13.1)
Sociodemographic Variables	
Parent employed:	
Yes	17 (36.9)
Household density, median (IQR)	5 (3 – 7)
Income:	
<R1000/m	19 (41.3)
R1000–R5000/m	24 (52.2)
>R5000/m	3 (6.5)
Education:	
Primary	4 (8.7)
Some secondary	29 (63)
Completed	12 (26.1)
Some tertiary	1 (2.2)
Perinatal factors	
Birth Season:	
Summer	13 (28.3)
Autumn	7 (15.2)
Winter	10 (21.7)

Spring	16 (34.8)
WAZ at birth, median (IQR)	0 (-0.51 – 0.91)
Gestational age:	
Premature (<37 weeks gestation)	10 (21.74)
Sex:	
Male	19 (41.3)
Delivery mode:	
Vaginal	39 (84.8)
Infant Exposures	
HIV-exposure:	
HIV-exposed, uninfected	8 (17)
Vaccine coverage:	
Birth (BCG, OPV)	46 (100)
6 weeks (PCV, RV, HepB, DTP)	46 (100)
10 weeks (HepB, DTP)	46 (100)
14 weeks (PCV, RV, HepB, DTP)	46 (100)
9 months	46 (100)

BCG, Bacille Calmette-Guérin; DTP, Diphtheria, tetanus, and acellular pertussis; HepB, Hepatitis B; HIV, Human Immunodeficiency virus; IQR, Interquartile range; OPV, Oral polio vaccine; PCV, Pneumococcal conjugate vaccine; R, South African rand; RV, rotavirus vaccine; SD, Standard deviation; WAZ, weight-for-age z-score.

3.2 Relative abundance of *O. hominis* 16S rRNA in samples with age data available

Samples with age data available were grouped by relative abundance of *O. hominis* reads. Age data was available for 540 *O. hominis*-positive samples. As seen in Figure 3.1, the relative abundances of *O. hominis* tend to increase after six months of age, with the highest relative abundances observed at around two years of age (24 months). Approximately 6.3% of samples with age data available have greater than 1% relative read abundance. For these samples, *O. hominis* read abundance ranged from 1.01% to 7.85% relative abundance.

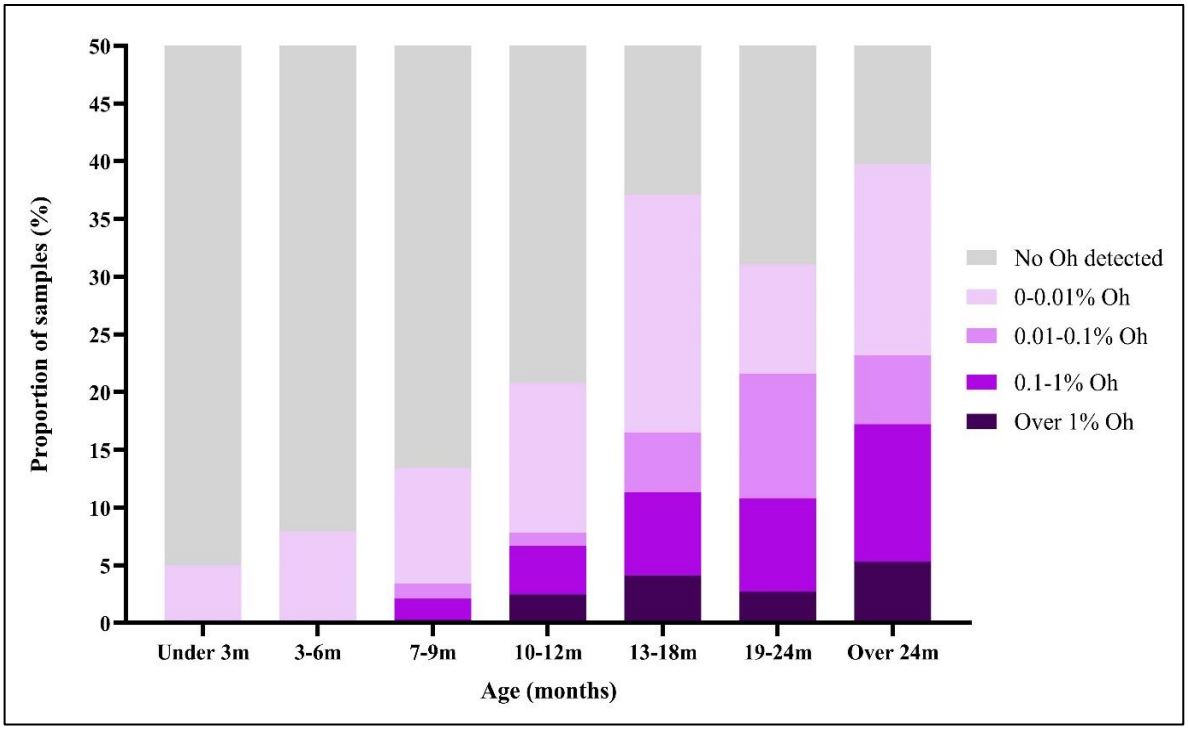

Figure 3.1. *Ornithobacterium hominis* 16S rRNA read abundance grouped by age. Oh, *O. hominis.*

Additionally, a nested case-control cohort, the longitudinal carriage of *O. hominis* in infants that developed LTRIs and age-matched controls was investigated as seen in Figure 3.2. The proportion of *O. hominis* abundance was largest in the period immediately preceding (T1) a respiratory illness event (T0) in cases (presented with LRTI symptoms); however, *O. hominis* relative abundance is more similar across timepoints in age-matched controls (no respiratory illness).

3.3 Isolating *O. hominis* from archived NP samples

Of the thirty-four 34 NP samples with *O. hominis* read abundance >1% that were selected for culture analysis, only 14 samples contained viable bacteria and could be cultivated under standard laboratory conditions. The sample collection data and 16S rRNA gene abundances of the 14 NP samples are shown in Table 3.2. *Ornithobacterium hominis* relative abundance was greater than 1% across all samples and ranged from 1.1% to as high as 7% relative *O. hominis* abundance. All samples were obtained during

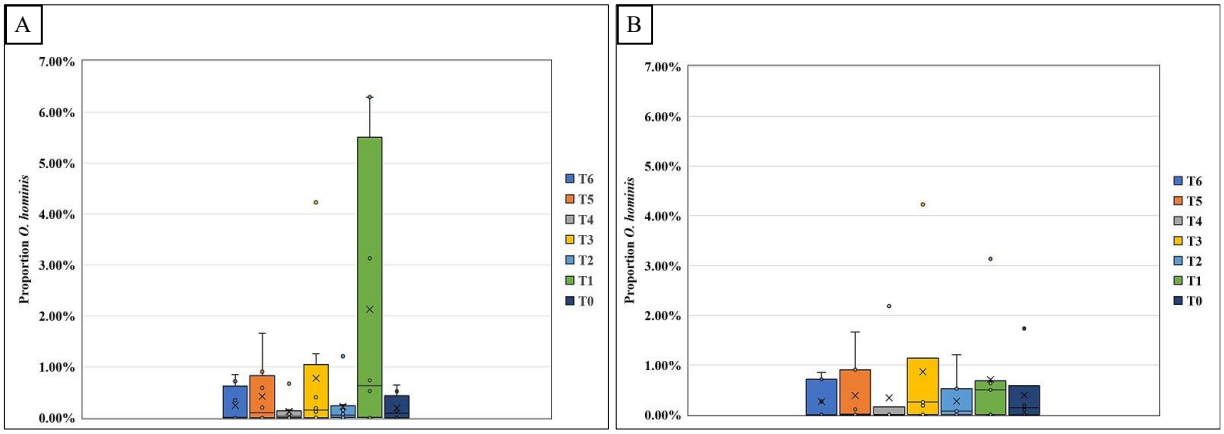

Figure 3.2. Boxplots summarising the proportional abundance of *O. hominis* in the nasopharynx of infants with lower respiratory tract infections (LRTI) and age-matched controls. A) LRTI cases (n = 7) and B) controls (n =7) at consecutive two-weekly timepoints prior to infection (T6–T1). T0, lower respiratory tract infection event.

routine sample collection, except for the sample from which SA-OH-C1 was isolated, which was collected during a respiratory illness episode. Each sample was obtained from a different individual. Ages of individuals ranged from 239 days to 767 days (~2 years of age) and the median age at sample collection was 350.5 days (IQR: 314 – 547.5 days). Limited metadata was available for samples DCHS-CDP1and DCHS-CDP2. Only year of sample collection was known for isolate SA-OH-C5. Antibiotic administration data at the time of sample collection was not available.

Table 3.2. *Ornithobacterium hominis* sample characteristics

Sample (recoded)	Proportion *O. hominis* (>1% abundance)	Age (days)	Year of collection	Routine/Illness	Recovered isolate (y, yes; n, no)	Assembly no.
DCHS-5SB1	0.014669	365	2013	Routine	n	
DCHS-F414	0.019905	545	2014	Routine	n	
DCHS-W3X9	0.019334	548	2014	Routine	n	
DCHS-L6W4	0.019751	783	2015	Routine	n	
DCHS-4YYY	0.013541	761	2016	Routine	y	SA-OH-C6
DCHS-4G3R	0.011137	336	2016	Routine	n	
DCHS-VQXL	0.013567	307	2013	Routine	n	

DCHS-49Y2	0.042626	239	2013	Routine	y	SA-OH-C2
DCHS-F19T	0.05339	336	2014	Routine	n	SA-OH-C3
DCHS-VK9D	0.018189	335	2015	Routine	n	
DCHS-EVC4	0.021848	254	2016	Routine	n	
DCHS-R0V1	0.062959	270	2014	Routine	y	SA-OH-C4
DCHS-43CM	0.070119748	546	2014	Routine	n	
DCHS-CDP1	NA	493	2014	Illness	n	SA-OH-C1
DCHS-CDP2	NA	NA	2015	NA	y	SA-OH-C5
DCHS-XKXK	0.016520894	767	2015	Routine	n	

NA, not available

Ornithobacterium hominis was successfully recovered from six samples that were viable for culturing and subsequent DNA extraction; however, isolates SA-OH-C1 and SA-OH-C3 could not be revived for subsequent phenotypic tests due to contamination. Isolation from primary culture was challenging due to substantial overgrowth by other taxa on CBA (Figure 3.3); however, colonies that resembled *O. hominis*

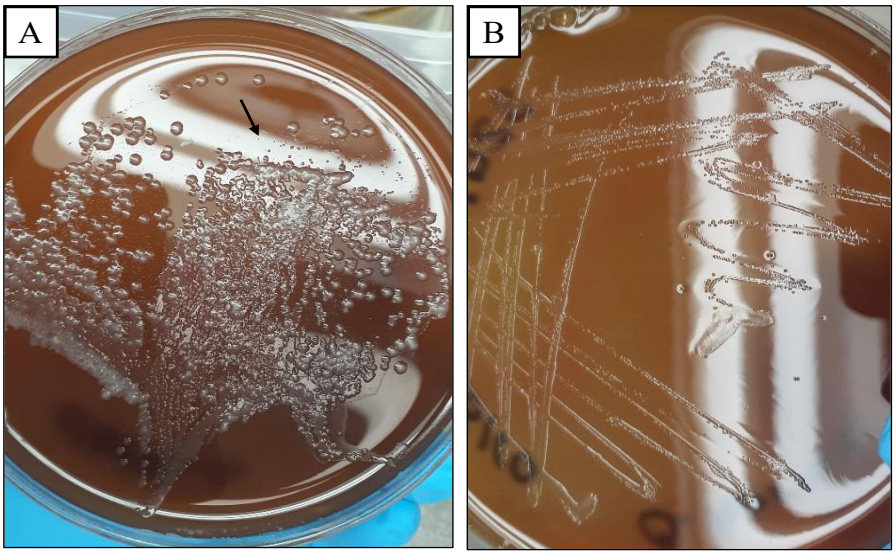

Figure 3.3. Primary isolation of *O. hominis* from archived NP swabs sample. A) Primary isolation of *O. hominis* isolate on CBA after 96 hours of incubation under microaerobic conditions. Black arrow indicates *O. hominis* colony selected for subculturing. B) Purified isolate after 72 hours of incubation on CBA.

as described by Lawrence *et al*. (2019), being pleomorphic, grey, and 1– 3 mm in diameter were sub-cultured on CBA until pure cultures were obtained[68]. Colonies typical of *O. hominis* morphology were greyish, glistening, convex, and circular. PCR targeting the TOXIN gene confirmed the presence of *O. hominis*, where an amplicon product of ~ 1.7 kb in size indicated a positive result, and there was no amplification in the negative control (Figure 3.4).

3.4 Morphological characterization

3.4.1 Optimal culture conditions for *O. hominis*

Ornithobacterium hominis isolates had not been recovered from South African samples before; thus, we evaluated the culturing conditions required to improve colony appearance and growth of the South African isolates. This is particularly important for nanopore sequencing which requires relatively large amounts of starting material per sample (> 400 ng of gDNA) for library preparation.

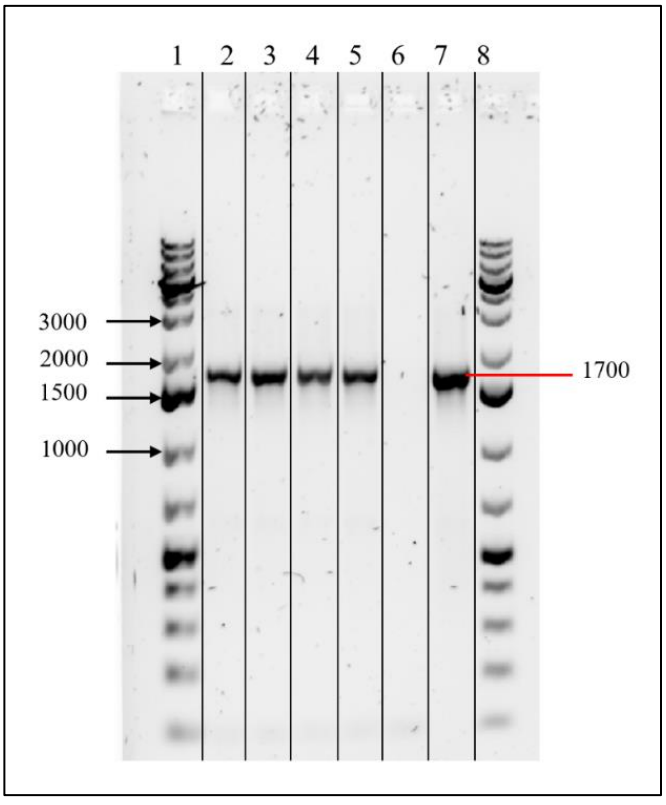

Figure 3.4. TOXIN PCR for *O. hominis* identification. Lane 1, GeneRuler 1 kb Plus DNA ladder. *O. hominis* isolates in lanes 2, 3, 4, 5. Lane 6, No template control. Lane 7, *O. hominis* positive DNA control (DNA extracted from type strain ATCC TSD-185).

Temperature

We assessed the impact of incubation temperature on the morphology of *O. hominis* colonies. Lawrence *et al*. (2019) described the colony morphologies of *O. hominis* isolates incubated at 35 °C; thus, we evaluated the colony morphologies at a higher temperature, 37 °C, and a lower temperature, 30 °C[68]. Incubation at 30 °C resulted in colonies exhibiting increased density and more robust growth (Figure 3.5). Specifically, denser growth was observed across the primary and secondary streaks. Conversely, incubation at 37 °C produced colonies with reduced density. The most notable difference in growth density between the temperature conditions was observed across the tertiary streaks, where growth was denser when incubated at 30 °C than when incubated at 37 °C, as indicated by the black ovals in Figures 3.5B and 3.5E. However, the colonies were similar in size between both temperature conditions as seen in Figure 3.5A and 3.5D. Notably, colonies incubated at 37 °C appeared drier and less convex, while those at incubated at 30 °C exhibited a more convex and smoother, shinier surface. Generally, more colonies were consistently observed on plates incubated at lower temperatures compared to those incubated at the higher

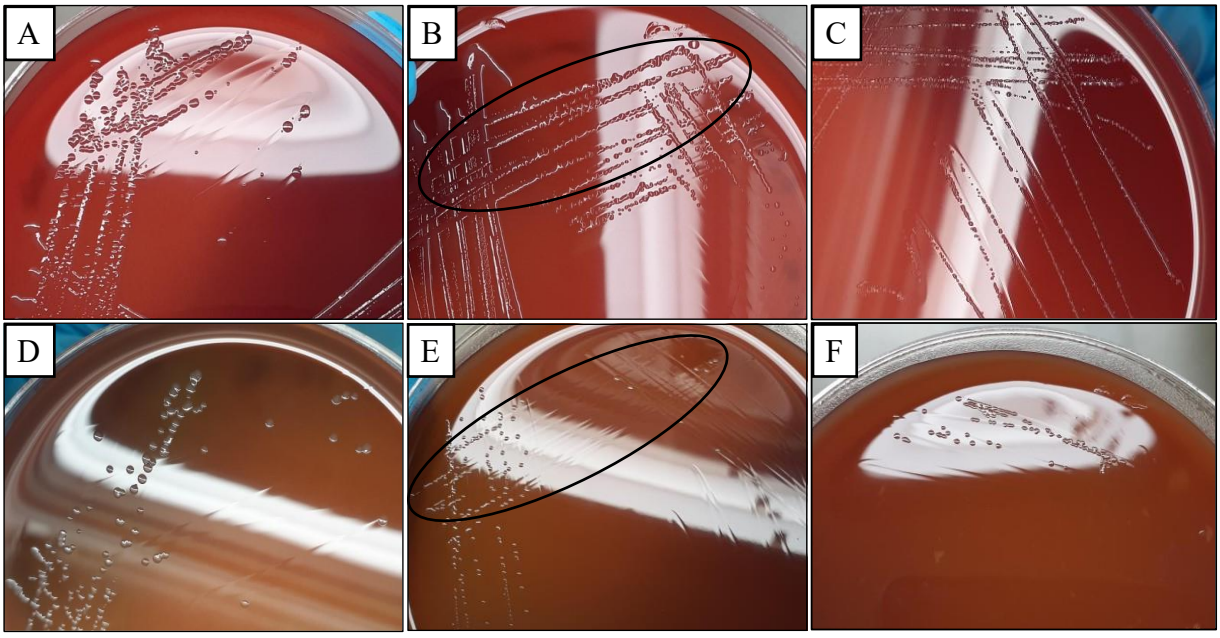

Figure 3.5. Evaluating impact of incubation temperature on O. *hominis* colony growth on Columbia blood agar (CBA). A–C) Colonies of *O. hominis* on CBA incubated at 30 °C. D–F) Colonies of *O. hominis* on CBA incubated at 37 °C. Black oval indicates tertiary streak on CBA.

temperature of 37 °C, as seen in Figure 3.5F. Thus, *O. hominis* grows more luxuriantly, and shows more consistent growth, at lower temperatures than higher temperatures.

Biphasic-media passage

Lower incubation temperatures improved colony growth on CBA; however, smaller punctiform colonies occasionally appeared, as shown in Figure 3.5C. This presents a challenge for subsequent DNA extraction, as high yields of DNA are necessary for ONT library preparation. Most DNA extraction protocols, which typically involve overnight liquid cultures, are unsuitable for *O. hominis* due to its slow growth and microaerobic nature. As a result, plate scrapes are the most effective method for obtaining sufficient cell material; however, the presence of punctiform colonies complicates this process, making it more difficult to obtain large amounts of material. Previous work from our lab suggested that passage through the liquid portion of biphasic media could enhance the growth of *O. hominis* and improve overall colony appearance. To address the issues of inconsistent growth and small colony size, we introduced a liquid-culture passage step. This method aimed to produce larger and more uniform colonies across various culture conditions.

Isolates cultured on CBA were inoculated into the 5 mm-deep liquid portion (Todd-Hewitts broth) of biphasic media. After 48 hours of static incubation, a loopful of the liquid culture was inoculated onto CBA. Concurrently, isolates from the same plates were sub-cultured directly onto the same CBA plate. Biphasic-media passage resulted in colonies that were larger and more uniform in size, while colonies from direct subculture were smaller, and often punctiform. As shown in Figure 3.6, colonies are smaller from direct sub-culturing (black circle) while colonies are larger, more mucoid, and uniform in shape after passage through biphasic media. Thus, biphasic media passage improved the overall appearance of *O. hominis* colonies versus direct subculture.

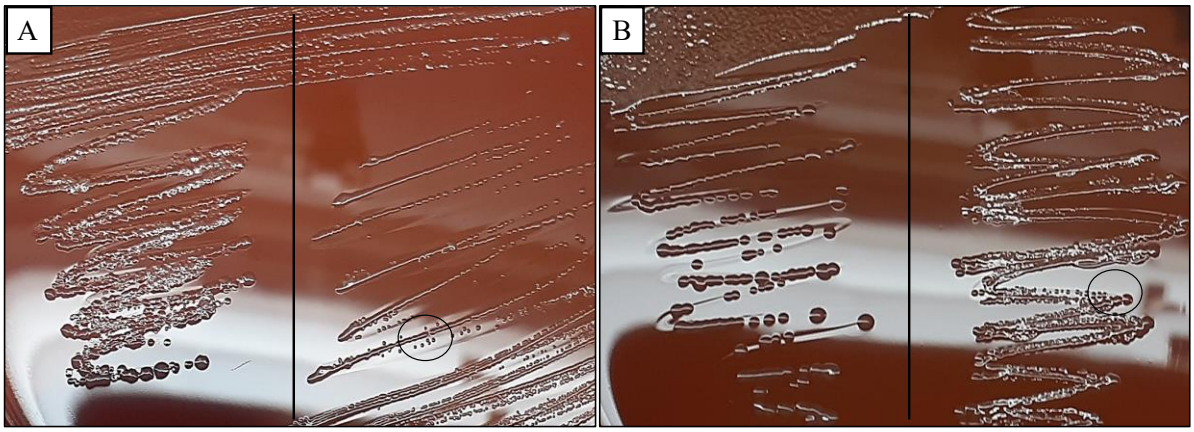

Figure 3.6. Biphasic media passage of *O. hominis* isolates on Columbia blood agar (CBA). A) *O. hominis* inoculated on CBA and incubated for 48 hours. Left, Passaged culture; Right, Direct sub-culture on CBA. Circle indicates punctiform colonies B) A second culture of another *O. hominis* strain inoculated on CBA and incubated for 72 hours Left, Passaged culture; Right, Direct sub-culture on CBA. Circle indicates punctiform colonies

3.4.2 Colony morphology

Although culturing attempts have been successful and isolate morphology has been previously elucidated, *O. hominis* has not been cultured from South African samples before; therefore, its microbiological characteristics should be determined. Characterization includes determining colony morphology, Gram-staining characteristics, antibiotic susceptibility testing and β-lactamase production. Consistent with findings by Lawrence *et al*. (2019), *O. hominis* colonies grow at different rates and can differ in colony morphology[68]. After 48 hours of incubation at 30 °C, there is some notable variation in growth rate. Smaller colonies (<1 mm diameter) can become apparent as demonstrated in Figure 3.7B, while other isolates appear to be slower growing, as distinct colony morphologies were not observed after 48 hours of incubation and streaks appear dry on the agar surface (Figure 3.7A).

