Development of the respiratory microbiome and resistome in preterm infants: Shotgun metagenomic analysis of nasopharyngeal aspirate samples

Polona Rajar

Ph.D. thesis

Institute of Oral Biology

Faculty of Dentistry

University of Oslo

Norway

Series of dissertations submitted to the
Faculty of Dentistry, University of Oslo

ISBN 978-82-8327-086-0

Cover: UiO.
Print production: Graphic center, University of Oslo.

Table of Contents

Acknowledgments

It is exactly five years ago that I began my position as a Ph.D. candidate at University in Oslo. After completing my medical degree in Ljubljana (Slovenia), during which I already found several excuses to travel and study in Norway, a clinically colored Ph.D. in the field of neonatology felt like a perfect puzzle piece for my professional carrier. It also justified me living in Oslo for the foreseeable future, which was attractive from several aspects, the main being having met David one year prior.

I still remember quite vividly the first introductory Ph.D. course I took, where they showed us a slide filled with statements from previous Ph.D. candidates. One of them was singled out as a very rare occasion, a candidate stating that everything went as planned, her work was completed within the designated time, articles published, thesis defended, and there were no additional problems whatsoever. I remember thinking that surely, this is achievable, as I do not mind working long hours, I've always studied hard and completed all exams early, the research group I was a part of was multidisciplinary with several professionals, the funding was in place and the outline of the project well written.

Today, I am a lot less naïve, and have gained a tremendous amount of experience. Foremost, I would like to thank my main supervisor, Kirsti Haaland, for guiding me through these last five years. All the obstacles we have met during this time have been educational on several levels and have made me grow as a researcher, clinician and as a person. She always went the extra mile to ensure that I received the support and help I needed, and at the same time allowed me to be independent, voice my concerns, and encouraged the development of my own ideas. She always took the time to celebrate our successes and was brave enough to agree to any activity I suggested for our annual team building sessions. Her doors were always open and her phone available, and she showed great understanding and empathy for any problems I was facing, for which I'm ever grateful.

There were periods of hard work, and even harder periods of not knowing what to work with. I would like to thank all my co-mentors for helping me with their vast experience and knowledge during all ups and down. Prof. Drude M. Fugelseth, Prof. Fernanda C. Petersen, Prof. Dag Berild and Prof. Ola D. Saugstad, thank you for attending every progress meeting, and all your fruitful discussions of possible ways forward every time it felt like the progress came to a halt. Thank you for encouraging me to attend congresses and conferences, and

introducing me to your colleagues, distinguished scientists of the field. And thank you Drude, for always knowing all relevant rules and making sure we never forgot them.

Next, I would like to thank all the parents who decided to participate in our study while going through an incredibly difficult period marked by preterm birth and intensive care of their child. I must admit that the days I got to see your toddlers at the last follow up have been of the most joyful ones. Thank you also to all the doctor and nursing staff at the neonatal intensive care unit Ullevål and all other hospitals to which some infants were transferred to (Ahus, Drammen, Elverum, Kalnes, Kristiansund, Lillehammer, Rikshospitalet, Trondheim, Tønsberg, Skien, Stavanger, Ålesund).

I was very fortunate to be in daily contact with a lot of great students, doctors and researchers, that motivated and encouraged me, inspired me with their own research and listened to all my frustrations. At Ullevål, sharing an office with Sissel made every day at work fun, and seeing Marlen, Kristina, Lisa, Tonje, Anne Karin and many others always lifted my spirits. At the Institute for Oral Biology, I would like to thank Gabriela, Heidi, Helene, Krystina, Navdeep, Sushma, Achal, Roger and Sudhanshu, for your help in the laboratory and for sharing your own Ph.D. experiences with me. I also always felt home at Pediatrisk forskningsinstituttet and would like to thank Lars, Runar, Rønnaug, Maria and Maria, Monika, Nina, and Benedicte for always including me in their activities, inviting me to seminars and putting any research problems I had in a wider perspective. I would also like to thank many others who in different ways helped me along the way; Dr. Sushma Nangia and all collaborators from India, Prof. Gorm Greisen, Prof. Ketil Størdal, Prof. Roger Simm, Terje Rootwelt, Tom Stiris, Ulrik Lausten-Thomsen and Ulf R. Dahle, Additionally, I would like to thank Lin Malin Ullern and Natalia Andronova for always having an answer to all of my many questions.

Research can be all consuming. I would like to thank everyone who made sure that my life did not revolve only around research. Nina and Martina, thank you for all the visits and trips between Sweden and Norway, they were always a welcome getaway. Petra, Evita, Nika, Kaja, Julie and Daniel, and all my high school and university friends, thank you for your encouragement along the way. Marie, thank you for all the hikes and sauna trips all winter long in Oslo. I would also like to thank David's best friend Oliver, for all the dinner evenings, movie nights and tours around Oslo. Claire, Aislinn, and all other instructors at Space Pilates Oslo, your reformer workouts were the best kind of therapy for a tired brain. I would also like to thank dr. Lilijana Kornhauser Cerar and the whole team at the Neonatology Section,

Department of Perinatology, University Medical Centre Ljubljana, Slovenia, for welcoming me back in the clinical work after years abroad, and for understanding the struggles of finishing my thesis at the same time.

As much as I enjoyed and felt at home in Oslo, it has been challenging to be away from home, especially during the covid pandemic. There are no words that can describe the support, love and help I have received from my family during this time. From clipping out my favorite puzzles from the newspaper and picking me up from any airport in the middle of the night, my parents Dušan and Jolanda, my brother Kristijan and soon to be sister-in-law Maša were always there for me. I am further eternally grateful for the way my parents raised me and the routines I picked up from them. I still look up to them to this day. I would also like to thank my family overseas, Iris and Michael, who have been waiting for me to defend this thesis for a while now, so that they can finally travel to Norway.

Last, but absolutely not least, I am most grateful for my two most faithful supporters. Our poodle Pika has studied with me since the national exams in high school, throughout all my medical exams and now also through my Ph.D. And the most important of all, my partner David. Thank you for all the support and for never letting me give up. I think you were more sure that I will complete this Ph.D. than I was quite a few times along the way. David, you always made sure I remembered to eat, sleep, and take a break when needed, and sacrificed your own time and energy to help me anytime I needed help. There have been some very long days and short nights during these last five years, but you made sure that we made the most out of the free time we had.

I have been looking forward to writing this section ever since I started writing my thesis. To be able to do that in my hometown, after coming home from a shift at the neonatology department, with Pika in my lap, David working from the other side of the living room and my family not a two-flight distance away - I could not have imagined a better way of placing the last dot on my thesis.

Polona Rajar

November 2023

Abbreviations

AMR: antimicrobial resistance

ARI: acute respiratory infection

ARG: antibiotic resistance gene

BPD: bronchopulmonary dysplasia

BW: birth weight

EONS: early onset neonatal sepsis

GA: gestational age

IAP: intrapartum antibiotic prophylaxis

LONS: late onset neonatal sepsis

MOB: mode of birth

NEC: necrotizing enterocolitis

NICU: neonatal intensive care unit

NPA: nasopharyngeal aspirate

PMA: post menstrual age

RDS: respiratory distress syndrome

sEONS: suspected early onset neonatal sepsis

VLBW: very low birth weight

WMS: whole metagenome sequencing

Thesis summary

Preterm infants are often exposed to microbiota modifying factors in early life and carry an increased risk of respiratory disease beyond infancy. Long hospital stays, intensive care procedures and use of antibiotics are essential for the preterm infant's survival but may leave lasting scars on their microbiota and their long-term health. Despite high prevalence of respiratory morbidities among preterm infants, the metagenome of respiratory microbiota has mostly remained uncharted. Methodological challenges of microbiota samples containing high host DNA and low microbial biomass have made the implementation of whole metagenomic sequencing for respiratory samples very demanding.

In this thesis, I first address the methodological challenges of upper respiratory samples analysis, using nasopharyngeal aspirates of very preterm infants. Our experimental set up compared several host DNA depletion and microbial DNA extraction methods to establish a standard operating protocol for metagenomic analysis. Using our protocol, we were able to achieve species and antibiotic resistance gene (ARG) resolution after whole metagenomic sequencing of both pooled and individual patient samples.

Further, we applied the protocol on 369 nasopharyngeal aspirate samples obtained from a cohort of 66 very preterm infants, sampled from birth until six months corrected age. Microbiome composition differed from that described in term infants. Interindividual variability, followed by postnatal age exerted the most significant impact on the overall microbiome composition. We observed a persistent effect of hospitalization until six months corrected age, and a transient effect of postnatal antibiotics that diminished until discharge from the hospital and was not visible at six months corrected age.

In the last part of the thesis, we investigated the nasopharyngeal antibiotic resistance profile in a sub-cohort of 36 infants, exposed to either ampicillin plus gentamycin for suspected early onset neonatal sepsis, or not receiving any postnatal antibiotics. The resistome composition strongly correlated with microbial communities. We found high-risk ARGs in preterm infants regardless of postnatal exposure and observed an increase in abundance and diversity of ARGs after discontinuation of antibiotics, which had a significant impact on the resistome composition but diminished until discharge. Additionally, we noticed a cumulative effect of pre- and postnatal antibiotics on the resistome of infants exposed to ampicillin plus gentamycin.

Serratia species associated resistance profile showed persistence across all sampling time points and was still visible at six months corrected age.

Sammendrag

For tidlig fødte barn (premature) blir ofte eksponert for mikrobiota-modifiserende faktorer tidlig i livet og har en økt risiko for luftveissykdom utover spedbarnsalderen. Langtidsopphold på sykehus, intensivbehandlingsprosedyrer og bruk av antibiotika er avgjørende for premature barns overlevelse, men kan etterlate varige arr på normalflora (mikrobiota) og påvirker langsiktige helse. Til tross for høy forekomst av respiratorisk sykdom blant premature spedbarn vet vi veldig lite om utviklingen av respiratorisk mikrobiota i denne populasjonen. I tillegg har metodiske utfordringer med prøver som inneholder høy andel verts-DNA og lav mikrobiell biomasse gjort implementering av helgenomsekvensering for luftveisprøver svært krevende.

I denne avhandlingen undersøkte vi først metodiske utfordringer ved bruk av nasofaryngealt aspirat fra svært premature spedbarn. Vårt eksperimentelle oppsett sammenlignet flere metoder for å fjerne verts-DNA og ekstrahere mikrobielt DNA for å etablere en standard driftsprotokoll for analyse av metagenom. Ved å bruke protokollen vår klarte vi å karakterisere bakterielle arter og antibiotikaresistensgen (ARG) etter helmetagenomsekvensering av både sammenslåtte og individuelle pasientprøver.

Videre brukte vi protokollen på 369 nasofaryngeale aspiratprøver fra en kohort på 66 svært premature spedbarn, samlet fra fødselen til seks måneders korrigert alder. Mikrobiomsammensetningen skilte seg fra den beskrevet for terminbarn. Interindividuell variasjon, etterfulgt av postnatal alder hadde mest signifikant effekt på den totale mikrobiomsammensetningen. Vi observerte en vedvarende effekt av sykehusinnleggelse, i form av tilstedeværelse av en typisk sykehusmikrobe, inntil seks måneders korrigert alder. I tillegg observerte vi en forbigående effekt av postnatale antibiotika som avtok frem til utskrivning fra sykehuset og ikke var synlig ved seks måneders korrigert alder.

I den siste delen av avhandlingen undersøkte vi den nasofaryngeal antibiotikaresistensprofil i en underkohort på 36 spedbarn, enten eksponert for ampicilin pluss gentamicin for mistenkt tidlig neonatal sepsis, eller ikke eksponert for noe antibiotika postnatalt. Resistom-sammensetningen korrelerte sterkt med sammensetningen av mikrobiomet. Vi fant høyrisiko ARG hos premature spedbarn uavhengig av postnatal eksponering. Det var en økning i mengde og mangfold av ARG etter seponering av antibiotika, som hadde en betydelig innvirkning på resistomsammensetningen, men var ikke lenger synlig ved utskrivning. I tillegg la vi merke til en kumulativ effekt av pre- og postnatale antibiotika på resistomet. *Serratia*-assosiert

resistensprofil viste persistens på tvers av alle prøvetakingstidspunkter og var fortsatt synlig ved seks måneders korrigert alder.

List of papers

Paper I: Rajar P, Dhariwal A, Salvadori G, Junges R, Åmdal HA, Berild D, Fugelseth D, Saugstad OD, Lausten-Thomsen U, Greisen G, Haaland K, Petersen FC. Microbial DNA extraction of high-host content and low biomass samples: Optimized protocol for nasopharynx metagenomic studies. Front Microbiol. 2022 Dec 21;13:1038120.

Paper II: Rajar P*, Dhariwal A*, Salvadori G, Åmdal HA, Berild D, Dahle UR, Fugelseth D, Greisen G, Lausten-Thomsen U, Saugstad OD, Petersen FC, Haaland K. Impact of antibiotics and hospitalization on the nasopharyngeal microbiome in very preterm infants. *Submitted to JAMA Pediatr.*

Paper III: Dhariwal A*, Rajar P*, Salvadori G, Åmdal HA, Berild D, Dahle UR, Fugelseth D, Greisen G, Saugstad OD, Haaland K, Petersen FC. The landscape of antibiotic resistance genes in the nasopharynx of preterm infants: Prolonged signature of hospitalization and effects by antibiotics. *Submitted to Nat. Commun.*

* These authors contributed equally

Relevant papers not included in this thesis

Rajar P, Saugstad OD, Berild D, Dutta A, Greisen G, Lausten-Thomsen U, Mande SS, Nangia S, Petersen FC, Dahle UR, Haaland K. Antibiotic Stewardship in Premature Infants: A Systematic Review. Neonatology. 2020;117(6):673-686.

1. Introduction

1.1 Microbiome

Microbiome refers to a community of microorganisms residing within a specified environment (microbiota) and their collective genome (1). Human microbiome studies aim to describe the composition of microbiomes residing in various body niches, and its interactions with host physiology. Perturbations are disruptions in the microbiota that cause a shift in microbial composition, effecting its diversity and functionality, i.e. dysbiosis. Dysbiosis is a broad term used for imbalance and dysfunction of the microbiota and is associated with increased risk of disease (2). Antibiotics are recognized as a major disruptor of the microbiota ecology (3). Other known factors shaping the microbiome are diet, host genetics and the environment (4).

1.2 Microbiome development in early life

Early life is a critical period for establishing human niche specific microbiota (5, 6). Perturbations and dysbiosis during this developmental window influence immune responses and may contribute to long term health outcomes (7-9).

Despite the premise of sterile in-utero environment being challenged by findings of placental microbiome, there is lack of firm evidence for prenatal microbial colonization (10). Prenatal exposure to microbial stimuli has been established, however it has been strongly suggested that colonization and acquisition of early-life microbiota originate from mother's fecal microbiota and the environment in non-pathogenic circumstances (10, 11). While it's possible that the prenatal exposure of fetal immune system to microbes could be triggered by viable microorganisms within the fetus (12), the prevailing consensus withing the scientific community is leaning towards the likelihood that maternal microbiota derived antigens and metabolites transverse the placenta by binding to IgG complexes and thus be presented to fetal immune cells (10, 13).

Rapid colonization after birth, occurring across different body sites, is influenced by different prenatal (such as maternal health, gestational age at birth, birth mode) and postnatal factors (such as antibiotic treatment, nutrition, hospitalization) (7, 14). These factors may cause disturbances of host and microbiota symbiosis in early life, which has been implicated in development of immunological, metabolic, and developmental disorders (15-19).

The largest, most diverse, and so far the most extensively studied microbiota is the gut (20). Microbiome data from samples obtained from the gut dominate largest online data repositories, followed by data originating from samples obtained from mouth, skin and vagina (21). Additionally, most knowledge regarding infant's microbiome development comes from studies of term infants (22). Less is known about the development of microbiome in other body niches and in preterm infants.

1.3 The respiratory microbiome

Bacteria colonize the mucosa along the whole length of the respiratory tract, ranging from anterior nostrils to alveoli. The premise of a sterile lung in healthy individuals has been refuted, and respiratory microbiome involvement in different respiratory pathologies (e.g. asthma and bronchopulmonary dysplasia (BPD)) has become increasingly investigated (17, 23, 24).

The respiratory tract is anatomically divided into compartments, each harboring their own niche microbiota. The focus of this thesis is on the microbiome colonizing the nasopharynx, a distinct compartment at the aero-digestive tract junction, with communication to sinuses, middle ear, oropharynx, and lower airways.

Following birth, microbes originating from the mothers' genito-rectal, oral, and skin flora, and the environment, rapidly colonize infants' upper respiratory tract mucosa. Bacterial abundance increases exponentially during the first weeks of life, after which the first patterns of colonization become stable enough to be recognized (19, 25, 26). In healthy term infants, the initial colonizers of the nasopharynx are most often *Staphylococcus* and *Streptococcus*, followed by *Corynebacterium* and *Dolosigranulum* before being overgrown with *Moraxella* by three months after birth (27). Further, stable microbial profiles can be recognized by six months of age, influenced by factors known to shape microbial communities (mode of delivery, nutrition, antibiotic therapy, seasonality) (19, 25, 28-30).

The role of nasopharyngeal microbiota has been described in relation to infectious and non-infectious disease of upper and lower respiratory tract (31, 32). Stable microbiome of nasopharynx contributes to respiratory health, as it is habituated by commensals that prevent the thrive and intrusion of opportunistic pathogens into the neighboring areas, and protect from occurrence of acute respiratory infections (ARI), e.g. otitis media, sinusitis and bronchitis (33).

Besides respiratory infection, dysbiosis of nasopharyngeal microbiome has been described in non-infectious respiratory disease, such as asthma (24, 27).

In early life, shaping of the microbiome modulates the functionality of the immune system (34). However, the pathophysiological mechanisms and causality linking dysbiosis and the development of respiratory disease still warrant further research (24, 30, 35). Some possible pathways have emerged from research in animal models. In a neonatal mice model, lack of lung colonization led to increased risk of allergic airway inflammation that persisted also into adulthood (36). Additionally, dysbiosis of the intestinal microbiota may influence immune modulation and inflation in the airways, through crosstalk between airway and gut microbiota, i.e. the gut-lung axis (37).

1.3.1 The gut – lung axis

Respiratory and intestinal tract are both lined with mucosa and share embryological origin. In the embryonic stage (3-6 weeks), the respiratory diverticulum develops from the ventral wall of the primitive foregut endoderm (38). Further, the respiratory and intestinal tract develop to become distinct compartments anatomically and in regard to their microbiome. Similar factors influence both the respiratory and gut microbiota, and metabolites produced by the gut microbiota may modulate the immune and inflammatory responses in the airways (39). A study by Binia et al. (30) found that variation in maternal human milk oligosaccharide secretor status in predominately breastfed infants influences the risk for ARIs in infants. Notably, this risk was not found to be modulated by changes in the nasopharyngeal microbiota, even when investigated at species level (30). Their evidence suggested that human milk oligosaccharides may have exerted a protective effect through immunomodulating metabolites produced by the gut microbiota, modulating the respiratory microbiota and the risk of respiratory infections through the gut-lung axis. Grier et al. investigated the development of gut and respiratory microbiota in term and preterm infants and found that community type of one niche was highly predictive of the community type at other sites (40). Further understanding of the mechanism involved in the gut-lung axis may reveal possibilities for therapeutic intervention in prevention or treatment of respiratory disease (41, 42).

1.4 Preterm birth

Preterm birth is defined as birth before 37 weeks gestational age (GA). Preterm infants are further classified by WHO as extremely preterm (born < 28 weeks GA), very preterm (from ≥ 28 to <32 weeks GA) and moderately to late preterm (from ≥ 32 to < 37 weeks GA) (43). Each year, 15 million infants are born worldwide and one million of premature infants die due to the complications of premature birth (43). Rates of preterm birth vary across countries and range from 5 – 18% of all births (43). In Norway, rates of preterm birth have been stable at around 6% (44).

Infection and inflammation are recognized among the most dominant causes of preterm birth. Changes from beneficial towards proinflammatory gut and vaginal microbiota during pregnancy have been described as possible contributors to the etiopathogenesis by influencing the pro-inflammatory cascade that may lead to preterm birth (18). *Lactobacillus* dominated vaginal microbiome has been linked to lower risk of preterm birth. However, a shift to anaerobic strains in the vaginal microbiome, and dominance of species with known inflammatory properties (*Enterobacter, Enterococcus*) in the gut microbiota, have been associated with increased risk of preterm delivery (18). Additionally, preterm infants whose mothers were diagnosed with chorioamnionitis had less abundant *Lactobacillus* in their airway (45). *Lactobacillus* was also less abundant in the airway microbiota of preterm infants that later developed BPD (45).

Other maternal or fetal medical conditions than infection account for less than one third of preterm births (e.g. abnormal placentation, multiple pregnancies, preterm rupture of membranes) (46). There is a lack of published studies investigating the association of different indications for preterm birth with microbiota colonization patterns. In term infants, a possible effect of elective versus emergency caesarean section (after onset of labor and rupture of membranes) on the gut microbiota has been described (47, 48).

Following birth, inflammatory and infectious diseases continue to threaten preterm infants. Early and late onset neonatal sepsis (EONS, LONS); respiratory distress syndrome (RDS); necrotizing enterocolitis (NEC); retinopathy of prematurity and BPD are affected by the preterm infants' immature immune system (49). Infants' immune and inflammatory responses are also modulated by microbiome and host epigenetics (50). The effects of prematurity alone are often further augmented by perinatal factors accompanying premature birth and the

characteristics of the neonatal period, such as medication, nutrition and hospitalization, which have been shown to modulate the development of the microbiome (22, 51).

The inflammation processes behind these common pathologies of prematurity lack effective target therapies. Association between microbial communities and inflammation processes has been established in NEC, and recent reviews have shown some evidence of probiotic use for lowering the risks of NEC in very low birth weight (VLBW) infants (52, 53). Further, Gallacher et al. hypothesized that a targeted treatment towards *Ureaplasma* and *Mycoplasma* might decrease rates of chronic lung disease in preterm infants, as these genera were present in infants developing BPD and are not affected by routinely used antibiotics (54).

1.4.1 Respiratory outcomes in preterm infants

The preterm infants included in this thesis were born at GA from 28 to 32 weeks. At this age of development, the lung has completed the canicular stage and entered the saccular stage, but lung growth is far from completed (55). Surfactant production and secretion by Type II pneumocytes increases but is not yet sufficient to prevent formation of atelectasis at 32 weeks GA (55). Atelectasis and baro-volutrauma due to treatment with positive pressure ventilation increase the risk of BPD (56). The incidence of BPD has fortunately decreased with advances in neonatal care, especially due to prenatal steroids to the mother with suspected premature delivery and surfactant administration in the newborn lungs, and now occurs mostly in extremely preterm infants (born at GA < 28 weeks) (57). Nonetheless, preterm infants without clinical lung injury still carry a higher risk of developing respiratory pathologies later in life than term infants do (58, 59). Compared with term infants, preterm infants have a higher risk of developing wheezing in childhood, treatment with bronchodilatations, hospitalization due to ARI and the development of asthma (59, 60).

Globally, consequences of preterm birth and respiratory infections are the leading causes of mortality until the age of five years (61). Exploring the interaction of pre- and post-natal environmental and host factors following preterm birth on the development of respiratory microbiome may uncover possible interventions that could improve the long-term respiratory outcomes of preterm infants.

1.4.2 Respiratory microbiome of preterm infants

Compared to term infants, the upper respiratory microbiota of preterm infants has been described to have a higher heterogeneity (within-group dissimilarity), increased abundance of *Proteobacteria* phylum and decreased abundance of *Firmicute* (62). These have shown to persist through episodes of viral infections of the upper airways, indicating a role in modulation of airway immune and inflammatory responses in preterm infants (62). Further, colonization of the upper respiratory tract with potential pathogenic strains increased the risk of BPD in a cohort of infants (n=102) born at GA < 32 weeks (63). Other taxonomic changes described in the nasopharyngeal microbiome of preterm infants were less abundant *Streptococcus* compared with term infants, earlier dominance with *Moraxella* and increased presence of Gram-negative representatives (such as *Neisseria*) (62).

Studies describing the development of the respiratory microbiome in term infants found a favorable effect of vaginal birth and breastfeeding. Unfortunately, preterm infants are more often delivered by caesarian section and initially unable to breastfeed (25). Additionally, preterm infants are often exposed to antibiotics prenatally and early in life, a known perturbant of the microbiome, and long hospitalization periods which have also been shown to influence the preterm infants' microbiome and resistome composition (29, 64).

Factors effecting microbiota of preterm infants often differ from factors with significant effect on term infant's microbiota. A recent study investigating the oral microbiome of preterm infants (GA < 29 weeks) found high early inter-individual variability (65). Further, preterm infants resembled full term infants oral microbiota by three months of age (65). The study also found that the oral microbiota after birth (GA < 29 weeks) and at 36 weeks postmenstrual age (PMA) was not significantly influenced by commonly recognized factors (such as birth mode, number of antibiotic doses, feeding practice), but rather by postnatal age, which has also been recognized as a main driver of preterm infants gut microbiota, together with gestational age and antibiotic exposure (66). Nasopharyngeal microbiota of preterm infants (GA < 32 weeks) was not influenced by antibiotic exposure, while mode of delivery and center location influenced the composition significantly (54). However, these preterm infants were only followed during their first four weeks of life (54), and other studies have found contrary effect of mode of delivery on the airway microbiota (45).

Conflicting conclusion of several respiratory microbiota studies regarding specific microbiome profiles and risk of respiratory infections could in part be a result of limited taxonomic resolution achieved with 16s rRNA amplicon sequencing, the most common method used for respiratory microbiota analysis. For example, microbiota profile dominated by *Moraxella* genus has been associated with both higher and lower risk of respiratory infections, and differentiation on species level might help distinguish these and other antagonistic associations of species within the same genus (30). To this date, there are no studies describing the longitudinal development of preterm infants' respiratory microbiome at species taxonomy resolution.

1.4.3 Early life antibiotics

Worldwide, infections contribute to 15% of neonatal deaths and treatment and survival of infants often depends on effective antibiotic (67). Incidence and mortality of EONS (defined as sepsis presenting at < 72 hours after birth) are reversely correlated with GA and birth weight (BW) (68-72). With higher risk for EONS in preterm infants, antibiotic treatment is often empirically prescribed due to uncertain clinical symptoms, and the delay and uncertainty of blood culture results (73), but practices vary greatly even across comparable NICUs (74, 75). A recent population-based register study in Norway of 5,296 infants born at GA < 32 weeks showed that 77% of included infants were treated with antibiotics within first week of life (76).

Efforts towards developing risk assessment protocols suited for preterm infants, together with other antibiotic stewardship actions to reduce unnecessary antibiotic exposure in early life, are in focus (68, 73), as antibiotic treatment has been shown to increases risk of NEC, BPD and death (77-80). Furthermore, antibiotic disruptions of the microbiome in early life may contribute to carriage of antibiotic resistance genes (ARG) (66, 81, 82).

Early life antibiotic therapy may also have a long-lasting effect on the microbiome and resistome. A 21-months follow-up of preterm infants exposed to antibiotics after birth showed an enriched antibiotic resistome and prolonged carriage of multidrug resistant organisms in the gut microbiome (66). Additionally, early life antibiotic treatment may predispose infants to antibiotic resistant infections later in life (83). There is a need for studies describing the effect of early antibiotic therapy on preterm infants' respiratory microbiome and resistome.

1.5 Antimicrobial resistance

Increase in antimicrobial resistance (AMR) is one of the main threats to global health. In addition to naturally developing AMR in the environment, the process has been sped up by excessive use of antibiotics in humans and animals. Health institutions are facing infections with microorganisms resistant to all available antimicrobials. Additionally, antibiotic resistant pathogens from hospitals may be introduced into the surrounding environment and become a threat to the general population (84). Data from a recent review showed that in Europe, approximately 670,000 infections with bacteria resistant to antibiotics occur each year, of which more than 33,000 end with death (85). Worldwide, infections caused by AMR could kill up to 10 million people annually by the year 2050 (86). In the neonatal population, it has been estimated that sepsis with resistant pathogens is responsible for 214,000 deaths each year, affecting especially low- and middle-income countries (87, 88).

The highest burden of death due to infections caused by resistant pathogens was due to respiratory infections, accounting for more than 1.5 million deaths in 2019 (89). Infections of the upper respiratory tract are also one of the most common reasons for over prescription of antibiotics in primary care (90).

1.6 Metagenomics

Since the first descriptions of host microbiota in the 17th century, microbiome research has advanced rapidly in the last decades after sequenced-based identification of human microbiota was described in 1996 (16S rRNA) (91). Advancing from initial culture-based, and later PCR-based identification methods broadened the specter of identified bacteria from host derived samples.

The introduction of whole metagenome sequencing methods (WMS; also referred to as shotgun metagenomics when generated using short-read sequencing technologies) improved resolution on inter- and intra-species level, and higher sensitivity for low abundance microbes was achieved (92). WMS allows for simultaneous identification of all microbes residing in the microbiome (also viral agents and fungi), but I have decided to focus solely on bacterial representatives in this thesis.

Similar to microbiome studies, WMS offers some great advantages in detection and identification of antibiotic resistance genes compared with culture-based and targeted methods

(PCR). Resistome characterization from shotgun data gives a more complete overview of antibiotic therapy effects on ARG burden in individuals than targeted methods and can help shape treatment strategies (93). However, in samples with high host DNA content, low numbers of obtained microbial reads present a major obstacle in both microbiome and resistome analysis (94).

The Human Microbiome Project was a significant milestone in host microbiome research, as it characterized microbiome composition across several body niches including skin, mouth, and nose (95). Advances in technology have also allowed for research of microbiome communities in sites with lower biomass, such as the nasopharynx (28). However, several pitfalls of microbiota studies based on low biomass samples (sensitivity for low abundance microbes, contamination, viability of bacterial cells) have shed a light of doubt on discoveries of fetal and placental microbiome (10).

1.6.1 Challenges in low microbial biomass sample analysis

1.6.1.1 Contamination

Low biomass microbiome studies are more exposed to bias due to contamination compared with high biomass samples (96-98). Contamination can occur during sampling (microbes from surrounding tissues, health personnel and environment) and processing (laboratory equipment and personnel, reagents). A study using nasal swabs found a significant effect of extraction kit lot number on microbiome composition, underlining the importance of contamination not only from the laboratory environment, but different contamination patterns within kits from the same producers (97).

Furthermore, cross contamination can occur from other samples processed and sequenced in same batches. As levels of contamination may vary across samples, this can be falsely interpreted as microbiome changes especially in low biomass samples (96).

The increase in interest for low biomass microbiota research has revealed the need for standardization of sampling, processing, and analysis procedures (99). While standard operating procedures are available for sampling, processing and analysis of high biomass stool samples, there is a lack of such golden standards for other types of samples (100-102). Recently, the RIDE checklist has been published, to help researchers mitigate the impacts of contamination and reduce bias in WMS studies in the future (96).

1.6.1.2 High Host DNA content

Human derived microbiome samples from low biomass sites (e.g. saliva, aspirates from upper and lower airways, skin scrapes, cerebrospinal fluid) are often characterized by high host DNA content which interferes with sequencing based techniques (103). Human genome, nearly 1,000-times larger than an average bacterial genome, may quickly conceal microbial reads in sequenced samples. The low ratio between bacterial and host DNA affects the sensitivity of WMS for low abundance microbes, is less cost efficient, and makes the analysis of sequencing data more demanding (104, 105). Studies have also shown that choice of DNA extraction method influences microbiome composition results, with heavier impact on low biomass samples (106).

Nasopharyngeal aspirate (NPA) samples are, in similarity with other samples obtained from the respiratory tract, characterized by low biomass and high host DNA content, with over 90% of obtained sequencing reads aligning to human genome (107, 108). In such samples, depletion of host DNA content is a crucial part of sample processing and sequencing data analysis.

1.6.2 Host DNA depletion and microbial DNA extraction

Approaches for host DNA depletion can be divided into pre-extraction and post-extraction procedures. Pre-extraction procedures focus on selectively removing host cells (e.g. filtration, selective cell lysis) followed by extracellular DNA degradation (e.g. chemically or enzymatically) prior to lysis of bacterial cells and DNA extraction (103, 104, 109, 110). Post-extraction methods focus on selectively removing host derived DNA after DNA extraction (e.g. by targeting host specific DNA sequences) (111).

For low biomass samples, pre-extraction host DNA depletion composed of selective cell lysis and extracellular DNA degradation has so far been described as more efficient, compared to other pre-extraction (e.g. filtration) or post-extraction methods (selective removal of CpG-methylated host DNA) (94, 104, 109, 112). An overview of some commonly used pre-extraction host DNA depletion protocols is illustrated in Figure 1.

Metagenomic DNA extraction methods differ according to the method used for bacterial cell lysis (chemical, enzymatical or mechanical). After the disruption of bacterial cell walls, DNA can be isolated using organic extraction (phenol–chloroform method), inorganic extraction (salting out, proteinase K treatment), or adsorption method (silica gel membrane) (113). Low

biomass samples are more sensitive to bias introduced by different DNA extraction methods compared with high biomass samples (106).

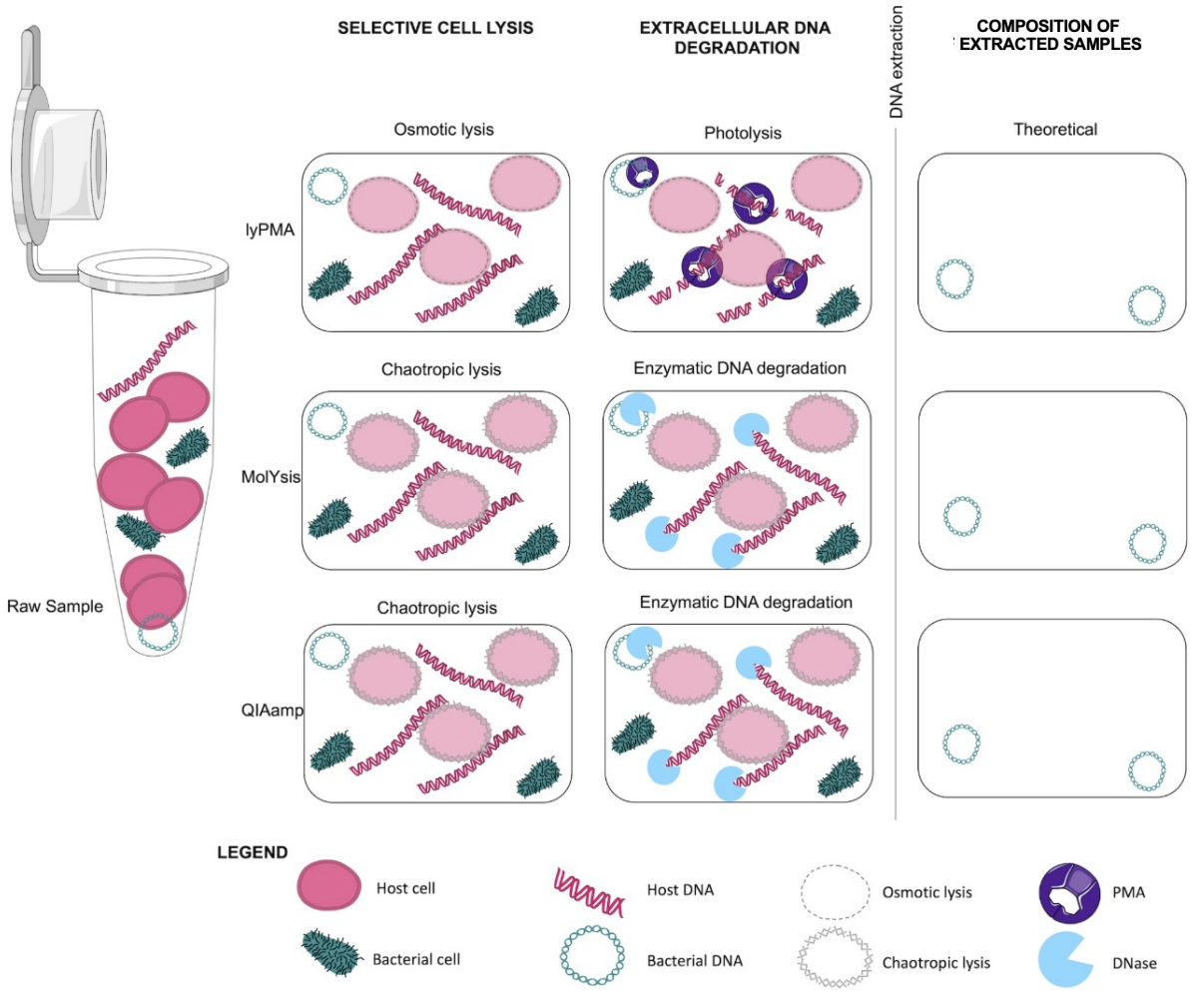

Figure 1: Host DNA depletion. Overview of the mechanism of action for commonly used pre-extraction host DNA depletion protocols. Figure is adapted from Paper I (108).

1.6.3 Sequencing depth for microbiome and resistome characterization

Even at greater sequencing depths (> 50 M reads), high host DNA content interferes with WMS sensitivity (108, 114). Relatively high cost of deep WMS and challenges of processing large volumes of data limit the number of sequencing reads (i.e. sequencing depth) that can be reasonably generated. The goal of host DNA depletion is increasing the ratio of microbial to host DNA reads, reducing the required sequencing depth needed for sufficient coverage of microbial sequences, allowing for a representation of the samples' true microbiome and resistome composition (115). Sequences that align with human genome are removed

computationally during downstream analysis as well, but as this is done after sequencing has been completed it cannot improve the sensitivity of WMS for low biomass samples.

Following host DNA depletion, host DNA content may still vary across samples, especially in samples obtained during the dynamic process of bacterial colonization in early life (108). In order to estimate needed sequencing depth prior to WMS, targeted qRT-PCR approaches estimating host and microbial DNA proportions in each sample can be used (116).

Additionally, after sequencing, a rarefactory analysis can be performed to estimate the required sequencing depth needed to characterize microbiome and resistome at various taxonomic levels (115). Richness increases with sequencing depth and ideally reaches the saturation level, observed by flattening of the curve after initial increase (asymptote). In samples where asymptote isn't reached at available sequencing depth, further increase in sequencing depth could influence richness of the sample and its composition (increasing identification of low abundance species). There is a lack of guidelines on how to analyze sample with variation in number of obtained raw and bacterial reads. Several approaches to data normalization have been described, each with their own limitations (117). Furthermore, the choice of bioinformatical tools used for metagenomic classification influence the composition results (118).

1.7 Rationale for the current thesis

Short- and long-term negative effects of antibiotic in early life are becoming increasingly more revealed. At the same time, we are losing the battle against antibiotic resistance. Preterm infants are often exposed to antibiotic perinatally and subjected to other microbiome modulating factors. The aim of this thesis was to gain insight into the development of early life respiratory microbiome and resistome, and to investigate the influences of early life antibiotics in a cohort of preterm infants using WMS. We included preterm infants born at GA 28 to 32 weeks, with an expected 50% antibiotic exposure rate. A longitudinal design was used to observe the effect of early life antibiotics during the first 8-9 months of life.

2 Aims and hypothesis

The main aim of my thesis was to investigate the development and especially the influence of early life antibiotics on nasopharyngeal microbiome and resistome in preterm infants using whole metagenomic sequencing for the analysis of upper respiratory samples.

Specific hypothesis

1. Nasopharyngeal aspirate samples have high host DNA content and low bacterial biomass. Host DNA depletion prior to metagenomic DNA extraction is required for downstream analysis of shotgun metagenomic data including taxonomic classification on species and gene (resistome) level. (Paper I)
2. The development of nasopharyngeal microbiome in preterm infants is influenced by exposure to antibiotic received in early life. Microbiome development of infants with no postnatal antibiotic treatment differs from infants receiving only early antibiotics following preterm birth, and infants receiving heterogenous antibiotic treatments between birth and six months corrected age. (Paper II)
3. Exposure to early antibiotics (combination of ampicillin plus gentamicin) initiated < 72 hours after birth due to increased risk or suspected EONS influences the composition of respiratory resistome compared to antibiotic-naive preterm infants. (Paper III)

3 Methods and subject

3.1 Born in the Twilight of Antibiotics

The “Born in the Twilight of Antibiotics” research project is a multi-center, international collaboration between several institutions located in Norway, Denmark, Sweden and India, led by the University of Oslo, with main focus on the effect of antibiotic on the respiratory microbiome and resistome development in preterm infants**.** The project has received financing from the Norwegian Research Council, the Olav Thon Foundation, University of Oslo and Oslo University Hospital.

The patient samples included in the works of this thesis were mainly collected from preterm infants included in the “Born in the Twilight of Antibiotic study”, a prospective observational cohort study conducted at the Neonatal Intensive Care Unit (NICU) at Ullevål Oslo University Hospital (OUS) (Oslo, Norway), and also some infants from Rigshospitalet (Copenhagen, Denmark). Patient inclusion, data collection and follow up appointments were performed at, or arranged by Ullevål OUS.

Optimization of downstream protocols, metagenomic DNA preparation and bioinformatic processing were conducted at the Institute of Oral Biology, University of Oslo (Oslo, Norway). Whole metagenomic sequencing using Illumina platform was conducted at the Norwegian Sequencing Centre (Oslo, Norway), a national technology platform facility hosted by the University of Oslo and supported by the “Functional Genomics” and “Infrastructure” programs of the Research Council of Norway and the South-Eastern Norway Regional Health Authorities. Overview of the work included in this thesis is illustrated in Figure 2.

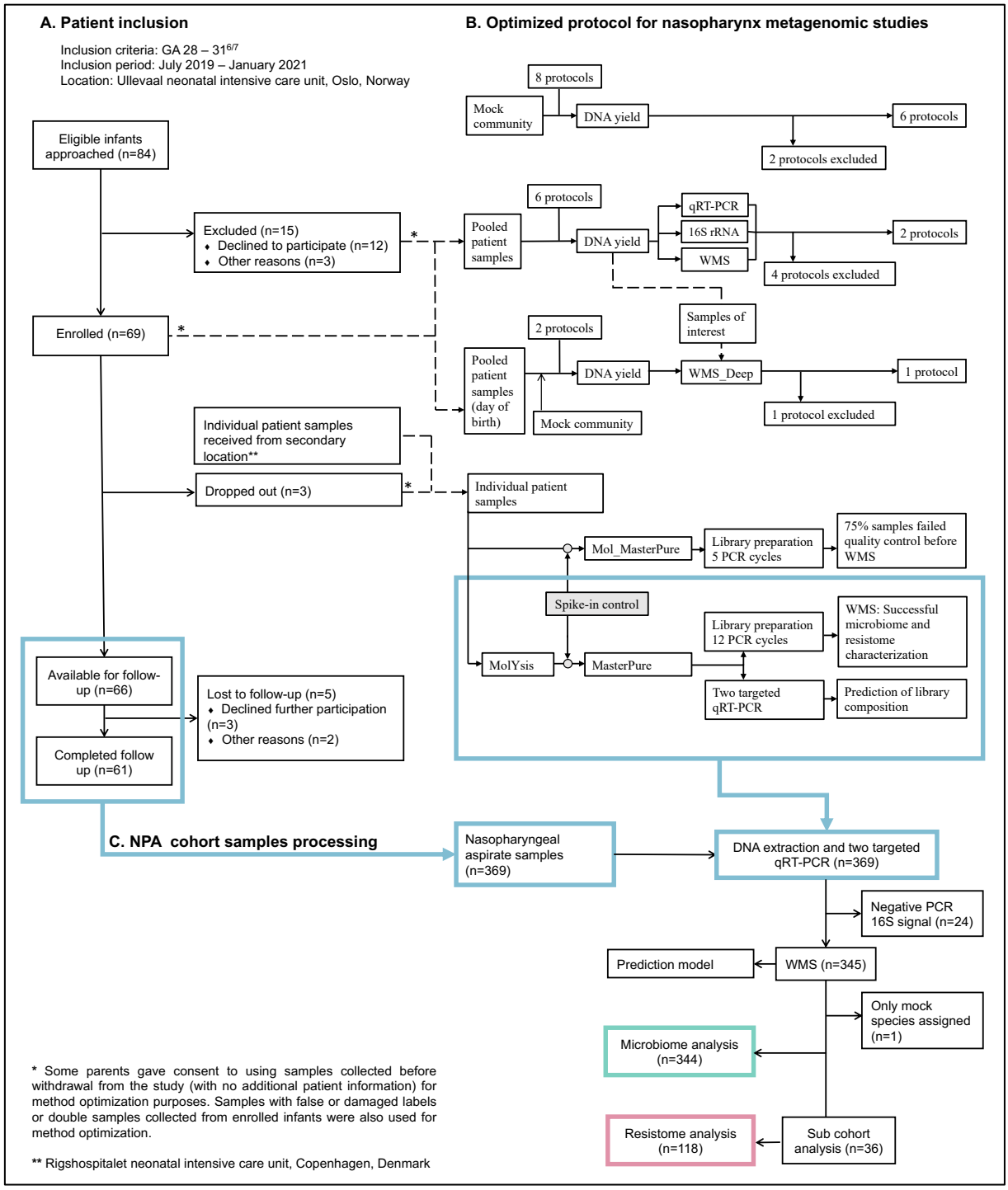

Figure 2: Overview of the work included in this thesis. A. Patient inclusion. **B.** Method optimization. **C.** Sample processing for microbiome and resistome analysis of cohort infants. Microbiome analysis (Paper II) included 344 samples from 66 infants. The resistome analysis included in this thesis and Paper III focused on a sub cohort of 181 samples obtained from 36 infants. GA: gestational age. NPA: nasopharyngeal aspirate sample. WMS: whole metagenome sequencing. ARG: antibiotic resistance gene.

3.1.1 Ethical statements

The study was performed in accordance with the Declaration of Helsinki and approved by the Oslo University Hospital's Data Protection Officer and the Regional Committee for Medical and Health Research Ethics - South East, Norway (2018/1381 REKD). Additionally, some of the samples used for the development of an optimized protocol were collected at NICU Rigshospitalet (Copenhagen, Denmark), where the study was approved by Danish National Committee for Health Research Ethics (H-180512193). Written informed consent was obtained from the infant's parents. The participants received no compensation.

3.1.2 Infant inclusion and sampling

This was a single-center longitudinal observational study at Ullevål NICU (OUS) and approached parents of all infants fulfilling inclusion criteria (preterm birth at GA between 28+0 and 31+6 weeks), born or transferred to Ullevål NICU within 48 hours after birth. Patient metadata was collected from mothers' and infants' electronic journals and stored in a secure platform (Services for Sensitive Data (TSD), University of Oslo).

Parents of all eligible infants (n=84) were approached prior to or shortly after birth between July 2019 and January 2021, and 69 infants were enrolled in the study (Figure 2A). Parents of three infants withdrew from further follow up before discharge from the NICU but consented to analysis of collected data and samples up to withdrawal date. Of the 66 infants available for the follow-up at 6 months corrected age, 61 infants completed their follow-up.

The sampling protocol is illustrated in Figure 3. Nasopharyngeal aspirate and feces samples were obtained within 48 hours of the following time points: day of life (DOL) 0 (day of birth), DOL 7, DOL 14, DOL 28, DOL 56 or discharge from the NICU, and at six months corrected age. Additional NPA and feces samples were obtained if antibiotic treatment was initiated or discontinued more than 48 hours before/after a scheduled sampling time point. Breast milk samples were collected at DOL 7 from mothers with sufficient milk production. If the samples could not be obtained at the designated DOL (e.g. unstable infant), sampling was performed as soon as feasible. Only NPA samples were analyzed for the purpose of this thesis.

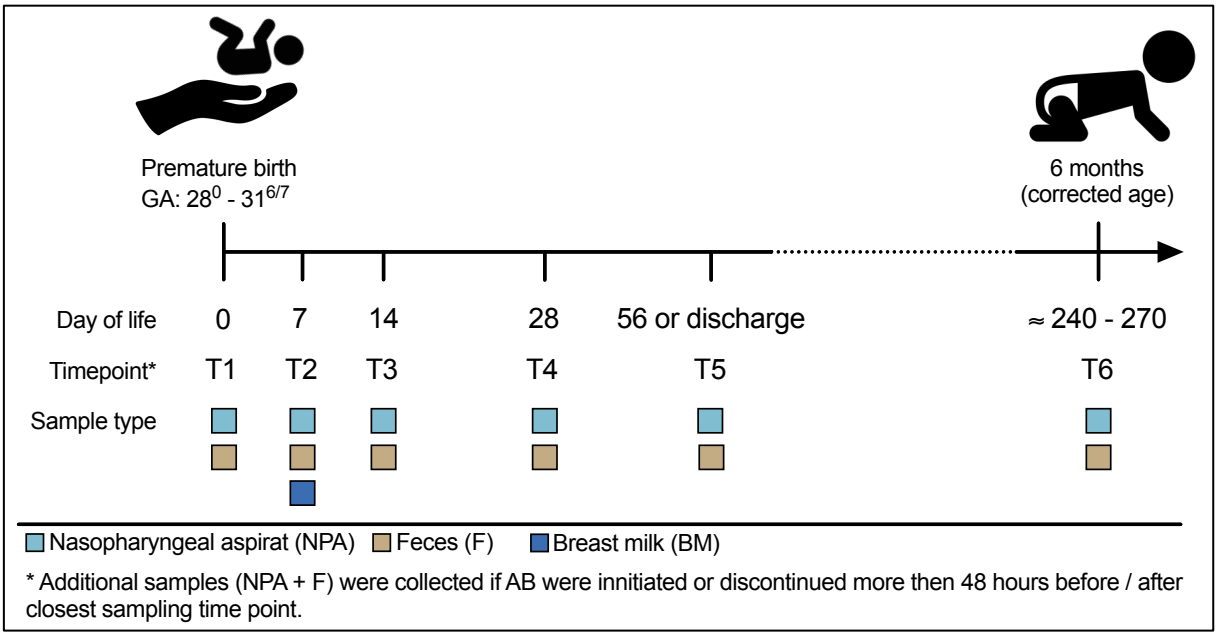

Figure 3: Sampling protocol. Preterm infants born at GA between 28+0 and 31+6 weeks were included in the study. Nasopharyngeal aspirate, feces and breast milk samples were obtained at designated time points. Additional samples were obtained within 48 hours of antibiotic treatment initiation or discontinuation.

A standardized protocol and a video tutorial for NPA sampling was created prior to the start of the study. The samples were obtained by special trained health personnel (intensive care nurses or medical doctors) using standard protection equipment to avoid contamination, and most samples were obtained by the same two researchers (PR and KH). Samples were obtained prior to feeding in infants not receiving continuous enteral tube feeding, by inserting a sterile suction catheter along the nasal wall into the nasopharynx, applying vacuum suction for 5 seconds, and removing the suction catheter without active suction. There was no pre-moisturizing of the suction catheter in infant's mouth. A sterile 2 ml 20% glycerol-saline solution was suctioned directly afterwards to rinse and recover any remaining material in the suction catheter and for cryopreservation of the sample. Infants transferred to another NICUs before discharge followed the study protocol. Sampling was performed by either the local health personnel or traveling personnel (PR) from NICU Ullevål.