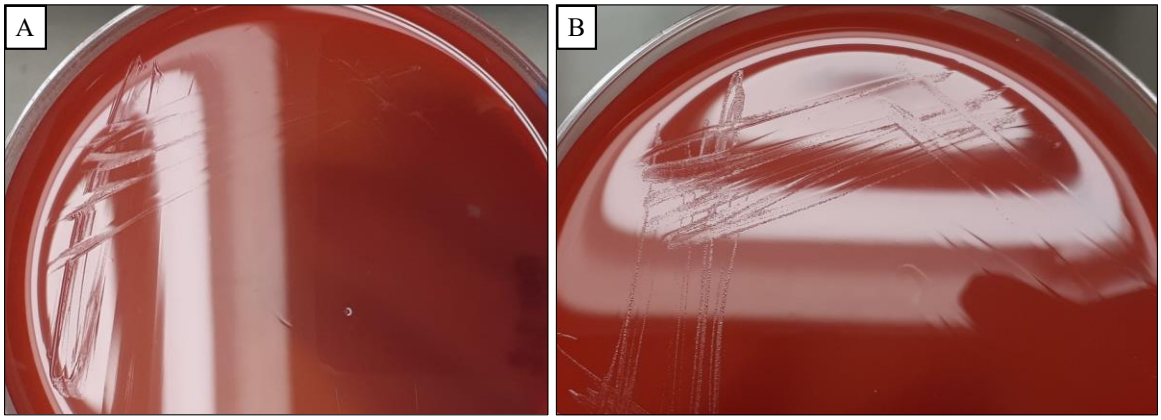

Figure 3.7. Growth of *O. hominis* on Columbia blood agar (CBA) at 30 °C after 48 hours. A) Growth of *O. hominis* isolate displaying dry streaks on CBA. B) Growth of *O. hominis* isolate displaying punctiform colonies on CBA.

After 96 hours of incubation at 30 °C, four colony morphologies may be predominant (Figure 3.8). Generally, colonies were greyish and translucent. Colonies were also glistening, rounded, and convex with smooth edges. Single-colony picks streaked onto CBA may produce uniformly sized and rounded colonies that are approximately 1 mm in diameter (Figure 3.8A). In addition, colonies can exhibit a more mucoid appearance, colonies are less distinct from each other and have less defined edges as seen in Figure 3.8B. Pure *O. hominis* cultures can display multiple colony morphologies. A single colony can give rise to a variety of colony sizes, ranging from punctiform (pinprick-sized, less than 1 mm in diameter) to larger colonies up to 2 mm in size, as shown in Figure 3.8C. Notably, punctiform colonies on mixed-morphology plates typically become apparent only after prolonged incubation, and their numbers tend to increase with extended incubation periods. Finally, Figure 3.8D shows uniformly punctiform colonies.

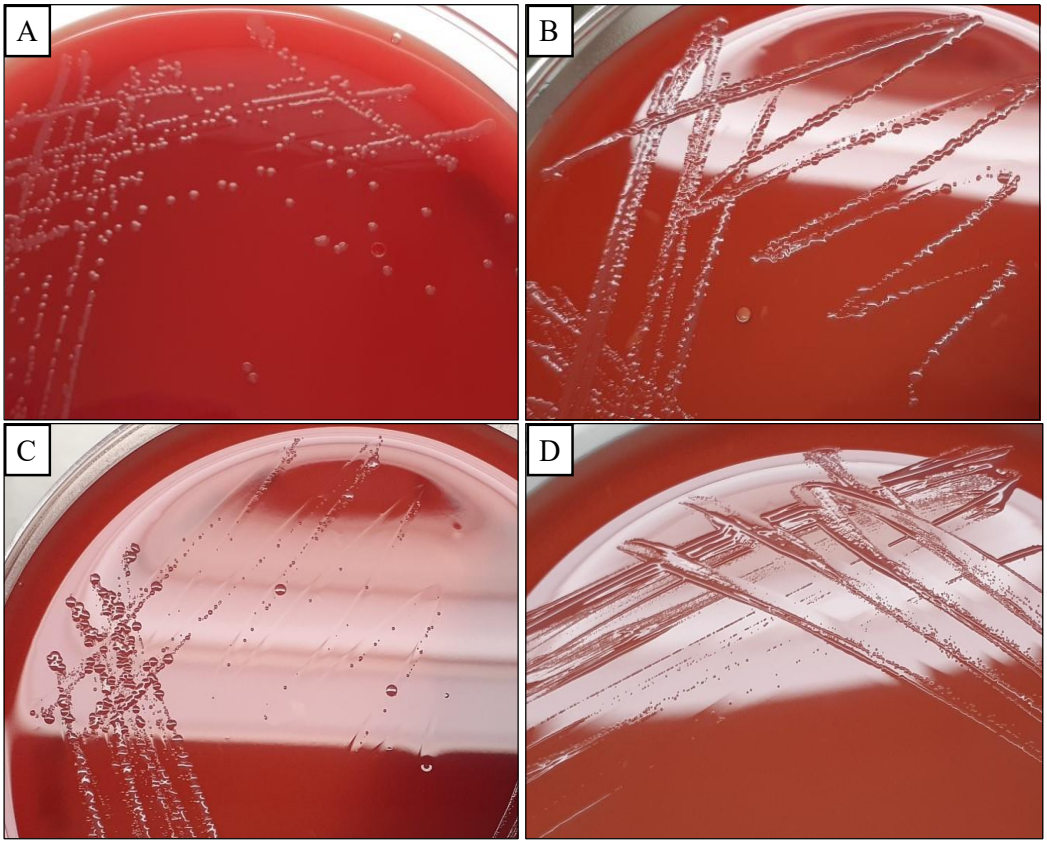

Figure 3.8. Main colony morphologies of South African *O. hominis isolates*. Isolates inoculated on Columbia blood agar and incubated at 30 °C for up to 72 hours. A) Uniformly sized colonies. B) Mucoid colony morphology. C) Mixed colony morphology. D) Punctiform *O. hominis* colonies.

After prolonged incubation (>120 hours), some larger colonies may collapse in the centre, forming a crater or 'bowl' shape with raised edges as seen in Figure 3.9A. Additionally, mucoid colonies generally increase in size with longer incubation durations and can grow to as large as 3–4 mm in diameter (Figure 3.9B). *Ornithobacterium hominis* colonies also exhibited partial haemolysis (also referred to as alpha haemolysis) on CBA plates after extended incubation (>96 hours) (Figure 3.10A). This haemolysis is most pronounced under the primary streak, as highlighted by the black-lined oval in Figure 3.10B but is not noticeable, or as pronounced, under individual colonies. Notably, haemolytic activity was not observed for smaller, punctiform colonies.

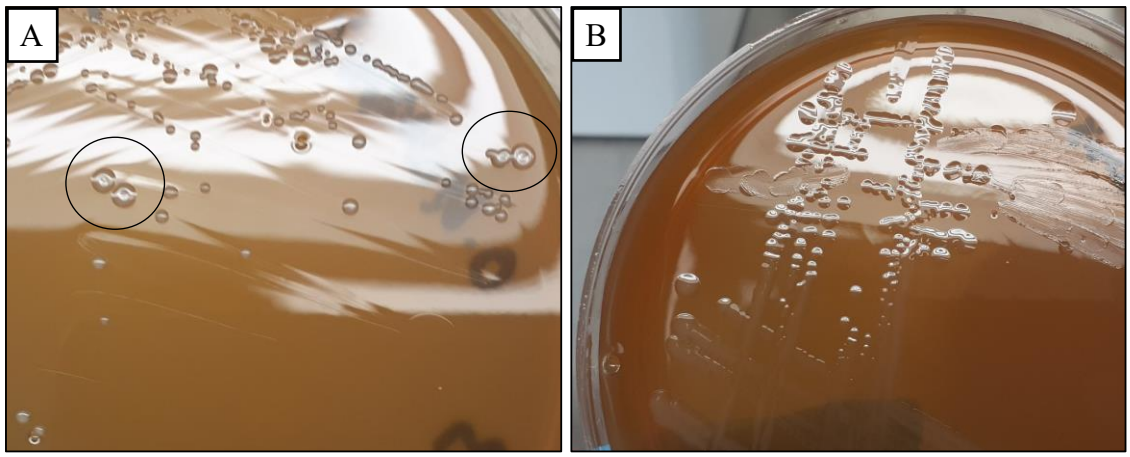

Figure 3.9. Colony morphologies of *O. hominis* incubated for longer durations (>120 hours) on Columbia blood agar. A) "Bowl"-shaped *O. hominis* colonies indicated by black circles. B) Large, mucoid *O. hominis* colonies.

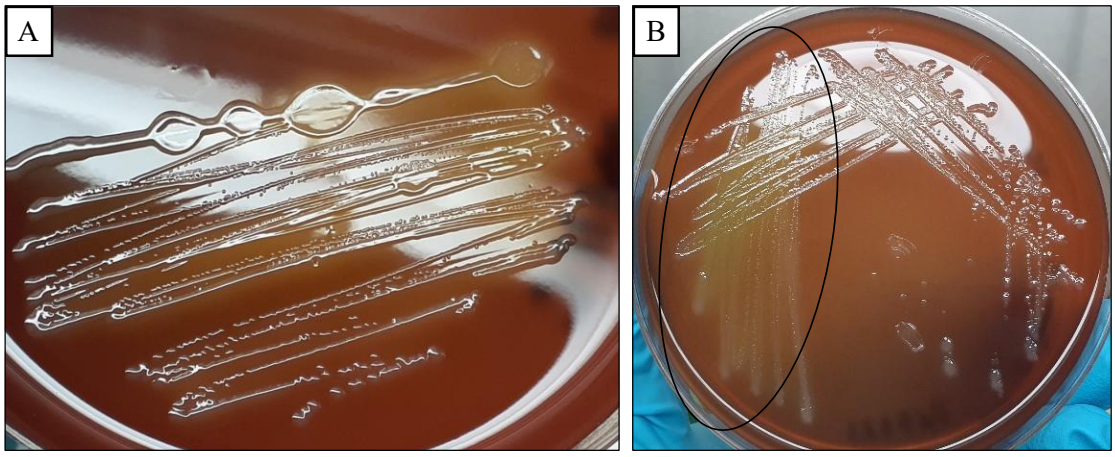

Figure 3.10. Partial haemolysis on Columbia blood agar (CBA) by *O. hominis*. A) *O. hominis* inoculated on CBA from liquid culture. B) *O. hominis* inoculated on CBA using 4-quadrant streaking method. Black oval indicates area of partial haemolysis.

3.4.3 Gram-stain characteristics

Although previous research has determined that *O. hominis* is a Gram-negative bacillus, research from collaborators suggested that *O. hominis* may exhibit relatively unique or distinguishing gram-stain characteristics. To confirm this with the South African isolates, gram-staining was performed using standard methodology as described by the American Society for Microbiology. A colony pick of a plate-cultured isolate produced short Gram-negative rods (Figure 3.11). These rods sometimes appear to be paired in an end-to-end shape. The bacilli can sometimes pair end-to-end at a slight angle to produce a "bent like" shape. The black circle shows this distinctive bent-shaped pair of bacilli. Likely, two bacilli

are joined at the ends to create this “U-shaped/boomerang” morphology. The blue circle indicates this “U-shaped” bacilli pair as well; however, the distal ends of the rods appear bulbous and rounded, giving the rods a dumbbell-like appearance.

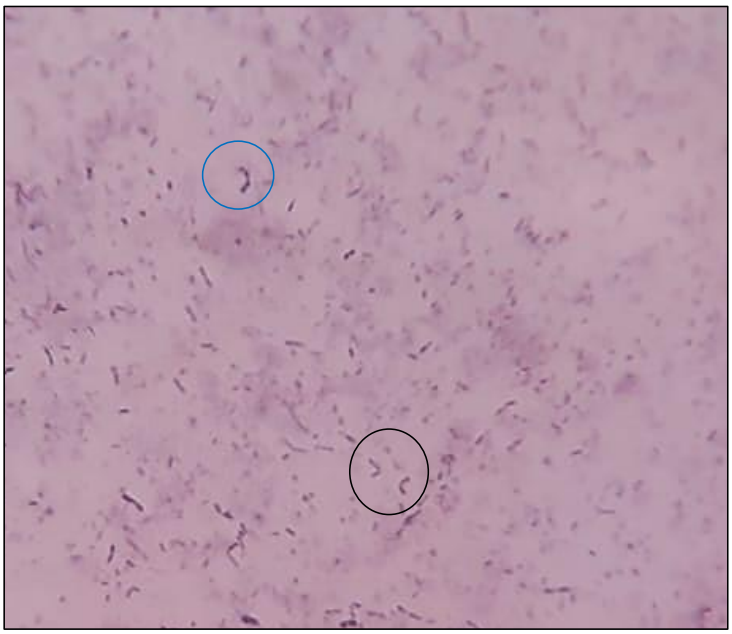

Figure 3.11. Image of gram-stained *O. hominis* isolate cultured on CBA (Magnification, 100×).

3.5 Antibiotic susceptibility testing and β-lactamase production

Antibiotic susceptibility to Polymyxin B, Vancomycin, Amoxicillin, and Ampicillin were evaluated. In conjunction, β-lactamase production was determined. Previous work conducted in our lab had indicated possible resistance to Polymyxins; thus, polymyxin B was added to the panel of antibiotics. *Ornithobacterium hominis* strains SA-OH-C1 and SA-OH-C3 were not subjected to antibiotic susceptibility testing due to issues with sample viability and contamination. Only zone diameters are reported for the isolates, while conclusive “susceptibility” or “resistance” categories have been excluded from Table 3.3. Vancomycin zones of inhibition were similar across the isolates, with zone size ranging from 13.3 mm to 16 mm indicating potential susceptibility to this antibiotic across the tested isolates (Table 3.3). Additionally, all isolates appeared to be susceptible to the β-lactam antibiotics, Amoxicillin

and Ampicillin; however, smaller zones of inhibition were noted for isolate SA-OH-C6 for these antibiotics. Only isolate SA-OH-C6 produced β-lactamases.

Table 3.3. Antibiotic susceptibility and β-lactamase production testing for *O. hominis* isolates

***O. hominis* strain**	**Antibiotic (µg)**	**Zone diameter ± SD (mm)**	**B-lactamase production (positive/negative)**
SA-OH-C2			Negative
	Polymyxin B (300)	8.33 ± 0.57	
	Vancomycin (5)	14 ± 1.73	
	Amoxicillin (25)	49.5 ± 3.53	
	Ampicillin (10)	45.5 ± 0.71	
SA-OH-C4			Negative
	Polymyxin B (300)	8.66 ± 0.57	
	Vancomycin (5)	16.3 ± 2.08	
	Amoxicillin (25)	43 ± 2.82	
	Ampicillin (10)	53.5 ± 0.71	
SA-OH-C5			Negative
	Polymyxin B (300)	7.67 ± 0.57	
	Vancomycin (5)	13.3 ± 2.08	
	Amoxicillin (25)	44 ± 2.64	
	Ampicillin (10)	46 ± 1.73	
SA-OH-C6			Positive

	Polymyxin B (300)	7.33 ± 1.15
	Vancomycin (5)	14.6 ± 2.08
	Amoxicillin (25)	30 ± 2.64
	Ampicillin (10)	34.7 ± 2.5

SD, Standard deviation

3.6 DNA extractions and quality assessment

DNA extraction is a critical aspect of genome sequencing from cultured samples. The purity and integrity of the extracted DNA significantly affect downstream applications. Precipitation based extraction methods were opted for instead of column-based methods to ensure high molecular weight (HMW) DNA with sufficient quality was obtained for downstream whole genome sequencing. Initially, PCI-extraction based methods were used for DNA extraction. However, PCI-based methods are time-consuming and involve hazardous organic solvents. Kit-based extraction methods are becoming more popular due to their ease-of-use and reproducibility; therefore, a HMW kit-based extraction method was used as an alternative. The Nanodrop spectrophotometer was used to assess for contamination using A260/230 and A260/280 ratios (Figure 3.12).

Figure 3.12A shows that the A260/A230 ratios for DNA samples extracted using the kit-based method is significantly higher than that for the PCI-extracted method. This higher ratio suggests lower levels of contamination by phenols, proteins, and salts in the kit-extracted DNA compared to the PCI-extracted DNA[89]. However, the A260/A280 ratios are not significantly different between the two extraction methods as seen in Figure 3.12B. Additionally, samples from both extraction methods exhibit A260/A280 ratios greater than 1.88, indicating RNA contamination. Although the difference is not significant, this ratio appears to be more elevated in the PCI-extracted samples compared to the kit-extracted samples.

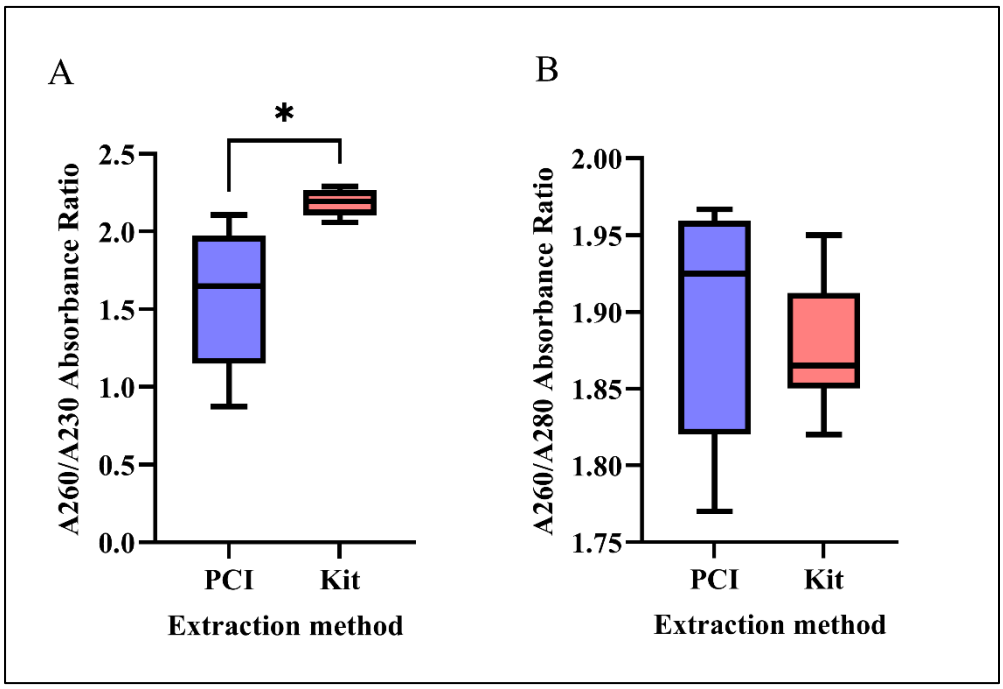

Figure 3.12. Box plots showing A260/230 ratios and A260/280 ratios for DNA samples extracted using PCI-based method (blue) or Kit-based method (orange). DNA is extracted from isolates 311747 and 534472 to allow for comparison of methods. Ratios determined using Nanodrop spectrophotometer. *P > 0.05. PCI, Phenol chloroform isoamyl alcohol.

Overall, these results suggest that RNA contamination is present in all the DNA samples extracted, regardless of the extraction method utilized. The qubit fluorometer was used to quantify DNA in samples extracted using PCI and Kit-based methods from isolate SA-OH-C4. Although the DNA yields cannot be compared between the extraction methods, the differences between the Nanodrop and Qubit measurements were compared as seen in Figure 3.13. The nanodrop measurements for PCI-extracted DNA are inflated to a greater extent than that of the kit-extracted DNA. The DNA concentrations for PCI-

extracted DNA, as measured by the Nanodrop, are significantly higher than those measured by the Qubit across all three samples. In contrast, the DNA concentrations for kit-extracted DNA do not differ significantly between Nanodrop and Qubit measurements, although Nanodrop measurements are slightly elevated compared to those obtained with the Qubit.

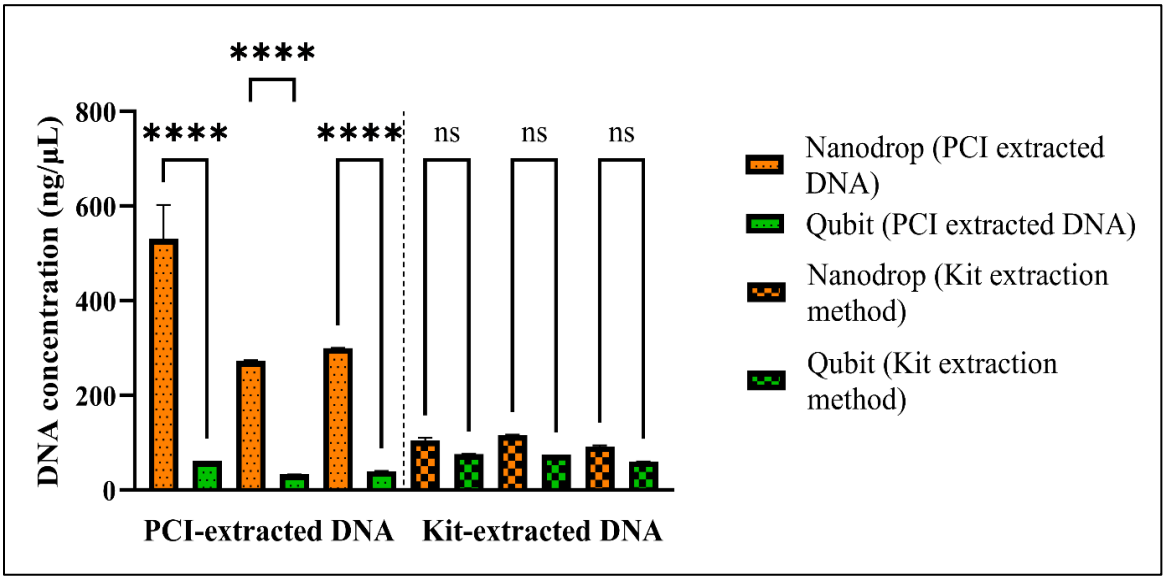

Figure 3.13. Comparison of Nanodrop and Qubit measurements from PCI and Kit-based DNA extraction methods. DNA extractions performed for three biological replicates of the same *O. hominis* strain. Nanodrop and Qubit measurements performed in triplicate. PCI, Phenol chloroform isoamyl alcohol; ns, not significant; ****P<0,000.1.

In addition to checking for contamination and quantifying DNA, it is recommended by ONT to visually assess the quality of the extracted DNA to ensure that the majority of the extracted DNA is intact and has undergone minimal shearing. To visually assess for DNA quality, the extracted DNA was electrophoresed on a 1.6% agarose gel stained with SYBER safe for DNA visualization (Figure 3.14). As illustrated in Figure 3.14, DNA extracted using the PCI method is highly fragmented as evidenced by pronounced smearing in the agarose gel. Furthermore, the extent of fragmentation varies significantly between individual extractions using the PCI-based method. In contrast, DNA extracted using the kit-based method appears to be more intact and of higher molecular weight than that using the PCI-based method, as indicated by minimal smearing on the gel and singular, pronounced DNA band.