At six months corrected age, infants were invited to an outpatient clinic at either Ullevål (OUS) or their local hospital, or visited at home (PR). NPA were obtained using transportable vacuum suction device, following the same procedure as described above. The samples were rapidly moved to -80°C where they were stored for up to three years before processing in the

laboratory. Samples obtained at local hospitals without a -80°C freezer or during home visits were transferred to -80°C on dry ice within 24 hours. Altogether, 369 NPA samples collected from 66 infants were available for cohort study analysis.

For the development of an optimized protocol for nasopharynx metagenomic studies (Paper I), we used samples obtained from preterm infants at Ullevål NICU (n=36) that were excluded from the main cohort study (details in Figure 2), and samples sent from Rigshospitalet (Copenhagen, Denmark) (n=6). Samples received from NICU at Rigshospitalet were obtained using the same protocol as described. They were stored at -80°C freezer and shipped to Oslo, Norway on dry ice.

3.1.3 Cohort characteristics

Of the 66 infants included in the study, 21 were not exposed to any postnatal antibiotics during the observational period (from birth until six months corrected age). Of the 45 infants exposed to postnatal antibiotics, 24 received “Only Early” antibiotics (ampicillin plus gentamicin), initiated < 72 hours after birth and discontinued after mean (SD) 2 (4) days. Furthermore, 21 infants were exposed to “Other antibiotics” and received different combinations of early antibiotics (initiated within 72 hours after birth); late antibiotics (initiated > 72 hours after birth and before discharge from NICU); and antibiotics after discharge (intravenous or per oral route). Detailed cohort characteristic comparing above groups are listed in Table 1.

Whole cohort (n=66)	**AB naive (n=21)**	**AB exposed (n=45)**	**Only early ABs (n=24)**	**Other ABs (n=21)**
Male, n (%)	12 (57%)	33 (73%)	18 (75%)	15 (71%)
GA, weeks (mean, SD)	31, 6/7	**29 4/7, 1 ****	**29 4/7, 1 ****	**29 3/7, 1 ****
BW, grams (mean, SD)	1562, 188	**1280, 262 ****	**1278, 255 ****	**1282, 275 ****
BW, group, n (%)		*	*	*
ELBW	0 (0%)	**6 (13%)**	**3 (12%)**	**3 (14%)**
VLBW	8 (38%)	**31 (69%)**	**17 (71%)**	**14 (67%)**
LBW	13 (62%)	**8 (18%)**	**4 (17%)**	**4 (19%)**
Vaginal delivery n (%)	4 (19%)	**16 (35.5%)**	**9 (37.5%)**	**8 (38%)**
Caesarian section, n (%)	17 (81%)	29 (64.5%)	15 (62.5%)	13 (62%)
ABs during pregnancy, n (%)				
None	21 (100%)	**32 (71%) ***	**13 (54%) ****	19 (90%)
< 10 days	0 (0%)	10 (22%)	10 (42%)	0 (0%)
≥10days	0 (0%)	3 (7%)	1 (4%)	2 (10%)
Intrapartum AB prophylaxis, n (%)				
None documented	2 (10%)	3 (7%)	1 (4%)	2 (10%)
Given	19 (90%)	42 (93%)	23 (96%)	19 (90%)
ROM, hours (means, SD)	3, 12	201, 491	**258, 422 ***	136, 563
Apgar at 10 min (mean, SD)	9, 1	9, 1	8, 1	9, 1
Nutrition, n (%)				
Fully breastfed	7 (39%)	13 (30%)	5 (22%)	8 (40%)
> 50% MOM	3 (17%)	11 (26%)	7 (30%)	4 (20%)
> 50% formula	6 (33%)	8 (18%)	5(22%)	3(15%)
Only formula	2 (11%)	11 (26%)	6 (26%)	5 (25%)
Lost to follow up	3 (14%)	2 (4%)	1 (4%)	1(5%)

Table 1. Cohort demographics (as published in Paper II). AB: antibiotic; GA: gestational age; BW: birth weight; LBW: low birth weight (from 1500g to < 2500g); VLBW: very low body weight (from 1000g to < 1500g); ELBW: extremely low body weight (< 1000g); ROM: rupture of membranes; MOM: mothers' own milk. * $p<0.05$, ** $p<0.001$.

3.2 Host DNA depletion and metagenomic DNA extraction

Due to lack of standard operating procedures for metagenomics analysis of low biomass samples with high host DNA, we first performed a method optimization study. Mock community samples, pooled and individual patient samples excluded from the cohort study (n=36), and additional individual patient samples received from Rigshospitalet (n=6) were included. Overview of the method optimization experiments is illustrated in Figure 2B. All tested protocols are listed in Table 2, as published in Paper I (108).

Protocol Name	Host DNA depletion kit	DNA extraction kit	Deviation from manufacturer′s protocol
MasterPure	None	MasterPure™ Gram Positive DNA Purification Kit (Epicentre, Madison, WI, USA)	Followed the manufacturer′s protocol (Available at: https://www.lucigen.com/docs/manuals/MA209E-MasterPure-Gram-Positive-DNA.pdf)
MagMAX	None	MagMAX™ Microbiome Ultra Nucleic Acid Isolation Kit (Applied Biosystems, Foster City, CA)	We followed the protocol for High throughput isolation of Nucleic Acid (RNA and DNA) from soil, biofluids, and other samples using Bead tubes and the KingFisher™ Duo Prime (Avaliable at: https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0018070_MagMAXMicrobiomeNuclAcidIsolatKit_SoilSalivaUrine_Automated_UG.pdf)
QIAamp	None	QiAmp DNA Microbiome Kit (Qiagen, Hilden, Germany)	For DNA extraction, we followed the protocol (Available at: https://www.qiagen.com/us/resources/resourcedetail?id=c403392b-0706-45ac-aa2e-4a75acd21006&lang=en), starting with step 6. (bacterial cells lysis).
PMA_MasterPure	lyPMA	MasterPure™ Gram Positive DNA Purification Kit	We followed the published method protocol for host DNA depletion with lyPMA (104). DNA extraction was performed as described in MasterPure protocol.
PMA_MagMax	lyPMA	MagMAX™ Microbiome Ultra Nucleic Acid Isolation Kit	We followed the published method protocol for host DNA depletion with lyPMA (104). DNA extraction was performed as described in MagMax protocol.
Mol_MasterPure	MolYsis™ Basic5 (Molzym, Bremen, Germany)	MasterPure™ Gram Positive DNA Purification Kit	We followed the manufacturer′s protocol (Available at: http://www.goffinmoleculartechnologies.com/wp-content/uploads/2012/01/MolYsis_Basic5_V3.0.pdf) for 1 ml samples and accordingly doubled the volume of reagents used in points 1. and 2. DNA extraction as described in MasterPure protocol.

Mol_MagMax	MolYsis™ Basic5	MagMAX™ Microbiome Ultra Nucleic Acid Isolation Kit	We followed the manufacturer's protocol (Available at: http://www.goffinmoleculartechnologies.com/wp-content/uploads/2012/01/MolYsis_Basic5_V3.0.pdf) for 1 ml samples, and accordingly doubled the volume of reagents used in steps 1. and 2. DNA extraction as described in MagMax protocol.
QIA_QIAamp	QIAamp	QIAamp DNA Microbiome Kit	We followed the protocol (Available at: https://www.qiagen.com/us/resources/resourcedetail?id=c403392b-0706-45ac-aa2e-4a75acd21006&lang=en) for 1 ml samples and accordingly doubled to volume of the reagent used in step 1.

Table 2. **Host DNA depletion and DNA extraction.** Protocols used tested for the development of an optimized protocol for nasopharynx metagenomic studies (as published in Paper I, (108)).

3.2.1 Positive and negative controls

Mock community samples used for positive control (Paper I) were prepared to match the low concentration of DNA found in nasopharyngeal aspirate samples from premature infants. We diluted 2 µl of mock community (Zymo, D6300) in 2 ml of sterile 20% glycerol saline solution, with an expected DNA yield of approximately 55 ng. The samples were placed on ice and processed immediately. All experiments with mock community samples were run in triplicates.

Spike-in control added to individual patient samples in Paper I, and all cohort samples (Paper II and III), was represented by 20 µl of Spike-in Control II Low Microbial Load (ZymoBIOMICS™, Catalog Nos. D6321 & D6321-10). The Spike-in Control standard is composed of three species not found in the human microbiome (*Truepera radiovictrix*, *Imtechella halotolerans* and *Allobacillus halotolerans*). Additionaly, six aliquots of Spike-in (20 µl, expected yield 0.4 ng) were extracted with MasterPure™ Gram Positive DNA Purification Kit (Epicentre, Madison, WI, USA) in two independent experiments and used as a positive control during analysis of cohort samples.

The cryoprotectant (2 ml sterile 20% glycerol saline solution, prepared in sterile conditions) used for clinical samples was vacuum suctioned into a sterile mucus trap at the NICU, under the same conditions as when obtaining samples from the infants, and later processed with each extraction method (Table 2), serving as a control for contamination during its production and the sampling procedure (Paper I). Five batches of glycerol saline solution were used for cohort sampling. One glycerol sample from each batch (n=5) as well as reagent blanks (n=5) were extracted and used as negative controls (Papers II and III). The negative controls were excluded from further processing due to concentrations too low to be measured with Qubits HS and < 0.1 ng 16S DNA yield according to qPCR.

Reagent controls (for each used extraction kit) were processed with each extraction method and served as controls for kit contamination (Papers II and III). The negative controls had too low concentration to be included in metagenomic libraries and were excluded from further processing.

3.2.2 Pooled and individual patient samples analyzed in Paper I

Processing of samples for the purpose of method optimization is illustrated in Figure 2B. Pooled samples were created from 18 unlabelled NPA samples, further divided into Pools A, B, C. Each pool was composed of samples from six infants and processed according to different protocols (Table 2). Six additional samples obtained within 24 hours of birth were pooled into Pool D. This experimental design was chosen so that each pool would have sufficient material to be tested with different protocols. Two aliquots from pool D were spiked with mock community (Zymo, D6300) prior to host DNA depletion and DNA extraction to create more diverse samples.

Eighteen patient samples were processed individually. They were spiked with Spike-in Control II for Low Microbial Load samples (Zymo, D6321 & D6321-10); twelve samples prior and three post host DNA depletion. Estimated DNA yield of the Spike-in was 0.4 ng.

The starting volume for all samples was approximately 2 ml, and the final DNA elution volume 30-50 µl. Samples underwent no additional freeze-thaw cycle prior to completed DNA extraction. The amount of extracted DNA was measured using Qubit™ dsDNA HS kit, on a Qubit 2.0 Fluorometer (Life Technologies, Darmstadt, Germany) and NanoDrop™ spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

3.2.1 Metagenomic DNA extraction of the cohort samples (Paper II and III)

Processing of samples obtained for the observational cohort study is illustrated in Figure 2C. Patient samples (n=369) were processed with MolYsis™ Basic5 (Molzym, Bremen, Germany) for host DNA depletion and MasterPure™ Gram Positive DNA Purification Kit (Epicentre, Madison, WI, USA) for DNA extraction, as previously described in our published protocol for handling nasopharyngeal aspirate samples from preterm infants (108). All samples belonging to the same infant were processed in the same batch.

Briefly, 2ml samples were thawed on ice and centrifuged at 10,000g for 10 minutes at 4°C. Pallets were re-suspended in 1ml PBS. Host cells were lysed with 250 µl chaotropic buffer CM (MolYsis™), vortexed at full speed for 15 seconds to mix and incubated at room temperature (+18 to +25°C) for 5 minutes. Extracellular DNA released from host cells was degraded by adding 250 µl buffer DB1 and 10 µl MolDNase B to the lysate, vortexing for 15 seconds and incubated at room temperature for 15 minutes. Lysate was centrifuged at ≥ 12,000 G for 10 minutes. Supernatant was removed and discarded, and the sediment was resuspended in 1 ml buffer RS. To remove residual MolDNase B activity, chaotropic salts and PCR inhibitors, samples were centrifuged at 10,000 G for 10 minutes at 4°C and supernatant was discarded. Extraction with MasterPure Gram Positive DNA Purification Kit (Epicentre, Madison, WI, USA) followed immediately after host DNA depletion. 150 µl of TE buffer was added to the pallet before the samples were spiked with 20 µl of Spike-in standard samples (Zymo, D6321 & D6321-10). Bacterial cells were lysed with 1µl Ready-Lyse Lysozyme during 30-minute incubation at 37°C. Next, 1 µl of Proteinase K was diluted in Gram Positive Lysis Solution and added to the sample, followed by 15 minutes incubation at 65°C while briefly vortexing the sample every 5 minutes, 5 minutes at 37°C and 5 minutes on ice. Debris were precipitated by adding 175 µl of MPC Protein Precipitation Reagent to the lysate, followed by vortexing and centrifugation at 13,000 G for 10 minutes at 4°C. Clean supernatant was transferred to High Speed Eppendorf tubes and incubated with 1µl RNase at 37°C for 30 minutes to degrade RNA. Next, 500 µl of isopropanol was added to the lysate. Tubes were inverted 40 times and centrifuged at 10,000 G for 10 minutes at 4°C to pellet the DNA. Supernatant was removed and pellets were washed with 70% ethanol and further air dried for 20-30 minutes. Finally, the DNA was resuspended in 35 µl elution buffer. DNA was quantified using Qubit HS and stored at -80°C for up to 4 weeks before further processing.

3.2.1.1 Real time PCR

Human and bacterial DNA in aliquots from patient pools A, B and C was quantified using the primer pair FP1065 5’ GCCCGTTCAGTCTCTTCGATT and FP1066 5’ CAAGGCAAAGCGAAATTGGT for the RPL30 gene, and bacterial DNA using the 16S rRNA universal primers FP1067 5’ CCATGAAGTCGGAATCGCTAG and FP 1068 5’ GCTTGACGGGCGGTGT (119)

Femto™ Quantification kits for host and bacterial DNA (Zymo, E2005 and E2006) were used according to the manufacturer's instructions for four individual patient samples and all cohort samples (n=369), according to the manufacturer's instructions. Used sample volume was 1 µl. Samples belonging to the same infant were analyzed on the same plate.

From the 16S rRNA qPCR results, the average yield of bacterial DNA from the Spike-in mock was deducted from bacterial DNA yields of the patient samples to determine samples' own bacterial load. Samples with negative bacterial (16S rRNA) yield according to qPCR were excluded from further processing (n=24).

3.2.2 Metagenomics library preparation

In paper I, samples with undetectable DNA concentration (Qubit™ dsDNA HS kit) were excluded prior to library preparation. Nextera DNA Flex Library Prep Kit (Illumina Inc., San Diego, CA, USA) was used for library preparation, following manufacturer's protocol. The only deviation was initially using five PCR amplification cycles for all library preparations (against producers' recommendations of 12 PCR cycles for low DNA input), to reduce possible PCR amplification bias (120) and enable comparison between samples. However, individual patient samples retrieved very low DNA yields and only three of the first 12 samples passed quality control and were sequenced. To optimize this step, we increased the PCR cycle number to 12 for six additional samples and used DNA input comparable to the DNA yield of the first 12 samples (6 ng). Library concentration and purity were measured with Qubit™ dsDNA HS kit on a Qubit 2.0 Fluorometer (Life Technologies, Darmstadt, Germany), NanoDrop™ spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA).

We used the same kit, Nextera Flex kit (Illumina Inc., San Diego, CA, USA) also for library preparation of cohort samples, according to manufacturer's instructions. Input yield was up to 10ng DNA according to Qubit and PCR results. Twelve PCR amplification cycles were used as per producer's recommendations. In efforts to reduce bias, all libraries were prepared by the same individual (PR), one PCR machine was used for amplification of all samples, and all samples belonging to the same infant were processed together. Library concentration and purity were measure with Qubit™ dsDNA HS kit on a Qubit 2.0 Fluorometer (Life Technologies, Darmstadt, Germany) and Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA).

3.2.3 Whole metagenome sequencing

WMS of samples included in Paper I was run on an NovaSeq SP platform (Illumina Inc., San Diego, CA, USA) using a paired-end sequencing approach with a targeted read length of 125 bp in high-output mode. The computations were performed on resources provided by Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.

Also WMS of cohort samples was conducted at the Norwegian Sequencing Centre (Oslo, Norway), on an NovaSeq S4 platform (Illumina Inc., San Diego, CA, USA) using a paired-end sequencing approach with a targeted read length of 150 base pairs in high-output mode. Libraries of all (n=345) samples proceeded to WMS in one pool with 5 nM concentration.

3.2.1 Bioinformatic processing

In Paper I, quality of raw reads was assessed using FASTQC (121). Adapter sequences and low-quality reads were removed with Trimmomantic (122). Thereafter, filtered quality reads were aligned to human reference genome using Bowtie2 (123) in order to remove human DNA contamination. The remaining high quality, clean reads were used for microbiome and resistome profiling. Microbiota profiling was done with MetaPhlAn3 (124). For the resistome analysis, the quality-filtered, clean reads were provided as input to Bowtie2 (123) alignment using default parameters to the ResFinder database (125). Reads were assigned to ARGs using an 80% gene coverage/fraction threshold. Counts of reads aligned to the ARGs were then used for downstream comparative analyses.

A similar pipeline was used for the analysis of cohort samples. Raw metagenomics sequencing data was processed inside a secure platform provided by Service for sensitive data (TSD, University of Oslo, Oslo, Norway). FASTQC (v.0.11.9) was used to assess the quality of raw and clean reads (121). Quality and adapter trimming, as well as quality control were performed with Trim galore (v.0.6.1)(126). Further, all reads were aligned with human reference genome (GRCh38) and human DNA sequences were removed using Bowtie2 (v.2.3.4.2)(123).

3.2.1.1 Rarefaction analysis

We performed rarefaction analysis to estimate the required sequencing depth needed to characterize microbiome and resistome at various taxonomic levels (Paper I). Seqtk tool (127) was used to sample clean reads into subsamples at various depths (10, 25, 30 50, 75% etc.), followed by taxonomic profiling using MetaPhlAn3 (124) to report the number of species

present within each subsample. RarefactionAnalyzer tool of the AMRPlusPlus pipeline (128) was used with 5% subsampling increments of the read data with 10 iterations at each level for resistome rarefaction analysis. The numbers of unique species, genes, mechanisms, and classes were plotted as a function of sampling depth using the ggplot2 package of R (129).

3.2.2 Microbiome profiling

MetaPhlAn 3.0 (124) was used for taxonomic microbiome profiling of the remaining clean, high-quality reads (default parameter: *q=0.05*), using a customized database, as we had to add the mock species. For the mock species detection, a custom parameter (*q=0.05*) was used. One sample was excluded before analysis as no reads were assigned to other than Spike-in species (Figure 2B).

3.2.3 Resistome profiling

Cleaned high-quality reads were mapped against the *nucleotide_fasta_protein_homolog_model* from the Comprehensive Antibiotic Resistance Database (CARD) (v.3.2.2) (130) using Bowtie2. The mapped reads from each sample were then filtered, sorted, and indexed using SAMtools. The number of reads mapped to each ARG was calculated using SAMtools *idxstats* and BEDTools, and only ARGs with a coverage of at least 80% were used in further downstream analyses. The mapped read counts were normalized by calculating reads per kilobase of reference gene per million bacterial reads (RPKM). Rarefaction analysis was performed using Rarefaction Analyzer (128) to determine the saturation of samples at various sequencing depths for recovery of ARGs. Further details regarding resistome profiling are described in Paper III (131).

3.3 Statistical analysis and data visualization

Values of 16S rRNA qPCR yield were log transformed prior to statistical analysis. Statistical analysis for cohort demographics and PCR results was performed in STATA SE 17, using unpaired t test, Pearson's chi-squared, Fisher's exact test (α=0.05) and one-way ANOVA (α=0.001).

Downstream analysis of WMS data was conducted in R (v.4.2.1) within RStudio (132, 133). Diversity analysis (α- and β-diversity) was performed using the vegan (v.2.5.7) (134) and phyloseq (v.1.34.0) (135) packages. Changes in α-diversity (Shannon index) were evaluated with Linear mixed-effects (LME), with individual as random factor and controlled for age (day

of life). β-diversity was performed on centered log-ratio (CLR) transformed species abundance data. Univariate and multivariate permutational multivariate analysis of variance (PERMANOVA) and principal component ordination analysis (PCA) were used to explore associations between metadata variables and the dispersion of samples' microbiome compositions. Statistical significance of β-diversity was tested with PERMANOVA, using adonis2 function (vegan R package) (136). Significance level was determined at $\alpha=0.001$, and $\alpha=0.05$ for the cross-sectional analysis. P values were corrected for false discovery rate where appropriate using Benjamini–Hochberg method. Community types were identified using Dirichlet multinomial mixture models approach with DirichletMultinomial (v.1.38.0) package (137). Kruskal–Wallis H test ($\alpha=0.05$) was used to investigate association of community type with sampling time point and antibiotic exposure. Figures were created using ggplot2 (v.3.3.5) R package (138) and GraphPad Prism 9.4.1 (458) © GraphPad Software, LLC (Boston, MA, USA).

3.4 Data Availability

The datasets generated during the development of the optimized protocol for nasopharynx metagenomic studies (Paper I) can be found in the BioProject (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA876384), accession number: PRJNA876384. The raw sequencing data (after removal of human DNA) for all cohort study samples (Papers II and III) are publicly available at NCBI SRA under BioProject ID: PRJNA1009231.

3.5 Ethical considerations

We have obtained the relevant ethical approval for our study. Despite its observational nature, with no modifications of infants' care and treatment introduced, ethical considerations regarding research involving preterm infants are extensive and should be addressed.

3.5.1 Sampling prior to consent

Due to the nature of our study and the importance of obtaining the first samples immediately after birth, we were granted permission by the Ethical Committee to obtain the first samples (feces and NPA) before obtaining informed consent from the parents in specific situations (e.g. mother admitted to ICU). We made great effort to talk with all parents already before birth, if they were admitted to the obstetrics department at eligible GA and expecting to deliver in the

following days. Some mothers were hospitalized immediately before or after preterm birth, and we attempted to speak with all regarding our study as soon as possible. In cases where samples were obtained prior to receiving informed consent, parents were notified of this as soon as it was possible to inform them of the study. None of the parents reacted negatively to this practise after they were informed, some were also relieved that the first samples were already taken, that it went without any problems, and could see that our study is manageable for the infant.

3.5.2 Patient harm and additional strain

We collected two different sample types from the infant (feces, NPA) and one from the mother (breast milk). Feces was obtained from the diaper and mothers' milk was obtained only if the mother had more than enough milk supply for the infant. I believe that this carried no harm or additional strain for the infants or their mothers.

Obtaining NPA samples was more complicated. We decided on using nasopharyngeal aspirates (and not swabs) also because preterm infants are often suctioned from mouth and nose during standard care, which would make sampling with a suction catheter more practically feasible. We tried to coordinate samples with times when infants needed aspiration due to mucus obstructing their breathing. The procedure is not regarded as invasive or harmful, and does not carry risk for the infant, but it can be bothersome for some, while others have managed to sleep through it. In some infants, respiratory support had to be temporarily removed during sampling. This was done for a very brief period (less than one minute), while the infant was closely monitored. We did not record any complications or negative side effects occurring from NPA sampling. Two families decided to drop out of the study as they felt that this mode of sampling is too much of a burden for the infant.

4 Summary of results / Summary of the papers

4.1 Microbial DNA extraction of high-host content and low biomass samples: Optimized protocol for nasopharynx metagenomic studies (Paper I)

We devised an optimal protocol for processing nasopharyngeal aspirate samples from preterm infants prior to whole metagenome sequencing, with focus on host DNA depletion, taxonomic profiling at species level and assignment of antibiotic resistance genes.

Three host DNA depletion and three DNA extraction protocols were compared, using RT-PCR and WMS. Protocols were initially tested using mock communities, followed by both pooled and individual patient samples.

Only MasterPure™ retrieved the expected DNA yield from mock community samples, which were adjusted to represent low biomass samples. This extraction protocol is based on a lytic method to improve Gram positive recovery without bead beating steps, which may cause loss of DNA yield.

Host DNA content in NPA samples prior to host DNA depletion was 99%. Additionally, we observed a large variation in total DNA yield, host DNA content, species and ARG richness in pooled and individual samples.

MolYsis™ host DNA depletion protocol, composed of selective cell lysis and enzymatic extracellular DNA degradation, achieved a reduction in host DNA content compatible with microbiome and resistome characterization after WMS. PCR results were indicative of achieved microbial enrichment.

Nasopharyngeal aspirate samples had high host DNA content and low biomass. By applying a protocol combining depletion with MolYsis™ and extraction with MasterPure™ for the purpose of whole metagenomic sequencing, we were able to achieve bacterial species and antibiotic resistance gene resolution.

4.2 Impact of antibiotics and hospitalization on the nasopharyngeal microbiome in very preterm infants (Paper II)

We performed a prospective observational cohort study including very preterm infants admitted to Oslo University Hospital, Ullevål. Sixty-six infants were included during a 19-month inclusion period, and 61 infants completed their follow up at six months corrected age. The cohort was divided in three groups according to postnatal antibiotic exposure. Twenty-one infants were not exposed to any postnatal antibiotics and had significantly higher gestational age and birth weight compared to antibiotic exposed infants. Twenty-four infants were exposed to ampicillin plus gentamycin, initiated immediately after birth due to increased risk of early onset neonatal sepsis, and 21 infants were exposed to other than only early antibiotics.

In total, 369 nasopharyngeal aspirate samples collected longitudinally at six sampling time points, ranging from birth until six months corrected age, were processed according to previously published protocol (Paper I). We obtained on average 6M bacterial reads from 344 nasopharyngeal aspirate samples included in microbiome analysis.