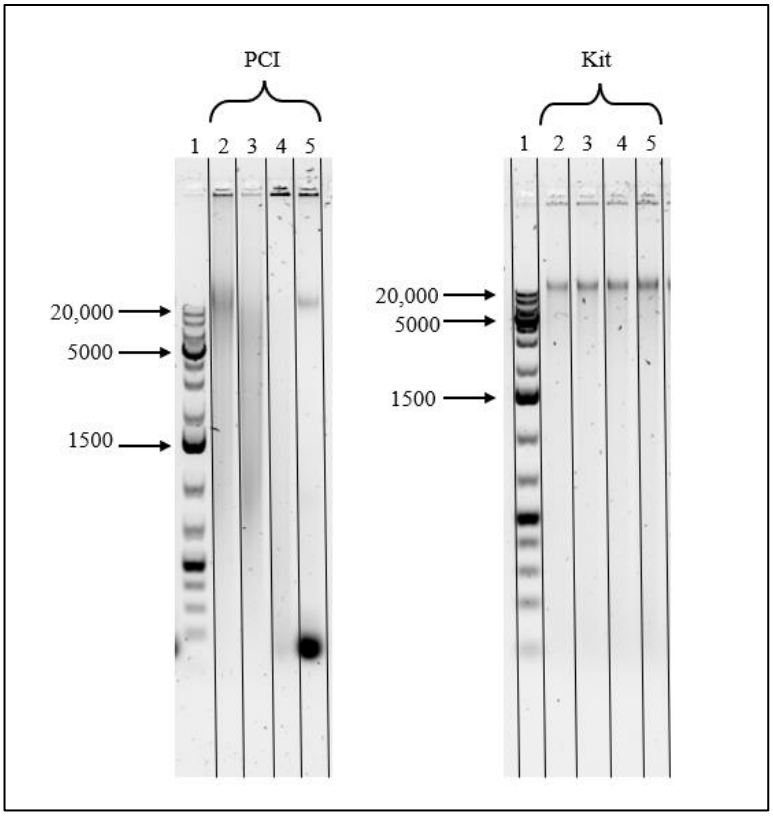

Figure 3.14. Assessing DNA integrity on agarose gel. DNA electrophoresed on 1.6 % agarose gel stained with SYBER safe dye. DNA loaded (50 ng) into each well. Lane 1, GeneRuler 1 kb plus DNA ladder. Lanes 2–5, DNA extracted using either PCI-extraction or Kit-extraction method. PCI, Phenol chloroform isoamyl alcohol.

3.7 Genomes

3.7.1 Genome assemblies

Following library preparation, sequencing was performed on the MinION flowcell. Post-hoc base calling was conducted, and genomic assemblies were generated with a 5kb read length cutoff for isolates SA-OH-C4, SA-OH-C6, SA-OH-C2, and SA-OH-C5, while a 1kb cutoff was used for isolates SA-OH-C1 and SA-OH-C3. The genome size was set to 2 Mb for all assemblies. As seen in Table 3.4, showing initial results (Assembly version 1) generated by the Flye assembler tool, there is notable variation in genome coverage, number of fragments (contigs), as well as fragment length and N50.

As seen for isolates SA-OH-C2 and SA-OH-C5, single contig assemblies were generated, while reads from SA-OH-C6 generated an unusually large genome (~3.5 kb) using a 5 kb read length cut-off. Conversely, SA-OH-C4 assembled into 29 contigs with a read cut-off of 5 kb (Table 3.4). Next, the

Table 3.4. *Ornithobacterium hominis* genome assemblies

Assembly Version	OH Sample	Min bases/read (Kb)	Approx. total length (Mb)	No. of Fragments	Fragments N50	Largest fragment (bp)	Mean coverage (×)
1	SA-OH-C4	5	2.08	29	123064	368491	10
1	SA-OH-C6	5	3.5	4	2050220	2050220	20
1	SA-OH-C2	5	2.05	1	2048047	2048047	23
1	SA-OH-C5	5	2.05	1	2056361	2056361	17
1	SA-OH-C1	1	2.08	5	1089709	1089709	41
1	SA-OH-C3	1	2.01	1	2007397	2007397	51
2	SA-OH-C4	1	1.99	8	478080	714412	50
2	SA-OH-C1	NA	2.05	3	2057767	2057767	55
2	SA-OH-C6	3	2.05	1	2057716	2057716	46

number of contigs in assemblies SA-OH-C4, SA-OH-C6, and SA-OH-C1 was reduced. In general, reducing the number of contigs involves improving coverage, which is achieved by increasing the pool of reads used for assembly generation. Thus, the trimming cutoff for these samples was lowered accordingly. SA-OH-C1 assembled into 5 contigs with a 1 kb read-length cut-off, thus, untrimmed reads were used to generate a more complete assembly of three contigs. This assembly exhibited an increased coverage of 55 ×, and the contigs could be manually curated and joined to generate a single, contiguous assembly.

Investigation of SA-OH-C6 assembly version 1 reveals a large assembly of approximately 3.5 Mb in size. A closer investigation of the individual contigs revealed 4 contigs, where contigs 5 and 2 appear to be repeats as seen in Table 3.5. Blast analysis of these contigs pairwise with *O. hominis* reference genome reveals that these are not likely to be contaminants and are likely sequencing artefacts. Contig 1 appears to be the complete *O. hominis* genome for this isolate based on the size (2 Mb) and circular nature. The characteristics for each contig are shown in Table 3.5.

Table 3.5. Contig characteristics of assembly version 1 for *O. hominis* isolate SA-OH-C6

Contig	Length (bp)	Coverage (×)	Circular	Repeat
contig_1	2050220	24	Y	N
contig_6	1416439	15	Y	N
contig_5	35636	29	N	Y
contig_2	7074	12	N	Y

Contig 6 was unusual for this assembly; the large size, circular nature, lower coverage (15 ×), and non-repeat nature indicated a contaminant, possibly from another bacterial species. Blast analysis of this contig revealed multiple hits within various *Helcococcus* species; thus, this contig is likely a whole or almost complete genome for a member of the *Helcococcus* species. To improve this assembly, the raw reads of SA-OH-C6 were mapped against the entire contig_1, taking into assumption that this is a complete *O. hominis* genome. Mapped reads were extracted and re-assembled with a 3kb cutoff to generate the SA-OH-C6-assembly version 2, a complete genome, as seen in Table 3.4.

3.7.2 Putative *O. hominis* plasmid

As seen for SA-OH-C4 generated using a minimum read length of 5 kb, the coverage was too low to generate a contiguous assembly (Table 3.4). Therefore, the read length cut-off was reduced to 1kb, which increased coverage to 50× and reduced the assembly contigs to eight contigs as seen for the assembly

version 2 in Table 3.4. The characteristics of these eight contigs are investigated in Table 3.6. Contigs were of similar coverage, linear, and not repeat sequences; however, compared to the other contigs, contig 11 exhibited unusually high coverage (647×). This contig was also short (4,5 kb), appeared to circularize, and was a repeat sequence. These are characteristics of a putative plasmid.

Table 3.6. Contig characteristics of assembly version 2 for *O. hominis* isolate SA-OH-C4

Contig	Length (bp)	Coverage (×)	Circular	Repeat
contig_5	714412	51	N	N
contig_8	478080	46	N	N
contig_9	231737	49	N	N
contig_2	211563	50	N	N
contig_6	170768	51	N	N
contig_7	126429	50	N	N
contig_3	59083	51	N	N
contig_11	4512	647	Y	Y

Contig 11 was subjected to Bakta annotation where two virulence associated protein D (*vapD*) encoding genes were identified as seen in Figure 3.15. Other larger open reading frames (>50 bp) were subjected to BLASTn analysis and functional identification using InterproScan for the putative identification of genes. In addition to the two copies of a *vapD* gene, *repB*, *mobV*, and transcriptional regulator genes were identified on the putative plasmid. The *vapD* genes encode for proteins with endoribonuclease activity and bind to RNA. The *repB* gene encodes for a Replication Initiation Protein B that has DNA-directed DNA polymerase activity. MOB relaxase genes are involved in DNA recombination and DNA binding. The transcription regulator gene function could not be determined based on gene-sequence data.

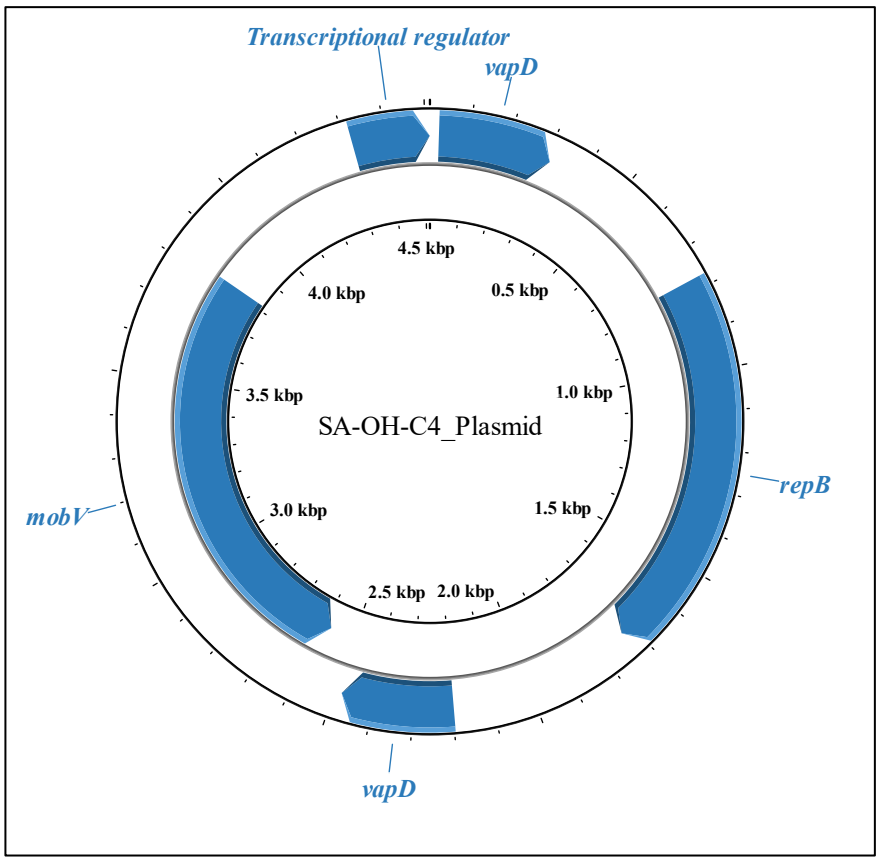

Figure 3.15. Annotated features of plasmid putatively identified from *O. hominis* isolate SA-OH-C4. Large open reading frames (blue arrows) annotated using Bakta and Interpro. Plot generated using Proksee.

Protein sequences of the plasmid genes were subjected to BLASTp analysis against the non-redundant protein sequences database. The analysis revealed significant alignments with virulence factor genes in public *O. hominis* genomes as seen in Table 3.7. Specifically, virulence factor genes were identified in contigs from Maela MAGs 22803 and 22767 from Thailand, as well as the Australian type strain TSD-185. Further alignments to virulence genes were observed in related species, including *Capnocytophaga* species (Order: *Flavobacteriales*, Family: *Flavobacteriaceae*), *Riemerella* species (Family: *Weeksellaceae*), and *O. rhinotracheale*. Additionally, BLASTp pairwise analysis showed a significant alignment to a chromosomal *vapD* gene in isolate SA-OH-C6 (98.92% identity).

Further BLASTp analysis of the other plasmid genes revealed that the *repB* gene was similar to replication initiation proteins in species from the *Bacteroides* genus (36–40% sequence identity). The *mobV* gene exhibited sequence similarity to *mobV* family relaxases in members of the Bacteroidota

phylum, including *Riemerella* species and *Elizabethkingia anophelis* (Family: Weeksellaceae) with 60.41–61% identity, and to *O. rhinotracheale* with 64% identity. Finally, the undefined transcriptional regulator was similar to transcriptional regulators in *O. rhinotracheale* (59–67% identity) and in the *O. hominis* MAG 22767 from the Maela cohort in Thailand (58.82% identity).

Table 3.7. BLASTp alignments for the *vapD* gene from SA-OH-C4 plasmid

Query	Alignments	% ID	% Coverage	E-value	Description
SA-OH-C4 Plasmid *vapD*	Virulence factor [*Ornithobacterium hominis*]	100%	100%	3.00E-72	Chromosomal *vapD* gene in Thai strain 22767
	Virulence factor [*Ornithobacterium hominis*]	100%	100%	3.00E-72	Chromosomal *vapD* gene in Thai strain 22803
	Virulence factor [*Ornithobacterium hominis*]	97.85%	98%	5.00E-62	Chromosomal *vapD* gene in Australian type strain

3.7.3 Hypothetical gene counts

After the genomes were assembled, the accuracy of the assemblies was further scrutinized using the hypothetical gene counts. Hypothetical gene counts versus total coding counts can assist with identifying increased frameshift errors, which inflate total gene counts and may interfere with the prediction and annotation of other genes. Total coding sequences and hypothetical gene counts for each genome are compared to the Australian type-strain genome as seen in Table 3.8. Compared to the Reference genome, the South African *O. hominis* genomes showed higher coding sequence counts and hypothetical gene counts, indicating an increase in frameshift errors for these genomes.

Table 3.8. *Ornithobacterium hominis* genome coding sequence counts and hypothetical gene counts

	Genome	Total CDS	Hypothetical genes
Reference genome	Aus TSD-185-ref	1873	167
South African Genomes	SA-OH-C5	1987	*238
	SA-OH-C2	2123	*364
	SA-OH-C6	2030	*248
	SA-OH-C1	1934	186
	SA-OH-C3	1840	169
	SA-OH-C4	1947	*246

CDS, Coding sequences

3.7.4. Assemblies generated using an alternative pipeline

Ideally, frameshifts would be corrected using higher accuracy reads. Unfortunately, Illumina® sequencing data was not available for these isolates thus a different approach was required to improve the assemblies. A different assembly pipeline was used to re-assemble the South African genomes, in an effort to produce assemblies with fewer errors. Pipeline 2 is the assembleBAC-ONT pipeline that was originally written by Andries van Tonder and is managed by Nextflow for ease of use. This pipeline also uses a different read trimming tool, Filtlong. As seen in Table 3.9, pipeline 2 generated assemblies of approximately 2 Mb, with variable coverage, and a read-length cutoff of 5 kb; however, the total coding sequences and hypothetical gene counts for the assemblies were greater than that for the assemblies generated using the first pipeline (Table 3.8), with approximately half of the total coding sequence counts annotated as hypothetical genes. Therefore, this alternative pipeline did not produce better quality assemblies than Pipeline 1.

Table 3.9. *Ornithobacterium hominis* assemblies generated using Pipeline 2

Sample	Min read length	Total length (Mb)	Fragments	Fragments N50	Largest fragment (bp)	Mean coverage (×)	Total CDS	Hypothetical genes
SA-OH-C2	5	2.05	1	2050116	2050116	24	2138	1142
SA-OH-C4	5	2.05	26	202486	329764	10	2511	1643
SA-OH-C5	5	2.06	1	2057407	2057407	18	2036	987
SA-OH-C6	5	2.05	3	2046909	2046909	25	2202	1165

3.7.5 Comparing pipeline usage

While these pipelines used similar tools to assemble and annotate, other factors are important to consider when deciding on using a different pipeline. Ease of use, accessibility, computational capacity and costs are important factors to consider when choosing pipelines. Table 3.10 compares select features of these in Pipeline 1, which is a manual workflow, and Pipeline 2. Both pipelines are designed for ONT -generated data and long reads. Additionally, both pipelines use the same tools: Flye to assemble, Medaka to polish assemblies and Bakta to annotate, except for the trimming tools. However, there are also key differences in their efficiency and ease-of use. The various tools, including Bakta annotation via a downloaded database, are integrated and comprehensive documentation is available via GitHub pages for Pipeline 2. In contrast, each tool for the manual pipeline must be installed, integrated and managed individually for Pipeline 1. Pipeline 2 benefits from strong community support, while support varies by tool for Pipeline 1. Importantly, both pipelines are open-source and use free tools.

Table 3.10. Comparison of pipelines used to generate assemblies

Category	Automated Pipeline (Nextflow)	Manual Pipeline
Sequencing Technology and Compatibility	Optimized for Oxford Nanopore and long reads	Optimized for Oxford Nanopore and long reads
Read QC	Integrated QC	Requires manual setup of QC
Speed	Optimized, parallel execution	Variable, depends on manual optimization
Memory Usage	Managed by Nextflow, efficient allocation	Manual management required, risk of inefficiency
Storage Requirements	Streamlined, handles intermediate files efficiently	Manual tracking and management of storage
Installation type	Can be installed and operated locally	Can be installed and operated locally
No. of installations	Single installation but requires installation of Nextflow	Multiple installations that are managed manually
Configuration	Configurable via Nextflow scripts	Each tool requires manual configuration
User Interface	Entirely via the command-line	Combined use of command-line, manual scripts, and GUI tools
Read Trimming Tool	Filtlong. Reads are filtered based on read length and overall read quality	Prowler. Reads are filtered using "window"-based approach
Assembly Accuracy	High, uses Flye and Medaka	High, uses Flye and Medaka
Key Metrics (N50)	Consistent, relies on Flye and integrated tools	Consistent, relies on Flye and integrated tools
Error Correction	Integrated polishing steps with Medaka	Manual integration of Medaka
Annotation Tool and Database accessibility.	Local Bakta, downloaded database, faster access	Web-based Bakta, manual upload and retrieval of files, potentially slower
Assembly Validation Tools and Downstream Analysis	Integrated, e.g., QUAST, CheckM2	Requires manual addition and setup
Documentation	Extensive and community-supported	Varies by tool. Most tools used are supported by documentation, but some may lack comprehensive documentation
Community Support	Active community, regular updates	Varies, tool-specific
Licensing	Open-source, free	Open-source, free
Learning Curve	Requires learning Nextflow DSL	Requires knowledge of each tool

3.8 *Ornithobacterium hominis* genome structure

After assemblies were annotated, the genomes were rotated to the origin of replication (oriC) using the position of the *mnmG* gene and GC skew, and landmarks were highlighted as illustrated in Figure 3.16.

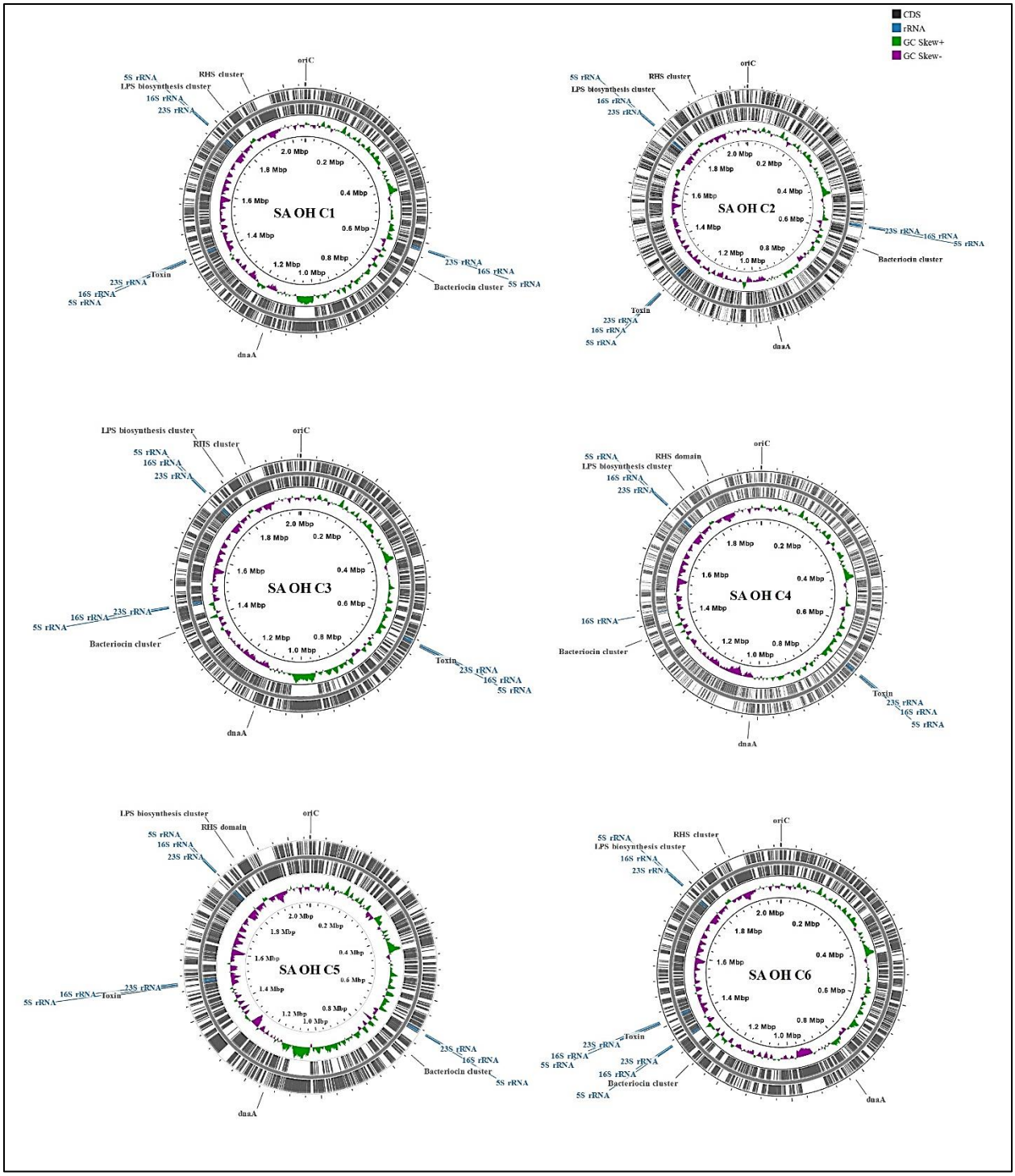

Figure 3.16. Genome plots of South African *O. hominis* strains SA-OH-C1 to SA-OH-C6. Plots generated using Proksee.