Six distinct nasopharyngeal microbial community types were determined by abundances of *Cutibacterium, Gemella, Serratia, Streptococcus* and *Staphylococcus*. Microbiome composition was strongly influenced by individual variation, followed by postmenstrual age. Postnatal antibiotic exposure exhibited a significant, but transient effect on the microbiome composition, that diminished before discharge from the hospital. The effect of hospitalization, represented by the abundance of *Serratia* endured beyond hospital discharge and was still observed at 8-9 months after birth.

4.3 Unravel the landscape of antibiotic resistance determinants in the nasopharynx and the impact of antibiotics: a longitudinal study of preterm infants (Paper III)

Preterm infants are often empirically exposed to antibiotics due to increased risk of early onset sepsis. We investigated the effect of ampicillin plus gentamycin, initiated immediately after birth, on a sub-cohort of 36 infants. The resistome composition of 15 antibiotic exposed infants was compared to the 21 infants not exposed to any postnatal antibiotics at six sampling time points, ranging from birth until six months corrected age. None of the blood cultures obtained prior to initiation of antibiotics were positive.

Antibiotic resistance genes, including high-risk ARGs were found in nearly all samples regardless of postnatal antibiotic exposure. We identified a core resistome composed of 13 ARGs that were highly prevalent and abundant across all samples. The resistome and microbiome composition were strongly correlated. Other variables significantly associated with overall resistome composition were individual variability and age.

We observed a transient increase in diversity and total abundance of ARGs, together with significant changes in the resistome composition after discontinuation of ampicillin plus gentamycin. These changes were even more profound in antibiotic exposed infants whose mothers also received antibiotic during pregnancy. Following termination of antibiotic treatment, no differences between the infant groups were observed at later sampling time points. Lastly, ARGs associated with *Serratia* species showed persistence from the first sampling time points until six months corrected age in 20 of 22 infants, regardless of antibiotic exposure.

5 General discussion

We performed an observational cohort study investigating the development of upper respiratory microbiome and resistome in very preterm infants. Nasopharyngeal aspirates of preterm infants had a high host DNA content and low biomass, warranting method optimization steps prior to WMS. Of the investigated protocols, Mol_MasterPure (composed of host DNA depletion with MolYsis™ Basic5 and DNA extraction with MasterPure™ Gram Positive DNA Purification Kit) produced most promising results allowing for downstream analysis of WMS data including species and ARG taxonomic resolution, in both pooled and individual patient samples. The protocol was further applied to all nasopharyngeal aspirate samples collected for the observational cohort study.

The composition of nasopharyngeal microbiota in preterm infants was diverse. *C. acnes, S. mitis* and *S. epidermidis* were the most prevalent species across all our samples. We recognized five community types, determined by various abundances of *Cutibacterium*, *Gemella, Serratia, Streptococcus* and *Staphylococcus* genera. Infants not exposed to any postnatal antibiotics differed from infants exposed to "Only Early" and "Other" postnatal antibiotics in absolute bacterial abundance and microbiome composition. While these effects were transient, the strongest significant effect influencing the microbiome composition between birth and six months corrected age was individual variation, followed by postmenstrual age. Additionally, long term presence of *S. marcescens* was detected in the nasopharyngeal microbiota of infants hospitalized after a presumably brief outbreak, underlining the significant effect of hospital environment on preterm infants' microbiota development.

Further, we looked more closely at the effect of ampicillin plus gentamycin for suspected EONS on the abundance and diversity of antibiotic resistant genes. Prevalence of antibiotic resistance genes was high also in samples obtained from infants not exposed to any postnatal antibiotics. The most significant effect on the resistome composition was dictated by the microbiome and followed by individual. After discontinuation of early antibiotics, we observed a significant, but short-lived increase in diversity and total abundance of antibiotic resistance genes. Additionally, a significant temporary effect of prenatal maternal antibiotic exposure was observed as a subgroup within the ampicillin plus gentamycin exposed group.

Despite the transient perturbation of antibiotics on microbiome and resistome, any aberrant changes in early life may exhibit long term consequence on respiratory health and increase the risk of antibiotic resistance genes carriage.

5.1 Methodological consideration

Low biomass and high host DNA content WMS microbiota studies are subjected to several confounding factors, mainly due to contamination and low sensitivity for sparse microbial sequences (94, 101). There are increasing efforts for implementation of general standards and criteria for low biomass WMS studies (96), which we adhered to as described below. While our methodology protocol was thoroughly investigated and published (108), there are still some limitations that need to be considered.

5.1.1 Experimental design

Our study was a prospective observational cohort study designed to include a cohort in which approximately half of the infants would be exposed to postnatal antibiotics after birth. The golden standard to compare an effect of an intervention (postnatal antibiotics) would be a randomized controlled trial, but this would not be ethically acceptable in our settings. We were able to include 79% of eligible infants and 92% of included infants completed the whole follow-up period. Exposure to postnatal antibiotics was higher than expected, as 67% of included infants received postnatal antibiotics. Infants that received postnatal antibiotics were (as expected) significantly more premature and smaller compared with antibiotic naïve infants. None of the antibiotic naïve infants were born at GA < 29 weeks or with extremely low birth weight, and none of the mothers of antibiotic naïve infants were exposed to any maternal antibiotics.

We corrected for this possible source of bias by adjusting for variables that significantly differed between the cohorts and had a significant influence on the microbiome composition in the diversity analyses.

Our longitudinal design extended until six months corrected age (8-9 months chronological age) to observe the development of the microbiome beyond the period of hospitalization and up until the age where stable microbiome profiles have been described in term infants (25).

Observational studies are subjected to several sources of possible bias. We attempted to reduce selection bias by approaching all available infants fitting the inclusion criteria across a period of 19 months. However, we did notice that parents of non-Norwegian descent were more likely to decline participation in the study. Further, we tried to minimize the effect of several environmental factors that may influence the nasopharyngeal microbiota. For most of all sampling time points, infants were hospitalized in a closed environment (NICU). Infants did also not start day care until the last sampling time point and followed the same vaccination program (35, 139).

Two special sets of circumstances occurred during our observation period. Seven months after initiation of our study, the COVID-19 pandemic reached Oslo. Day-care centers for any older siblings of infants included in the study were closed, and families also minimized contact with non-household members. Different regulations during the rest of the study period resulted in visitation restriction at the NICU and obligatory use of face masks by the health staff. Secondly, a *S. marcescens* outbreak occurred at the NICU during our observation period, which had a significant effect on the microbiome composition.

Samples used in the method optimization study (Paper I) were collected from the same Norwegian cohort of infants, with some additional samples received from a second location (Rigshospitalet, Denmark). To compare the efficiency of tested protocols on host DNA depletion and DNA extraction, samples were pooled to minimize the effect of samples' individual variability. The number of parallel runs for each protocol was limited by the number of samples we had available for method optimization purposes. No meaningful statistical tests could be performed due to the large variation between patient pools and the small number of samples processed with each protocol sent to WMS. Further, we had no patient metadata available for these samples and have perhaps not included samples representative of our whole cohort (regarding antibiotic exposure and age of sampling). Including a larger number of samples with known metadata would allow us to statistically compare the efficiency of different protocols and produce more sturdy results.

Additionally, despite a large database of collected variables, the size of our cohort limited the number of variables that could be meaningfully included in the statistical analysis. We included variables previously described to have the largest effect on the microbiome and resistome composition. Other recorded variables and possible unrecorded confounders might be significantly correlated with microbiome composition and health outcomes.

5.1.2 Sampling and infant health data collection

Information and measurement bias were minimized by using a limited number of personnel for both sampling, patient data collection and analysis. Two researchers (PR and KH) collected the majority of all NPA samples, even of infants transferred to a different hospital. Infants' health data was collected from electronic journal records (registered by nurses and physicians). One researcher (PR) collected all infant health data. Personnel at the department was aware of the ongoing study, but since there was no change in clinical practice or in recording of health data, we do not believe this could have influenced our results.

Standard protection equipment (face mask, sterile gloves, protection gown) was used during sampling, and the sampling was performed inside the infant's incubator or bed. All sampling equipment used was sterile (suction catheter, connecting tube, mucus specimen trap).

We decided on using nasopharyngeal aspirates in contrast to nasal swabs, and used active suction only when the tip of the suction catheter was located in the nasopharynx, in order to minimize contamination with the microbiome of the anterior nostril (140). The catheter was not moisturized in the infants' mouth, to avoid contamination with oral microbiota. Despite these precautions, we cannot exclude contamination from neighboring sites (anterior nostrils, infants' skin), or mothers' skin and the environment (NICU), as those were not sampled and investigated.

Our intention was to obtain samples at initiation and discontinuation of antibiotic treatment. However, since a deterioration of infants' clinical status was often the cause for antibiotic initiation, infants were sometimes not stable enough to undergo NPA sampling. Collecting samples at the time of discontinuation of antibiotics was challenging especially in infants that transferred to other hospitals, as we were not always notified of their antibiotic treatment in real time. Antibiotic class, timing and duration varied substantially as well, and 22 samples of antibiotic exposed infants were excluded prior to WMS analysis due to low biomass. These factors influenced the ability for a more detailed analysis of influence of antibiotics. It is possible that having all samples obtained immediately prior and post antibiotic treatment might capture antibiotics induced changes in the microbiome and resistome more precisely and allowed for a better comparison of observed effects between infants. However, the composition at the end of the follow up (six months corrected age) was not influenced by deviations from the sampling protocol.

Six samples used in Paper I were received from a secondary location. They were obtained using the same sampling protocol, but by different health personnel than all other samples included in the method optimization study. Despite close geographic position and comparable medical practice between Denmark and Norway, differences between the two sites are expected. However, the extracted DNA concentration and host DNA content of samples from both locations was similar enough for the purpose of method optimization study.

5.1.3 Resistome sub-cohort analysis

Due to the variation in postnatal antibiotic treatment and inability to collect samples immediately prior to initiation of antibiotics, the analysis of resistome included only a sub-cohort of antibiotic exposed infants, and all infants without postnatal antibiotic exposure. Altogether, resistome composition of 181 samples belonging to 36 infants was analysed. This cohort was composed of 21 antibiotic-naïve infants and 15 infants treated with ampicillin plus gentamycin due to suspected EONS, where samples were obtained on the same day as antibiotics were initiated. Also in this sub-cohort, antibiotic exposed infants had significantly lower GA and BW, and their mothers had a longer time since ROM prior to delivery compared to antibiotic-naive infants. Detailed characteristic of the sub-cohort, compared to antibiotic naïve infants, and an overview of collected samples are available in Paper III.

Selection of the sub-cohort might introduce some bias towards healthier infants. Most often, samples were not obtained immediately prior to antibiotic initiation due to instability of preterm infants, meaning that more sick infants might not be included. These results are thus not representative for the whole study cohort but provide insight into the effect of ampicillin plus gentamycin on resistome development of a sub-cohort.

5.1.4 Host DNA depletion and metagenomic DNA extraction

Microbial DNA may be lost during procedures intended for host DNA depletion and DNA extraction. This is particularly critical in samples with low microbial biomass. Obtaining a stringent positive control to address possible bias introduced with the combination of host DNA depletion and DNA extraction methods proved to be challenging.

Initially, we used a defined mock community as a positive control and processed it in triplets with all tested protocols, but we did not use the optimal amount of stock (as recommended by the producer) to better represent the low biomass property of our samples of interest. We

obtained the expected DNA yield after DNA extraction with the MasterPure protocol. As the mock in stored in RNA/DNA shield, some of the bacterial cells will be lysed prior to processing and further lost during host DNA depletion steps (removal of extracellular DNA from lysed cells). Hence, the retrieved DNA yield or the retrieved microbial composition could not be compared with the non-depleted reference.

Using clinical samples as their own positive control (comparing the composition in reference and depleted aliquots from the same pool) was challenging as well, due to the low number of bacterial reads obtained from reference samples compared with depleted aliquots. Additionally, chemical or osmotic lysis followed by extracellular DNA removal might eliminate damaged or unviable bacteria from clinical samples. This is a wanted effect, as the observed microbial composition then reflects that of a viable microbiota but restricts further investigation of a possible taxonomic bias between depleted and reference samples.

There is however a concern that besides removing extracellular DNA from unviable or damaged bacterial cells, DNA from bacteria with a thin or missing cell wall (*Mycoplasma, Ureaplasma*) could also be lost during host DNA removal procedures (103), potentially introducing a taxonomic bias as a loss of Gram-negative species (94, 103, 109). The upper respiratory microbiota of term infants is mainly represented by Gram-positive bacteria, but Gram-negative bacteria including *Moraxella* have also been described as important colonizers (24). We observed a possible bias towards the loss of Gram-negative species (Paper I), but also due to the limitations described above, our study was not designed for taxonomic bias analysis. Implementing a positive control with known amounts of viable bacterial cells, mimicking colonization of the upper respiratory tract, was not feasible in our study.

Despite similar mechanism of tested host DNA depletion protocols, they performed quite differently also compared with published findings. The lyPMA protocol has been described as more effective than MolYsis and QIAamp in low biomass and high host content saliva samples (104). However, in our samples, host DNA content remained as high as in the reference samples, in line with results seen in bovine milk sample analysis (141). Further, QIAamp has outperformed MolYsis in host DNA removal in saliva in infected tissue samples (104, 109), while analysis of samples processed with QIAmp protocol in our study was not possible due to too low DNA yields. MolYsis performed best regarding host DNA depletion in our samples. It was initially developed to improve sepsis diagnostics (142, 143) and has since proven efficient in host DNA removal across a variety of samples (e.g. milk, cerebrospinal fluid) (94, 110, 144).

Our findings and published literature so far indicate that performance of host DNA depletion methods varies between sample types. In also highlights the challenge of comparing microbiota findings in studies using different sample processing protocols. Finding the optimal protocol for investigated sample might be challenging but is essential especially in low biomass and high host DNA samples used for microbiota studies.

Beside different host DNA depletion methods, we also tested three DNA extraction kits with different DNA extraction methods: semi-automated DNA extraction with bead beating (MagMax), spin column based (QIAamp), and lysosome and proteinase K based DNA extraction with ethanol precipitation (MasterPure). The tested bacterial cell lysis approaches have different advantages and limitations. Bead beating might improve the recovery of Gram-positive bacteria compared with other cell wall lysis procedures and decreases the possibility of an extraction bias (145), but it may also in lower DNA yield and result in DNA shredding (146). The MasterPure protocol performed superiorly, as it was the only one retrieving the expected DNA yield from low biomass adjusted mock community samples. Due to the loss of microbial DNA observed with the other two kits, they were excluded from further processing.

After establishing the optimized protocol for our type of samples, this was further used for the analysis of 369 cohort samples. We also added an additional control step prior to metagenomic library preparation, to estimate the host DNA and bacterial DNA ration in our samples using qRT PCR. All samples belonging to the same infant were processed in the same batch. Metagenomic DNA extraction and PCR of all cohort samples were performed in a span of six weeks, and majority of samples was processed by one researcher (PR).

5.1.5 Negative controls

Five regent kits were used for host DNA depletion and DNA extraction of all cohort samples, and we extracted a negative control for each used reagent kit. Due to very low DNA concentrations extracted from negative controls, these were excluded prior to sequencing. Despite very low 16S signal from the negative controls, we cannot exclude the possibility of contamination.

5.1.6 Metagenomic library preparation

During metagenomic library preparation, unique barcodes were used for each sample. We additionally controlled this process with a ZymoBIOMICS Microbial Community DNA

Standard (Zymo, D6305) and produced results very similar to the theoretical composition of the standard (data not published, Figure 4).

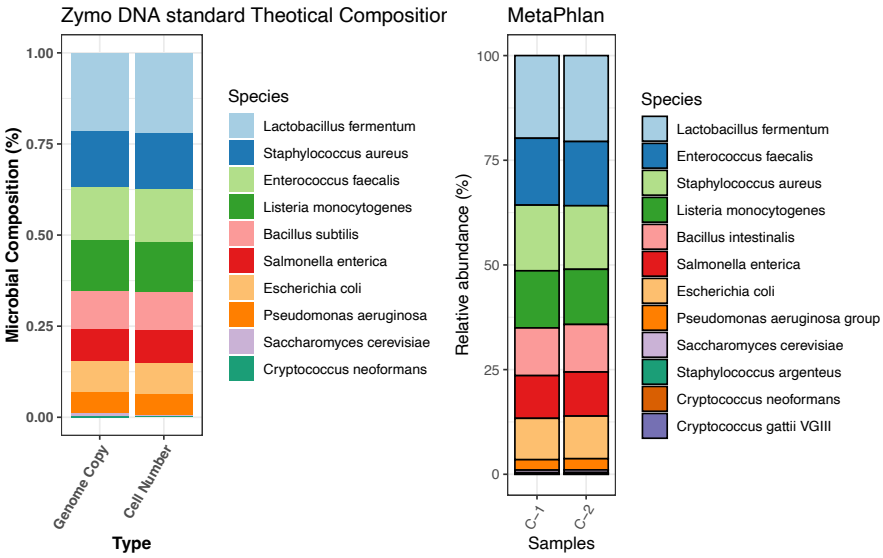

Figure 4. DNA standard for metagenomic library preparation control. We used 50 and 100 ng of Microbial Community DNA Standard (Zymo, D6305) to test the library preparation protocol. The obtained composition (samples C-1 and C-2) matched the theoretical composition of the standard.

Number of PCR amplification cycles during library preparation may present a source of bias (147). Due to the low DNA yields, we had to use 12 amplification cycles and took other steps to minimize possible bias. In an effort to reduce possible bias, all libraries were prepared by the same individual (PR) in a period of four weeks. One PCR machine was used for amplification of all samples and all samples belonging to the same infant were processed in the same batch.

5.1.7 Whole metagenomic sequencing

Samples used for the development of the optimized protocol for nasopharynx metagenomic studies were sequenced using WMS in three different rounds. All sequencing was performed at the same institution in the span of nearly three years. Composition and contamination bias could have been introduced by different sequencing rounds, but as we mainly evaluated host DNA content from these samples, we do not believe this factor carried a significant effect on our results. For the cohort analysis, 345 libraries were pooled in equimolar concentrations into

one sequencing pool, to obtain comparable sequencing depths between samples (148), and were sequenced simultaneously on one sequencing platform.

WMS provides only information regarding relative microbial composition. Using a spike with known absolute abundance can serve as a positive control, enables normalization of composition data to absolute values and determines total microbial load in the samples (149). Further, it can reduce bias in the interpretation of microbiome dynamics and interactions (150). We used a Spike-in with known absolute abundance composed of three species not found in the human microbiome (*Truepera radiovictrix*, *Imtechella halotolerans* and *Allobacillus halotolerans*).

In Paper I, the retrieved abundance of *I. haloterans* was used as a reference for quantification of total microbial load. The same Spike-in was also added to all cohort samples after host DNA depletion, prior to metagenomic DNA extraction. Due to the Spike-in being stored in DNA shield, some bacterial cells could already be lysed, and their DNA would be lost during host DNA depletion procedures. After WMS, we identified all three species belonging to the Spike-in control in 93% of all (n=345) samples, despite one of Spike-in species being resistant to lysosome lysis (*T. radiovictrix)*. As we did not use an optimal extraction method for all Spike-in species (mechanical lysis, bead beating), could not control for the host DNA depletion step, and did not retrieve the expected DNA yield (0.4 ng) in 27% of samples included in the cohort study (evaluated with 16S rRNA, qPCR), normalization of microbial composition for the cohort study was not appropriate. We were thus not able to quantify the bacterial abundance after WMS but were able to describe it using qPCR.

5.1.8 Data analysis

Despite equimolar library pooling, samples varied greatly in number of obtained raw and bacterial reads. There is a lack of established guidelines regarding data normalization. We focused our analysis on methods not sensitive for the variation in sequencing depth and did not exclude any sample based on the number of obtained total or bacterial reads. Briefly, within sample (α) diversity was described using Shannon index, which is less sensitive to the differences in library sizes and describes the diversity within the samples while accounting for both richness and evenness (151). For between sample (β) diversity, we used CLR transformation as a normalization method followed by PCA (152, 153).

The bioinformatics tool may have a greater effect on the composition than the extraction method (144). Prior to analysing WMS data with MetaPhlAn, we attempted using Kraken2 for taxonomic classification (unpublished data), but faced a problem during rarefactory analysis, as none of the samples reached the asymptote and increasing sequencing depth did not improve this. However, as the relative taxonomic composition of samples sequenced shallow and deep remained similar, and repeating the analysis with an even larger sequencing depth was not feasible, we decided to analyse our sequencing data with MetaPhlAn3. The two software use different approaches for taxonomic assignment. As a result, Kraken2 will often assign a higher number of especially low abundant species, while MetaPhlAn only assigns sequences to specific taxa when certain marker genes are present in the sequences. Even though a lower number of individual taxa on phylum, genera and species level was assigned with MetaPhlAn, this was mostly a reflection of higher number of low abundance taxa (< 1% relative abundance) assigned with Kraken, while the abundant taxa (≥ 1% relative abundance) remained similar.

Bioinformatical analysis of raw shotgun sequencing data and diversity analyses at the University of Oslo were performed by another PhD candidate, biostatistician, and present the base of his thesis (AD). These are briefly presented in the sections above but are otherwise outside the scope of this thesis.

5.2 Main findings

5.2.1 Nasopharyngeal aspirate samples of very preterm infants have high host DNA content and low biomass

Host DNA content of NPA samples analyzed with WMS without host DNA depletion steps was 99%. Samples collected in the first week of life had a higher host DNA content compared with samples obtained later in life.

In a study by Bosch et al., nasopharyngeal microbiota was sampled in 102 term born infants during their first six months of life and analysed using 16S rRNA amplicon sequencing (154). Host DNA content cannot be evaluated with this approach. However, only 15% of samples obtained on the day of birth had sufficient DNA concentration for further analysis, which increased exponentially in the next days. Of samples obtained on day 1, 45% were successfully analysed, followed by a success rate of over 97% for all remaining sampling time points.

A lower success rate (due to low biomass) has been described for preterm infants. A study by Gallacher et al. (54) included 55 infants born at GA ≤ 32 weeks and collected nasopharyngeal and tracheal aspirates, bronchoalveolar lavage and feces samples at five sampling time points during infants' first 30 days of life. Only 6.7% of nasopharyngeal aspirate samples obtained during the first three days after birth, and < 40% of samples obtained at later time points had a high enough bacterial load for 16S amplicon sequencing. All infants were exposed to some postnatal antibiotics after birth but no further specification on timing or duration of treatment was described.

In our cohort, we also observed an increase in bacterial abundance of nasopharyngeal microbiome from birth on, in line with other published findings (19, 25, 54, 63, 102). Of samples obtained between DOL 0 and 3, 88% were successfully analysed with WMS, with success rate at other sampling time points ≥ 92%. Despite including a similar age group of infants as Gallacher et al., their higher fail rate might be a result of a different threshold (> 0.1 ng/mcgL after PCR amplification) and antibiotic exposure of their whole cohort. Within our cohort, we observed a higher failure rate (due to lower biomass) for samples obtained from antibiotic exposed infants, indicating general suppression of bacterial colonization in nasopharyngeal microbiota of infants exposed to antibiotics in early life.

5.2.1.1 Two-targeted predictive qPCR

Estimating the needed sequencing depth in low biomass samples with high host DNA content is challenging. Information regarding host-microbial ratio in samples prior to sequencing may help customize sequencing protocol and optimize time-cost efficiency of WMS (116).

In Paper I, we used both one (16S rRNA gene) and two targeted (16s rRNA and host DNA) qPCR to estimate bacterial enrichment in samples prior to sequencing. Microbial enrichment observed with one target approach was indicative of host DNA depletion seen after WMS but required a control sample (without depletion) for every tested sample, which might not be feasible in practice. Due to low DNA yield from individual patient samples, the two-targeted approach could be tested only in four (out of six) individual patient samples. We observed a correlation between the ratio of bacterial / host WMS reads and bacterial / host DNA quantity (qPCR).

In the analysis of cohort samples, we used two-targeted PCR approach on all 369 NPAs. We were able to observe a correlation between PCR and WMS results in 345 samples (data not published). DNA yield and sequences belonging to mock species were included in the analysis. We found strong linear correlation between log ratios of bacterial / host DNA quantity (qPCR) and bacterial / host WMS reads (R^2=0.87), and between PCR and WMS relative host DNA content (%) (R^2=0.90) (Figure 5).

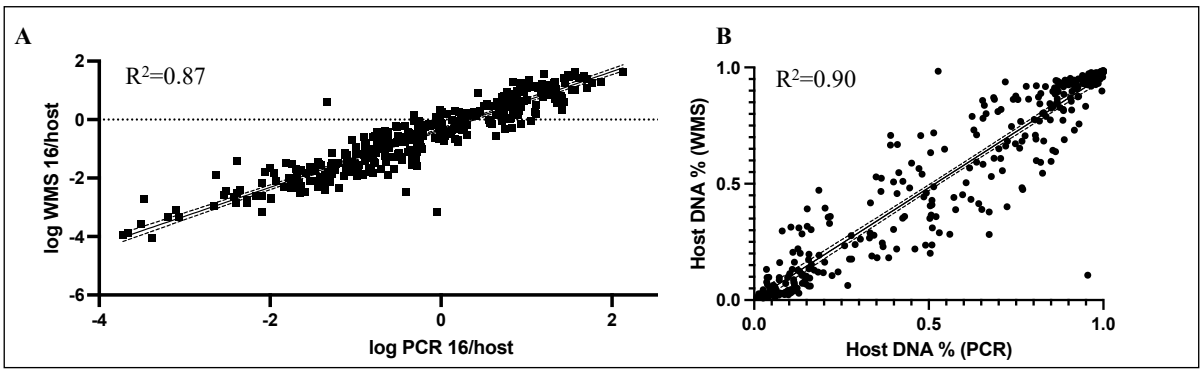

Figure 5: WMS composition predictive model based on two targeted PCR. A. Linear correlation between log ratios of bacterial / host DNA quantity (qPCR) on the x axis and bacterial / host WMS reads on y axis. B. Linear correlation between host DNA content (%) obtained with PCR (on the x axis) and WMS (on the y axis).