On average, the genomes are about 2.05 Mb in size, and the GC content of the genomes ranged from 35%–36%. All genomes have three rRNA operons except for SA-OH-C4 which is missing the 5S and 23S rRNA genes for one of its rRNA operons. All genomes encode for the toxin (annotated as the *toxA*) gene, an RHS cluster, LPS biosynthesis cluster, and a bacteriocin cluster. These features are highlighted for all genomes. The position of the *dnaA* gene is noteworthy, as this gene is position near the terminus of the *O. hominis* genomes across the isolates. Next, the overall structure of the South African *O. hominis* genomes was investigated and compared to the type-strain genome (Figure 3.17). Single, circular contig assemblies allow for insights into genomic structure and position of genes of interest. Whole genome alignment results revealed large symmetrical, inversions around the termini of the genomes, and these

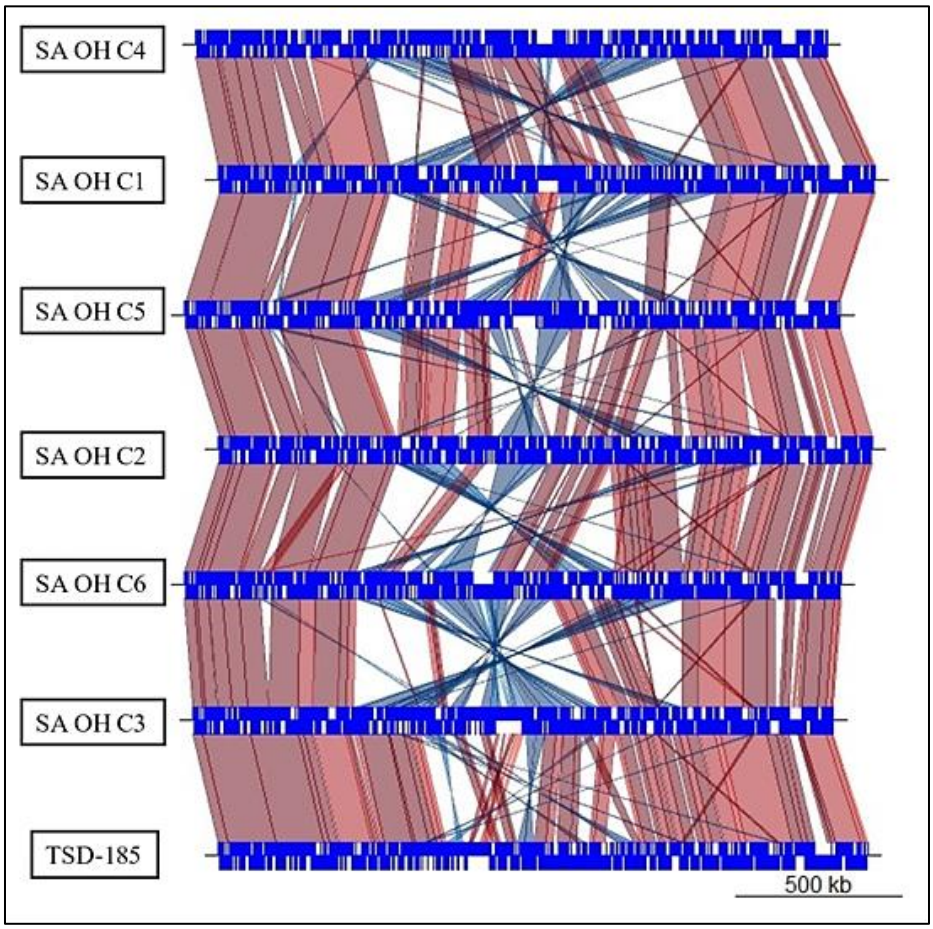

Figure 3.17. Whole genome alignment of South African *O. hominis* SA-OH-C1 to SA-OH-C6 and type strain (TSD-185) genomes. Whole genome alignments were performed using BLASTn and visualised using ggplot2 in R. Red alignments indicate regions of identity in the same orientation while blue alignments indicate regions of identity with the opposite orientation.

inversions appear to be consistent among the South African isolates compared to the type strain genome (TSD-185) as seen in Figure 3.17.

3.9 Genomic features of interest

3.9.1 Screening for AMR genes

AMR finder and CARD resistance gene identifier tools were used to screen for and identify AMR genes in the genomes. Table 3.11 details the findings of the AMR search. CARD further identified the same two genes associated with vancomycin resistance, *vanT* and *vanY*, in all South African strains. Two strains, SA-OH-C1 and SA-OH-C6, were found to have a gene encoding for a β-lactamase, *cfxA*, that might confer broad activity against β-lactam antibiotics.

Table 3.11. Antimicrobial resistance genes identified in *O. hominis* genomes

***O. hominis* strains**	**Gene**	**AMR gene family**	**Function**	**Antibiotic**
SA-OH-C1 SA-OH-C2 SA-OH-C3 SA-OH-C4 SA-OH-C5	*vanT*	Glycopeptide resistance gene cluster	Antibiotic target alteration	Vancomycin
SA-OH-C1 SA-OH-C2 SA-OH-C3 SA-OH-C4 SA-OH-C5	*vanY*	Glycopeptide resistance gene cluster	Antibiotic target alteration	Vancomycin
SA-OH-C1 SA-OH-C2	*cfxA*	CfxA family broad-spectrum class a β-lactamase	Antibiotic inactivation	B-lactam targets (broad)

3.9.2 Virulence genes and mechanisms

Virulence factors are properties or, more specifically, gene products that enable a bacterial species to colonize the host, potentially enhance its disease-causing potential, and evade host immune responses. Although the *O. hominis* is not believed to be pathogenic itself, its capacity as a persistent colonizer implies that it produces factors that may aid in its ability to colonize for protracted periods of time[14]. Virulence factors include toxins, cell-surface proteins for attachment (adhesins), and cell-surface molecules, such as the LPS, that aid in protecting the bacterium.

The PMT-like toxin (*toxA*)

As seen in Figure 3.16, the toxin or *toxA* gene is present in all South African *O. hominis* genomes and it exhibits high-sequence conservation across the South African genomes.

Gliding motility-associated genes (*gld*) and *sprA*

All South African genomes housed a complement of *gld* genes (*gldB, gldC, gldD, gldG, gldH, gldJ, gldK, gldL, gldM, gldN*) and the *sprA* encoding a cell-surface protein is present in all isolates. These genes form the T9SS, a protein export pathway linked to adhesin export and is considered a virulence factor. The *gld* genes, *gldB, gldH,* and *gldJ* are a gliding-motility lipoproteins, while *gldG* is a gliding motility-associated substrate-binding component of an ABC transporter.

Adhesin genes: *Fibrobacter succinogenes* major paralogous domain (FSD) containing proteins

In the South African *O. hominis* genomes, we identified genes annotated as FSD-containing proteins. These FSD genes are present in varying numbers across the genomes, typically ranging from 15 to 20 copies. A phase variation repeat sequence that is positioned immediately upstream of the FSD genes, as well as genes encoding for Hep Hag and Fimbrillin adhesins, was used to identify unannotated FSD genes.

The type strain of *O. hominis* contains approximately 20 FSD genes, whereas *O. rhinotracheale* has six FSD genes according to the UniProt database. Of note, FSD genes are predominantly found in species of the Fibrobacterota and Bacteroidota phyla. Further comparison with other related species reveals

variability in FSD gene numbers. Members of the genus *Chryseobacterium* (Family: Weeksellaceae) have higher counts of FSD genes: *C. fistulae* (Taxonomy ID: 2675058) has 16 FSD genes, *C. nemetophagum* (Taxonomy ID: 2305228) possesses 43 FSD genes, and *C. potablequae* (Taxonomy ID: 2675057) has 23 FSD genes.

The FSD genes in South African *O. hominis* range in size from approximately 350 to 600 amino acids (aa), with the FSD domain spanning about 175 to 200 aa. InterPro predictions indicate that these proteins are membrane-associated and positioned in the extracellular space. UniProt structural predictions of the longer (~500–600 aa) *O. hominis* FSD proteins feature a β-barrel structure, specifically an immunoglobulin domain, which is characterised by two antiparallel β-sheets forming a sandwich-like structure, as shown in Figure 3.18A.

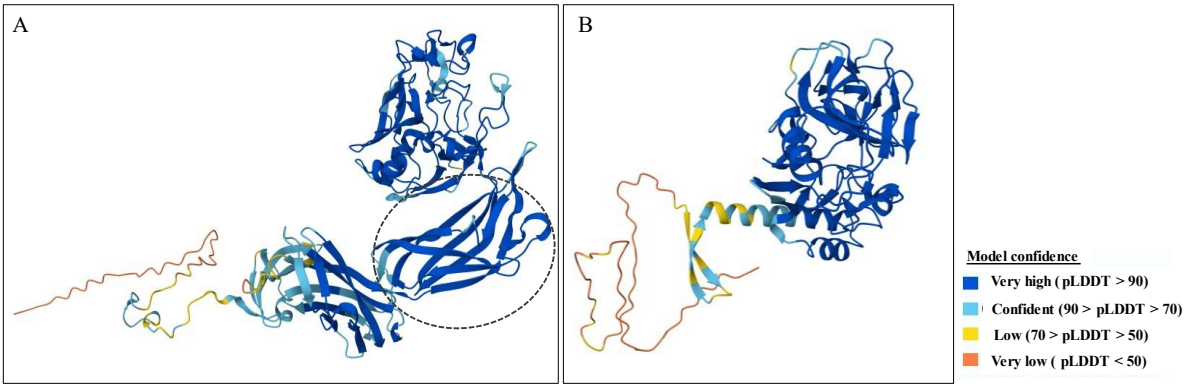

Figure 3.18. AlphaFold predicted structures of the *O. hominis Fibrobacter succinogenes* major paralogous domains. A) Uniprot ID: A0A383TVG3 (579 aa). B) Uniprot ID: A0A383TWE2 (353 aa). AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100.

Adhesin genes: Hep Hag Genes

Each South African *O. hominis* genome contains two genes annotated as Hep Hag proteins, which encode proteins with a YadA-like, left-handed β-roll domain (CDD: cd12820). This domain is characteristic of the collagen-binding virulence factor YadA, indicating that these are adhesin proteins.

Adhesin genes: Fimbrillin Family Proteins

Fimbrillin family proteins are present in isolates SA-OH-C1, SA-OH-C2, and SA-OH-C6, with only one Fimbrillin family protein identified per genome. Interpro predictions also indicate that this is a membrane-bound protein and positioned in the extracellular space.

Lipopolysaccharide biosynthesis cluster

The structure and composition of the LPS biosynthesis locus was investigated in the South African genomes and compared to LPS clusters from other *O. hominis* strains. BLASTn pairwise analysis revealed that SA-OH-C1, SA-OH-C2 and SA-OH-C6 are similar to the LPS cluster from a Cambodian *O. hominis* MAG provided by S. Salter from data ERR4181619 while SA-OH-C3, SA-OH-C4 and SA-OH-C5 are more similar to the LPS cluster from strain 22767 from the Maela cohort in Thailand. The LPS locus alignment in Figure 3.19 demonstrates the LPS cluster composition the alignments.

The LPS cluster exhibits conservation in gene sequences at the beginning (~14 kb) of the locus and at the final two sugar transferase genes. The LPS clusters from SA-OH-C1, -C2, and -C6 strains are similar to the Cambodian LPS type, with most observable differences situated in the middle of the cluster. SA-OH-C1 and -C6 feature a novel transport gene, possibly a novel flippase, along with distinct glycosyltransferases and a unique acyltransferase. SA-OH-C2 differs from C1 and C6, showing greater similarity to the Cambodian LPS type in the first ~20 kb, with shared transport genes; however, C2 also includes an additional acyltransferase and lacks an aminotransferase. Conversely, SA-OH-C3, -C4, and -C5 have an LPS cluster with high similarity to the LPS cluster from Maela strain 22767 from Thailand.

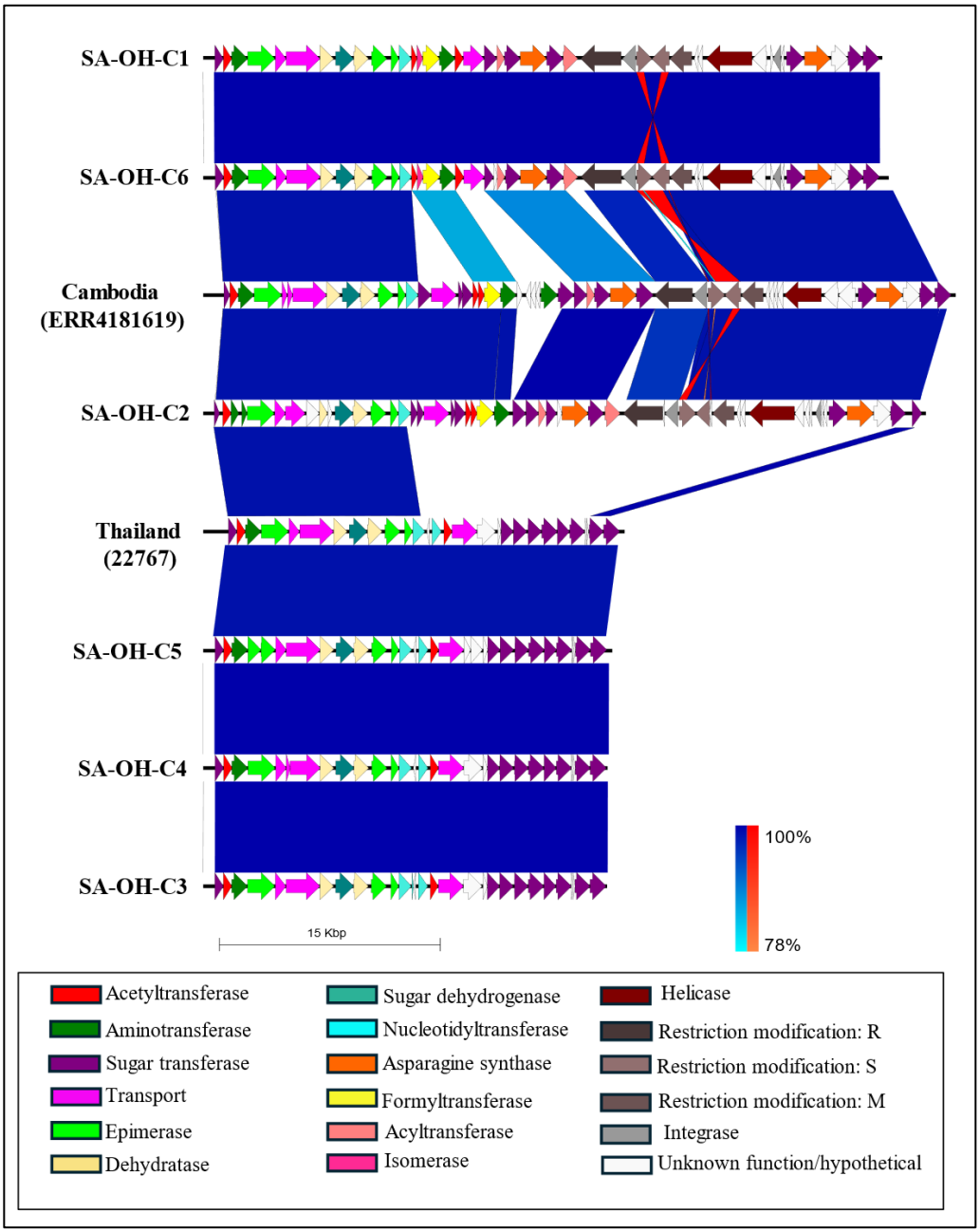

Figure 3.19. *Ornithobacterium hominis* lipopolysaccharide (LPS) biosynthesis locus BLASTn alignments. South African *O. hominis* LPS clusters from isolates SA-OH-C1 to SA-OH-C6 aligned to the Cambodian LPS cluster from metagenome assembled genome ERR4181619 and LPS cluster from strain 22767 using BLAStn. Coding sequence features are marked by arrows and genes were annotated in Bakta. Vertical blocks between sequences indicate regions of shared similarity (dark blue for matches in the same direction or bright orange for inverted matches). Image generated using EasyFig.

The structure of the RHS cluster

The RHS locus is potential virulence factor in *O. hominis*. All *O. hominis* isolates have an RHS locus and the use of longer sequencing reads, spanning the length of this locus, aided in resolving its structure. The RHS locus is locus is ~ 45 kb and situated towards the end of the genome at approximately 1.9 Mb. The *rhs* gene and its concomitant gene fragments are typically identified on the same strand. In the case of the South African isolates, all the genes of the RHS locus sit on the reverse strand with flanking integrase or transposase genes on the opposite strand as seen in Figure 3.20.

Figure 3.20 displays the general structure of the RHS locus in *O. hominis* genomes. Upstream of the *rhs* gene are two genes encoding for outer membrane protein β-barrel domain 2 (OMP-b-brl-2) containing proteins (IPR025665) followed by the T9SS C-terminal domain (CTD). The full-length *rhs* gene encodes for a protein of approximately 3000 aa and is composed of distinct domains that were identified using InterproScan: a Salmonella virulence plasmid 65 kDa B protein (sprB) domain (IPR003284), followed by an Integrin α N-terminal domain (IPR028994), and an insecticidal TcdB toxin domain (IPR022045)[90]. This is followed by the highly conserved RHS core domain (IPR031325) and a variable C-terminal tip, which may exhibit toxin activity. Following the C-terminal tip is believed to be a cognate immunity gene.

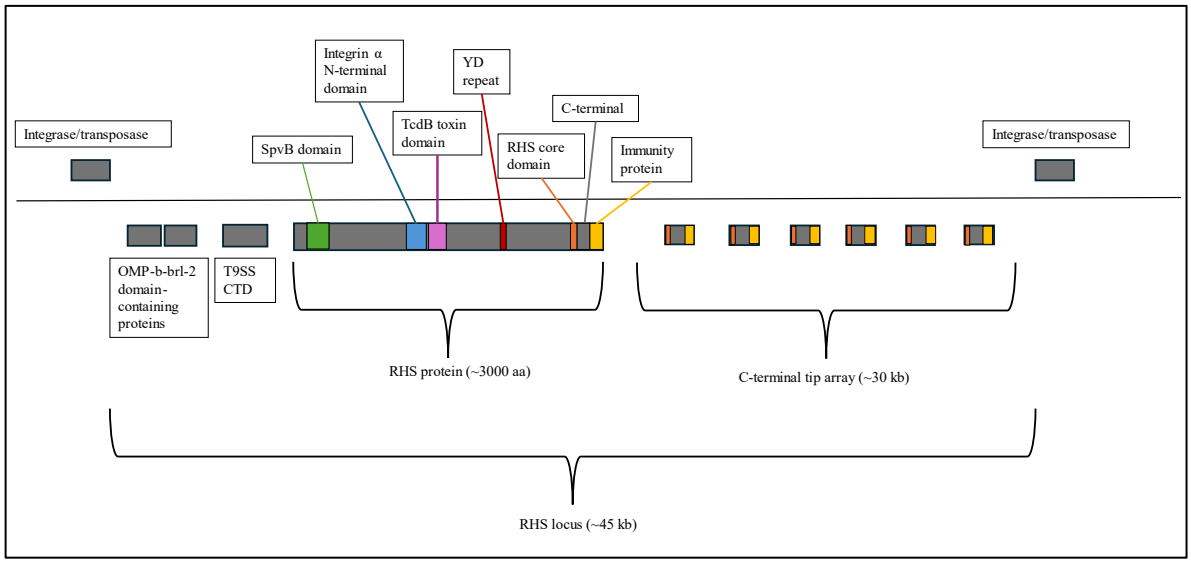

Figure 3.20. Structure of the *O. hominis* rearrangement hotspot (RHS) locus. OMP-b-brl-2, Outer membrane protein β-barrel domain 2; T9SS CTD, Type 9 secretion system C-terminal domain; SpvB, Salmonella virulence plasmid 65 kDa B protein; TcdB, toxin complex domain B.

Downstream of this *rhs* gene is a ~ 30 kb array of gene fragments, called the C-terminal tip array, containing RHS core and C-terminal tip sequences with a possible immunity protein, as shown in Figure 3.20. The number of C-terminal tips in the array varies between genomes, as does the length of the individual C-terminal tips.

Next, we investigated the variation in the C-terminal toxin tips found in the full-length *rhs* gene of the South African genomes. The fragmentary C-terminal tips are likely to be functionally silent, hence the analysis was focused on the full-length *rhs* gene. The C-terminal sequences were compared across the South African isolates using multiple sequence alignment (MSA)[91], and analysis started from the conserved RHS-core domain, denoted by the conserved animo acid sequence, YYYGARYY (Figure 3.21). The MSA demonstrates that SA-OH-C1 and SA-OH-C6 have identical C-terminal toxin tips. While SA-OH-C4 and SA-OH-C5 also share C-terminal toxin tips. However, SA-OH-C2 and SA-OH-C3 have unique tips not observed in the other South African strains. This suggests that C-terminal tips can be conserved across strains. Next, InterProScan was used to predict domains within the C-terminal toxin tips to identify the potential function[90]. The SA-OH-C2 *rhs* C-terminal tip encodes for a putative protein

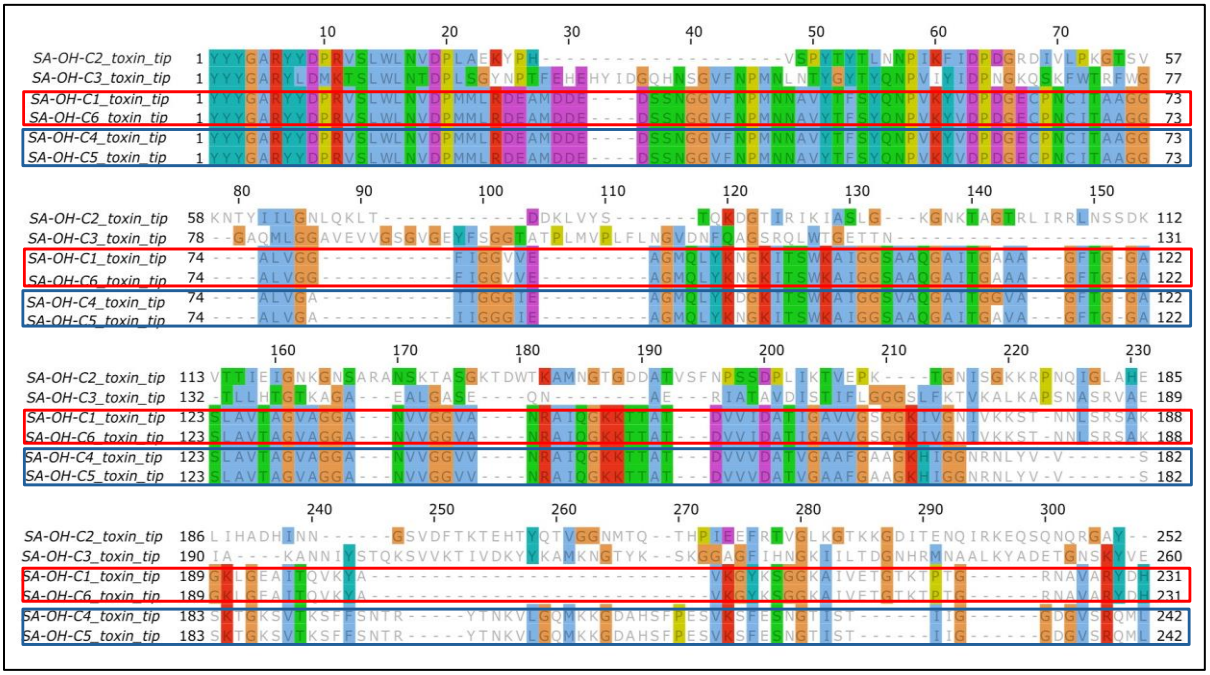

Figure 3.21. Multiple sequence alignment (MSA) of *O. hominis* rearrangement hotspot (RHS) C-terminal tips from full-length *rhs* genes. MSA performed using ClustalW and visualized in Jalview v10.0.7. Red rectangle indicates shared toxin tips between isolates SA-OH-C1 and -C6, while blue block indicates shared toxin tips between isolates SA-OH-C4 and -C5.

belonging to the NleD-like effector-protein family (IPR028208), possibly with a metallopeptidase function, and the SA-OH-C3 *rhs* C-terminal might code for a toxin domain with similarity to the N-terminal domain of ParB (IPR039380). However, protein or domain matches in the C-terminal tips of the other isolates were not identified.

Additionally, a BLASTp analysis of the C-terminal tips of the full-length *rhs* genes was performed to investigate potential C-terminal conservation across *O. hominis* strains and related species. Table 3.12 shows that the South African *O. hominis rhs* C-terminal tips share high sequence similarity to displaced tips from the Australian type-strain and Maela strain 22767 from Thailand RHS clusters. Moreover, isolate SA-OH-C3 shares an identical C-terminal tip with a displaced tip from the type strain. Additionally, the C-terminal tip for SA-OH-C2 was identical to a displaced tip from the strain 22767 and a displaced tip in *Bergeyella cardium*, with high sequence similarity to a C-terminal tip in *Flavobacterium filum*. C-terminal tips from SA-OH-C4 and -C5 also shared high sequence similarity with a displaced tip in *O. rhinotracheale*. There were also several hits ranging from 62–69% identity to other C-terminal tips from other species in the *Flavobacterium* genus that were not included in the table. This suggests that C-terminal tips are conserved across closely related species.

Table 3.12. BLASTp alignments for intact rearrangement hotspot (RHS) C-terminal tips in South African *O. hominis* genomes

Query	Alignments	% ID	% Coverage	E-value	Description
SA-OH-C1, SA-OH-C6	RHS repeat-associated core domain-containing protein [*Ornithobacterium hominis*]	89.5	66	2.00E-123	Displaced tip from Australian type strain
SA-OH-C5, SA-OH-C4	RHS repeat-associated core domain-containing protein [*Ornithobacterium hominis*]	99.27	100	0	Displaced tip from Australian type strain
	RHS repeat-associated core domain-containing protein [*Ornithobacterium hominis*]	96.51	62	1.00E-11	Displaced tip from Thai strain (22767)
	RHS repeat-associated core domain-containing protein [*Ornithobacterium rhinotracheale*]	78.49	62	1.00E-75	Displaced tip
SA-OH-C3	RHS repeat-associated core domain-containing protein [*Ornithobacterium hominis*]	100	100	0.00E+00	Displaced tip from Australian type strain
	RHS repeat-associated core domain-containing protein [*Ornithobacterium hominis*]	86.91	66	1.00E-115	Displaced tip from Thai strain (22767)
SA-OH-C2	RHS repeat-associated core domain-containing protein [*Ornithobacterium hominis*]	100	100	0.00E+00	Displaced tip from Thai strain (22767)
	RHS repeat-associated core domain-containing protein [*Bergeyella cardium*]	100	100	0.00E+00	Displaced tip
	RHS repeat-associated core domain-containing protein [*Flavobacterium filum*]	73.15	100	2.00E-127	Displaced tip

3.10 Pangenome analysis

The pangenome includes the entire repertoire of genes found within a group of genomes, usually belonging to a specific taxon, and serves as a tool for assessing their genetic diversity. The pangenome analysis was performed to identify the core genome of *O. hominis*, which are genes present across all isolates, and the accessory genome that represents more variable genes that genes that are present in some strains. A pangenome analysis of *O. hominis* included the South African and Australian genomes as well as the MAGS from the Maela cohort. The thresholds that distinguish the core genes and accessory genes can be user-defined.

In this study, the core genome was defined as the genes present in over 99% of the strains. As seen in Table 3.13, of 2843 total genes included in the analysis, 1172 genes were identified as core genes, while 1671 genes make up the accessory genome. Accessory genes were further categorized into shell and cloud groups, which are Panaroo-specific terms, based on their frequency within the analysed genomes. Shell genes are shared across 15%–95% of the genomes, while cloud genes are found in less than 15% of the genomes and include those present in only one strain; thus, cloud genes can also be referred to as "unique" genes.

Table 3.13. Distribution of genes from pangenome analysis

Pangenome categories	Threshold	Number of genes
Core genes	(99% <= strains <= 100%)	1172
Shell genes	(15% <= strains < 95%)	1249
Cloud genes	(0% <= strains < 15%)	422
Total genes	(0% <= strains <= 100%)	2843

3.10.1 Core genome repertoire

Among the core genes identified, 10 *gld* genes (*gldC*, *gldH*, *gldM*, *gldL*, *gldN*, *gldJ*, *gldB*, *gldK*, *gldG*, *gldD*) were present. Among these, *gldG* is an ABC (ATP-binding cassette) transporter ATP-binding subunit implicated in virulence. Additionally, a gene encoding for a T9SS type A sorting domain-

containing protein and genes encoding for PorV (the T9SS outer membrane channel protein) and SprA (the translocon of the T9SS) were also detected among the core genes. Other virulence-associated and antibiotic-resistance genes included components of ABC transporters associated with multi-drug resistance, such as the ABC-type transporter ATP-binding protein EcsA, a putative ABC transporter ATP-binding/permease protein YheI, and an ABC-2 family transporter protein. Genes encoding components of the ABC transporter complex LolCDE were also identified. Penicillin-binding proteins (PBP), as well as a gene encoding a Metallo-β-lactamase superfamily protein, were also identified as core genes. In addition, six genes encoding for different OMP-b-brl-2 domain-containing proteins, which is a fully integrated outer membrane, were identified among the core genes. OMP-b-brl-2 domain-containing proteins may serve a role as virulence factors and are involved in cell adhesion, nutrient uptake, and cell signalling.

A bacteriocin type signal sequence was also identified in all the *O. hominis* genomes, while three bacteriocin-encoding genes were identified in 12/13 genomes and two Class IIb bacteriocin lactobin A/cerein 7B family proteins were also found in 12/13 genomes. These genes were missing from the strain 22803. Bacteriocins are antimicrobial peptides and are indicators of microbial interactions and competition. Additionally, the *toxA* gene, annotated as PMT in the pangenome analysis output, was also identified among the core genes, reinforcing the conserved nature of this gene.

The presence of these core genes suggests that they are conserved and, thus, might be essential for the survival and colonization of *O. hominis*.

3.10.2 Accessory genome repertoire

Approximately 26% of the accessory genome consists of hypothetical proteins. The accessory genome includes phage genes, mobilization proteins, FSD genes, LPS biosynthesis genes, RHS repeat-associated core domain-containing proteins, bacteriocin genes, and genes involved in conjugal transfer such as *traG* and *traD*. Additionally, relaxases, transposases, recombinases, and integrases are frequently identified in the accessory genome.

Among the accessory genes, 52 FSD genes were identified. Of these, only one FSD gene was conserved across all 13 genomes, and six were shared by at least 10 genomes. In contrast, nine unique FSD genes were found, with four identified in SA-OH-C5 and three in SA-OH-C3. Thus, despite the abundance of FSD genes in *O. hominis* genomes, there is considerable variability of these genes across the strains.

Among the *O. hominis* genomes analysed, 13 bacteriocin-related genes, including those encoding for bacteriocin-signal sequences and bacteriocin proteins, were identified. Strain 22803 from the Maela cohort contains only a bacteriocin-type signal sequence and one hypothetically annotated bacteriocin gene. In contrast, strains 22298, 22767, and Australian strains 902C1, 903C1, and 916C1 encode for all 13 bacteriocin-related genes. Notably, two bacteriocin-encoding genes were absent from all South African genomes. This indicates significant variation in bacteriocin content across the different strains.

Notably, the Co/Zn/Cd efflux system component *czcD*, which is a heavy metal ion transporter involved in heavy metal resistance, was found in the accessory genomes and is present in all South African strains. Furthermore, unique genes such as a cation efflux protein and a HNH endonuclease were identified in isolate SA-OH-C5, while isolate SA-OH-C2 has an uncharacterized protein with ParB-like and HNH nuclease domains and a protein belonging to the Borrelia ORF-A superfamily. Additionally, isolate SA-OH-C4 has a unique N-acetylneuraminate lyase, which is involved in sialic acid metabolism where sialic acid from the host can be used as an energy source. A minor outer membrane protein, Omp16, also a potential virulence factor was identified in SA-OH-C4. Most unique genes were annotated as hypothetical proteins, but some FSD genes, transcriptional regulators, mobilization proteins, and ABC transporter proteins were uniquely present in the South African genomes. The presence of these unique genes suggests strain-specific adaptations.

3.11 Core genome alignment and phylogenetic analysis

After identifying the core genes, a core genome alignment was performed for the South African and Australian isolate genomes and the MAGs from the Maela camp in Thailand. As shown in Figure 3.22, with the tree rooted using the MAG 22803 as the outgroup, the core genes from *O. hominis* strains form three distinct clades based on geographic location. Additionally, the South African *O. hominis* strains are more closely related evolutionarily to the Australian strains than to the strains from the Maela cohort from Thailand. However, the long branch lengths as indicated by the tree scale suggest significant evolutionary distance between the strains.

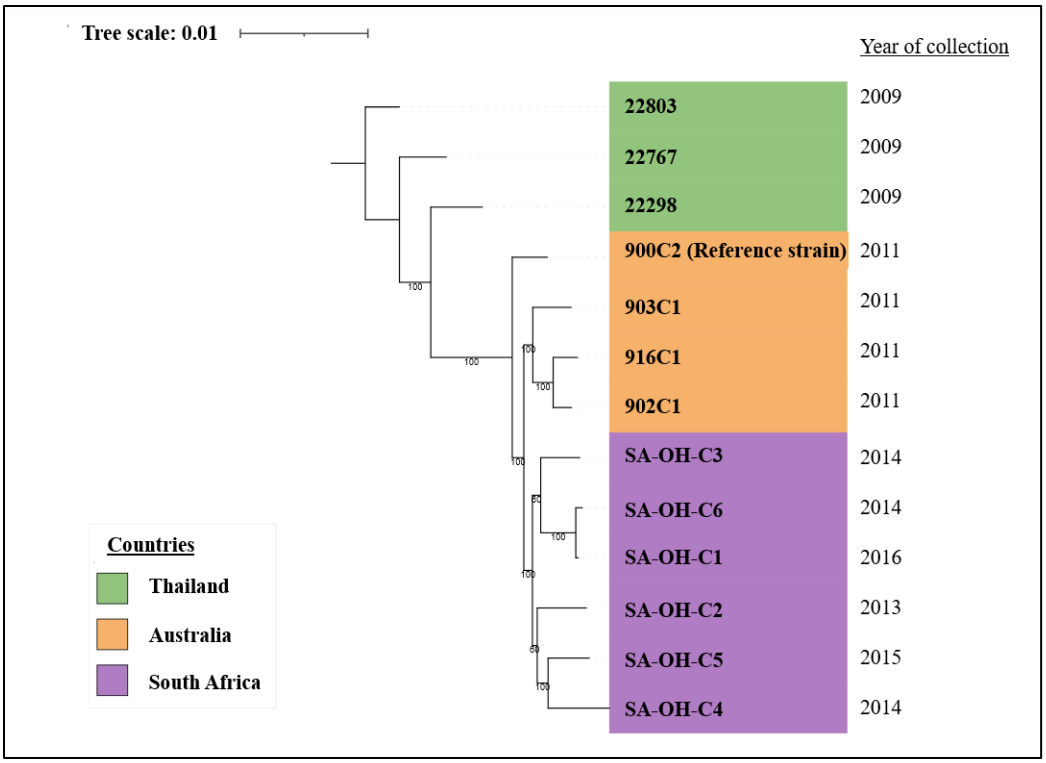

Figure 3.22. Phylogenetic tree of *O. hominis* reconstructed from the core genome alignment of 13 genomes from Thailand (Maela cohort), Australia, and South Africa. The evolutionary history was inferred using the maximum likelihood method and Tamura-Nei model. The bootstrap consensus was inferred from 10 replicates. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to branches. Scale bar indicates 0.01 substitutions per position. The collection year for the sample from which the genomes are derived is shown on the right side of the tree. Evolutionary analyses were conducted in Mega11 and tree visualization using iTOL tree visualizer.

Additionally, the sample collection year for the South African isolates varies across 2013 to 2016, while the Australian and Maela samples were collected within the same years, 2011 and 2009, respectively. Interestingly, South African isolates SA-OH-C1 and SA-OH-C6, although collected in different years and from different children, are closely related strains based on the shorter branch lengths.

3.12 COG functional prediction of core genes

After core genes were identified, we sought gain a better understanding of the functions of these genes. Protein sequences of core genes were subjected to EggNOG-mapper (v2.0.1), which can map protein-coding genes to the orthologous families in the eggNOG database, thereby providing a functional assignment for the genes[88]. Clusters of orthologous groups (COGs) were assigned to the core genes and gene counts for the following COG categories are shown in Figure 3.23. Approximately 900 COGs were found in the core genome. A large proportion of the core genes could not be functionally assigned (179/900, 19.89%); however, out of the COGs that could be identified, these were dominated by genes

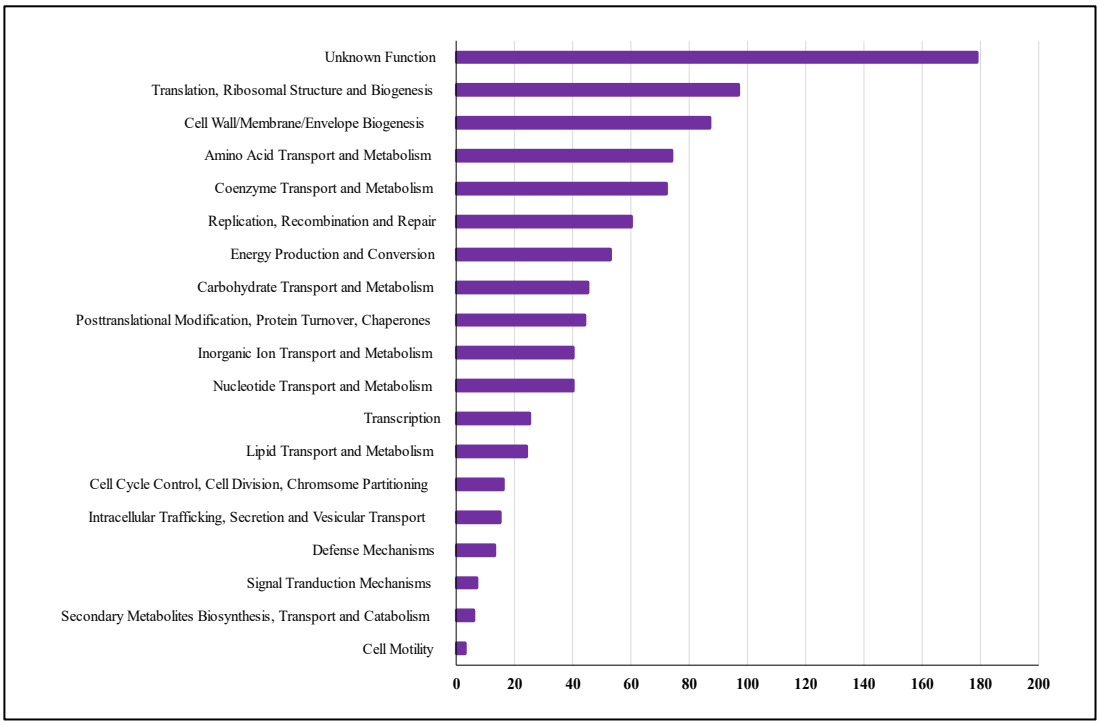

Figure 3.23. Summary of *O. hominis* core genes COG functional assignments. Core genes identified using Panaroo pangenome analysis of South African and Australian *O. hominis* genomes as well as draft genomes from the Maela cohort in Thailand. Core genes subjected to COG functional assignment and categorized into COG categories. Counts of genes per COG category are shown in the bar plot.

involved in translation and ribosomes (97/900, 10.78%), followed by genes involved in Cell membrane biogenesis (87/900, 9.67%) and metabolic genes in Amino acid metabolism (74/900, 8.22%), and Coenzyme metabolism (72/900, 8.00%). In comparison, genes involved in Cell motility (3/900, 0.33%), Secondary metabolite Biosynthesis genes (6/900, 0.67%), signal transduction (7/900, 0.78%) and defence mechanisms (13/900, 1.44%) are more underrepresented in *O. hominis*.

Chapter 4: Discussion

This study reveals the presence of *O. hominis* in infants from a South African birth cohort and presents the first successful culturing and sequencing of African isolates. Notably, these South African isolates represent the only known repository of *O. hominis* strains outside of Australia[68]. This is a significant development as most of the current literature regarding *O. hominis* has been based on metagenomic sequencing data from cohorts based in Thailand, Australia, and more recently, Egypt[14,71]. The identification of *O. hominis* in South African infants creates new opportunities for research into its ecological role in the NP microbiome of infants in sub-Saharan Africa, a region that is disproportionally impacted my respiratory illness[92]. Moreover, the successful culturing and sequencing of these isolates offers new prospects for studies on the pathogenic potential and genetic diversity of *O. hominis*. The increase in *O. hominis* relative abundance with infant age, and notable increase just before pneumonia events suggests that *O. hominis* plays a role in on disease-risk outcomes, warranting further investigation into the role of *O. hominis* within the NP microbiome and its potential for pathogenicity.

The identification of virulence-associated genes, AMR genes, and several adhesins provides insights into potential mechanisms of virulence and host interactions[93-95]. Furthermore, the identification of a putative plasmid housing genes involved in virulence and possibly horizontal gene transfer warrants further investigation into HGT mechanisms in *O. hominis* and the impact on its evolution[96]. The virulence factors identified in this study should be further explored to understand their potential impact on the host and NP microbiome.

Additionally, the identification of three distinct LPS biosynthesis clusters in the South African isolates provides insights into potential mechanisms of *O. hominis* interactions with other bacterial cells and the host immune system[97]. The resolution of the RHS cluster arrangement using ONT long-read sequencing suggests that *O. hominis* utilised a RHS tip displacement mechanism, which may further influence its virulence[98,99].

Notably, pangenome analysis revealed that South African *O. hominis* strains form a distinct clade, more closely related to Australian strains than to those from the Maela cohort in Thailand. This suggests the presence of transmission patterns or environmental adaptations influencing the evolution of this species. Altogether, these findings reveal new insights into the clinical and ecological relevance of *O. hominis*.

The 16S sequencing dataset from archived DCHS samples was initially screened for *O. hominis* to identify candidate samples for subsequent culture attempts; however, several interesting aspects regarding the carriage and prevalence of *O. hominis* in this cohort were noted and this led to a more extensive study of *O. hominis* carriage in this cohort. *Ornithobacterium hominis* was detected in approximately 32% of participants in this cohort, which is slightly lower than the carriage prevalence observed in the Maela cohort, which was approximately 45%[14]. Furthermore, the relative read abundances of *O. hominis* in samples from our cohort were much lower than that reported for the Maela and Australian samples. The highest *O. hominis* read abundance in this cohort was at 7.8%, while relative abundance in the Maela and Australian cohorts were as high as 71% and 55%, respectively[12,54]. Differences in environment and climate settings or host characteristics could explain this deviation in *O. hominis* relative abundance in NP samples.

We further observed that the relative abundances of *O. hominis* in our cohort increased after 6 months of age and continued to increase till 24 months of age. This corroborates with findings by Salter *et al*. (2017), where the unclassified Flavobacterium increased in relative abundance at around 4–6 months of age in the Maela cohort[12]. This aligns with the understanding that the infant NP bacteriome gradually shifts during the first year of life, with a major shift occurring during the first three months of life[12].

Furthermore, *O. hominis* was identified as a persistent colonizer in 46 children with at least 0.1% relative abundance. This corroborates findings by Salter *et al*. (2017), where this species was found to persistently colonize the infant NP at high proportional abundances[12]. Interestingly, when investigating whether the characteristics of this study's participants are representative of the broader 16S rRNA amplicon

sequencing cohort, the current study cohort shows a slightly higher percentage of pre-term births (21.74% vs. 17%) and a slightly larger median household size (5 people vs. 4 people), compared to those of the broader cohort study. The current study cohort also exhibits a slightly higher rate of vaginal births (84.8% vs. 80%) and a lower percentage of HIV-exposed but uninfected infants (17% vs. 27%) than that of the broader cohort study. Conversely, the rates of maternal smoking are comparable between the current study cohort and the broader cohort study (26% vs. 27%)[100], and both cohorts also demonstrate high immunization coverage (100% vs. >98%). Additionally, the WAZ scores at birth were consistent across the cohorts. These findings suggest potential associations between *O. hominis* carriage and increased household density, birth outcomes, and maternal health factors; however, the differences between the cohort with *O. hominis* longitudinal carriage and the broader cohort were not subject to statistical analysis to determine significance.

Interestingly, in a sub-set of infants with respiratory illness events- and their age- matched controls, *O. hominis* abundance appears to increase in the two-week period immediately preceding a respiratory illness. In contrast, *O. hominis* abundance was more evenly spread across timepoints in age-matched controls. Overall, this suggests that *O. hominis* colonisation might increases disease risk, possibly by modulating the NP microbiome. However, the sample size for this analysis is small and this limits the strength of the associations that can be drawn from this observation, necessitating further validation. Additionally, the aetiology of the pneumonia cases was not determined, so it remains unknown whether the infections were due to a bacterial or viral pathogen and which species was implicated. Therefore, further associations with *O. hominis* and specific pathogens cannot be elucidated based on this analysis. Overall, it would be worthwhile investigating the mechanisms behind this persistent colonization observed in the 46 children and the potential role *O. hominis* may play in modulating the NP environment.

Our initial attempts to culture *O. hominis* on CBA were based on samples with higher proportional abundance of *O.* hominis (>1%). Despite prioritizing high-abundance samples, primary isolation was a

lengthy process as *O. hominis* is a slow growing organism that could take up to 5 days for colonies to appear, consistent with findings by Lawrence *et al*. (2019)[68]. Slow growth suggests that *O. hominis* is less metabolically active, which is indicative of nutrient deprivation or possible deficient metabolic pathways. Although Columbia agar enriched with sheep's blood was used, additional nutrients and/or growth factors might need to be added to support the growth of *O. hominis*; however, it is not known which specific nutrients are essential for optimal growth[101]. Slow growth rates are also characteristic of *O. rhinotracheale* isolates cultured on TSA under microaerobic conditions; however, supplementation of broth medium with 5% FBS and 2% IsoVitaleX has shown to improve growth rates of *O. rhinotracheale isolates*[101]. Thus, growth factor supplementation of either broth or solid media may improve *O. hominis* growth in a similar manner. Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG) databases could be used to reconstruct the metabolic pathways in *O. hominis*, and subsequently identify the missing pathways or deficient pathways that are essential for *O. hominis* metabolism and optimal growth[102]. The output from KEGG analyses could be combined with pangenome analysis data to identify conserved metabolic pathways or pathways that may be deficient or missing across isolates.