5.2.2 Inter-individual variation and age had the largest effect on microbiome and resistome composition

Effect of individual was the largest significant variable influencing microbiome composition, in line with published literature on term and preterm infants. The resistome composition is strongly correlated to the microbiome composition and similarly exhibited high inter-individual variability. Binia et al. (2021) used WMS for analysis of nasopharyngeal microbiota in 422 samples from term born infants at 2 and 4 months of age (30). Their mean sequencing depth was 38.1 M reads, with an average of 4 M high quality host-filtered reads per sample, which is comparable with our study. The individual differences accounted for 62% of variance. In a small study comparing microbiota of nasal washes using 16S rRNA amplicon sequencing of term (n=6) and preterm (n=7) infants at two time points (between six to 24 months of life), preterm infants exhibited a higher individual variation and within group heterogeneity compared with term infants (62).

In our cohort, PMA was the second strongest variable influencing the overall composition of microbiome and resistome. GA and PMA have been described to inflict a significant effect on gut microbiome of preterm infants, more so than e.g. MOB, which has an important effect on the gut and respiratory microbiome of term infants (25, 66, 155). We did not observe any significant effect of MOB on the microbiome composition in our cohort. This could in part be due to immature mucosal barriers that host intestinal and respiratory microbiota of preterm infants (40).

5.2.3 Nasopharyngeal microbiome and resistome composition

Mixed microbial flora was observed across all sampling time points. Figure 3 in Paper II illustrates mean relative microbial abundance at different sampling time points and according to antibiotic exposure. Nasopharyngeal microbiota composition, host DNA content and absolute bacterial abundance on species taxonomy level for each individual infant is illustrated in Figures 6-8, where infants are grouped according to antibiotic exposure and gestational age. Resistome composition according to sampling time point and postnatal antibiotic exposure is illustrated in Figure 1 of Paper III.

The development of nasopharyngeal microbiome in healthy, vaginally born, breastfeed term infants with minimal exposure to antibiotics has been investigated using 16S rRNA amplicon sequencing. Mixed flora was observed immediately after birth, further evolving to *S. aureus*

dominated profiles within weeks. Differentiation towards *Corynebacterium* and *Dolosigranulum* dominated profiles was seen within the first months of life, and by the age of six months, stabile microbiota profiles dominated by *Moraxella, Streptococcus* or *Hemophilus* have been recognized (25, 28, 35). Term infants exposed to antibiotics displayed lower abundances of beneficial *Corynebacterium* and *Dolosigranulum* genera, and increased colonization with gram-negative bacteria (19, 35, 156). *Dolosigranulum* has been described as a respiratory health associated genera in older children as well (157). Further, *Moraxella* and *Dolosigranulum* depleted nasal microbiota between the ages from two to 24 months in infants with two or more antibiotic treatment courses in early life have been linked to increased risk of asthma development (24).

In our cohort of very preterm infants, recognition of microbiota profiles according to dominating taxa was not as clear due to high heterogeneity, as earlier described in preterm infants (62). We were however able to identify six statistically significant microbiome clusters, i.e. community types. Community types were determined by abundances of five genera (*Cutibacterium, Gemella, Serratia, Streptococcus, Staphylococcus*). We observed a higher number of changes between community types in infants exposed to "Other antibiotics", which could reflect exposure to different classes of antibiotics compared with only ampicillin plus gentamycin or no postnatal antibiotics (158).

We found no significant differences in abundance of different taxa between antibiotic exposed infants but observed some patterns of purely descriptive nature. This may however be specific to the time and space of our observations.

The dominant phyla across our samples were *Proteobacteria* and *Firmicutes*, followed by *Actinobacteria*, *Fusobacteria* and *Bacteroides*. *Staphylococcus*, *Streptococcus* and *Cutibacterium* were the most prevalent genera (found in > 90 % of all samples) and were dominant (> 50% abundance) in > 10% of all samples. Immediately after birth, the dominating genus was *Serratia* (mean relative abundance 81%), followed by *Streptococcus* (7%) and *Staphylococcus* (5%). Other, less abundant genera during infants' stay at the NICU were *Corynebacterium*, *Dolosigranulum*, *Enterobacter*, *Escherichia*, *Gemella*, *Klebsiella*, *Neisseria*, *Rothia* and *Veillonella*. At six months corrected age, mean relative abundance of *Staphylococcus* accounted for < 1%, and *Moraxella* emerged as one of the more abundant genera (22%), together with *Streptococcus* (26%) and *Serratia* (25%). On the species level, *C.*

acnes, S. mitis and *S. epidermidis* were the most prevalent (found in > 88% of samples), but dominant (> 50% abundance) in less than 10% of all samples, regardless of antibiotic exposure.

WMS allowing for species recognition is an important tool for differentiation of pathogens from commensals. We identified species that are common pathogens for late onset sepsis (LOS) in preterm infants (*E. coli*, *K. pneumoniae*). We observed a higher mean abundance of these potentially pathogen genera (*Enterobacter*, *Escherichia*, *Klebsiella*) at time point 2 of "Only Early" antibiotics group. For most infants, this were the first samples obtained after discontinuation of ampicillin plus gentamycin for suspected EONS. These species decreased over time and accounted for < 1% mean abundance at six months corrected age. However, presence of pathogens during early colonization has been shown to potentially increase the risk of ARI and chronic inflammation by triggering an inflammatory immune response (159, 160).

A range of species present antagonistic associations with respiratory pathogens (30). We observed patterns of known antagonistic relationship between *S. epidermidis* and *S. aureus* (161), and observed that both species were correlated with carriage of high risk ARG. Despite possible pathogenic potential of *S. aureus*, it appears it is a beneficial colonizer in early life due to production of antimicrobial molecules, antagonistic effect with *S. pneumoniae*, and is present in most infants without causing disease in early neonatal life (co-occurs with *Corynebacterium* and reduces *S. aureus* virulence) (139). We still observed a high mean relative abundance of *S. pneumoniae* in the "Other" antibiotic group at six months corrected age, possibly due to consequence of low mean *Corynebacterium* abundance that have been shown to promote *S. aureus* growth (162). Further, early colonization with Streptococcus has been described as an asthma predictor, and targeting pathogenic bacteria within the NP microbiome could be a prophylactic approach to asthma (35).

In our cohort, respiratory-health associated microbiome profiles with dominant *Corynebacterium* or *Dolosigranulum* genera (as described in term infants) were sparse, found in only 2% of all samples, regardless of antibiotic exposure or nutrition after discharge. This could be a result of delayed latching and oral feeding (163, 164), and lower rates of exclusive breastfeeding after discharge (29).

Early colonization with Moraxella (at one month) and low abundances of *Corynebacterium* and *Dolosigranulum* could predispose infants to an increased risk of chronic inflammation in airway mucosa and reversible airway obstruction (19, 159). *Moraxella* had a higher prevalence

across all time points and became dominant (> 50% of relative abundance) at six months corrected age in 16% of our samples.

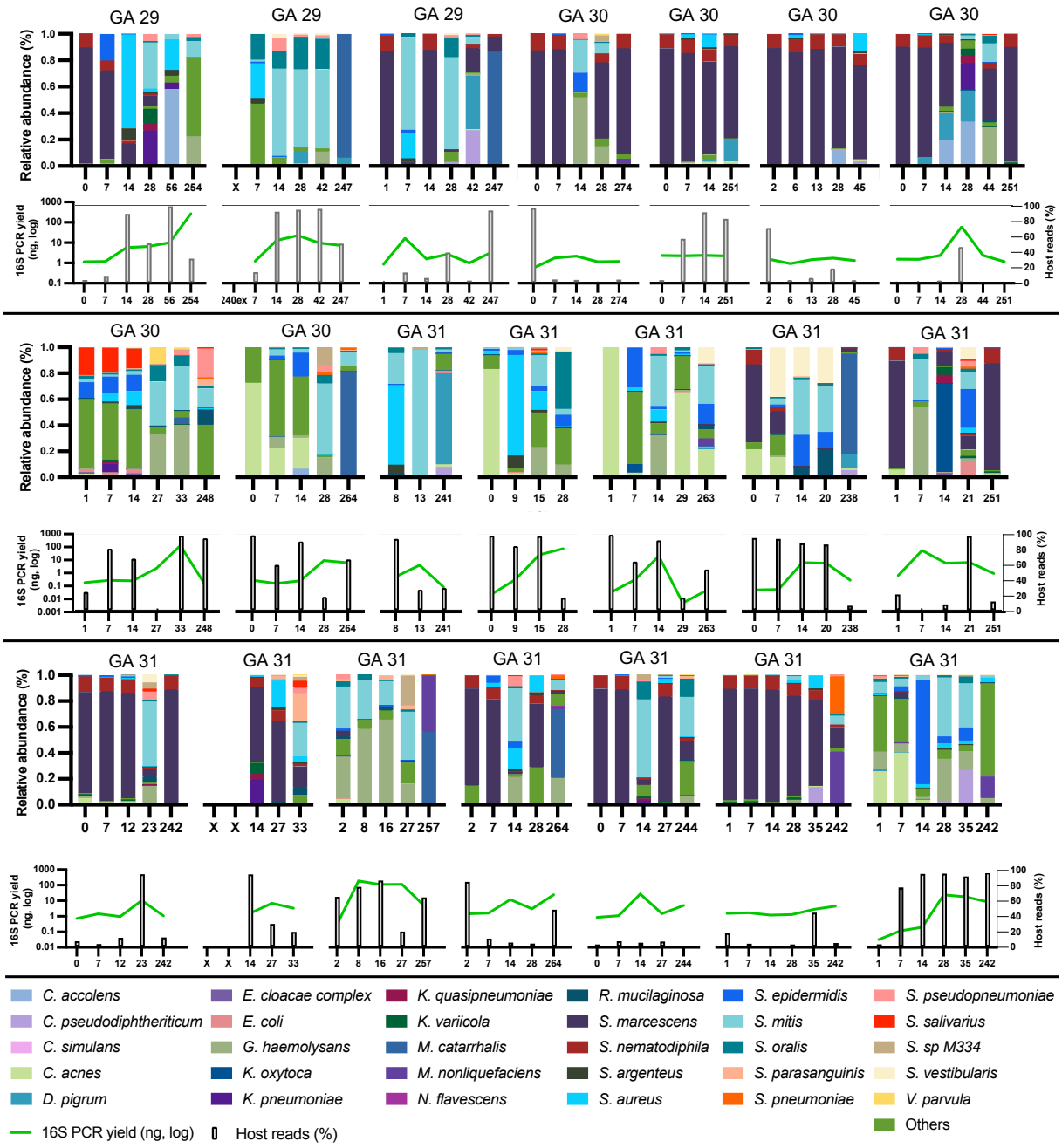

Figure 6: Antibiotic naïve infants. Microbiome composition, absolute bacterial abundance (ng) and host DNA content (%) of samples obtained from infants not exposed to any postnatal antibiotics. Gestational age (GA) for each infant is written above corresponding bar plots representing relative bacterial abundance on species taxonomic level. The x axis represents day of life at sampling. "X" on x axis marks samples excluded prior to whole metagenomic sequencing or analysis.

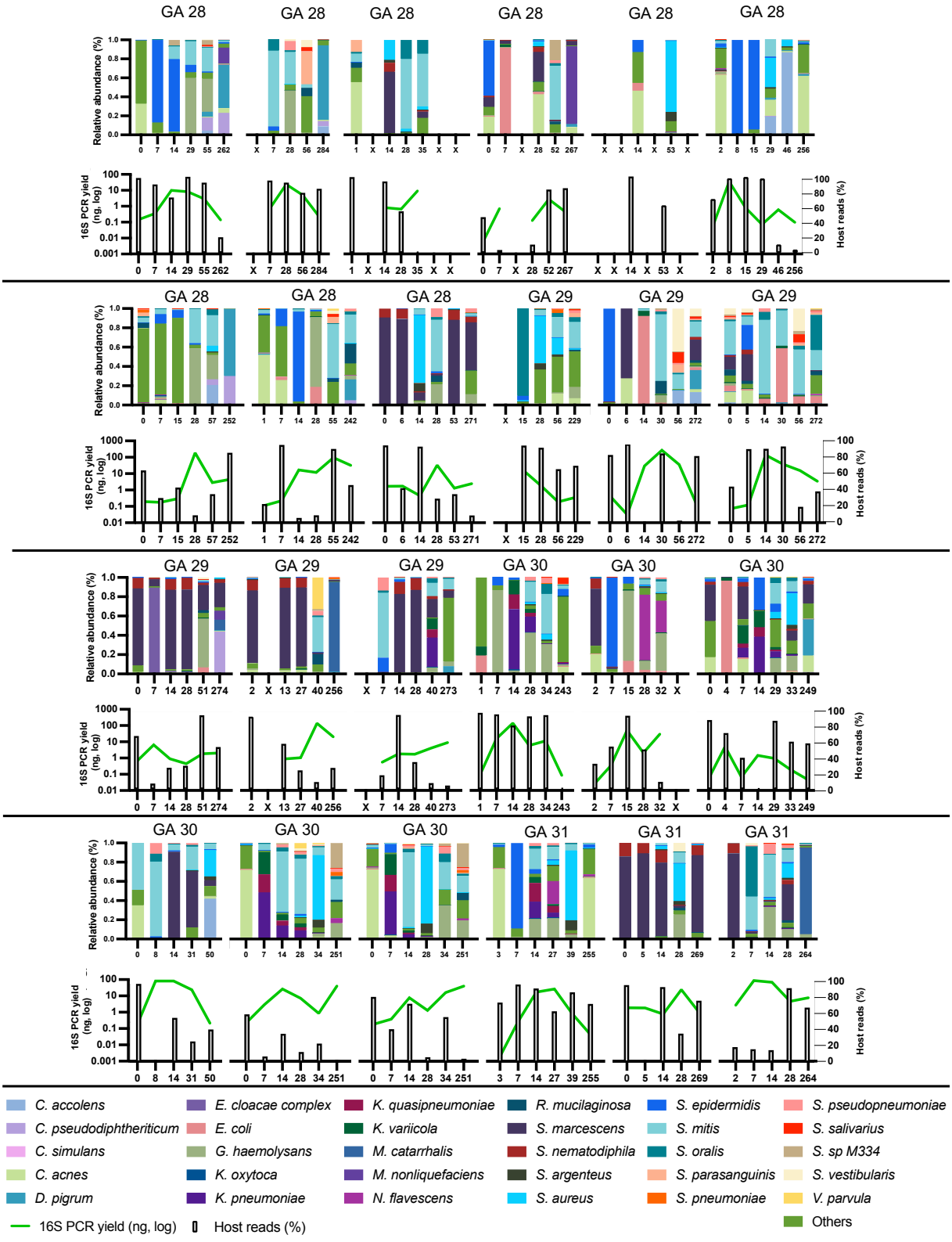

Figure 7: "**Only Early" antibiotics.** Microbiome composition, absolute bacterial abundance (ng) and host DNA content (%) of samples obtained from infants exposed to ampicillin plus gentamycin for suspected early onset neonatal sepsis. Gestational age (GA) for each infant is written above corresponding bar plots representing relative bacterial abundance on species

taxonomic level. The x axis represents day of life at sampling. “X” on the x axis marks samples excluded prior to whole metagenomic sequencing or analysis.

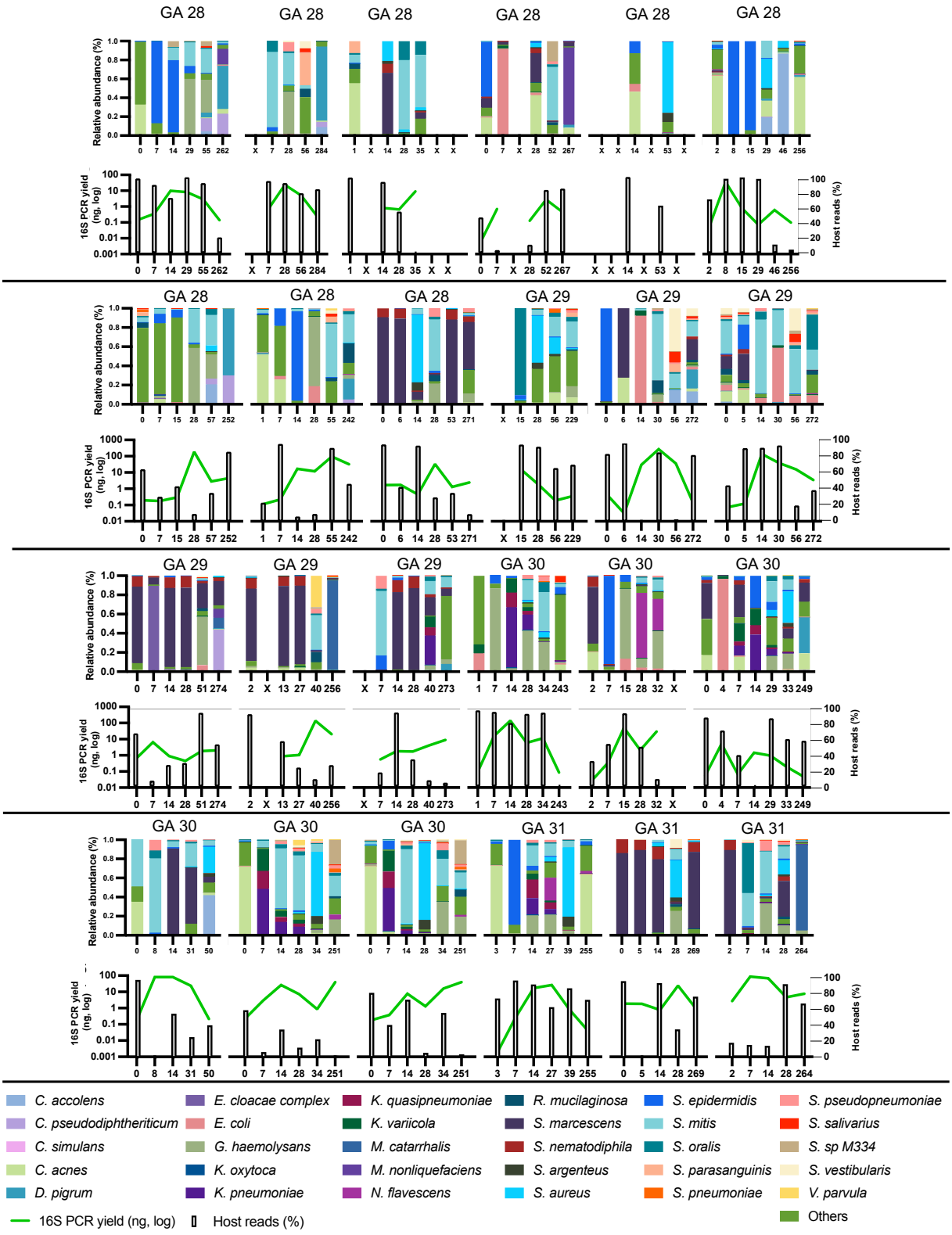

Figure 8: **"Other" antibiotics**. Microbiome composition, absolute bacterial abundance (ng) and host DNA content (%) of samples obtained from infants exposed to other antibiotics than only ampicillin plus gentamycin. Gestational age (GA) for each infant is written above corresponding bar plots representing relative bacterial abundance on species taxonomic level. The x axis represents day of life at sampling. "X" on the x axis marks samples excluded prior to whole metagenomic sequencing or analysis.

The resistome composition was investigated in 36 infants (Figure 1, Paper III). As expected, it was strongly associated with microbiome composition. We found ARGs in nearly all samples, regardless of antibiotic exposure. High ARG richness has been described in correlation to low microbiota maturity at 1 year of age and a higher risk of developing asthma later in life (35, 165).

In total, 373 ARGs belonging to 15 ARG classes and relating to five antibiotic resistance mechanisms were recovered. Multidrug was the most abundant ARG class, followed by fluoroquinolone, beta-lactam, tetracycline and aminoglycoside. Further, we were able to identify a core resistome of 13 ARGs that were both most prevalent and highly abundant in our cohort (Table 3). We observed some high-risk ARG (rank I) (166), also in preterm infants not exposed to any postnatal antibiotics, but their presence diminished until six months corrected age (Supplementary Fig. 4, Paper III).

Antibiotic resistance class	**Resistance genes**
Multidrug	*acrB, oqxB, mexI*
Macrolide-lincosamide-streptogramin	*mel, pmrA, rlmA(II)*
Beta-lactam	*blaZ, SST-1*
Fluoroquinolone	*patA, patB*
Tetracycline	*tet(41), tetM*
Aminoglycoside	*AAC(6')-Ic.*

Table 3: The core resistome of preterm nasopharyngeal microbiota. Thirteen antibiotic resistance genes belonging to six different antibiotic class groups were most prevalent and highly abundant across all investigated samples.

Premature infants are exposed to a combination of variables that modify microbiota across all body niches (167, 168). The differences in colonization patterns between term and preterm infants might also be due to underwent intensive care, that may e.g., lead to disrupted anatomical barriers in the intestinal and respiratory tract (such as intubation, nasogastric feeding tube placement). In addition, preterm infants endure long hospital stays and are exposed to potentially pathogen bacteria from the hospital environment while their immune system is immature (169), which has been shown to influence the infants skin and gut microbiome and resistome composition (170).

5.2.4 Prolonged colonization with Serratia marcescens

Serratia marcescens is a low virulent opportunistic pathogen often found in NICU environment and can cause local and systemic nosocomial infections (171, 172). It has been associated with several epidemic events, often caused by multi drug resistant strains, with difficulties in identifying the source of infection (172, 173). *Serratia* has also been described as an inhabitant of the nasal microbiota (174).

A few months into our study, an outbreak with *S. marcescens* was detected in the NICU. Of the infants included in our study, one had a positive nasopharyngeal culture (swab) and two had *S. marcescens* isolated from local infection sites. Three other infants hospitalized during this period, not included in our study, tested positive in different cultured samples. The department carried out required infection control measures. Two screenings of all hospitalized infants were performed, and environmental samples were collected. The origin of the outbreak was not identified. The second screening retrieved no positive nasopharyngeal cultures, and the outbreak was declared to be over.

Nevertheless, WMS analysis of our samples revealed long term implications of the *S. marcescens* outbreak. After organizing the samples chronologically according to sampling date (Figure 5, Paper II), we identified *S. marcescens* in very low abundance (< 0.01%) in 7% of samples obtained before the start of the outbreak, and in a higher abundance in two samples (1% and 84% abundance) obtained from the infant who was first identified as infected with *S. marcescens* (these two NPA samples were obtained 10 and three days before the first isolation of *S. marcescens* from the same infant). In the NPA samples obtained for our study during the duration of the outbreak, *S. marcescens* was found in 50% of samples, with abundance ranging from < 0.001% to 85%.

After the screening which retrieved no positive nasopharyngeal cultures, the rates of *S. marcescens* reads in our samples remained high, with 65% of samples had some *S. marcescens* reads. WMS could be a promising tool in early detection and infection control during an outbreak, as we recovered *S. marcescens* from infants with negative screening cultures (175).

Timing of hospitalization at the NICU in regard to *Serratia* outbreak (hospitalized prior to, during or after *Serratia* outbreak) was modestly but significantly associated with microbiome composition (PERMANOVA, R^2 1.3%, p=0.001). Ten infants were discharged or moved to a different hospital before the outbreak. We analyzed samples of nine of these infants at 6 months corrected age, and none had >0.01% *S. marcescens* reads (three had *S. marcescens* reads at < 0.001% abundance). Of the infants discharged after the outbreak, 67% of infants had any *S. marcescens* reads, of which 56% had > 1% abundance, and 42% > 10% abundance. Effect of hospitalization before, during or after Serratia outbreak was the only significant variable influencing the microbiome composition at six months corrected age (PERMANOVA, R^2 4.7%, p=0.003).

Preterm infants are often hospitalized for longer periods of time, and exposure to NICU flora while their immune system is immature and defense mechanism insufficient can leave a long-lasting signature on infants' microbiome and resistome (169, 170). We only analyzed the resistome of a part of our cohort and could thus not explore the influence of hospitalization on the resistome. We did however observe a resistome community type (i.e. resistotype) specific to the *Serratia* community type, and found strong correlation between some ARGs (*SST-1, AAC(6')-Ic, tet(41)*, *MexI*) and *Serratia* species. These ARGs persisted from time point 1 until six months corrected age in 20 infants, indicating a possible long-term effect of hospitalization on the resistome.

5.2.5 Antibiotic exposure in early life has transient effect on microbiome and resistome composition

Disruptions of microbiota development caused by antibiotic treatment in early life has been implied in pathologies of several diseases (24, 176, 177). We included 21 preterm infants not exposed to any prenatal or postnatal antibiotics in our cohort. However, their mothers received antibiotics during delivery, which could influence the nasopharyngeal microbiota of infants (178).

A general suppression of bacterial growth with no effect on α-diversity has previously been observed after prophylactic antibiotic treatment with benzylpenicillin and gentamycin in preterm infants (54). Similarly, we did not observe a significant difference in α-diversity during the time of administration of ampicillin plus gentamycin (between sampling time points 1 and 2), however most infants have already received antibiotics prior to sampling at time point 1, which could influence the baseline samples.

Further, we observed a transient effect of different antibiotic treatments on microbiome composition at two sampling time points, which diminished before six months corrected age. In early life, microbiota provides signals to the host immune system that are essential for training of immune tolerance and immune maturation. Changes in early life respiratory microbiota have more effect on respiratory health than changes later in life (139). Even transient disruptions during this window might influence long term respiratory health (139). Additionally, antibiotics influence the microbiota of other niches as well, and the combined effect of early life antibiotic disruptions has been described to increase the risk of the development of several pathologies such as wheeze, asthma, and eczema (24, 176, 177).