Altogether, the slow growth of *O. hominis* on CBA combined with the low culture success rate from primary samples (6 isolates out of 34 NP samples, success rate = 17.6%) indicate that current isolation protocols for *O. hominis* are sub-optimal. This is often the case for fastidious, slow-growing organisms as most routine culture methods and agar-based media have been developed for faster-growing organisms[49]. Moreover, the success of primary culture from NP samples is hindered by the presence of faster-growing organisms, such as *Streptococcal* and *Staphylococcal* species, which quickly dominate the agar surface. Thus, the development of a *O. hominis* selective media would be beneficial for primary isolation. Antibiotic susceptibility testing revealed that South African *O. hominis* isolates were possibly resistant to polymyxin B; thus, to overcome overgrowth by other fast-growing microbes, supplementing growth media with polymyxin or colistin might aid in isolating the species from polymicrobial samples; however, polymyxin B resistance will have to be evaluated for other *O. hominis* strains from diverse geographic

regions to confirm if this trait is conserved in this species. This, and other strategies to optimise isolation of *O. hominis* in clinical samples, needs to be explored further.

Given that *O. hominis* is difficult to isolate from primary samples, PCR targeting the TOXIN (*toxA*) gene is a useful method for confirming *O. hominis* identity during isolation attempts. The *toxA* target PCR was used to estimate prevalence in NP swabs in the Maela cohort, where it was concordant with the 16S target[14]. All isolates in this study tested positive for the *toxA* gene, similarly Lawrence *et al*. (2019) successfully screened for *O. hominis* from Australian samples using PCR targeting *O. hominis*-specific regions of the *toxA* gene. This suggests that PCR screening for the presence of the *toxA* gene is a viable option for the detection of *O. hominis* and can be used for strains isolated from different geographic regions. The ubiquitous and conserved nature of the *toxA* gene in *O. hominis* stimulates investigations into its potential roles. Despite the fact that PMT and the *O. hominis* putative protein encoded by *toxA* share similar active sites, the expression or function of these toxins possibly differ since there are no reports of PMT-like toxin mediated disease in the Maela cohort[14]. However, PMT has a wide-ranging impact, and the modulation of immune cell activation is thought to be the primary function of this protein[103]. Therefore, it's likely that that *O. hominis* toxin encoded by *toxA* might bind to host cells and exhibit immunomodulatory effects. The potential impact of the *O. hominis* toxin encoded by the *toxA* gene should be investigated and future studies could investigate the expression of this toxin, determine its crystal structure, investigate its binding to host cells, and confirm its immunomodulatory effects.

The colony morphologies of our *O. hominis* isolates are consistent with those described by Lawrence *et al.* (2019), who reported the first successful culturing of *O. hominis* from Australian samples[68]. *Ornithobacterium hominis* colonies appeared circular, grey, glistening, and ranged in size from punctiform colonies to 1–3 mm in diameter after 48–120 hours of incubation. Notably, single-colony sub-cultures varied in size, with punctiform colonies appearing after prolonged incubation and sporadically. This is indicative of small colony variants (SCV), which is linked to environmental stress[104]. SCVs were also observed by Lawrence *et al*. (2019), suggesting that this might be a characteristic feature of *O.*

hominis growth under these culture conditions[68]. Additionally, SCVs are characteristic of the *O. rhinotracheale* isolates and exhibited higher minimum inhibitory concentration (MIC) values than that of the wild-type isolates[101]. Although SCVs are associated with antibiotic resistance and persistent/chronic infections, the clinical relevance of SCVs is yet to be thoroughly investigated[104,105].

We also observed alpha haemolysis in our isolates, a phenomenon which was not originally reported by Lawrence *et al*., but has been observed for *O. rhinotracheale*[101]. Regarding gram-stain characteristics, while typical short gram-negative bacilli were observed, distinct "U-shaped" bacilli were also noted. Although the significance and reason for this morphology was not elucidated, it may be a useful feature for further differentiation of *O. hominis* from other short, gram-negative bacilli.

Results of the nitrocefin test revealed that only one of the isolates produced β-lactamases, whereas all Australian *O. hominis* isolates produced β-lactamase; thus, β-lactamase production is not a universal characteristic of *O. hominis* as previously thought[68]. Interestingly, this β-lactamase producing isolate displayed reduced susceptibly to β-lactams, while the other isolates appeared to be more susceptible. Although antibiotic administration data in this study was not available, it is possible that infants in this cohort have been exposed to β-lactam antibiotics. B-lactam antibiotics are frequently prescribed in public sector hospitals in South Africa[106]. B-lactams, specifically, Amoxicillin, are commonly used to treat upper respiratory tract infections as well as pneumonia in infants[106]. Given the high incidence of respiratory illness in the cohort, β-lactams are likely one of the most prescribed antibiotic classes in hospital settings and ambulatory care[73,106]. Therefore, it is likely that the exposure to β-lactam antibiotics in this cohort may be relatively high. Despite South African *O. hominis* isolates being susceptible to β lactams, the high exposure to β-lactams coupled with the confirmed presence of β-lactamase genes and the β-lactam resistance observed in Australian *O. hominis* isolates[68], indicates the potential for β-lactam resistance in South African *O. hominis* strains.

In addition, antibiotic susceptibility testing revealed that all South African *O. hominis* isolates were likely resistant to polymyxin B. Polymyxin B is used to treat severe infections caused by gram-negative bacteria[107]. Colistin (Polymyxin E) is also used in a similar manner and exhibits similar antimicrobial activity to Polymyxin B *in vitro*[107]. Polymyxins are believed to exert their primary mode of action by disrupting the integrity of the outer membrane of gram-negative bacterial cells, although it may have other cellular targets[108]. Polymyxin B electrostatically binds the outer membrane via the Lipid A component of the negatively charged LPS and alters the outer membrane structure, thereby destabilising the permeability barrier, and ultimately resulting in cell death[109,110]. Modification of the LPS, specifically the lipid A portion of the LPS, is believed to be the main resistance mechanism against polymyxin B[109]. Mutations that lead to modifications of the lipid A structure, alter the net charge of the lipid A and this leads to decreased affinity for the antibiotic class[109]. The *mcr* and *eptA* genes are involved in the transfer of phosphoethanolamine to lipid A, and mutations in *PhoP*/*PhoQ* and *PmrA*/*PmrB* genes lead to the addition of 4-amino-L-arabinose to the lipid A component[109,111]. These modify the lipid A molecule with positively charged moieties, thereby reducing the electrostatic affinity with polymyxin, and thus attachment is reduced. However, the exact mechanisms by which *O. hominis* modifies its lipid A structure, consequently conferring resistance to Polymyxin B is unknown, and determining structural changes to the LPS based exclusively on genetic data is complex.

Resistance to polymyxins has also been observed in *O. rhinotracheale* isolates cultured from infections in broiler chickens from Egypt[112] thus, polymyxin resistance is not unusual for this genus. Although all isolates were resistant to polymyxin B, this do not conclusively indicate that *O. hominis* is intrinsically resistant, as LPS modification mechanisms can differ between isolates. Furthermore, the observed polymyxin B resistance in *O. hominis* isolates from infants in this cohort is noteworthy. Although antibiotic data is not available, it is unlikely these infants were exposed to the polymyxin B as it is a last-resort antibiotic used to treat severe, multidrug-resistant infections in South Africa[113]. Additionally, due to its toxicity, it is generally reserved for critically ill patients[113]. Therefore, routine use of Polymyxin B in

this cohort is not likely, this prompts investigation or interest into the potential mechanisms behind the acquisition of polymyxin resistance in *O. hominis*.

Although the antibiotic susceptibility testing results are insightful, these results should be interpreted with caution. Categorical reporting of "susceptibility" or "resistance" for *O. hominis* is not recommended as there are no established breakpoints for *O. hominis*. Additionally, the antibiotic susceptibility testing was performed using a modified version of the disk-diffusion method, where susceptibility testing was performed on CBA instead of the standard Muller-Hinton agar and cultures were incubated for 48 hours to allow for sufficient growth; therefore, these initial antibiotic susceptibility results should be interpreted with caution. However, to obtain a more comprehensive antibiotic susceptibility profile of *O. hominis*, a panel of other antibiotics should be evaluated, and more isolates included. Ideally, MIC values, particularly for polymyxins, should also be determined.

Ornithobacterium hominis colony size increased and appeared more luxuriant on CBA when sub-cultured using biphasic media, likely due to the enhanced nutrient availability provided by the liquid broth overlaying the solid agar base. This approach, initially suggested by Susannah Salter, was inspired by the successful culturing of *Campylobacter pylori* in biphasic media[114]. Considering that this method is simple to adopt, biphasic media culturing represents a viable strategy for optimizing *O. hominis* growth to obtain sufficient cell material for DNA extractions. It is worth mentioning that *O. hominis* grows in static liquid cultures, while most culturing methods require agitation to dissolve gases; therefore, the volume of the liquid culture is critical for maintaining appropriate oxygen levels. It should be noted that cell densities were not monitored during biphasic media culturing, thus, the effects of broth passage on cell proliferation cannot be accurately determined. Moreover, the success of this method is heavily influenced by the viability of the initial inoculum from the plate. Beyond improving *O. hominis* growth, the use of biphasic media presents new opportunities to study *O. hominis* in broth culture, which is essential for future research into secreted products, such as the *toxA* gene product, and for conducting future metabolic tests.

Although these strategies to improve *O. hominis* growth are notable, a more comprehensive understanding of this species necessitates further phenotypic testing. Other phenotypic tests are yet to be conducted or reported for this species. These include, but are not limited to, physiological tests for pH, salt and temperature tolerance, and biochemical tests such as substrate utilization and sugar fermentation. Additionally, while these strategies are beneficial for growth optimization, the development of minimal media would allow for more precise testing of growth requirements and investigating metabolic profiles.

DNA extraction quality and quantity varied greatly between samples, but generally, the kit-based method produced better quality DNA that was more intact and with less contamination than the phenol-chloroform method. High-molecular weight DNA with sufficient purity is essential for the success of ONT sequencing as ONT relies on long fragments and the carry-through of chemical contamination negatively affects flow-cell performance. The phenol-chloroform method involves multiple tube transfers, thereby increasing the risk of contamination and shearing, possibly resulting in lower-quality DNA. A major limitation of this comparison between the extraction methods was that plate scrapes were used instead of broth cultures; therefore, the input cell material quantities were not standardized, possibly accounting for the yield variation observed.

Furthermore, RNA contamination remained an issue across all samples, despite the addition of RNase A. RNA contamination may inflate DNA quantities when quantifying using the Nanodrop, this ultimately results in insufficient amounts of input for library preparation. Thus, accurately quantifying DNA using a Qubit fluorometer, which is not impacted by the presence of RNA, should be incorporated as an essential quality control step in ONT sequencing. Overall, RNA removal steps need to be re-evaluated for these methods. Additionally, other extraction protocols should be evaluated for *O. hominis* to determine the best DNA extraction method.

All genomes assembled into single, circular contigs of ~2 Mb in size except for SA-OH-C6, which generated an unusually large genome of approximately ~3.5 Mb due to the presence of *Helcococcus*

contig. The *Helcococcus* contamination as not observed for the other isolates, suggesting that the SA-OH-C6 culture plate was contaminated with *Helcococcus*. This finding emphasises the impact of contamination on genome assembly. It further underscores the importance of culturing *O. hominis* on more selective media to reduce the risk of contamination and validating genome assemblies. Mapping and extracting reads to exclude contaminants is a useful measure to ensure the accuracy and reliability of genome assemblies.

Notably, a putative plasmid was identified in SA-OH-C4. One of the contigs for the SA-OH-C4 genome exhibited high coverage and was a circular, repeat sequence. Further investigation of the ORFs in this contig revealed a a *mobV*-family relaxase gene and a *repB* gene, which are indicative of a plasmid. To the best of our knowledge, this is the first reported evidence of a plasmid in *O. hominis*. The presence of a *mobV* relaxase, which is involved in plasmid mobilization and conjugation, is evidence for horizontal gene transfer in *O. hominis*[96]. Additionally, the plasmid may contribute the *O. hominis* virulence as the *vapD* gene encodes for a protein with endoribonuclease activity[115]. The exact role of this *vapD* in *O. hominis* is not known, but *vapD* genes have been identified in pathogenic bacteria and are believed to encode for virulence-associated factors[116]. These genes have been identified in non-encapsulated *H. influenzae*, and encode for a ribonuclease toxins that promote the persistence of *H. influenzae* in epithelial cells[95]. *H. influenzae* vapD is believed to degrade mRNA transcripts related to metabolism, resulting in cell stasis, and causing persistent infections in the respiratory tract of the host[95]. This is in response to stress stimuli such as antibiotics or nutrient deprivation[95]. *VapD* genes are also found in some strains of *Helicobacter pylori*, a human pathogen that colonizes gastric epithelial cells[116]. This vapD gene in *H. pylori* encodes for a toxin with an endoribonuclease function, and its expression is implicated in chronic infections and may contribute to persistent colonisation in epithelial cells[117]. However, despite the association of *vapD* expression with increased virulence, not much else is known about is biological role [116]. The *vapD* gene in *O. hominis* might encode for similar toxins, possibly contributing to the persistent

colonization of *O. hominis*; however, the expression and function of the *O. hominis* vapD will have to be confirmed.

Furthermore, the *vapD* gene has also been identified in *O. rhinotracheale* isolates, indicating possible conservation of function across related species[118]. In addition, a *vapD* gene was also identified in draft genomes of *O. hominis* from Egypt, further suggesting a functional role in *O. hominis* virulence[71]. BLASTp analysis showed that the *vapD* gene on the plasmid was identical to the *vapD* genes in MAGS from the Maela cohort and Australian *O. hominis* genomes and was highly similar to the chromosomal *vapD* gene in isolate SA-OH-C6. The shared sequences between the plasmid *vapD* and chromosomal *vapD* genes in different *O. hominis* strains suggests a potential for plasmid integration into chromosomes. The presence of the *vapD* genes on the plasmid is evidence for the potential virulence role of plasmids in *O. hominis* and may additionally, contribute to antibiotic resistance, niche adaptation, and genetic diversification. However, further validation is required to confirm the presence and function of the plasmid.

The 5kb read length cutoff is primarily used to prevent assemblies from breaking at the rRNA operons, which are common assembly break sites. Each rRNA operon is slightly over 5kb in length and consists of three genes repeated three times in the *O. hominis* genome. Ensuring reads exceed this length (>5kb) mitigates genome fragmentation that is often observed in short-read assemblies. Fragmented assemblies, due to the inability to resolve repeat regions, is a major limitation of using short-read technology to assemble whole genomes[119]. In addition to rRNA operons, FSD genes and the RHS cluster are sites that disrupt short-read assemblies and lead to fragmented genomes[13]. The RHS cluster includes an array of repeats with varying copy numbers. This is another concern of short-read usage as these repeat regions will exceed the length of short reads (typically <300 bp), resulting in the collapse of repeated sequences and inaccurate copy numbers as a consequence[119]. Long reads that can span the entire repeat region are a simple solution, emphasizing the value of long-read sequencing technology in achieving more complete and accurate whole genome assemblies.

Before further genome analysis was conducted, hypothetical gene counts were used to evaluate the assemblies. Hypothetical gene counts are indicative of frameshift errors, which inflate total gene counts and impact downstream analyses. Compared to the reference genome, the South African *O. hominis* genomes exhibited inflated gene counts, attributable to the increase in the number of hypothetical genes, indicating a greater number of frameshift errors. ONT sequencing exhibits relatively high error rates (10–20%), mostly attributed to indels (insertions and deletions) leading to frameshift errors[120]. Ideally, high-accuracy short reads (error rates are often <1%) should be used to correct errors in long reads, provided that the same sample/isolate was sequenced using short-read and long-read technology as seen with the complete genome sequence of the *O. hominis* type strain[13]. However, short read sequences where not available for these strains, thus, the errors could not be corrected.

The assembleBAC-ONT pipeline, developed by Dr Andries van Tonder, was used as an alternative workflow for assembling and annotating these genomes[81]. Although this pipeline produced assemblies with more hypothetical genes than that of the manual pipeline, the usage of an automated pipeline with integrated tools highlights the importance of developing bioinformatic tools/pipelines that are more streamlined, efficient and user-friendly. It is important to note that only six genomes were assembled and annotated but considering that some ONT kits are intended to sequence either a maximum of 24 or 96 samples, one must consider the efficiency and scalability of these assembly pipelines. While the manual pipeline produced better annotated assemblies, the management of this workflow, including file handling and configuration, would increase in complexity with increasing numbers of genomes/samples; thus, this approach is potentially inefficient. In contrast, file management is automated and more consistent with Nextflow-managed pipelines; thus, automated pipelines are more scalable than manual pipelines. Conversely, the manual pipeline is more "beginner-friendly" and may be a better option for the assembly of fewer genomes.

The genome characteristics corelate with those reported for the *O. hominis* reference strain (MSHR-COH1) genome that consisted of a single circular chromosome of 2,036,909 bp and with a GC content of

35.72% [13]. The type-strain genome also houses 1,899 genes (with 1,830 coding sequences) and three rRNA operons[13]. Similar to the type-strain, the origin of replication (oriC) was determined using the position of the *mnmG* gene and GC skew, since the *dnaA* gene is positioned near the genome terminus rather than the origin, which is standard[13].

Symmetrical inversions around the terminus in the South African *O. hominis* genomes are characteristic of replication-associated structural rearrangements (RARS), which is the most prevalent type of structural variation in bacterial chromosomes[121]. These inversions are often mediated by homologous recombination, as suggested by the presence of repetitive elements at the breakpoints, and their symmetry is influenced by the stage of replication[121]. Notably, RARS are associated with altered virulence/pathogenicity phenotypes as observed in *Staphylococcus aureus* and *Streptococcus pyogenes*[122,123]. Beyond the functional role of RARS, this type of structural variation could be used in epidemiological studies to track strain evolution and complement traditional gene or nucleotide-based phylogenetic analyses. However, their utility in elucidating bacterial diversity is not yet fully explored and is limited by the availability of complete bacterial genomes.

The presence of AMR genes *VanT* and *VanY* indicate potential resistance to glycopeptides such as vancomycin. Glycopeptides are clinically used to treat infections caused by Gram-positive organisms, often through binding to the C-terminal D-alanyl-D-alanine of the lipid II bacterial cell-wall precursor, thereby preventing peptidoglycan crosslinking and inhibiting cell wall synthesis[124].
However, vancomycin resistance mechanisms typically employ multiple enzymes that alter the peptidoglycan, thereby rendering the antibiotic ineffective[125]. The *Van* genes responsible for these modifications are organized into resistance cassettes or operons, consisting of multiple *Van* genes[125]. Since screening tools only identified two *Van* genes *O. hominis*, it is unlikely that this resistance mechanism is utilized, and this annotation may not be accurate.

Nevertheless, antibiotic susceptibility testing results suggest that *O. hominis* may be susceptible, or at least weakly susceptible, to vancomycin. This finding is unexpected as *O. hominis* is a Gram-negative bacterium. Glycopeptides are generally ineffective against Gram-negative species because the molecules are too large to traverse the lipid bilayer, thus preventing access to the cell wall[126]. Therefore, Gram-negative bacteria, including *O. hominis*, are typically considered intrinsically resistant to glycopeptides. However, vancomycin susceptibility has been observed in a closely related species, *Elizabethkingia meningoseptica* (Family: *Weeksellaceae*), a Gram-negative hospital-associated pathogen[127]. It should be noted that discrepancies in vancomycin susceptibility testing have been reported for *E. meningoseptica* when using the disk-diffusion method[128]. Therefore, while vancomycin susceptibility in *O. hominis* is not entirely implausible, the antibiotic susceptibly testing results should be interpreted with caution, and alternative methods should be employed to confirm susceptibility to glycopeptides such as vancomycin.

Additionally, the *CfxA* gene, *CfxA2* variant, was identified in *O. hominis*, encoding for a CfxA β-lactamase that hydrolyses penicillins and cephalosporins. The presence of *CfxA* is usually associated with Gram-negative oral bacteria, mainly *Veillonella* and *Prevotella* spp; however, the presence of this gene does not always confer β-lactamase resistance. Kalali *et al*. (2022) observed susceptibility to amoxicillin, ampicillin, and cefixime among *CfxA*-carrying isolates, primarily *Prevotella* and *Veillonella* species from periodontal samples[129]. Conversely, Yokoyama *et al*. (2023) reported that *Prevotella* clinical isolates with the *CfxA2* gene exhibited higher MICs against ampicillin compared to negative isolates; however, MIC values for cefmetazole and meropenem were not significantly different[130]. Thus, the presence of the cfxA2 gene does not necessarily indicate resistance to β-lactam antibiotics.

Despite the identification of three AMR genes, it is possible that unexplored or novel resistance genes remain undetected in *O. hominis*. Current AMR databases are inherently limited, as they primarily rely on known resistance genes identified in other organisms, which restricts the effectiveness and accuracy of AMR detection in under-researched species. Therefore, alternative methods should be considered for screening antimicrobial resistance genes in poorly characterized organism. Furthermore, the presence of a

AMR gene is not evidence for a functional gene or the expression of an AMR phenotype and the predictive value of phenotypic resistance determined with genotypic data varies greatly[129]. Therefore, the expression of AMR genes should ideally be validated using RT-PCR. This further highlights the importance of cultured isolates as predicted AMR should be phenotypically validated. More *O. hominis* isolates should undergo antibiotic-susceptibility testing to gain a more comprehensive understanding of antibiotic resistance patterns. Additionally, a broader spectrum of antibiotics, including various penicillins and cephalosporins, should be tested against *O. hominis* to further elucidate its resistance profile.

The presence of *gld* genes dispersed throughout the genomes suggests that *O. hominis* utilises the T9SS for the export of large proteins. Additionally, Additionally, the *porV* and *sprA* genes were identified as core *O. hominis* genes in the pangenome analysis, which further supports the potential use of the T9SS in *O. hominis*[131]. The T9SS translocates large proteins, including virulence factors, across the bacterial outer membrane, where they can be secreted into the extracellular space or remain attached to the bacterial surface[63]. Therefore, this system could facilitate the export of virulence factors in *O. hominis*, such as toxins and adhesins. The T9SS is considered a pathogenicity factor as it is implicated in human oral pathogens associated with periodontitis, such as *Porphyromonas* and *Prevotella* species[132]. For example, *Porphyromonas gingivalis* uses the T9SS to export proteases, known as gingipains, that degrade host proteins, disrupt host defence mechanisms, and modulate inflammatory responses[133]. Although it is not known which specific large proteins are secreted using the T9SS in *O. hominis*, further research is required to confirm the presence of a functional T9SS and to identify its cargo proteins. Given that the T9SS is associated with virulence and host immune system modulation, it is important to investigate this system in *O. hominis*.

The RHS locus is a feature found in all *O. hominis* genomes, of approximately 45 kb in size, located near the end of the genome. Upstream of the RHS protein is a T9SS CTD, a T9SS secretion signal. The presence of the T9SS CTD suggest that this RHS gene is a T9SS effector that is secreted to the extracellular space or attached to the cell surface[131]. The RHS protein itself is composed of several

distinct domains: a sprB domain, followed by an Integrin α domain, and an insecticidal Tc toxin domain. This is followed by the highly conserved RHS domain and a variable C-terminal tip. Interestingly, conserved RHS-repeat-containing N-terminal regions with variable C-terminal regions are also found in the insecticidal Tc toxins of entomopathogenic bacteria, such as *Yersinia* spp. and *Serratia entomophilia*[134]. However, Tc-like genes have been identified in bacteria with no known insect association, such as *Fibrobacter succinogenes*, as well as in human pathogens like *Salmonella enterica* and *Yersinia pestis*[134]. Thus, it is possible that the *O. hominis* RHS gene might be a Tc-like gene, potentially encoding for a Tc toxin.

Tc toxins are pore-forming toxins that bind to target cell membranes via glycans on the host-cell surface, exerting cytotoxic effects by allowing the passage of toxic effectors into the target cell[135,136]. However, the role of Tc-like genes in bacteria without known insect associations is not well understood, though Tc toxins have been shown to induce cytotoxic effects in human cell lines[99]. Therefore, it is possible that the *O. hominis* RHS protein is host targeted. Future research could involve expressing and purifying this RHS toxin to evaluate its binding to host cells and associated cytotoxicity.

Downstream of the RHS gene are fragmentary RHS core and tip sequences. SA-OH-C1, SA-OH-C6 and SA-OH-C4, SA-OH-C5 shared identical RHS tips, respectively, while unique tips were identified in SA-OH-C2 and SA-OH-C3. Additionally, there was high sequence similarity to other displaced tips in *O. hominis* RHS clusters from Australia and the Maela cohort in Thailand. This indicates that RHS tips are conserved across *O. hominis* strains from different geographic regions, including Australia, Thailand, and South Africa. This is consistent with the RHS tip displacement mechanism originally suggested by Jackson *et al.* (2009), whereby, C-terminal tips are displaced when new tips are inserted onto the RHS gene, this explains the apparent variation in C-terminal tips[98]. Therefore, since orthologous tip types are observed across different *O. hominis* strains, this variation in RHS tips is likely due to the existence of a large pool of well-conserved tips that are exchanged between strains. In addition, significant alignments to fragmentary tips from other *Flavobacterium* species including *O. rhinotracheale*, *Flavobacterium*

filum, and *Bergeyella cardium* is consistent with further observations by Jackson *et al*. (2009), predicting that the pool of available RHS tips is slow evolving and not hypervariable, and that tip exchange is restricted to members of the same clade[98]. The dynamic exchange of RHS tips likely contributes to genetic diversity and may influence the pathogenicity and environmental adaptability of *O. hominis*. These findings imply that the RHS system plays a critical role in the adaptability and evolution of *O. hominis*.

Lipopolysaccharide biosynthesis clusters from South African *O. hominis* consists of three main LPS types. SA-OH-C3, SA-OH-C4, and SA-OH-C5 were similar to the LPS cluster of strain 22767 from the Maela cohort, and the LPS sequence is conserved across the length of the cluster for these strains. While SA-OH-C1, SA-OH-C2, and SA-OH-C6 were similar to Cambodian LPS type; however, LPS biosynthesis genes in SA-OH-C1 /SA-OH-C6 were different to that of SA-OH-C3. Genetic alterations in the LPS biosynthesis cluster can result in variations in the composition and structure of LPS, thereby impacting the function of its components. These changes could affect colonization and pathogenicity of *O. hominis* strains. The lipid A component is generally well-conserved but subject to variation[137]. Changes in acyltransferases can lead to modification of the types of acyl chains added to lipid A, affecting membrane stability[97]. Additionally, alterations in lipid A structure can influence the host immune response, potentially making it more or less immunogenic, as lipid A is recognized by host immune receptors like TLR4[97]. Changes in glycosyltransferases suggest variations in sugar composition, which can significantly affect the hypervariable O-antigen. The O-antigen can vary in size and can differ in sugar composition and arrangement[138]. Different glycosyltransferases can add or modify sugars, thereby altering length and chemical composition[138]. Increases in O-antigen length is associated with hindered complement-mediated killing, while changes in sugar composition can alter antimicrobial susceptibility[66,137]. Moreover, the presence of novel transport genes may influence how LPS components are shuttled across the inner membrane and assembled on the cell surface. Therefore, changes in transporters could affect the efficiency of LPS transport, impacting the overall assembly of LPS. In

general, LPS structure and composition variations can influence the adaptation and fitness of a particular strain, which is important in dynamic host environments such as the NP.

The presence of FSD, Hep Hag, and Fimbrillin family proteins, in the South African *O. hominis* genomes suggest a potential mechanism for host cell/surface adhesion and colonization. Adhesins are cell-adhesion molecules that are among the first bacterial molecules to interact with the host[93]. These are cell-surface proteins that facilitate the recognition and attachment to target host cells and are structurally and functionally diverse[93]. Bacterial adhesion plays a key role in persistence by allowing colonization of the host and preventing clearance, initiating signalling pathways that influence bacterial invasion, and trigger host inflammatory responses. Attachment to host cells also facilitates the secretion of effector proteins[93]. Therefore, adhesins are important virulence factors.

FSD genes are found in high numbers per genome, ranging from 17 to 20 genes, which is comparable to the type-strain genome that has 20 FSD genes[13]. Although high copy numbers of FSD genes are observed in every *O. hominis* genome, the function of the FSD-containing proteins is not known; however, these FSD genes possibly encode for adhesins. These FSD genes encode for proteins that range in size with longer FSD genes typically containing an immunoglobulin-like fold. Immunoglobulin folds are found in bacterial adhesins involved in host colonisation, supporting their role in adhesion[94]. Additionally, the β-barrel structure observed in longer FSD-containing proteins suggests a stable membrane integration, potentially facilitating adhesion to host cells or other surfaces. The membrane localization of these proteins, as predicted by InterPro, further supports their role in cell-surface interactions and adhesion. Therefore, the presence of genes encoding for FSD-containing proteins in our *O. hominis* genomes is evidence of a potential adhesion mechanisms.

Furthermore, poorly annotated FSD genes were identified using a conserved upstream phase variation repeat sequence. This sequence is not only found upstream of most the FSD genes, but also the Hep Hag and Fimbrillin genes; this is further evidence to support the potential role of FSD-containing proteins as

adhesins. Hep Hag and Fimbrillin family proteins were other adhesins identified. The Hep-Hag gene encodes for proteins with a YadA (Yersinia adhesin)-like collagen-binding domain. The YadA is an outer membrane protein encoded by a virulence plasmid in enteropathogenic *Y. enterocolitica* and *Y. pseudotuberculosis.* YadA is believed to be involved in adhesion to epithelial cells and extracellular matrix proteins, such as collagen, by forming a matrix on the bacterial outer membrane[139]. These genes enhance the virulence of *Y. enterocolitica* virulence[139,140]. A study by Holm *et al*. (2003) sought to identify *M. catarrhalis* gene products involved in adherence and found that Hag expression is important for adherence of *M. catarrhalis* to human lung and middle ear cells in vitro[141]. Fimbrillin family proteins were present in single copies in three genomes and InterPro analysis suggests that these genes are adhesins.

Overall, the presence of FSD, Hep Hag, and Fimbrillin family proteins, in the South African *O. hominis* genomes suggest a potential mechanism for host cell-surface adhesion and colonization. The variation in FSD numbers across strains and species indicate a role in adaptation to environmental niches or host environments. Core genome analysis revealed that FSD genes are not conserved between strains and are highly variable with each strain likely housing different complements of FSD genes. Furthermore, putative phase-variable regions upstream of these adhesion genes indicate adaptation to dynamic host environments, such as the NP. The presence of multiple types of adhesins is further evidence of *O. hominis* adherence to host surfaces as the use of multiple adhesins would allow for effective adhesion to host surfaces. Therefore, these adhesin genes likely play essential roles in colonisation and virulence of *O. hominis*. Future studies should confirm that the expression of these genes and characterise the functional roles of these adhesins in *in vitro* assays.

Core genome alignment and phylogenetic analysis revealed distinct clustering of isolates based on their geographic origin. Specifically, core genomes from South Africa, Australia and Thailand form separate clades. The genomes from the Maela cohort in Thailand appear more divergent than the Australian and South African core genomes, while the Australian and South African genomes are more closely related.

This suggests that the South African genomes share a closer evolutionary relationship to the Australian clade than the Maela clade from Thailand. Interestingly, two isolates, SA-OH-C1 and SA-OH-C6, are highly similar strains found in two different infants according to phylogenetic analysis. Given that these isolates were cultured from samples collected two years apart, this is unlikely to indicate recent transmission. Instead, it is possible that the infants were colonized by the same or highly similar strains around the same time, providing evidence for the persistent carriage of *O. hominis* strains. Salter *et al*. (2019) demonstrated that strains obtained from the same children at different time points have similar genomes, indicating that the same or closely related strains are carried for protracted periods of time in individuals[12,14]. Earlier oligotyping data from the Maela cohort had also shown that individuals tends to be persistently colonised by one *O. hominis* oligotype even though several strains existed within the cohort[12]. Thus, evidence form the Maela cohort indicates that individuals are colonised by a single strain of *O. hominis* that tends to persistently colonise the host. Therefore, the carriage of closely related strains in infants that were sampled two years apart is further evidence for this continuous carriage of *O. hominis* strains.

Importantly, pangenome analyses are impacted by the availability and quality of genome assemblies. It should be noted that the core genomes from the Maela strains were generated from metagenomic sequence data, while the Australian and South African core genomes are generated from genomic data from cultured isolates, which might introduce biases in the observed divergence. Furthermore, while the South African genomes are more complete as they are generated with ONT sequence data, this also increases the risks of frameshift errors, thereby impacting the annotations and consequently the pangenome analysis. In addition, only 13 core genomes from three geographic regions were included in the analysis and the evolutionary relationships between the strains are distant; thus, this phylogenetic tree may not be an accurate reflection of the evolutionary relationships across the *O. hominis* strains. A more comprehensive sampling of *O. hominis* isolates would provide better insights into the genetic diversity and evolutionary relationships of this species. Although *O. hominis* has been identified in samples from

various countries, including Kenya and Gambia, most sequence data available is based on fragmentary contigs. Thus, increasing the number of isolates cultured from diverse geographic regions will contribute to a more comprehensive understanding of the evolutionary dynamics of *O. hominis*. Furthermore, expanded sampling would enhance the robustness of phylogenetic analyses and improve our understanding of the genetic diversity within this species.

The COG functional assignment of *O. hominis* core genes provides insights into the conserved biological processes for this species. A significant proportion of the core genes were assigned to functions related to translation, ribosomal structure, and biogenesis, underscoring the importance of protein synthesis machinery in these bacteria. However, this finding is not surprising and is corroborated by other analyses of core gene function where translation/ribosomal genes and genes involved in metabolism are often significantly enriched in core genomes across various taxa [142]. In contrast, core genes involved in secondary metabolite biosynthesis and defence mechanisms are more limited and reflect more niche functioning; therefore, these are likely to be more enriched in the accessory genomes[142]. However, the high proportion of core genes with unknown functions complicates the interpretation of the COG functional analysis, impacting the accuracy of the biological contributions. Overall, the core genomes COG functional assignment provides an initial insight into the conserved biological processes in *O. hominis*. However, further studies are needed to validate the unknown genes and investigate their role in *O. hominis* biology.

Conclusions

This study aimed to isolate, characterise, and sequence the genomes of *O. hominis* from a South African birth cohort. Six *O. hominis* isolates were cultured using standard microbiological techniques and sequenced using ONT, representing the first *O. hominis* isolates successfully cultured from African samples and some of the first complete genomes for this species. Analysis of the genomes revealed key genetic features that may contribute to the virulence of *O. hominis*. Considering that *O. hominis* persistently colonises the NP and often at high relative abundances, the identification of virulence features emphasizes the need for ongoing research into the clinical significance of this species. Further research is needed to investigate the potential impact that *O. hominis* may have on the NP microbiome and the host. This could include investigating *O. hominis* interactions with respiratory pathogens, such as the pneumococcus, and commensals. Future research should also investigate the expression and function of virulence features, including the toxA toxin, the rhs protein, and adhesins such as the FSD-containing proteins. This research not only expands our understanding of *O. hominis* but also underscores the importance of investigating lesser-studied or uncharacterized species and their potential contribution to respiratory diseases and broader health outcomes.

References

1 Naghavi, M. e. a. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* **403**, 2100-2132 (2024). https://doi.org/10.1016/S0140-6736(24)00367-2

2 Bender, R. G. e. a. Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990-2021: a systematic analysis from the Global Burden of Disease Study 2021. *Lancet Infect Dis* **24**, 974-1002 (2024). https://doi.org/10.1016/S1473-3099(24)00176-2

3 Bogaert, D., De Groot, R. & Hermans, P. W. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect Dis* **4**, 144-154 (2004). https://doi.org/10.1016/S1473-3099(04)00938-7

4 Cleary, D. W. & Clarke, S. C. The nasopharyngeal microbiome. *Emerg Top Life Sci* **1**, 297-312 (2017). https://doi.org/10.1042/ETLS20170041

5 Teo, S. M. *et al.* The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host Microbe* **17**, 704-715 (2015). https://doi.org/10.1016/j.chom.2015.03.008

6 Biesbroek, G. *et al.* Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am J Respir Crit Care Med* **190**, 1283-1292 (2014). https://doi.org/10.1164/rccm.201407-1240OC

7 Xu, L., Earl, J., Bajorski, P., Gonzalez, E. & Pichichero, M. E. Nasopharyngeal microbiome analyses in otitis-prone and otitis-free children. *International Journal of Pediatric Otorhinolaryngology* **143**, 110629 (2021). https://doi.org/https://doi.org/10.1016/j.ijporl.2021.110629

8 Shak, J. R., Vidal, J. E. & Klugman, K. P. Influence of bacterial interactions on pneumococcal colonization of the nasopharynx. *Trends Microbiol* **21**, 129-135 (2013). https://doi.org/10.1016/j.tim.2012.11.005

9 Cook, L. C., LaSarre, B. & Federle, M. J. Interspecies communication among commensal and pathogenic streptococci. *mBio* **4** (2013). https://doi.org/10.1128/mBio.00382-13

10 Stevens, E. J., Bates, K. A. & King, K. C. Host microbiota can facilitate pathogen infection. *PLoS Pathog* **17**, e1009514 (2021). https://doi.org/10.1371/journal.ppat.1009514

11 Joubert, I. A., Otto, M., Strunk, T. & Currie, A. J. Look Who's Talking: Host and Pathogen Drivers of Staphylococcus epidermidis Virulence in Neonatal Sepsis. *Int J Mol Sci* **23** (2022). https://doi.org/10.3390/ijms23020860

12 Salter, S. J. *et al.* A longitudinal study of the infant nasopharyngeal microbiota: The effects of age, illness and antibiotic use in a cohort of South East Asian children. *PLOS Neglected Tropical Diseases* **11**, e0005975 (2017). https://doi.org/10.1371/journal.pntd.0005975

13 Salter, S. J., Marsh, R. L. & Parkhill, J. Complete Genome Sequence of Ornithobacterium hominis Type Strain MSHR-COH1 (ATCC TSD-185). *Microbiology Resource Announcements* **12**, e00379-00323 (2023). https://doi.org/doi:10.1128/mra.00379-23

14 Salter, S. J. *et al.* 'Candidatus Ornithobacterium hominis': insights gained from draft genomes obtained from nasopharyngeal swabs. LID - 10.1099/mgen.0.000247 [doi] LID - e000247. (2019).

15 Flynn, M. & Dooley, J. The microbiome of the nasopharynx. . *J Med Microbiol* **70(6)** (2021). https://doi.org/10.1099/jmm.0.001368

16 Kumpitsch, C., Koskinen, K., Schöpf, V. & Moissl-Eichinger, C. The microbiome of the upper respiratory tract in health and disease. *BMC Biology* **17**, 87 (2019). https://doi.org/10.1186/s12915-019-0703-z
17 Esposito, S. & Principi, N. Impact of nasopharyngeal microbiota on the development of respiratory tract diseases. *Eur J Clin Microbiol Infect Dis* **37**, 1-7 (2018). https://doi.org/10.1007/s10096-017-3076-7
18 Stearns, J. C. *et al.* Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age. *ISME J* **9**, 1246-1259 (2015). https://doi.org/10.1038/ismej.2014.250
19 Bosch, A. A. T. M. *et al.* Maturation of the Infant Respiratory Microbiota, Environmental Drivers, and Health Consequences. A Prospective Cohort Study. *American Journal of Respiratory and Critical Care Medicine* **196**, 1582-1590 (2017). https://doi.org/10.1164/rccm.201703-0554OC
20 Claassen-Weitz, S. *et al.* Succession and determinants of the early life nasopharyngeal microbiota in a South African birth cohort. *Microbiome* **11**, 127 (2023). https://doi.org/10.1186/s40168-023-01563-5
21 Biesbroek, G. *et al.* The Impact of Breastfeeding on Nasopharyngeal Microbial Communities in Infants. *American Journal of Respiratory and Critical Care Medicine* **190**, 298-308 (2014). https://doi.org/10.1164/rccm.201401-0073OC
22 Bosch, A. A. T. M. *et al.* Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. *EBioMedicine* **9**, 336-345 (2016). https://doi.org/https://doi.org/10.1016/j.ebiom.2016.05.031
23 Wahl, B. *et al.* Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000-15. *Lancet Glob Health* **6**, e744-e757 (2018). https://doi.org/10.1016/S2214-109X(18)30247-X
24 Fitzwater, S. P., Chandran, A., Santosham, M. & Johnson, H. L. The worldwide impact of the seven-valent pneumococcal conjugate vaccine. *Pediatr Infect Dis J* **31**, 501-508 (2012). https://doi.org/10.1097/INF.0b013e31824de9f6
25 Biesbroek, G. *et al.* Seven-valent pneumococcal conjugate vaccine and nasopharyngeal microbiota in healthy children. *Emerg Infect Dis* **20**, 201-210 (2014). https://doi.org/10.3201/eid2002.131220
26 Mika, M. *et al.* Influence of the pneumococcal conjugate vaccines on the temporal variation of pneumococcal carriage and the nasal microbiota in healthy infants: a longitudinal analysis of a case-control study. *Microbiome* **5**, 85 (2017). https://doi.org/10.1186/s40168-017-0302-6
27 Manenzhe, R. I. *et al.* Nasopharyngeal Carriage of Antimicrobial-Resistant Pneumococci in an Intensively Sampled South African Birth Cohort. *Frontiers in Microbiology* **10** (2019).
28 Collaro, A. J. *et al.* The effect of early childhood respiratory infections and pneumonia on lifelong lung function: a systematic review. *The Lancet Child & Adolescent Health* (2023).
29 Grimwood, K. & Chang, A. B. Long-term effects of pneumonia in young children. *Pneumonia* **6**, 101-114 (2015).
30 Mthembu, N., Ikwegbue, P., Brombacher, F. & Hadebe, S. Respiratory Viral and Bacterial Factors That Influence Early Childhood Asthma. *Front Allergy* **2**, 692841 (2021). https://doi.org/10.3389/falgy.2021.692841
31 Wensel, C. R., Pluznick, J. L., Salzberg, S. L. & Sears, C. L. Next-generation sequencing: insights to advance clinical investigations of the microbiome. *J Clin Invest* **132** (2022). https://doi.org/10.1172/JCI154944
32 Peterson, D. *et al.* Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes. *Front Microbiol* **12**, 670336 (2021). https://doi.org/10.3389/fmicb.2021.670336

33 Kelly, M. S. *et al.* Non-diphtheriae Corynebacterium species are associated with decreased risk of pneumococcal colonization during infancy. *The ISME Journal* **16**, 655-665 (2022). https://doi.org/10.1038/s41396-021-01108-4
34 Kwambana, B. A., Barer, M. R., Bottomley, C., Adegbola, R. A. & Antonio, M. Early acquisition and high nasopharyngeal co-colonisation by Streptococcus pneumoniae and three respiratory pathogens amongst Gambian new-borns and infants. *BMC Infectious Diseases* **11**, 175 (2011). https://doi.org/10.1186/1471-2334-11-175
35 Shiri, T. *et al.* Interrelationship of Streptococcus pneumoniae, Haemophilus influenzae and Staphylococcus aureus colonization within and between pneumococcal-vaccine naïve mother-child dyads. *BMC infectious diseases* **13**, 1-9 (2013).
36 Lewnard, J. A. *et al.* Epidemiological Markers for Interactions Among Streptococcus pneumoniae, Haemophilus influenzae, and Staphylococcus aureus in Upper Respiratory Tract Carriage. *The Journal of Infectious Diseases* **213**, 1596-1605 (2015). https://doi.org/10.1093/infdis/jiv761
37 Ramsey, M. M., Freire, M. O., Gabrilska, R. A., Rumbaugh, K. P. & Lemon, K. P. Staphylococcus aureus Shifts toward Commensalism in Response to Corynebacterium Species. *Frontiers in Microbiology* **7** (2016). https://doi.org/10.3389/fmicb.2016.01230
38 Xu, L., Earl, J. & Pichichero, M. E. Nasopharyngeal microbiome composition associated with Streptococcus pneumoniae colonization suggests a protective role of Corynebacterium in young children. *PLOS ONE* **16**, e0257207 (2021). https://doi.org/10.1371/journal.pone.0257207
39 Bottery, M. J. *et al.* Inter-species interactions alter antibiotic efficacy in bacterial communities. *The ISME Journal* **16**, 812-821 (2022). https://doi.org/10.1038/s41396-021-01130-6
40 Perez, A. C. *et al.* Residence of Streptococcus pneumoniae and Moraxella catarrhalis within polymicrobial biofilm promotes antibiotic resistance and bacterial persistence in vivo. *Pathogens and Disease* **70**, 280-288 (2014). https://doi.org/10.1111/2049-632X.12129
41 Armbruster Chelsie, E. *et al.* Indirect Pathogenicity of Haemophilus influenzae and Moraxella catarrhalis in Polymicrobial Otitis Media Occurs via Interspecies Quorum Signaling. *mBio* **1**, e00102-00110 (2010). https://doi.org/10.1128/mBio.00102-10
42 Xu, Q., Almudervar A Fau - Casey, J. R., Casey Jr Fau - Pichichero, M. E. & Pichichero, M. E. Nasopharyngeal bacterial interactions in children.
43 Lynch, M. D. & Neufeld, J. D. Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* **13**, 217-229 (2015). https://doi.org/10.1038/nrmicro3400
44 Lagier, J. C. *et al.* Culturing the human microbiota and culturomics. *Nat Rev Microbiol* **16**, 540-550 (2018). https://doi.org/10.1038/s41579-018-0041-0
45 Liu, S. *et al.* Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. *Microbiome* **10**, 76 (2022). https://doi.org/10.1186/s40168-022-01272-5
46 Vartoukian, S. R. Cultivation strategies for growth of uncultivated bacteria. *J Oral Biosci* **58**, 142-149 (2016). https://doi.org/10.1016/j.job.2016.08.001
47 Zengler, K. & Zaramela, L. S. The social network of microorganisms — how auxotrophies shape complex communities. *Nature Reviews Microbiology* **16**, 383-390 (2018). https://doi.org/10.1038/s41579-018-0004-5
48 Murugkar, P. *et al.* Identification of a growth factor required for culturing specific fastidious oral bacteria. *J Oral Microbiol* **15**, 2143651 (2023). https://doi.org/10.1080/20002297.2022.2143651
49 Dedysh, S. N. Describing difficult-to-culture bacteria: Taking a shortcut or investing time to discover something new? *Syst Appl Microbiol* **46**, 126439 (2023). https://doi.org/10.1016/j.syapm.2023.126439
50 Gutleben, J. *et al.* The multi-omics promise in context: from sequence to microbial isolate. *Critical Reviews in Microbiology* **44**, 212-229 (2018). https://doi.org/10.1080/1040841X.2017.1332003

51 Salter, S. J. *et al.* A longitudinal study of the infant nasopharyngeal microbiota: The effects of age, illness and antibiotic use in a cohort of South East Asian children. *PLOS Neglected Tropical Diseases* **11**, e0005975. Figure 0005971, Aggregate of all routine swabs from all infants, proportional abundance of the 0005915 OTUs that account for >0005998% of the cohort microbiota, by age in months (0005971–0005924); doi: https://doi.org/0005910.0001371/journal.pntd.0005975.g0005001. (2017). https://doi.org/10.1371/journal.pntd.0005975

52 Kwambana, B. A. *et al.* Differential effects of frozen storage on the molecular detection of bacterial taxa that inhabit the nasopharynx. *BMC Clinical Pathology* **11**, 1-8 (2011).