Antibiotic stewardship programs for preterm infants have resulted in a reduction in antibiotic prescriptions over the last 10 years in Norway, but the rates of antibiotic use in not-infected preterm infants are still high (76). In our cohort, none of the infants treated with empiric antibiotics after birth had a positive blood culture (obtained prior to the start of antibiotics). The more prevalent stewardship actions are aimed at shortening the duration of empiric antibiotics and have reduced antibiotic use across several NICUs (74). In the gut microbiome of preterm infants, a significant difference was observed when comparing long versus short treatment with amoxicillin/ceftazidime (179). We did not observe a significant effect of ampicillin and gentamycin duration, but rather an effect of any early antibiotics on the

respiratory microbiome and resistome. Approaches to reduce initiation of empiric antibiotic after preterm birth might be beneficial (e.g. series of clinical observations) but are challenging to implement in the smallest preterm patients (180, 181).

5.2.6 Impact of ampicillin plus gentamycin on preterm infants resistome

Antibiotic given due to increased risk of EONS are usually started immediately after birth, and often before we were able to obtain study samples. This is why we decided to investigate the effect of antibiotics on the resistome only in infants, where we were able to obtain the sample on the same day as antibiotics were initiated. Even in these baseline samples, a visible increase in heterogeneity was observed compared with resistome composition of antibiotic naïve infants (Figure 2B, Paper III).

We observed transient perturbations after treatment with ampicillin plus gentamycin, which were more pronounced in infants exposed to prenatal antibiotics as well, pointing to a possible cumulative effect of pre- and postnatal antibiotic therapy on the infants resistome (Supplementary figure 6B, Paper III). Interestingly, we found no significant effect of IAP, which has been described to transiently increase ARGs in the gut and respiratory resistome (182, 183). The effect of ampicillin plus gentamycin for suspected early onset neonatal sepsis was most pronounced immediately after discontinuation of antibiotics. We did not observe any significant differences between antibiotic exposed and naïve infants at later time points. Duration of antibiotic regimes has been found to significantly influence the abundance of ARGs, but this was not evident from our data (93).

In our study, no differences in resistome composition or carriage of high-risk ARGs were observed between antibiotic exposed and antibiotic naïve infants at six months corrected age. However, the large inter-individual variability, high abundance of ARGs and presence of high-risk ARG in preterm infants illustrate a thin line of battle between available antibiotics and emergence of antibiotic resistant infections, also in the neonatal population (184).

6 Conclusions and future research

6.1 Main conclusions

This thesis comprehensively describes the methodology for metagenomic analysis of nasopharyngeal aspirate samples, and developmental trajectories of upper respiratory microbiome and resistome in a cohort of very preterm infants. To our knowledge, this is the most detailed description of longitudinal upper respiratory microbiota development in preterm infant so far.

Standard operation procedures for low biomass and high host DNA samples are needed in the scientific community to reduce bias and allow for better comparability between studies focusing on exploring extraintestinal microbiota niches. Our published methodology protocol is a step towards optimal, standardizing sample processing for upper respiratory microbiota analysis.

Further, the findings of the study contribute to uncovering the effects of prematurity and prematurity related factors, such as use of antibiotics and the hospital environment flora on infants' respiratory microbiome and resistome development. Even in the absence of postnatal antibiotics, preterm infants displayed aberrant colonization patterns compared to those described in term infants, with a tendency to microbiome signatures described in proinflammatory settings, and presence of high risk ARGs. The persisting fingerprint of hospitalization on the infants' microbiome and resistome indicates the susceptibility of immature mucosal barriers to environmental nosocomial pathogens. Despite the transient nature of antibiotic induced perturbations observed in the microbiome and resistome, these may increase the risk for later respiratory pathologies through immune system modulation, and possibly expose infants to antibiotic resistant infections.

6.2 Future perspectives

Despite low incidence of bronchopulmonary dysplasia in infants born at GA > 28 weeks in high income settings, they still face a high burden of respiratory disease in early childhood, resulting in increased risk for hospitalization and antibiotic prescriptions. Dysbiosis in the upper respiratory microbiome during the critical window for immune system modulation is a possible contributor to the development of respiratory pathologies later in life. We followed our cohort until six months corrected age but have the possibility to collect information regarding their respiratory health outcomes for later time points as well. This might help us discover early life microbiota signatures in infants at risk for e.g. asthma development.

Further, it has been proposed that some respiratory health outcomes (e.g. protective effect of breastfeeding and probiotics on risk of ARI development) are mediated by intestinal microbiota rather than nasopharyngeal (29, 30, 42). We collected fecal samples at the same time points as nasopharyngeal aspirates. These samples have also been sequenced by WMS, and we plan on analyzing the intestinal microbiome and resistome in the near future. The parallel correlation of respiratory and intestinal microbiota development may further advance our findings.

Nasopharynx can serve as an easily accessible compartment for determining ARG carriage. However, the intestinal microbiota has been described to harbour a higher ARG diversity compared to oral microbiota in adults (185). Intestinal resistome analysis of our samples would contribute to the knowledge regarding the relationship of nasopharyngeal and intestinal ARG pools and their development in early life, and the possible difference in persistence of resistome perturbations following antibiotic treatment between the two compartments.

Shotgun metagenomic sequencing of our samples produced a vast amount of sequencing data of the nasopharyngeal metagenome. So far, we have used these data to characterize the microbiome and resistome. These analyses depend on available reference databases and repeating them later in time might uncover additional ARGs from our samples. Metagenomic sequencing data may also be used beyond describing the composition of the microbiota, e.g. to predict its metabolic function (186). This can add an additional dimension to the understanding of the interplay between the host, the microbiota, and the development of respiratory pathologies.

7 References

1. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome. 2015;3(1):31.
2. Wei S, Bahl MI, Baunwall SMD, Hvas CL, Licht TR. Determining Gut Microbial Dysbiosis: a Review of Applied Indexes for Assessment of Intestinal Microbiota Imbalances. Applied and environmental microbiology. 2021;87(11).
3. Ramirez J, Guarner F, Bustos Fernandez L, Maruy A, Sdepanian VL, Cohen H. Antibiotics as Major Disruptors of Gut Microbiota. Frontiers in Cellular and Infection Microbiology. 2020;10.
4. Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, et al. The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. Microbiol Mol Biol Rev. 2017;81(4).
5. Costello EK, Carlisle EM, Bik EM, Morowitz MJ, Relman DA. Microbiome assembly across multiple body sites in low-birthweight infants. mBio. 2013;4(6):e00782-13.
6. Parfrey LW, Knight R. Spatial and temporal variability of the human microbiota. Clinical Microbiology and Infection. 2012;18:5-7.
7. Tamburini S, Shen N, Wu HC, Clemente JC. The microbiome in early life: implications for health outcomes. Nature Medicine. 2016;22(7):713-22.
8. Gschwendtner S, Kang H, Thiering E, Kublik S, Fösel B, Schulz H, et al. Early life determinants induce sustainable changes in the gut microbiome of six-year-old children. Scientific reports. 2019;9(1):12675.
9. Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, et al. Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function. Science. 2012;336(6080):489-93.
10. Kennedy KM, de Goffau MC, Perez-Muñoz ME, Arrieta MC, Bäckhed F, Bork P, et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. Nature. 2023;613(7945):639-49.
11. Hornef MW, Torow N. 'Layered immunity' and the 'neonatal window of opportunity' – timed succession of non-redundant phases to establish mucosal host–microbial homeostasis after birth. Immunology. 2020;159(1):15-25.
12. Mishra A, Lai GC, Yao LJ, Aung TT, Shental N, Rotter-Maskowitz A, et al. Microbial exposure during early human development primes fetal immune cells. Cell. 2021;184(13):3394-409.e20.
13. Ganal-Vonarburg SC, Hornef MW, Macpherson AJ. Microbial-host molecular exchange and its functional consequences in early mammalian life. Science. 2020;368(6491):604-7.
14. Fouhy F, Watkins C, Hill CJ, O'Shea C-A, Nagle B, Dempsey EM, et al. Perinatal factors affect the gut microbiota up to four years after birth. Nature communications. 2019;10(1):1517-.
15. Scholtens PAMJ, Oozeer R, Martin R, Amor KB, Knol J. The Early Settlers: Intestinal Microbiology in Early Life. Annual Review of Food Science and Technology. 2012;3(1):425-47.
16. Wopereis H, Oozeer R, Knipping K, Belzer C, Knol J. The first thousand days – intestinal microbiology of early life: establishing a symbiosis. Pediatric Allergy and Immunology. 2014;25(5):428-38.
17. Gallacher DJ, Kotecha S. Respiratory Microbiome of New-Born Infants. Front Pediatr. 2016;4.
18. Staude B, Oehmke F, Lauer T, Behnke J, Gopel W, Schloter M, et al. The Microbiome and Preterm Birth: A Change in Paradigm with Profound Implications for Pathophysiologic Concepts and Novel Therapeutic Strategies. BioMed research international. 2018;2018:7218187.
19. Bosch A, de Steenhuijsen Piters WAA, van Houten MA, Chu M, Biesbroek G, Kool J, 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. 2017;196(12):1582-90.
20. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science. 2005;307(5717):1915-20.
21. Abdill RJ, Adamowicz EM, Blekhman R. Public human microbiome data are dominated by highly developed countries. PLOS Biology. 2022;20(2):e3001536.

22. Groer MW, Luciano AA, Dishaw LJ, Ashmeade TL, Miller E, Gilbert JA. Development of the preterm infant gut microbiome: a research priority. Microbiome. 2014;2:38.
23. Dickson RP, Huffnagle GB. The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease. PLOS Pathogens. 2015;11(7):e1004923.
24. Toivonen L, Schuez-Havupalo L, Karppinen S, Waris M, Hoffman KL, Camargo CA, Jr., et al. Antibiotic Treatments During Infancy, Changes in Nasal Microbiota, and Asthma Development: Population-based Cohort Study. Clinical Infectious Diseases. 2021;72(9):1546-54.
25. Bosch A, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, et al. Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. EBioMedicine. 2016;9:336-45.
26. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(26):11971-5.
27. Flynn M, Dooley J. The microbiome of the nasopharynx. Journal of medical microbiology. 2021;70(6).
28. Biesbroek G, Tsivtsivadze E, Sanders EA, Montijn R, Veenhoven RH, Keijser BJ, et al. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. American journal of respiratory and critical care medicine. 2014;190(11):1283-92.
29. Biesbroek G, Bosch AA, Wang X, Keijser BJ, Veenhoven RH, Sanders EA, et al. The impact of breastfeeding on nasopharyngeal microbial communities in infants. American journal of respiratory and critical care medicine. 2014;190(3):298-308.
30. Binia A, Siegwald L, Sultana S, Shevlyakova M, Lefebvre G, Foata F, et al. The Influence of FUT2 and FUT3 Polymorphisms and Nasopharyngeal Microbiome on Respiratory Infections in Breastfed Bangladeshi Infants from the Microbiota and Health Study. mSphere. 2021;6(6):e0068621.
31. Mizgerd JP. The infant nose. Introducing the respiratory tract to the world. American journal of respiratory and critical care medicine. 2014;190(11):1206-7.
32. Teo SM, Tang HHF, Mok D, Judd LM, Watts SC, Pham K, et al. Airway Microbiota Dynamics Uncover a Critical Window for Interplay of Pathogenic Bacteria and Allergy in Childhood Respiratory Disease. Cell host & microbe. 2018;24(3):341-52.e5.
33. Kang HM, Kang JH. Effects of nasopharyngeal microbiota in respiratory infections and allergies. Clin Exp Pediatr. 2021;64(11):543-51.
34. Saglani S, Custovic A. Childhood Asthma: Advances Using Machine Learning and Mechanistic Studies. American journal of respiratory and critical care medicine. 2019;199(4):414-22.
35. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, et al. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. Cell host & microbe. 2015;17(5):704-15.
36. Gollwitzer ES, Saglani S, Trompette A, Yadava K, Sherburn R, McCoy KD, et al. Lung microbiota promotes tolerance to allergens in neonates via PD-L1. Nat Med. 2014;20(6):642-7.
37. Liu C, Makrinioti H, Saglani S, Bowman M, Lin LL, Camargo CA, Jr., et al. Microbial dysbiosis and childhood asthma development: Integrated role of the airway and gut microbiome, environmental exposures, and host metabolic and immune response. Frontiers in immunology. 2022;13:1028209.
38. de Santa Barbara P, van den Brink GR, Roberts DJ. Molecular etiology of gut malformations and diseases. Am J Med Genet. 2002;115(4):221-30.
39. Dang AT, Marsland BJ. Microbes, metabolites, and the gut-lung axis. Mucosal Immunol. 2019;12(4):843-50.
40. Grier A, McDavid A, Wang B, Qiu X, Java J, Bandyopadhyay S, et al. Neonatal gut and respiratory microbiota: coordinated development through time and space. Microbiome. 2018;6(1):193.
41. Zhang D, Li S, Wang N, Tan HY, Zhang Z, Feng Y. The Cross-Talk Between Gut Microbiota and Lungs in Common Lung Diseases. Frontiers in microbiology. 2020;11:301.
42. Luoto R, Ruuskanen O, Waris M, Kalliomaki M, Salminen S, Isolauri E. Prebiotic and probiotic supplementation prevents rhinovirus infections in preterm infants: a randomized, placebo-controlled trial. The Journal of allergy and clinical immunology. 2014;133(2):405-13.

43. WHO. Preterm birty - Key facts 2018.[cited 2023 12th February].Available from: https://www.who.int/news-room/fact-sheets/detail/preterm-birth.
44. Obstetrisk guidelinemøte april 2022. Preterm fødsel 2022 [cited 2023 12th February]. Available from: https://www.legeforeningen.no/contentassets/7a83bea0df7e47ae8f8bfb11ea929381/preterm-fodsel-tm-11302021-clean-version-.pdf.
45. Lal CV, Travers C, Aghai ZH, Eipers P, Jilling T, Halloran B, et al. The Airway Microbiome at Birth. Scientific reports. 2016;6:31023.
46. Purisch SE, Gyamfi-Bannerman C. Epidemiology of preterm birth. Seminars in perinatology. 2017;41(7):387-91.
47. Azad MB, Konya T, Maughan H, Guttman DS, Field CJ, Chari RS, et al. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. Cmaj. 2013;185(5):385-94.
48. Cai M, Loy SL, Tan KH, Godfrey KM, Gluckman PD, Chong YS, et al. Association of Elective and Emergency Cesarean Delivery With Early Childhood Overweight at 12 Months of Age. JAMA Netw Open. 2018;1(7):e185025.
49. Sharma AA, Jen R, Butler A, Lavoie PM. The developing human preterm neonatal immune system: a case for more research in this area. Clin Immunol. 2012;145(1):61-8.
50. Lu L, Claud EC. Intrauterine Inflammation, Epigenetics, and Microbiome Influences on Preterm Infant Health. Curr Pathobiol Rep. 2018;6(1):15-21.
51. Zou ZH, Liu D, Li HD, Zhu DP, He Y, Hou T, et al. Prenatal and postnatal antibiotic exposure influences the gut microbiota of preterm infants in neonatal intensive care units. Annals of clinical microbiology and antimicrobials. 172018.
52. Dermyshi E, Wang Y, Yan C, Hong W, Qiu G, Gong X, et al. The "Golden Age" of Probiotics: A Systematic Review and Meta-Analysis of Randomized and Observational Studies in Preterm Infants. Neonatology. 2017;112(1):9-23.
53. Agha L, Staiger D, Brown C, Soll RF, Horbar JD, Edwards EM. Association of Hospital Adoption of Probiotics With Outcomes Among Neonates With Very Low Birth Weight. JAMA Health Forum. 2023;4(5):e230960.
54. Gallacher D, Mitchell E, Alber D, Wach R, Klein N, Marchesi JR, et al. Dissimilarity of the gut-lung axis and dysbiosis of the lower airways in ventilated preterm infants. The European respiratory journal. 2020;55(5):1901909.
55. Rehman S., Bacha D. Embryology, Pulmonary. [Updated 2022 Aug 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK544372/.
56. Bhandari V. Bronchopulmonary Dysplasia/Chronic Lung Disease of the Newborn. In: Buonocore G, Bracci R, Weindling M, editors. Neonatology: A Practical Approach to Neonatal Diseases. Cham: Springer International Publishing; 2018. p. 887-911.
57. JOBE AH, BANCALARI E. Bronchopulmonary Dysplasia. American journal of respiratory and critical care medicine. 2001;163(7):1723-9.
58. Tracy MC, Cornfield DN. Bronchopulmonary Dysplasia: Then, Now, and Next. Pediatr Allergy Immunol Pulmonol. 2020;33(3):99-109.
59. Harju M, Keski-Nisula L, Georgiadis L, Räisänen S, Gissler M, Heinonen S. The burden of childhood asthma and late preterm and early term births. J Pediatr. 2014;164(2):295-9.e1.
60. Priante E, Moschino L, Mardegan V, Manzoni P, Salvadori S, Baraldi E. Respiratory Outcome after Preterm Birth: A Long and Difficult Journey. Am J Perinatol. 2016;33(11):1040-2.
61. Perin J, Mulick A, Yeung D, Villavicencio F, Lopez G, Strong KL, et al. Global, regional, and national causes of under-5 mortality in 2000-19: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet Child Adolesc Health. 2022;6(2):106-15.
62. Perez GF, Pérez-Losada M, Isaza N, Rose MC, Colberg-Poley AM, Nino G. Nasopharyngeal microbiome in premature infants and stability during rhinovirus infection. J Investig Med. 2017;65(6):984-90.
63. Lauer T, Behnke J, Oehmke F, Baecker J, Gentil K, Chakraborty T, et al. Bacterial Colonization within the First Six Weeks of Life and Pulmonary Outcome in Preterm Infants <1000 g. J Clin Med. 2020;9(7).

64. Hourigan SK, Moutinho TJ, Jr., Berenz A, Papin J, Guha P, Bangiolo L, et al. Gram-negative Microbiota Blooms in Premature Twins Discordant for Parenteral Nutrition-associated Cholestasis. J Pediatr Gastroenterol Nutr. 2020;70(5):640-4.
65. Selway CA, Collins CT, Makrides M, Sullivan TR, Weyrich LS. Variable preterm oral microbiome stabilizes and reflects a full-term infant profile within three months. Pediatric research. 2023.
66. Gasparrini AJ, Wang B, Sun X, Kennedy EA, Hernandez-Leyva A, Ndao IM, et al. Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. Nat Microbiol. 2019;4(12):2285-97.
67. Lucia Hug DS, and Danzhen You. Levels & Trends in Child Mortality. UNICEF; 2017.
68. Puopolo KM, Benitz WE, Zaoutis TE. Management of Neonates Born at </=34 6/7 Weeks' Gestation With Suspected or Proven Early-Onset Bacterial Sepsis. Pediatrics. 2018;142(6).
69. Weston EJ, Pondo T, Lewis MM, Martell-Cleary P, Morin C, Jewell B, et al. The burden of invasive early-onset neonatal sepsis in the United States, 2005-2008. The Pediatric infectious disease journal. 2011;30(11):937-41.
70. Stoll BJ, Hansen NI, Sanchez PJ, Faix RG, Poindexter BB, Van Meurs KP, et al. Early onset neonatal sepsis: the burden of group B Streptococcal and E. coli disease continues. Pediatrics. 2011;127(5):817-26.
71. Stoll BJ, Hansen NI, Bell EF, Walsh MC, Carlo WA, Shankaran S, et al. Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993-2012. Jama. 2015;314(10):1039-51.
72. Schrag SJ, Farley MM, Petit S, Reingold A, Weston EJ, Pondo T, et al. Epidemiology of Invasive Early-Onset Neonatal Sepsis, 2005 to 2014. Pediatrics. 2016;138(6).
73. Puopolo KM, Benitz WE, Zaoutis TE. Management of Neonates Born at >/=35 0/7 Weeks' Gestation With Suspected or Proven Early-Onset Bacterial Sepsis. Pediatrics. 2018;142(6).
74. Rajar P, Saugstad OD, Berild D, Dutta A, Greisen G, Lausten-Thomsen U, et al. Antibiotic Stewardship in Premature Infants: A Systematic Review. Neonatology. 2020;117(6):673-86.
75. Giannoni E, Dimopoulou V, Klingenberg C, Navér L, Nordberg V, Berardi A, et al. Analysis of Antibiotic Exposure and Early-Onset Neonatal Sepsis in Europe, North America, and Australia. JAMA Network Open. 2022;5(11):e2243691-e.
76. Vatne A, Hapnes N, Stensvold HJ, Dalen I, Guthe HJ, Støen R, et al. Early Empirical Antibiotics and Adverse Clinical Outcomes in Infants Born Very Preterm: A Population-Based Cohort. J Pediatr. 2023 Feb;253:107-114.e5.
77. Cotten CM, Taylor S, Stoll B, Goldberg RN, Hansen NI, Sanchez PJ, et al. Prolonged duration of initial empirical antibiotic treatment is associated with increased rates of necrotizing enterocolitis and death for extremely low birth weight infants. Pediatrics. 2009;123(1):58-66.
78. Kuppala VS, Meinzen-Derr J, Morrow AL, Schibler KR. Prolonged initial empirical antibiotic treatment is associated with adverse outcomes in premature infants. J Pediatr. 2011;159(5):720-5.
79. Ting JY, Synnes A, Roberts A, Deshpandey A, Dow K, Yoon EW, et al. Association Between Antibiotic Use and Neonatal Mortality and Morbidities in Very Low-Birth-Weight Infants Without Culture-Proven Sepsis or Necrotizing Enterocolitis. JAMA pediatrics. 2016;170(12):1181-7.
80. Esaiassen E, Fjalstad JW, Juvet LK, van den Anker JN, Klingenberg C. Antibiotic exposure in neonates and early adverse outcomes: a systematic review and meta-analysis. The Journal of antimicrobial chemotherapy. 2017;72(7):1858-70.
81. Gasparrini AJ, Crofts TS, Gibson MK, Tarr PI, Warner BB, Dantas G. Antibiotic perturbation of the preterm infant gut microbiome and resistome. Gut Microbes. 2016;7(5):443-9.
82. Langdon A, Crook N, Dantas G. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. Genome medicine. 2016;8(1):39-.
83. Thänert R, Sawhney SS, Schwartz DJ, Dantas G. The resistance within: Antibiotic disruption of the gut microbiome and resistome dynamics in infancy. Cell Host & Microbe. 2022;30(5):675-83.
84. Salazar C, Giménez M, Riera N, Parada A, Puig J, Galiana A, et al. Human microbiota drives hospital-associated antimicrobial resistance dissemination in the urban environment and mirrors patient case rates. Microbiome. 2022;10(1):208.
85. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant

bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. The Lancet Infectious Diseases. 2019;19(1):56-66.
86. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations. 2016.
87. Laxminarayan R, Matsoso P, Pant S, Brower C, Røttingen J-A, Klugman K, et al. Access to effective antimicrobials: a worldwide challenge. The Lancet. 2016;387(10014):168-75.
88. Laxminarayan R, Bhutta ZA. Antimicrobial resistance-a threat to neonate survival. Lancet Glob Health. 2016;4(10):e676-7.
89. Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet. 2022;399(10325):629-55.
90. McDonagh MS, Peterson K, Winthrop K, Cantor A, Lazur BH, Buckley DI. Interventions to reduce inappropriate prescribing of antibiotics for acute respiratory tract infections: summary and update of a systematic review. The Journal of international medical research. 2018;46(8):3337-57.
91. Pariente N. A field is born. . Nature Milestones Human Microbiota Research. 2019
92. Peterson D, Bonham KS, Rowland S, Pattanayak CW, RC, Klepac-Ceraj V, et al. Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes. Frontiers in microbiology. 2021;12.
93. Pettigrew MM, Kwon J, Gent JF, Kong Y, Wade M, Williams DJ, et al. Comparison of the Respiratory Resistomes and Microbiota in Children Receiving Short versus Standard Course Treatment for Community-Acquired Pneumonia. mBio. 2022;13(2):e0019522-e.
94. Rubiola S, Chiesa F, Dalmasso A, Di Ciccio P, Civera T. Detection of Antimicrobial Resistance Genes in the Milk Production Environment: Impact of Host DNA and Sequencing Depth. Frontiers in microbiology. 2020;11:1983-.
95. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett C, Knight R, Gordon JI. The human microbiome project: exploring the microbial part of ourselves in a changing world. Nature. 2007;449(7164):804-10.
96. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends in Microbiology. 2019;27(2):105-17.
97. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87-.
98. Douglas CA, Ivey KL, Papanicolas LE, Best KP, Muhlhausler BS, Rogers GB. DNA extraction approaches substantially influence the assessment of the human breast milk microbiome. Scientific reports. 2020;10(1):123-.
99. Greathouse KL, Sinha R, Vogtmann E. DNA extraction for human microbiome studies: the issue of standardization. Genome Biol. 2019;20(1):212-.
100. Cardona S, Eck A, Cassellas M, Gallart M, Alastrue C, Dore J, et al. Storage conditions of intestinal microbiota matter in metagenomic analysis. BMC microbiology. 2012;12(1):158.
101. Hasrat R, Kool J, de Steenhuijsen Piters WAA, Chu MLJN, Kuiling S, Groot JA, et al. Benchmarking laboratory processes to characterise low-biomass respiratory microbiota. Scientific reports. 2021;11(1):17148.
102. Theodosiou AA, Dorey RB, Read RC, Jones CE. The infant pharyngeal microbiomes: origin, impact and manipulation. Current opinion in infectious diseases. 2020;33(6).
103. Horz HP, Scheer S, Vianna ME, Conrads G. New methods for selective isolation of bacterial DNA from human clinical specimens. Anaerobe. 2010;16(1):47-53.
104. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome. 2018;6(1):42-.
105. Shi Y, Wang G, Lau HC-H, Yu J. Metagenomic Sequencing for Microbial DNA in Human Samples: Emerging Technological Advances. International Journal of Molecular Sciences. 2022;23(4):2181.
106. Sui HY, Weil AA, Nuwagira E, Qadri F, Ryan ET, Mezzari MP, et al. Impact of DNA Extraction Method on Variation in Human and Built Environment Microbial Community and Functional Profiles Assessed by Shotgun Metagenomics Sequencing. Frontiers in microbiology. 2020;11:953.

107. Nelson MT, Pope CE, Marsh RL, Wolter DJ, Weiss EJ, Hager KR, et al. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. Cell Reports. 2019;26(8):2227-40.e5.
108. Rajar P, Dhariwal A, Salvadori G, Junges R, Åmdal HA, Berild D, et al. Microbial DNA extraction of high-host content and low biomass samples: Optimized protocol for nasopharynx metagenomic studies. Frontiers in microbiology. 2022;13:1038120.
109. Heravi FS, Zakrzewski M, Vickery K, Hu H. Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. J Microbiol Methods. 2020;170:105856.
110. Hasan MR, Rawat A, Tang P, Jithesh PV, Thomas E, Tan R, et al. Depletion of Human DNA in Spiked Clinical Specimens for Improvement of Sensitivity of Pathogen Detection by Next-Generation Sequencing. Journal of clinical microbiology. 2016;54(4):919-27.
111. Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, et al. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. PloS one. 2013;8(10):e76096.
112. Thoendel M, Jeraldo PR, Greenwood-Quaintance KE, Yao JZ, Chia N, Hanssen AD, et al. Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. J Microbiol Methods. 2016;127:141-5.
113. Gupta N. DNA Extraction and Polymerase Chain Reaction. J Cytol. 2019;36(2):116-7.
114. Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn LJ, et al. Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. Frontiers in microbiology. 2019;10:1277.
115. Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, et al. Impact of sequencing depth on the characterization of the microbiome and resistome. Scientific reports. 2018;8(1):5890.
116. Cho MY, Oliva M, Spreafico A, Chen B, Wei X, Choi Y, et al. Two-Target Quantitative PCR To Predict Library Composition for Shallow Shotgun Sequencing. mSystems. 2021;6(4):e00552-21.
117. McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. Methods for normalizing microbiome data: An ecological perspective. Methods in Ecology and Evolution. 2019;10(3):389-400.
118. Rajan SK, Lindqvist M, Brummer RJ, Schoultz I, Repsilber D. Phylogenetic microbiota profiling in fecal samples depends on combination of sequencing depth and choice of NGS analysis method. PloS one. 2019;14(9):e0222171.
119. Yigit E, Feehery GR, Langhorst BW, Stewart FJ, Dimalanta ET, Pradhan S, et al. A Microbiome DNA Enrichment Method for Next-Generation Sequencing Sample Preparation. Current Protocols in Molecular Biology. 2016;115(1):7.26.1-7..14.
120. Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific reports. 2017;7(1):17668.
121. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
122. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
123. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.
124. Beghini, Francesco, et al. "Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3." Elife 10 (2021): e65088.
125. Florensa AF, Kaas RS, Clausen P, Aytan-Aktug D, Aarestrup FM. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. Microb Genom. 2022;8(1).
126. Krueger, F. "Trim Galore: a wrapper tool around Cutadapt and FastQC." Trim Galore (2012).
127. Li, Heng. "seqtk Toolkit for processing sequences in FASTA/Q formats." GitHub 767 (2012): 69.
128. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res. 2020;48(D1):D561-D9.

129. Wickham, H. (2016). Data analysis. In ggplot2 (pp. 189-201). Springer, Cham.
130. Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic acids research. 2020;48(D1):D517-D25.
131. Dhariwal AR, P. , Salvadori G, Åmdal H, Berild D, Dahle U, Fugelseth D, et al. Unravelling the landscape of antibiotic resistance determinants in the nasopharynx and the impact of antibiotics: a longitudinal study of preterm infants. . Submitted manuscript.
132. Allaire J. RStudio: integrated development environment for R. Boston, MA. 2012;770(394):165-71.
133. R Core Team R, Team RC. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2020. 2021.
134. Oksanen, Jari, et al. "Package 'vegan'." Community ecology package, version 2.9 (2013): 1-295.
135. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PloS one. 2013;8(4):e61217.
136. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, et al. Vegan: Community Ecology Package (R Package Version 2.6-2). 2022.
137. Holmes I, Harris K, Quince C. Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics. PloS one. 2012;7(2):e30126.
138. Wickham H. ggplot2. Wiley interdisciplinary reviews: computational statistics. 2011;3(2):180-5.
139. Man WH, de Steenhuijsen Piters WA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. Nature reviews Microbiology. 2017;15(5):259-70.
140. Luna PN, Hasegawa K, Ajami NJ, Espinola JA, Henke DM, Petrosino JF, et al. The association between anterior nares and nasopharyngeal microbiota in infants hospitalized for bronchiolitis. Microbiome. 2018;6(1):2.
141. Ganda E, Beck KL, Haiminen N, Silverman JD, Kawas B, Cronk BD, et al. DNA Extraction and Host Depletion Methods Significantly Impact and Potentially Bias Bacterial Detection in a Biological Fluid. mSystems. 2021;6(3):e0061921.
142. Horz H-P, Scheer S, Huenger F, Vianna ME, Conrads G. Selective isolation of bacterial DNA from human clinical specimens. Journal of Microbiological Methods. 2008;72(1):98-102.
143. Gebert S, Siegel D, Wellinghausen N. Rapid detection of pathogens in blood culture bottles by real-time PCR in conjunction with the pre-analytic tool MolYsis. The Journal of infection. 2008;57(4):307-16.
144. Yap M, Feehily C, Walsh CJ, Fenelon M, Murphy EF, McAuliffe FM, et al. Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome. Scientific Reports. 2020;10(1):21665.
145. Salonen A, Nikkilä J, Jalanka-Tuovinen J, Immonen O, Rajilić-Stojanović M, Kekkonen RA, et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. J Microbiol Methods. 2010;81(2):127-34.
146. Abusleme L, Hong BY, Dupuy AK, Strausbaugh LD, Diaz PI. Influence of DNA extraction on oral microbial profiles obtained via 16S rRNA gene sequencing. J Oral Microbiol. 2014;6.
147. Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 2011;12(2):R18.
148. Muller BH, Mollon P, Santiago-Allexant E, Javerliat F, Kaneko G. In-depth comparison of library pooling strategies for multiplexing bacterial species in NGS. Diagnostic microbiology and infectious disease. 2019;95(1):28-33.
149. Stämmler F, Gläsner J, Hiergeist A, Holler E, Weber D, Oefner PJ, et al. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. Microbiome. 2016;4(1):28.
150. Wang X, Howe S, Deng F, Zhao J. Current Applications of Absolute Bacterial Quantification in Microbiome Studies and Decision-Making Regarding Different Biological Questions. Microorganisms. 2021;9(9).
151. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nature Reviews Microbiology. 2018;16(7):410-22.

152. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Frontiers in microbiology. 2017;8.
153. McMurdie PJ. Normalization of Microbiome Profiling Data. In: Beiko RG, Hsiao W, Parkinson J, editors. Microbiome Analysis: Methods and Protocols. New York, NY: Springer New York; 2018. p. 143-68.
154. Bosch AATM, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, et al. Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. EBioMedicine. 2016;9:336-45.
155. Aguilar-Lopez M, Dinsmoor AM, Ho TTB, Donovan SM. A systematic review of the factors influencing microbial colonization of the preterm infant gut. Gut Microbes. 2021;13(1):1-33.
156. Prevaes SM, de Winter-de Groot KM, Janssens HM, de Steenhuijsen Piters WA, Tramper-Stranders GA, Wyllie AL, et al. Development of the Nasopharyngeal Microbiota in Infants with Cystic Fibrosis. American journal of respiratory and critical care medicine. 2016;193(5):504-15.
157. Penela-Sánchez D, Rocafort M, Henares D, Jordan I, Brotons P, Cabrerizo M, et al. Impact of the bacterial nasopharyngeal microbiota on the severity of genus enterovirus lower respiratory tract infection in children: A case-control study. Pediatric pulmonology. 2023;58(6):1728-37.
158. Reyman M, van Houten MA, Watson RL, Chu M, Arp K, de Waal WJ, et al. Effects of early-life antibiotics on the developing infant gut microbiome and resistome: a randomized trial. Nature communications. 2022;13(1):893.
159. Følsgaard NV, Schjørring S, Chawes BL, Rasmussen MA, Krogfelt KA, Brix S, et al. Pathogenic bacteria colonizing the airways in asymptomatic neonates stimulates topical inflammatory mediator release. American journal of respiratory and critical care medicine. 2013;187(6):589-95.
160. Vissing NH, Chawes BL, Bisgaard H. Increased risk of pneumonia and bronchiolitis after bacterial colonization of the airways as neonates. American journal of respiratory and critical care medicine. 2013;188(10):1246-52.
161. Iwase T, Uehara Y, Shinji H, Tajima A, Seo H, Takada K, et al. Staphylococcus epidermidis Esp inhibits Staphylococcus aureus biofilm formation and nasal colonization. Nature. 2010;465(7296):346-9.
162. Yan M, Pamp Sünje J, Fukuyama J, Hwang Peter H, Cho D-Y, Holmes S, et al. Nasal Microenvironments and Interspecific Interactions Influence Nasal Microbiota Complexity and S. aureus Carriage. Cell host & microbe. 2013;14(6):631-40.
163. Biagi E, Aceti A, Quercia S, Beghetti I, Rampelli S, Turroni S, et al. Microbial Community Dynamics in Mother's Milk and Infant's Mouth and Gut in Moderately Preterm Infants. Frontiers in microbiology. 2018;9:2512.
164. Biagi E, Quercia S, Aceti A, Beghetti I, Rampelli S, Turroni S, et al. The Bacterial Ecosystem of Mother's Milk and Infant's Mouth and Gut. Frontiers in microbiology. 2017;8:1214.
165. Al'Abri IS, Durmusoglu D, Crook N. What E. coli knows about your 1-year-old infant: Antibiotic use, lifestyle, birth mode, and siblings. Cell host & microbe. 2021;29(6):854-5.
166. Zhang A-N, Gaston JM, Dai CL, Zhao S, Poyet M, Groussin M, et al. An omics-based framework for assessing the health risk of antimicrobial resistance genes. Nature communications. 2021;12(1):4765.
167. Grier A, Qiu X, Bandyopadhyay S, Holden-Wiltse J, Kessler HA, Gill AL, et al. Impact of prematurity and nutrition on the developing gut microbiome and preterm infant growth. Microbiome. 2017;5(1):158.
168. Brooks B, Firek BA, Miller CS, Sharon I, Thomas BC, Baker R, et al. Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. Microbiome. 2014;2(1):1.
169. Hartz LE, Bradshaw W, Brandon DH. Potential NICU Environmental Influences on the Neonate's Microbiome: A Systematic Review. Adv Neonatal Care. 2015;15(5):324-35.
170. Hourigan SK, Subramanian P, Hasan NA, Ta A, Klein E, Chettout N, et al. Comparison of Infant Gut and Skin Microbiota, Resistome and Virulome Between Neonatal Intensive Care Unit (NICU) Environments. Frontiers in microbiology. 2018;9(1361).
171. Moles L, Gómez M, Moroder E, Jiménez E, Escuder D, Bustos G, et al. Serratia marcescens colonization in preterm neonates during their neonatal intensive care unit stay. Antimicrobial Resistance & Infection Control. 2019;8(1):135.

172. Cristina ML, Sartini M, Spagnolo AM. Serratia marcescens Infections in Neonatal Intensive Care Units (NICUs). Int J Environ Res Public Health. 2019;16(4).
173. Dessì A, Puddu M, Testa M, Marcialis MA, Pintus MC, Fanos V. Serratia marcescens infections and outbreaks in neonatal intensive care units. Journal of chemotherapy (Florence, Italy). 2009;21(5):493-9.
174. Liu CM, Price LB, Hungate BA, Abraham AG, Larsen LA, Christensen K, et al. Staphylococcus aureus and the ecology of the nasal microbiome. Science Advances. 2015;1(5):e1400216.
175. Muyldermans A, Crombé F, Bosmans P, Cools F, Piérard D, Wybo I. Serratia marcescens outbreak in a neonatal intensive care unit and the potential of whole-genome sequencing. J Hosp Infect. 2021;111:148-54.
176. Kelderer F, Mogren I, Eriksson C, Silfverdal SA, Domellöf M, West CE. Associations between pre- and postnatal antibiotic exposures and early allergic outcomes: A population-based birth cohort study. Pediatr Allergy Immunol. 2022;33(9):e13848.
177. Lu Y, Wang Y, Wang J, Lowe AJ, Grzeskowiak LE, Hu YJ. Early-Life Antibiotic Exposure and Childhood Asthma Trajectories: A National Population-Based Birth Cohort. Antibiotics. 2023;12(2):314.
178. Sanyang B, de Silva TI, Kanteh A, Bojang A, Manneh J, Piters W, et al. Effect of intra-partum azithromycin on the development of the infant nasopharyngeal microbiota: A post hoc analysis of a double-blind randomized trial. EBioMedicine. 2022;83:104227.
179. Zwittink RD, Renes IB, van Lingen RA, van Zoeren-Grobben D, Konstanti P, Norbruis OF, et al. Association between duration of intravenous antibiotic administration and early-life microbiota development in late-preterm infants. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology. 2018;37(3):475-83.
180. Vatne A, Klingenberg C, Øymar K, Rønnestad AE, Manzoni P, Rettedal S. Reduced Antibiotic Exposure by Serial Physical Examinations in Term Neonates at Risk of Early-onset Sepsis. The Pediatric infectious disease journal. 2020;39(5):438-43.
181. Salas Garcia S, Valcarcel Ruescas B, Aristoy Zabaleta M, León García S, Parra-Llorca A, Toledo JD. Early-onset sepsis: a cohort study comparing serial clinical observation with laboratory testing. Journal of perinatology : official journal of the California Perinatal Association. 2021;41(10):2482-7.
182. Arboleya S, Saturio S, Gueimonde M. Impact of intrapartum antibiotics on the developing microbiota: a review. Microbiome Research Reports. 2022;1(4):22.
183. Bojang A, Baines SL, Camara B, Guerillot R, Donovan L, Marqués RS, et al. Impact of Intrapartum Oral Azithromycin on the Acquired Macrolide Resistome of Infants' Nasopharynx: A Randomized Controlled Trial. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2020;71(12):3222-5.
184. GARDP study reveals that babies are increasingly dying of neonatal sepsis caused by drug-resistant bacterial infections [press release]. 06.04.2022 2022.
185. Carr VR, Witherden EA, Lee S, Shoaie S, Mullany P, Proctor GB, et al. Abundance and diversity of resistomes differ between healthy human oral cavities and gut. Nature communications. 2020;11(1):693.
186. Mallick H, Franzosa EA, McLver LJ, Banerjee S, Sirota-Madi A, Kostic AD, et al. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. Nature communications. 2019;10(1):3136.

ERRATALISTE

Navn kandidat: Polona Rajar

Avhandlingstittel: Development of the respiratory microbiome and resistome in preterm infants: Shotgun metagenomics analysis of nasopharyngeal aspirate samples

Forkortelser for type rettelser:

Cor – korrektur

Celtf – endring av sidelayout eller tekstformat

Side	Linje	Fotnote	Originaltekst	Type rettelse	Korrigert tekst
65	Figure 7 legend		.. other than only ampicillin plus gentamycin..	cor	..ampicillin plus gentamycin ..
67	Figure 8 legend		.. ampicillin plus gentamycin for suspected early onset neonatal sepsis.	cor	.. other antibiotics than only ampicillin plus gentamycin.

I

I

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 21 December 2022
DOI 10.3389/fmicb.2022.1038120

OPEN ACCESS

EDITED BY
Gaofeng Jiang,
Wuhan University of Science and Technology, China

REVIEWED BY
Fatemah Sadeghpour Heravi,
Macquarie University, Australia
Stephen Welty,
University of Washington, United States

*CORRESPONDENCE
Fernanda Cristina Petersen
f.c.petersen@odont.uio.no

†These authors share last authorship

SPECIALTY SECTION
This article was submitted to Microbial Immunology, a section of the journal Frontiers in Microbiology

RECEIVED 06 September 2022
ACCEPTED 28 November 2022
PUBLISHED 21 December 2022

CITATION
Rajar P, Dhariwal A, Salvadori G, Junges R, Åmdal HA, Berild D, Fugelseth D, Saugstad OD, Lausten-Thomsen U, Greisen G, Haaland K and Petersen FC (2022)
Microbial DNA extraction of high-host content and low biomass samples: Optimized protocol for nasopharynx metagenomic studies.
Front. Microbiol. 13:1038120.
doi: 10.3389/fmicb.2022.1038120

Microbial DNA extraction of high-host content and low biomass samples: Optimized protocol for nasopharynx metagenomic studies

Polona Rajar[1,2], Achal Dhariwal[2], Gabriela Salvadori[2], Roger Junges[2], Heidi Aarø Åmdal[2], Dag Berild[3,4], Drude Fugelseth[1,4], Ola Didrik Saugstad[5], Ulrik Lausten-Thomsen[6], Gorm Greisen[6], Kirsti Haaland[1†] and Fernanda Cristina Petersen[2*†]

[1]Department of Neonatal Intensive Care, Division of Paediatric and Adolescent Medicine, Oslo University Hospital Ullevål, Oslo, Norway, [2]Institute of Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway, [3]Department of Infectious Diseases, Oslo University Hospital, Oslo, Norway, [4]Institute of Clinical Medicine, Faculty of Medicine, Oslo University, Oslo, Norway, [5]Department of Paediatric Research, University of Oslo, Oslo, Norway, [6]Department of Neonatology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

Introduction: Low microbial biomass and high human DNA content in nasopharyngeal aspirate samples hinder comprehensive characterization of microbiota and resistome. We obtained samples from premature infants, a group with increased risk of developing respiratory disorders and infections, and consequently frequent exposure to antibiotics. Our aim was to devise an optimal protocol for handling nasopharyngeal aspirate samples from premature infants, focusing on host DNA depletion and microbiome and resistome characterization.

Methods: Three depletion and three DNA extraction protocols were compared, using RT-PCR and whole metagenome sequencing to determine the efficiency of human DNA removal, taxonomic profiling and assignment of antibiotic resistance genes. Protocols were tested using mock communities, as well as pooled and individual patient samples.

Results: The only extraction protocol to retrieve the expected DNA yield from mock community samples was based on a lytic method to improve Gram positive recovery (MasterPure™). Host DNA content in non-depleted aliquots from pooled patient samples was 99%. Only samples depleted with MolYsis™ showed satisfactory, but varied reduction in host DNA content, in both pooled and individual patient samples, allowing for microbiome and resistome characterisation (host DNA content from 15% to 98%). Other depletion protocols either retrieved too low total DNA yields, preventing further analysis, or failed to reduce host DNA content. By using Mol_MasterPure protocol on aliquots from pooled patient samples, we increased the number of bacterial reads by 7.6 to 1,725.8-fold compared to non-depleted reference samples. PCR results were indicative of achieved microbial enrichment. Individual patient samples processed with Mol_MasterPure protocol varied greatly in total DNA yield, host DNA content (from 40% to 98%), species and antibiotic resistance gene richness.

Discussion: Despite high human DNA and low microbial biomass content in nasopharynx aspirates of preterm infants, we were able to reduce host DNA content to levels compatible with downstream shotgun metagenomic analysis, including bacterial species identification and coverage of antibiotic resistance genes. Whole metagenomic sequencing of microbes colonizing the nasopharynx may contribute to explaining the possible role of airway microbiota in respiratory conditions and reveal carriage of antibiotic resistance genes.

KEYWORDS

microbiome, whole metagenomic sequencing, host DNA depletion, low biomass, respiratory microbiome, resistome, premature infant, antimicrobial resistance

Introduction

Sequencing technologies have given us insight into the detailed structure of microbial communities inhabiting various niches of the human body. Evidence that microbiome composition and interactions with host cells influence human physiology and pathology are being increasingly reported in the literature. So far, most studies have focused on the gut microbiome, partly due to its abundance and accessibility. Microbial communities in sites with low microbial biomass such as the nasopharynx, are more challenging and less investigated (Biesbroek et al., 2014). The nasopharyngeal microbiome has the potential to carry implications for disease of upper and lower respiratory tract (Mizgerd, 2014; Teo et al., 2018). Common colonizers of the nasopharynx include species with high pathogenic potential (e.g., *Streptococcus pneumoniae*, *Klebsiella pneumoniae, Staphylococcus aureus*), as well as colonizers that seldom cause diseases, but can serve as a reservoir of antibiotic resistance genes (ARG; Morley et al., 2019; Manenzhe et al., 2020).

Microbiome develops most dynamically in the first 2–3 years of life (Milani et al., 2017). Factors influencing early colonization can carry serious health implications early and later in life (Rautava et al., 2012). Premature infants have immature immune system and are often early exposed to antibiotics, disrupting the developing microbiome, increasing the presence of ARG and thus also contributing to increased antimicrobial resistance in general, one of the main threats to global health (World Health Organization, 2014; Gasparrini et al., 2019). Application of next generation sequencing technology broadens knowledge of the effect of the different variables on the microbiome and resistome, and the discovery of ARG.

The methodological and financial obstacles are substantial in low microbial biomass and high host DNA samples (Shi et al., 2022). Efficient removal of host DNA is necessary before sequencing, to allow for a cost and time efficacy and precise analysis of the samples (Nelson et al., 2019). Furthermore, due to its low microbial biomass, such samples are more subjected to biases or false positives due to contamination during sampling and processing (Salter et al., 2014; Eisenhofer et al., 2019; Douglas et al., 2020). Recently published minimal standards requirements for microbiome studies (Greathouse et al., 2019) should be followed while striving further towards the establishment of generally accepted and applied standard operating procedures for different sample types for human microbiome studies. To date, there is a lack of established standard operating procedures for low biomass samples (Theodosiou et al., 2020; Hasrat et al., 2021).

The aim of this study was to compare the efficiency of different protocols combining host DNA depletion and microbial DNA extraction from nasopharyngeal aspirates of premature infants, for the purpose of microbiome and resistome profiling using whole metagenomic sequencing (WMS).

Materials and methods

Ethical statement

The study was performed in accordance with the Declaration of Helsinki and approved by the Hospital's Data Protection Officer and the Regional Committee for Medical and Health Research Ethics - South East, Norway (2018/1381 REKD), and by the Danish National Committee for Health Research Ethics (H-180512193). Written informed consent was obtained from infant's parents. The participants received no compensation.

Samples and experimental design

Patient samples

Nasopharyngeal aspirate samples (n = 42) were obtained from premature infants born between 28^{+0} and 31^{+6} weeks gestational age during their stay at the Neonatal Intensive Care units at Ullevaal, Oslo University Hospital, Oslo, Norway and Rigshopitalet, Copenhagen, Denmark. Samples were obtained by inserting a sterile suction catheter along the nasal wall into the nasopharynx, applying vacuum suction for 5 s and removing the suction catheter without active suction. Standard protection equipment to avoid contamination was used. There was no pre-moisturizing of the suction catheter. A sterile 2 ml 20% glycerol solution was suctioned directly afterwards through the suction catheter to rinse any mucus remains and for cryopreservation of the sample.

The samples were rapidly moved to −80°C, where they were stored for up to 10 months. In the laboratory, 18 random samples were divided into Pools A, B, C, each comprising samples from six infants and processed according to different protocols (Table 1). Six samples obtained within 24 h of birth were pooled into Pool D. This experimental design was chosen so that each pool would have sufficient material to be tested with different protocols. Two aliquots from pool D were spiked with mock community (Zymo, D6300) prior to host DNA depletion and DNA extraction to create more diverse samples. Eighteen patient samples were processed individually. They were spiked with Spike-in Control II for Low Microbial Load samples (Zymo, D6321 & D6321-10); 12 samples prior and three post host DNA depletion. Estimated DNA yield of the spike-in was 0.4 ng. The Spike-in Control standard is composed of three species not found in the human microbiome. (*Truepera radiovictrix*, *Imtechella halotolerans* and *Allobacillus halotolerans*). From these, we used *I. haloterans* counts as a reference for quantification of total microbial load. The other two species in the spike-in were excluded from the analysis as they were either non-susceptible to chemical lysis (*T. radiovictrix*) or not in the MetaPhlan database used in the downstream taxonomic analysis (*A. haloterans*). Experimental design is illustrated in Figure 1.

Positive and negative control

Mock community samples used for positive control were prepared to match the low concentration of DNA found in nasopharyngeal aspirate samples from premature infants measured in a pilot study (data not shown). Two microliter of mock community (Zymo, D6300) with expected DNA yield of approximately 55 ng were diluted in 2 ml of sterile 20% glycerol solution, to resemble the preparation of patient samples. The

TABLE 1 Host DNA depletion and DNA extraction protocols.