53 Feazel, L. M. *et al.* Effects of vaccination with 10-valent pneumococcal non-typeable Haemophilus influenza protein D conjugate vaccine (PHiD-CV) on the nasopharyngeal microbiome of Kenyan toddlers. *PloS one* **10**, e0128064 (2015).

54 Marsh, R. *et al.* The microbiota in bronchoalveolar lavage from young children with chronic lung disease includes taxa present in both the oropharynx and nasopharynx. *Microbiome* **4**, 1-18 (2016).

55 Salter, S. J. *et al.* 'Candidatus Ornithobacterium hominis': insights gained from draft genomes obtained from nasopharyngeal swabs. LID - 10.1099/mgen.0.000247 [doi] LID - e000247.

56 Lapierre, P. & Gogarten, J. P. Estimating the size of the bacterial pan-genome. *Trends in Genetics* **25**, 107-110 (2009). https://doi.org/10.1016/j.tig.2008.12.004

57 Kubatzky, K. F. Pasteurella multocida toxin - lessons learned from a mitogenic toxin.

58 Wilson, B. A. & Ho, M. in *Pasteurella multocida: Molecular Biology, Toxins and Infection* (eds Klaus Aktories, Joachim H. C. Orth, & Ben Adler) 93-111 (Springer Berlin Heidelberg, 2012).

59 Vincent, M. S. *et al.* Dynamic proton-dependent motors power type IX secretion and gliding motility in Flavobacterium. *PLOS Biology* **20**, e3001443 (2022). https://doi.org/10.1371/journal.pbio.3001443

60 Johnston Joseph, J., Shrivastava, A. & McBride Mark, J. Untangling Flavobacterium johnsoniae Gliding Motility and Protein Secretion. *Journal of Bacteriology* **200**, e00362-00317 (2017). https://doi.org/10.1128/JB.00362-17

61 Salter, S. J. *et al.* 'Candidatus Ornithobacterium hominis': insights gained from draft genomes obtained from nasopharyngeal swabs. LID - 10.1099/mgen.0.000247 [doi] LID - e000247. Figure 4, Structure of PMT C-terminal domain, and a model of the equivalent region of 'Candidatus O. hominis' ToxA. (2019).

62 McBride Mark, J. Bacteroidetes Gliding Motility and the Type IX Secretion System. *Microbiology Spectrum* **7**, 10.1128/microbiolspec.psib-0002-2018 (2019). https://doi.org/10.1128/microbiolspec.psib-0002-2018

63 Gorasia, D. G., Veith, P. D. & Reynolds, E. C. The Type IX Secretion System: Advances in Structure, Function and Organisation. *Microorganisms* **8** (2020). https://doi.org/10.3390/microorganisms8081173

64 Jamet, A. *et al.* A widespread family of polymorphic toxins encoded by temperate phages. *BMC Biology* **15**, 75 (2017). https://doi.org/10.1186/s12915-017-0415-1

65 Jamet, A. & Nassif, X. New Players in the Toxin Field: Polymorphic Toxin Systems in Bacteria. *mBio* **6**, 10.1128/mbio.00285-00215 (2015). https://doi.org/10.1128/mbio.00285-15

66 Reyes, R. E. *et al.* Mechanisms of O-antigen structural variation of bacterial lipopolysaccharide (LPS). *The complex world of polysaccharides*, 71-98 (2012).

67 Lawrence, K. A. *et al.* Method for culturing Candidatus Ornithobacterium hominis. *Journal of Microbiological Methods* **159**, 157-160. Figure 153, Comparison of Ca. Ornithobacterium hominis lipopolysaccharide biosynthesis loci; p. 159 (2019). https://doi.org/https://doi.org/10.1016/j.mimet.2019.03.006

68 Lawrence, K. A. *et al.* Method for culturing Candidatus Ornithobacterium hominis. *Journal of Microbiological Methods* **159**, 157-160 (2019). https://doi.org/https://doi.org/10.1016/j.mimet.2019.03.006

69 Marsh, R. L. *et al.* The microbiota in bronchoalveolar lavage from young children with chronic lung disease includes taxa present in both the oropharynx and nasopharynx. *Microbiome* **4**, 37 (2016). https://doi.org/10.1186/s40168-016-0182-1
70 Coleman, A. *et al.* Upper Respiratory Tract Microbiome of Australian Aboriginal and Torres Strait Islander Children in Ear and Nose Health and Disease. *Microbiol Spectr* **9**, e0036721 (2021). https://doi.org/10.1128/Spectrum.00367-21
71 Ahmed, N., Azab, M., Enany, S. & Hanora, A. Draft genome sequence of novel Candidatus Ornithobacterium hominis carrying antimicrobial resistance genes in Egypt. *BMC Microbiol* **24**, 47 (2024). https://doi.org/10.1186/s12866-023-03172-6
72 Zar, H. J., Barnett, W., Myer, L., Stein, D. J. & Nicol, M. P. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. *Thorax* **70**, 592 (2015). https://doi.org/10.1136/thoraxjnl-2014-206242
73 Zar, H. J., Barnett, W., Myer, L. & Nicol, M. P. *Childhood pneumonia – the Drakenstein Child Health Study*. Vol. 106 (2016).
74 Zar, H. J. *et al.* Klebsiella pneumoniae Lower Respiratory Tract Infection in a South African Birth Cohort: a Longitudinal Study. *Int J Infect Dis* **121**, 31-38 (2022). https://doi.org/10.1016/j.ijid.2022.04.043
75 Smith, A. C. & Hussey, M. A. Gram Stain Protocols. *American Society for Microbiology* 3-4 (2005).
76 Hudzicki, J. Kirby-Bauer Disk Diffusion Susceptibility Test Protocol. *American Society for Microbiology*, 5-18 (2009).
77 Wizard® HMW DNA Extraction Kit Technical Manual TM604. (Revised July 2022). https://doi.org/https://worldwide.promega.com/products/nucleic-acid-extraction/genomic-dna/high-molecular-weight-dna-extraction-kit/?gclid=Cj0KCQjw6cKiBhD5ARIsAKXUdyY30vW6O5mfYUezAp-qFlTSjy8SCNtKDVQSfgTgJSJHH2qvDg5ciYIaAlMkEALw_wcB&catNum=A2920&tabset0=1
78 Lee, S., Nguyen, L. T., Hayes, B. J. & Ross, E. M. Prowler: a novel trimming algorithm for Oxford Nanopore sequence data. *Bioinformatics* **37**, 3936-3937 (2021). https://doi.org/10.1093/bioinformatics/btab630
79 Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology* **37**, 540-546 (2019).
80 Schwengers, O. *et al.* Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial Genomics* **7** (2021). https://doi.org/https://doi.org/10.1099/mgen.0.000685
81 Van Tonder, A. *assembleBAC-ONT v1.1*, <https://github.com/avantonder/assembleBAC-ONT> (2023).
82 Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nature Biotechnology* **35**, 316-319 (2017). https://doi.org/10.1038/nbt.3820
83 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
84 Grant, J. R. *et al.* Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research* **51**, W484-W492 (2023). https://doi.org/10.1093/nar/gkad326
85 Tonkin-Hill, G. *et al.* Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology* **21**, 180 (2020). https://doi.org/10.1186/s13059-020-02090-4
86 Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* **38**, 3022-3027 (2021). https://doi.org/10.1093/molbev/msab120
87 Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **52**, W78-W82 (2024). https://doi.org/10.1093/nar/gkae268

88 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825-5829 (2021). https://doi.org/10.1093/molbev/msab293

89 Koetsier, G. & Cantor, E. A practical guide to analyzing nucleic acid concentration and purity with microvolume spectrophotometers. *New England Biolabs Inc*, 1-8 (2019).

90 Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Research* **51**, D418-D427 (2022). https://doi.org/10.1093/nar/gkac993

91 Madeira, F. *et al.* The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Research* **52**, W521-W525 (2024). https://doi.org/10.1093/nar/gkae241

92 Diseases, G. B. D. & Injuries, C. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204-1222 (2020). https://doi.org/10.1016/S0140-6736(20)30925-9

93 Kline, K. A., Falker, S., Dahlberg, S., Normark, S. & Henriques-Normark, B. Bacterial adhesins in host-microbe interactions. *Cell Host Microbe* **5**, 580-592 (2009). https://doi.org/10.1016/j.chom.2009.05.011

94 Chatterjee, S., Basak, A. J., Nair, A. V., Duraivelan, K. & Samanta, D. Immunoglobulin-fold containing bacterial adhesins: molecular and structural perspectives in host tissue colonization and infection. *FEMS Microbiol Lett* **368** (2021). https://doi.org/10.1093/femsle/fnaa220

95 Ren, D., Walker, A. N. & Daines, D. A. Toxin-antitoxin loci vapBC-1 and vapXD contribute to survival and virulence in nontypeable Haemophilus influenzae. *BMC Microbiology* **12**, 263 (2012). https://doi.org/10.1186/1471-2180-12-263

96 Coluzzi, C., Garcillán-Barcia, M. P., de la Cruz, F. & Rocha, E. P. C. Evolution of Plasmid Mobility: Origin and Fate of Conjugative and Nonconjugative Plasmids. *Molecular Biology and Evolution* **39**, msac115 (2022). https://doi.org/10.1093/molbev/msac115

97 Di Lorenzo, F. *et al.* A Journey from Structure to Function of Bacterial Lipopolysaccharides. *Chem Rev* **122**, 15767-15821 (2022). https://doi.org/10.1021/acs.chemrev.0c01321

98 Jackson, A. P., Thomas, G. H., Parkhill, J. & Thomson, N. R. Evolutionary diversification of an ancient gene family (rhs) through C-terminal displacement. *BMC Genomics* **10**, 584 (2009). https://doi.org/10.1186/1471-2164-10-584

99 Song, N. *et al.* N-Glycans and sulfated glycosaminoglycans contribute to the action of diverse Tc toxins on mammalian cells. *PLoS Pathog* **17**, e1009244 (2021). https://doi.org/10.1371/journal.ppat.1009244

100 Claassen-Weitz, S. a. X., Yao and Workman, Lesley and Hannan, Luke and Gardner-Lubbe, Sugnet and Mwaikono, Kilaza Samson and Harris Mounaud, Stephanie and Nierman, William C. and Africa, Samantha and Patel, Fadheela and Dube, Felix and Allen, Veronica and Ah Tow Edries, Lemese and Zar, Heather and Nicol, Mark P. *The Nasopharyngeal Microbiome in South African Children with Lower Respiratory Tract Infection: A Nested Case-Control Study of the Drakenstein Child Health Study* (SSRN, 2024).

101 Zahra, M. *et al.* Isolation and Characterization of Small-Colony Variants of Ornithobacterium rhinotracheale. *Journal of Clinical Microbiology* **51**, 3228-3236 (2020). https://doi.org/10.1128/jcm.01337-13

102 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* **428**, 726-731 (2016). https://doi.org/https://doi.org/10.1016/j.jmb.2015.11.006

103 Kubatzky, K. F. Pasteurella multocida toxin - lessons learned from a mitogenic toxin. *Front Immunol* **13**, 1058905 (2022). https://doi.org/10.3389/fimmu.2022.1058905

104 Johns, B. E., Purdy, K. J., Tucker, N. P. & Maddocks, S. E. Phenotypic and Genotypic Characteristics of Small Colony Variants and Their Role in Chronic Infection. *Microbiol Insights* **8**, 15-23 (2015). https://doi.org/10.4137/MBI.S25800

105 Proctor, R. A. *et al.* Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nature Reviews Microbiology* **4**, 295-305 (2006). https://doi.org/10.1038/nrmicro1384
106 Skosana, P. P. *et al.* A national, multicentre, web-based point prevalence survey of antimicrobial use and quality indices among hospitalised paediatric patients across South Africa. *J Glob Antimicrob Resist* **29**, 542-550 (2022). https://doi.org/10.1016/j.jgar.2021.12.003
107 Falagas, M. E. *et al.* Clinical use of intravenous polymyxin B for the treatment of patients with multidrug-resistant Gram-negative bacterial infections: An evaluation of the current evidence. *J Glob Antimicrob Resist* **24**, 342-359 (2021). https://doi.org/10.1016/j.jgar.2020.12.026
108 Trimble, M. J., Mlynarcik, P., Kolar, M. & Hancock, R. E. Polymyxin: Alternative Mechanisms of Action and Resistance. *Cold Spring Harb Perspect Med* **6** (2016). https://doi.org/10.1101/cshperspect.a025288
109 Moffatt, J. H., Harper, M. & Boyce, J. D. Mechanisms of Polymyxin Resistance. *Adv Exp Med Biol* **1145**, 55-71 (2019). https://doi.org/10.1007/978-3-030-16373-0_5
110 Chen, X. *et al.* Cell Membrane Remodeling Mediates Polymyxin B Resistance in Klebsiella pneumoniae: An Integrated Proteomics and Metabolomics Study. *Front Microbiol* **13**, 810403 (2022). https://doi.org/10.3389/fmicb.2022.810403
111 Hussein, N. H., Al-Kadmy, I. M. S., Taha, B. M. & Hussein, J. D. Mobilized colistin resistance (mcr) genes from 1 to 10: a comprehensive review. *Mol Biol Rep* **48**, 2897-2907 (2021). https://doi.org/10.1007/s11033-021-06307-y
112 Mohamed ES, H. A. a. E. E. M. Current status of multidrug resistance of Ornithobacterium rhinotracheale from avian host. *International Journal of Veterinary Science* **11**, 539-543 (2022). https://doi.org/https://doi.org/10.47278/journal.ijvs/2021.127
113 Labuschagne, Q. *et al.* COLISTIN: adult and paediatric guideline for South Africa, 2016. *Southern African Journal of Infectious Diseases* **31**, 3-7 (2016). https://doi.org/10.1080/23120053.2016.1144285
114 Shadowen, R. D. & Sciortino, C. V. Improved growth of Campylobacter pylori in a biphasic system. *Journal of Clinical Microbiology* **27**, 1744-1747 (1989). https://doi.org/10.1128/jcm.27.8.1744-1747.1989
115 Kwon, A.-R. *et al.* Structural and biochemical characterization of HP0315 from Helicobacter pylori as a VapD protein with an endoribonuclease activity. *Nucleic Acids Research* **40**, 4216-4228 (2012). https://doi.org/10.1093/nar/gkr1305
116 Delgado-Sapién, G. *et al.* Evolutionary dynamics of the vapD gene in Helicobacter pylori and its wide distribution among bacterial phyla. *SOJ Microbiology & Infectious Diseases* **8**, 1-20 (2020). https://doi.org/10.15226/sojmid/8/1/001105
117 Morales-Espinosa, R. *et al.* High expression of Helicobacter pylori VapD in both the intracellular environment and biopsies from gastric patients with severity. *PLOS ONE* **15**, e0230220 (2020). https://doi.org/10.1371/journal.pone.0230220
118 Yusefinejad, S., Gharibi, D., Khosravi, M., Mayahi, M. & Shapouri, M. R. S. A. Isolation, identification, antibiotic resistance profile and molecular analysis of Ornithobacterium rhinotracheal isolates from turkeys. *Veterinary Medicine and Science* **10**, e1490 (2024). https://doi.org/https://doi.org/10.1002/vms3.1490
119 Koren, S. & Phillippy, A. M. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol* **23**, 110-120 (2015). https://doi.org/10.1016/j.mib.2014.11.014
120 Zhang, H., Jain, C. & Aluru, S. A comprehensive evaluation of long read error correction methods. *BMC Genomics* **21**, 889 (2020). https://doi.org/10.1186/s12864-020-07227-0
121 D'Iorio, M. & Dewar, K. Replication-associated inversions are the dominant form of bacterial chromosome structural variation. *Life Sci Alliance* **6** (2023). https://doi.org/10.26508/lsa.202201434

122 Bao, Y. J. *et al.* Novel genomic rearrangements mediated by multiple genetic elements in Streptococcus pyogenes M23ND confer potential for evolutionary persistence. *Microbiology (Reading)* **162**, 1346-1359 (2016). https://doi.org/10.1099/mic.0.000326
123 Bowden Katherine, E. *et al.* Genome Structural Diversity among 31 Bordetella pertussis Isolates from Two Recent U.S. Whooping Cough Statewide Epidemics. *mSphere* **1**, 10.1128/msphere.00036-00016 (2016). https://doi.org/10.1128/msphere.00036-16
124 Blaskovich, M. A. T. *et al.* Developments in Glycopeptide Antibiotics. *ACS Infectious Diseases* **4**, 715-735 (2018). https://doi.org/10.1021/acsinfecdis.7b00258
125 Stogios, P. J. & Savchenko, A. Molecular mechanisms of vancomycin resistance. *Protein Sci* **29**, 654-669 (2020). https://doi.org/10.1002/pro.3819
126 Fernandes, M. M. *et al.* Nanotransformation of Vancomycin Overcomes the Intrinsic Resistance of Gram-Negative Bacteria. *ACS Applied Materials & Interfaces* **9**, 15022-15030 (2017). https://doi.org/10.1021/acsami.7b00217
127 Shailaja, V. V., Reddy, A. K., Alimelu, M. & Sadanand, L. N. Neonatal Meningitis by Multidrug Resistant Elizabethkingia meningosepticum Identified by 16S Ribosomal RNA Gene Sequencing. *Int J Pediatr* **2014**, 918907 (2014). https://doi.org/10.1155/2014/918907
128 Jean, S. S., Hsieh, T. C., Ning, Y. Z. & Hsueh, P. R. Role of vancomycin in the treatment of bacteraemia and meningitis caused by Elizabethkingia meningoseptica. *Int J Antimicrob Agents* **50**, 507-511 (2017). https://doi.org/10.1016/j.ijantimicag.2017.06.021
129 Kalali, N., Kadkhoda, Z., Amid, R., Ghourchian, S. & Douraghi, M. Identification of oral anaerobic bacteria and the beta-lactamase resistance genes from Iranian patients with periodontitis. *Anaerobe* **75**, 102515 (2022). https://doi.org/https://doi.org/10.1016/j.anaerobe.2022.102515
130 Yokoyama, S., Hayashi, M., Goto, T., Muto, Y. & Tanaka, K. Identification of cfxA gene variants and susceptibility patterns in β-lactamase-producing Prevotella strains. *Anaerobe* **79**, 102688 (2023). https://doi.org/https://doi.org/10.1016/j.anaerobe.2022.102688
131 Paillat, M., Lunar Silva, I., Cascales, E. & Doan, T. A journey with type IX secretion system effectors: selection, transport, processing and activities. *Microbiology* **169** (2023). https://doi.org/https://doi.org/10.1099/mic.0.001320
132 Veith, P. D., Glew, M. D., Gorasia, D. G., Cascales, E. & Reynolds, E. C. The Type IX Secretion System and Its Role in Bacterial Function and Pathogenesis. *Journal of Dental Research* **101**, 374-383 (2022). https://doi.org/10.1177/00220345211051599
133 Lunar Silva, I. & Cascales, E. Molecular Strategies Underlying Porphyromonas gingivalis Virulence. *Journal of Molecular Biology* **433**, 166836 (2021). https://doi.org/https://doi.org/10.1016/j.jmb.2021.166836
134 ffrench-Constant, R. & Waterfield, N. in *Advances in Applied Microbiology* Vol. 58 (eds Allen I. Laskin, Joan W. Bennett, Geoffrey M. Gadd, & Sima Sariaslani) 169-183 (Academic Press, 2005).
135 Roderer, D. *et al.* Glycan-dependent cell adhesion mechanism of Tc toxins. *Nat Commun* **11**, 2694 (2020). https://doi.org/10.1038/s41467-020-16536-7
136 Aleksandrova, N. A., Roche, S. G., Low, Y. S. & Landsberg, M. J. Recent insights into mechanisms of cellular toxicity and cell recognition associated with the ABC family of pore-forming toxins. *Biochem Soc Trans* **51**, 1235-1244 (2023). https://doi.org/10.1042/BST20221409
137 Bertani, B. & Ruiz, N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus* **8** (2018). https://doi.org/10.1128/ecosalplus.ESP-0001-2018
138 Whitfield, C., Williams, D. M. & Kelly, S. D. Lipopolysaccharide O-antigens-bacterial glycans made to measure. *J Biol Chem* **295**, 10593-10609 (2020). https://doi.org/10.1074/jbc.REV120.009402
139 Roggenkamp, A., Neuberger, H.-R., Flügel, A., Schmoll, T. & Heesemann, J. Substitution of two histidine residues in YadA protein of Yersinia enterocolitica abrogates collagen binding, cell

adherence and mouse virulence. *Molecular Microbiology* **16**, 1207-1219 (1995). https://doi.org/https://doi.org/10.1111/j.1365-2958.1995.tb02343.x

140 Tamm, A. *et al.* Hydrophobic domains affect the collagen-binding specificity and surface polymerization as well as the virulence potential of the YadA protein of Yersinia enterocolitica. *Molecular Microbiology* **10**, 995-1011 (1993). https://doi.org/https://doi.org/10.1111/j.1365-2958.1993.tb00971.x

141 Holm Melissa, M., Vanlerberg Serena, L., Sledjeski Darren, D. & Lafontaine Eric, R. The Hag Protein of Moraxella catarrhalis Strain O35E Is Associated with Adherence to Human Lung and Middle Ear Cells. *Infection and Immunity* **71**, 4977-4984 (2003). https://doi.org/10.1128/iai.71.9.4977-4984.2003

142 Hyun, J. C., Monk, J. M. & Palsson, B. O. Comparative pangenomics: analysis of 12 microbial pathogen pangenomes reveals conserved global structures of genetic and functional diversity. *BMC Genomics* **23**, 7 (2022). https://doi.org/10.1186/s12864-021-08223-8