Protocol name	Host DNA depletion kit	DNA extraction kit	Deviation from manufacturer's protocol
MasterPure	None	MasterPure™ Gram Positive DNA Purification Kit (Epicentre, Madison, WI, United States)	Followed the manufacturer's protocol (Available at: https://www.lucigen.com/docs/manuals/MA209E-MasterPure-Gram-Positive-DNA.pdf)
MagMAX	None	MagMAX™ Microbiome Ultra Nucleic Acid Isolation Kit (Applied Biosystems, Foster City, CA)	We followed the protocol for High throughput isolation of Nucleic Acid (RNA and DNA) from soil, biofluids, and other samples using Bead tubes and the KingFisher™ Duo Prime (Avaliable at: https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0018070_MagMAXMicrobiomeNuclAcidIsolatKit_SoilSalivaUrine_Automated_UG.pdf)
QIAamp	None	QiAmp DNA Microbiome Kit (Qiagen, Hilden, Germany)	For DNA extraction, we followed the protocol (Available at: https://www.qiagen.com/us/resources/resourcedetail?id=c403392b-0706-45ac-aa2e-4a75acd21006&lang=en), starting with step 6. (bacterial cells lysis).
PMA_MasterPure	lyPMA	MasterPure™ Gram Positive DNA Purification Kit	We followed the published method protocol for host DNA depletion with lyPMA (Marotz et al., 2018). DNA extraction was performed as described in MasterPure protocol.
PMA_MagMax	lyPMA	MagMAX™ Microbiome Ultra Nucleic Acid Isolation Kit	We followed the published method protocol for host DNA depletion with lyPMA (Marotz et al., 2018). DNA extraction was performed as described in MagMax protocol.
Mol_MasterPure	MolYsis™ Basic5 (Molzym, Bremen, Germany)	MasterPure™ Gram Positive DNA Purification Kit	We followed the manufacturer's protocol (Available at: http://www.goffinmoleculartechnologies.com/wp-content/uploads/2012/01/MolYsis_Basic5_V3.0.pdf) for 1 ml samples and accordingly doubled the volume of reagents used in points 1. and 2. DNA extraction as described in MasterPure protocol.
Mol_MagMax	MolYsis™ Basic5	MagMAX™ Microbiome Ultra Nucleic Acid Isolation Kit	We followed the manufacturer's protocol (Available at: http://www.goffinmoleculartechnologies.com/wp-content/uploads/2012/01/MolYsis_Basic5_V3.0.pdf) for 1 ml samples, and accordingly doubled the volume of reagents used in steps 1. and 2. DNA extraction as described in MagMax protocol.
QIA_QIAamp	QIAamp	QIAamp DNA Microbiome Kit	We followed the protocol (Available at: https://www.qiagen.com/us/resources/resourcedetail?id=c403392b-0706-45ac-aa2e-4a75acd21006&lang=en) for 1 ml samples and accordingly doubled to volume of the reagent used in step 1.

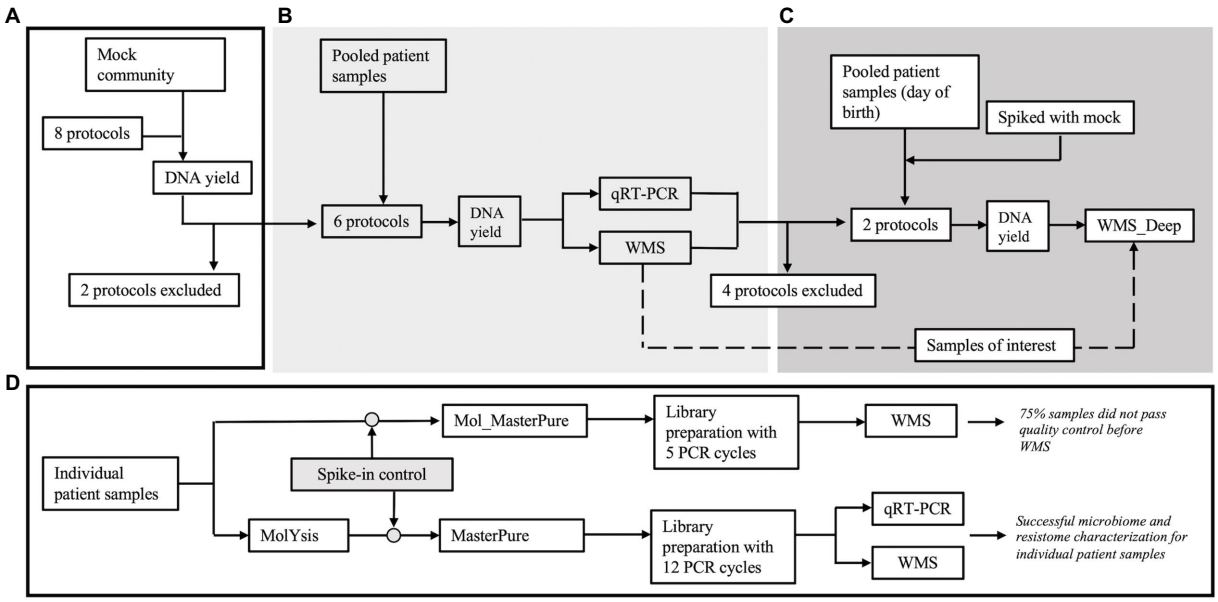

FIGURE 1
Overview of the experiments. **(A)** Eight protocols were initially tested with Mock community samples (Zymo, D6300) to evaluate loss of DNA during processing. Two protocols were excluded prior to processing clinical samples. **(B)** Three pools (A, B, and C) were created by blindly pooling six 2ml samples from premature infants for each pool. Six aliquots (2ml each) from each pool were processed according to the remaining six protocols. **(C)** After reviewing results, four protocols were excluded and an additional pool was created using six samples from premature infants, obtained right after birth (Pool D). Six aliquots (2ml) from this pool were processed according to the remaining two protocols. Two of the aliquots were spiked with mock community. **(D)** The most promising protocol (Mol_MasterPure) was further tested on individual patient samples.

samples were placed on ice and processed immediately. All experiments with mock community samples were run in triplicates.

The cryoprotectant (2 ml sterile 20% glycerol solution, prepared in sterile conditions) used for clinical samples was vacuum suctioned into a sterile mucus trap at the NICU, under the same conditions as when obtaining samples from the infants, and later processed with each extraction method, serving as a control for contamination during its production and the sampling procedure. Reagent controls (for each used extraction kit) were extracted with each extraction method and served as controls for kit contamination. The negative controls had too low concentration to be used for library prep and were excluded from further processing.

Host DNA depletion and DNA extraction methods

Samples were processed with combinations of different depletion and extraction methods (Table 1). The starting volume for all samples was approximately 2 ml, and the final DNA elution volume 30–50 µl. Samples underwent no additional freeze–thaw cycle prior to completed DNA extraction. The amount of extracted DNA was measured using Qubit™ dsDNA HS kit, on a Qubit 2.0 Fluorometer (Life Technologies, Darmstadt, Germany) and NanoDrop™ spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States).

Real time PCR

Human DNA was amplified using the primer pair FP1065 5' GCCCGTTCAGTCTCTTCGATT and FP1066 5' CAAGGCAAAGCGAAATTGGT for the RPL30 gene, and bacterial DNA using the 16S rRNA universal primers FP1067 5' CCATGAAGTCGGAATCGCTAG and FP 1068 5' GCTTGACGGGCGGTGT (Yigit et al., 2016). All reactions were performed in duplicates. The final PCR reaction volume was 25 µl, comprising 12.5 µl Maxima SYBR Green/ROX qPCR Master Mix (2×) containing Maxima Hot Start Taq DNA Polymerase, dNTPs and SYBR Green I in an optimized PCR buffer with ROX passive reference dye, 1 µl DNA template (up to 70 ng), 0.4 µM forward and reverse primers, 1× SYBR green (Life Technologies), and the remainder nuclease-free water. The amplification was carried out over 40 cycles (30 s at 98°C, 60 s at 55°C, 60 s at 72°C) with an initial 10 min hot start at 95°C. Bacterial enrichment was calculated as relative values after normalizing all the data against human DNA and comparing it to non-enriched samples.

Additionally, Femto™ Quantification kits for host and bacterial DNA (Zymo, E2005 and E2006) were used according to the manufacturer's instructions for four individual patient samples. Used sample volume was 1 µl.

Library preparation

Nextera DNA Flex Library Prep Kit (Illumina Inc., San Diego, CA, United States) was used for library preparation, following manufacturer's protocol. The only deviation was initially using five PCR amplification cycles for all library preparations (against producers' recommendations of 12 PCR cycles for low DNA input), to reduce bias and enable comparison between samples. However, individual patient samples retrieved very low DNA yields and only three of the first 12 samples passed quality control and were sequenced. To optimize this step, we increased the PCR cycle number to 12 for six additional samples and used DNA input comparable to the DNA yield of the first 12 samples (6 ng). Library concentration and purity were measured with Qubit™ dsDNA HS kit on a Qubit 2.0 Fluorometer (Life Technologies, Darmstadt, Germany), NanoDrop™ spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States) and Bioanalyzer 2,100 (Agilent, Santa Clara, CA, United States).

Sequencing

WMS was conducted at the Norwegian Sequencing Centre (Oslo, Norway). WMS was run on an NovaSeq SP platform (Illumina Inc., San Diego, CA, United States) using a paired-end sequencing approach with a targeted read length of 125 bp in high-output mode.

Data analysis

Quality of raw reads was assessed using FASTQC (Andrews, 2010). Adapter sequences and low-quality reads were removed with Trimmomatic (Bolger et al., 2014). Further on, filtered quality reads were aligned to human reference genome using Bowtie2 (Langmead and Salzberg, 2012) in order to remove human DNA contamination. The remaining high quality clean reads were used for microbiome and resistome profiling. Microbiota profiling was done with Metaphlan3 (Beghini et al., 2021). For the resistome analysis, the quality-filtered, clean reads were provided as input to Bowtie2 (Langmead and Salzberg, 2012) alignment using default parameters to the ResFinder database (Florensa et al., 2022). Reads were assigned to ARGs using an 80% gene coverage/fraction threshold. Counts of reads aligned to the ARGs were then used for downstream comparative analyses.

Rarefaction analysis

We performed rarefaction analysis to estimate the required sequencing depth needed to characterize microbiome and resistome at various taxonomic levels. Seqtk tool (Li, 2012) was used to sample clean reads into subsamples at various depths (10, 25, 30 50, 75% etc.), followed by taxonomic profiling using Metaphlan 3 (Beghini et al., 2021) to report the number of species present within each subsample. RarefactionAnalyzer tool of the AMRPlusPlus pipeline (Doster et al., 2020) was used with 5% subsampling increments of the read data with 10 iterations at each level for resistome rarefaction analysis. The numbers of unique species, genes, mechanisms, and classes were plotted as a function of sampling depth using the ggplot2 package in R (Wickham, 2016).

Results

DNA recovery from low biomass microbial mock samples

Some microbial DNA may be lost during both host DNA depletion and DNA extraction procedures. We explored this using low concentrations of a defined microbial mock community with an expected DNA yield of 55 ng (2 µl, Zymo D6300). Of the three different commercial kits for microbial DNA extraction, the approximate expected yield was recovered only with MasterPure protocol (Figure 2A). We continued with testing protocols that deplete host DNA pre-extraction using 2 µl of Zymo mock community. Among the five protocols with DNA depletion, recovery was highest with Mol_MasterPure protocol, which retrieved on average (SD) 23.5 (9.7) ng DNA. Four other protocols with depletion retrieved on average less than 25% of DNA yield, relative to the yield obtained with Mol_MasterPure protocol (Figure 2B).

DNA recovery from patient samples and evaluation of microbial enrichment through real-time qPCR

To test the performance of the different combinations of depletion and extraction methods in human low biomass samples, nasopharynx aspirates from premature infants were blindly grouped into three pools (A, B, C), each comprising samples from 6 different infants. MasterPure, the protocol that retrieved the highest DNA amount using mock communities (Figure 2A) served as a non-depletion reference protocol to which the other protocols were compared to. All protocols that included a host DNA depletion step showed a reduction in total DNA recovery compared to the no depletion reference (Figure 3A). Sixteen samples from pools A, B and C were evaluated with real time qPCR to determine whether the proportion of microbial DNA/host DNA increased following DNA depletion. Two samples (QA_QIAamp from pool A and Mol_MagMax from pool C) were excluded due to low DNA yield. Protocols using lyPMA for depletion recovered on average more than 19% of DNA yield (Figure 3B) and showed the poorest performance in microbial DNA enrichment (Figures 3C–E) under tested conditions. Lower recovery of total DNA and higher enrichment were seen with protocols using MolYsis or QIAamp for depletion.

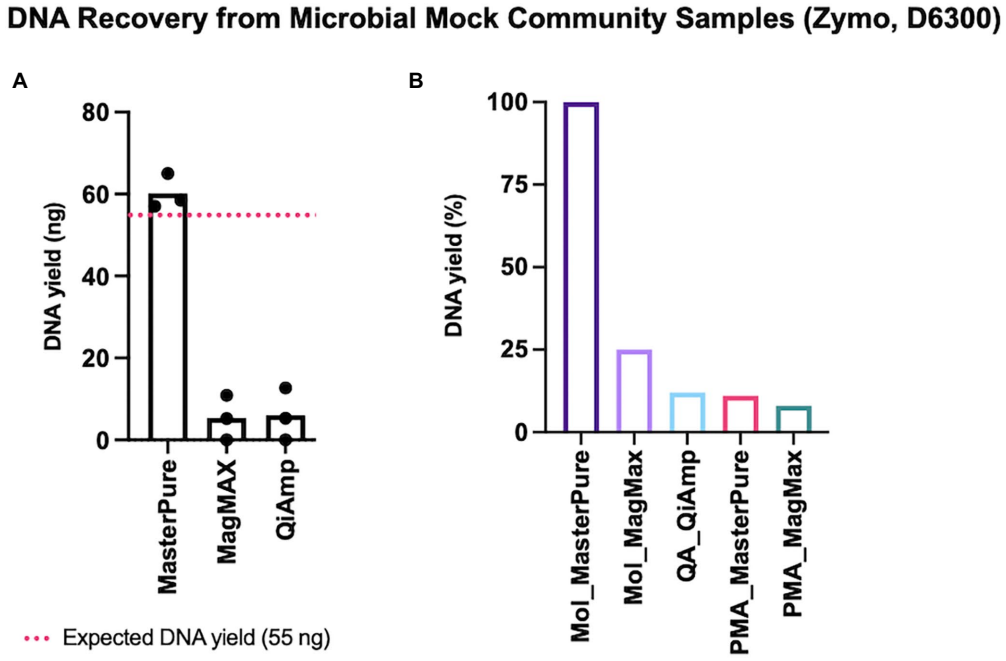

FIGURE 2
DNA recovery from mock samples. **(A)** Three commercial DNA extraction kits without depletion steps were tested. Estimated input was 55ng (dashed red line). Values for individual samples are presented as dots. Bars correspond to mean values from three independent experiments. **(B)** Retrieved DNA yield using protocols composed of host DNA depletion and DNA extraction steps. Values were normalized to the mean yield obtained with Mol_MasterPure protocol. Bars represent retrieved mean DNA yield relative to the mean yield obtained with Mol_MasterPure protocol.

Evaluation raw reads with whole metagenome sequencing at increasing sequencing depth

Sixteen samples from patient pools A, B and C with DNA concentration above 0.01 ng/μl were further processed for library preparation. Eight of the 16 samples produced libraries with concentration > 10 nM, minimum threshold recommended by the sequencing provider, and underwent WMS performed at an average depth of 15 M reads per sample. Detailed information regarding library preparation and raw reads are listed in Table 2. Information regarding excluded samples is listed in the Online Supplement (Supplementary Table 1). In reference samples (no depletion, MasterPure protocol) from all three pools the percentage of reads belonging to host DNA was 99%. All six samples processed with QIA_QIAamp and PMA_MagMax were excluded prior to WMS (Supplementary Table 1). Five samples processed with other host DNA depletion protocols passed the criteria to proceed to WMS (Table 2). Their host DNA content is shown in Figure 3D. From Pool D, all six aliquots were further processed for library preparation. One sample was excluded prior to WMS (Supplementary Table 1). Sequencing depth was an average of 54 M reads per sample (Protocol Name_Deep). Detailed information regarding library prep and raw reads are listed in Table 2.

Impact of different methods and sequencing depth on microbiome profile characterization

Three samples processed with the protocol Mol_MasterPure, which showed promising results with regards to host DNA removal, and one reference sample (Pool C) were additionally sequenced with increased depth (approximately 54 M reads per sample; Protocol Name_Deep, Figure 1C) to explore the influence of sequencing depth on the recovered bacterial reads and ARGs (Table 2).

Rarefaction analysis was performed to investigate whether enough bacterial reads were obtained to represent the species richness in each of the samples (Figure 4). Mol_MasterPure protocol preformed best across all patient pools, compared with other protocols. Sufficient sequencing depth to characterize the microbiome on species level was obtained also from two samples processed with PMA_MasterPure (Pool A) and Mol_MagMax protocol (Pool B; Figure 4). From pool D, only the spiked aliquots (Mol_MasterPure, spiked) showed a greater reduction in the host

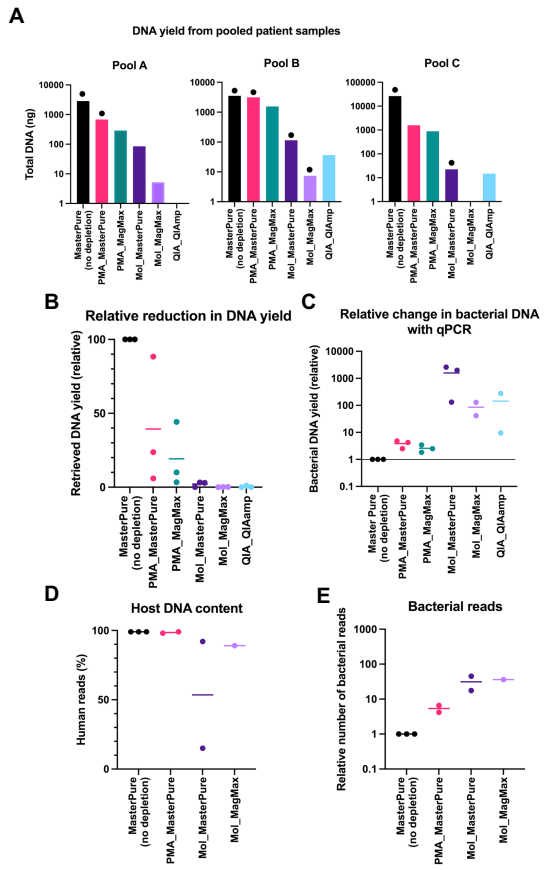

FIGURE 3
Pooled patient samples. **(A)** DNA yield extracted from pools A, B, and C according to different protocols. Each bar represents DNA yield of one sample. Samples that proceeded to WMS are marked with black dots (library concentration>10nM). **(B)** Comparison of relative reduction in total DNA yield with protocols for DNA depletion. MasterPure (no depletion) was used as a reference. **(C)** Relative change in the proportion of microbial DNA/host DNA, evaluated with real-time qPCR. **(D)** Host DNA content evaluated with WMS in samples from patients' pools A, B, and C. MasterPure protocol served as a reference. **(E)** Increase in number of bacterial reads in depleted samples, relative to non-depleted reference samples. **(B–E)** Values for individual samples are presented as dots. Bars correspond to mean values from samples from different patient pools, processed with the same protocol.

DNA content allowing to capture full diversity of species according to the rarefaction analysis (Figure 4). Merged data from both WMS rounds, comparing reference (MasterPure) and Mol_MasterPure protocols, showed that the Mol_MasterPure protocol resulted in (mean, range) 495.6 (7.6 to 1,725.8) -fold increase in the number of bacterial reads (Figure 5).

Relative abundance (Supplementary Figure 1) remained similar when the same samples were analysed after shallow and deep WMS, on genus and species level. No visible differences in the taxonomic composition were observed between aliquots processed with different protocols. From pool D, only the two aliquots spiked with Zymo mock (D6300) passed the rarefaction analysis (Supplementary Figure 1).

Effect of host DNA content and sequencing depth on resistome characterization

From the 17 samples sent to WMS, ARG could be assigned from datasets of nine samples. Eight of these samples were depleted with MolYsis, and two (Mol_MasterPure sample from pool B and C) were sequenced at two different sequencing depths.

All three samples originating from pool A, processed with different protocols, had a similar number of total reads (11–16 M reads), but differed in host DNA content and consequently number of bacterial reads (Table 2). The Mol_MasterPure aliquot had the lowest host DNA content (19%) and the highest number of unique antimicrobial resistance (AMR) determinants on all annotation levels (Figure 6). The two samples from pool B had equal number of total reads (16 M) and similar host DNA content (Mol_MP 92%, Mol_MM 89%). Increasing the sequencing depth of Mol_MasterPure to 67 M reads resulted in detection of unique AMR determinants. Sample Mol_MasterPure from pool C had the lowest host DNA content (15%) and was sequenced at two sequencing depths. Obtaining 68 M vs. 16 M total reads resulted in detection of 29 unique AMR determinants on allele level, with no changes on mechanism, class and ARG level. Increase of AMR determinants at various resistome classification levels resulting from both reduction in host DNA content and increase in sequencing depth is shown in Figure 6. The resistome composition at different annotation levels is shown in Supplementary Figure 2.

To determine the necessary sequencing depth for resistome characterisation we further performed a rarefaction analysis at different annotation levels for two samples (Mol_MasterPure) originating from different patient pools (B and C) sequenced at two sequencing depths. Samples from pool B still presented high host DNA content (92%) after depletion with MolYsis. The increase in sequencing depth improved resistome characterization, reaching saturation at mechanism and class levels, but the rarefaction curves at gene and allele levels still did not appear to have reached the plateau (Figure 7). Samples from pool C had a lower host DNA content. The rarefaction analysis for mechanism, class and gene reached the saturation plateau already at sequencing depth of 16 M. Increasing the sequencing depth to 68 M improved the resolution on allele level as well, but the rarefaction analysis suggests further increase in sequencing depth might increase the number of characterized alleles (Figure 7).

Individual patient samples

We further tested the Mol_MasterPure protocol on 18 patient samples obtained at timepoints from birth to 6 months corrected age, to explore how the protocol performs despite the variations expected in individual samples (Figure 1D). Twelve samples prepared identical to the pooled samples, using five PCR cycles during library preparation, had very low yield after library preparation (median, range: 6.48 ng, 5.88–123 ng). Nine

TABLE 2 Information regarding library preparation, host DNA content and number of bacterial, and ARG associated reads.

Pool	Protocol name	Library DNA input (ng)	Library conc. (ng/μl)	Molarity pool sent for seq. (nM)	Total reads	Preprocessed reads	Human reads		Remaining reads		Bacterial reads	Bacterial reads (% of Remaining reads)	ARG richness (No of ARG)	ARG associated reads	ARG associated reads (% of Remaining reads)
Pool A	MasterPure (no depletion)	499.4	14.6	10	15.83 M	12.57 M	12.44 M	99%	0.13 M	1.0%	0.02 M	14.8%	3	178	0.10%
	PMA_MasterPure	408	4.7	10	15.63 M	9.01 M	8.84 M	98%	0.17 M	1.9%	0.11 M	62.0%	8	612	0.40%
	Mol_MasterPure_Deep*	50.7	0.1	0.4 (low)	11.12 M	7.46 M	1.43 M	19%	6.04 M	80.9%	5.49 M	90.9%	20	90,326	1.50%
Pool B	MasterPure (no depletion)	501.3	16.6	10	15.86 M	12.38 M	12.27 M	99%	0.11 M	0.9%	0.00 M	1.4%	0	0	0.00%
	PMA_MasterPure	499.2	4	10	13.31 M	9.87 M	9.75 M	99%	0.12 M	1.2%	0.01 M	9.5%	0	0	0.00%
	Mol_MasterPure	68.4	5.7	10	17.74 M	12.88 M	11.79 M	92%	1.09 M	8.5%	0.71 M	65.2%	8	780	0.10%
	Mol_MasterPure_Deep	68.4	5.7	6	66.89 M	51.11 M	46.68 M	91%	4.43 M	8.7%	2.87 M	64.7%	12	4,699	0.10%
	Mol_MagMax	4.4	3.9	10	15.30 M	13.22 M	11.80 M	89%	1.42 M	10.7%	0.74 M	52.2%	11	1,318	0.10%
Pool C	MasterPure (no depletion)	530	15.6	10	14.43 M	12.77 M	12.66 M	99%	0.11 M	0.9%	0.00 M	4.2%	0	0	0.00%
	MasterPure_Deep (no depletion)	530	15.6	6	66.71 M	61.59 M	61.07 M	99%	0.52 M	0.8%	0.02 M	4.7%	0	0	0.00%
	Mol_MasterPure	13.4	3.2	10	15.62 M	13.17 M	2.03 M	15%	11.14 M	84.6%	8.19 M	73.5%	32	49,578	0.40%
	Mol_MasterPure_Deep	13.4	3.2	6	67.95 M	58.12 M	8.73 M	15%	49.38 M	85.0%	36.91 M	74.7%	32	225,237	0.50%
Pool D	MasterPure.1_Deep (no depletion)	502.2	22.8	6	54.27 M	49.44 M	49.04 M	99%	0.41 M	0.8%	0.02 M	4.1%	0	0	0.00%
	MasterPure.2_Deep (no depletion)	177.6	17.8	6	54.29 M	49.43 M	49.02 M	99%	0.41 M	0.8%	0.01 M	2.1%	0	0	0.00%
	Mol_MasterPure.1_Deep	18.4	0.8	6	44.53 M	29.51 M	29.01 M	98%	0.49 M	1.7%	0.13 M	25.5%	13	1,350	0.30%
	Mol_MasterPure.1_Deep, spiked	19.9	0.5	6	53.49 M	28.06 M	9.97 M	36%	18.10 M	64.5%	17.02 M	94.0%	8	67,411	0.40%
	Mol_MasterPure.2_Deep, spiked	103.8	1.4	6	63.94 M	54.21 M	23.39 M	43%	30.82 M	56.9%	28.02 M	90.9%	17	55,989	0.20%

Rows in grey correspond to parallel samples that have been sequenced deeper. The initial threshold for sending libraries for sequencing was reduced from 10 to 6 nM for the samples with the lowest biomass, pool D, to enable inclusion of all five samples at equal quantities in the sequencing pool. *Sample Mol_MasterPure from pool A had too low concentration to be included in the equimolar pool but was added to the second sequencing round at a lower concentration.

10.3389/fmicb.2022.1038120

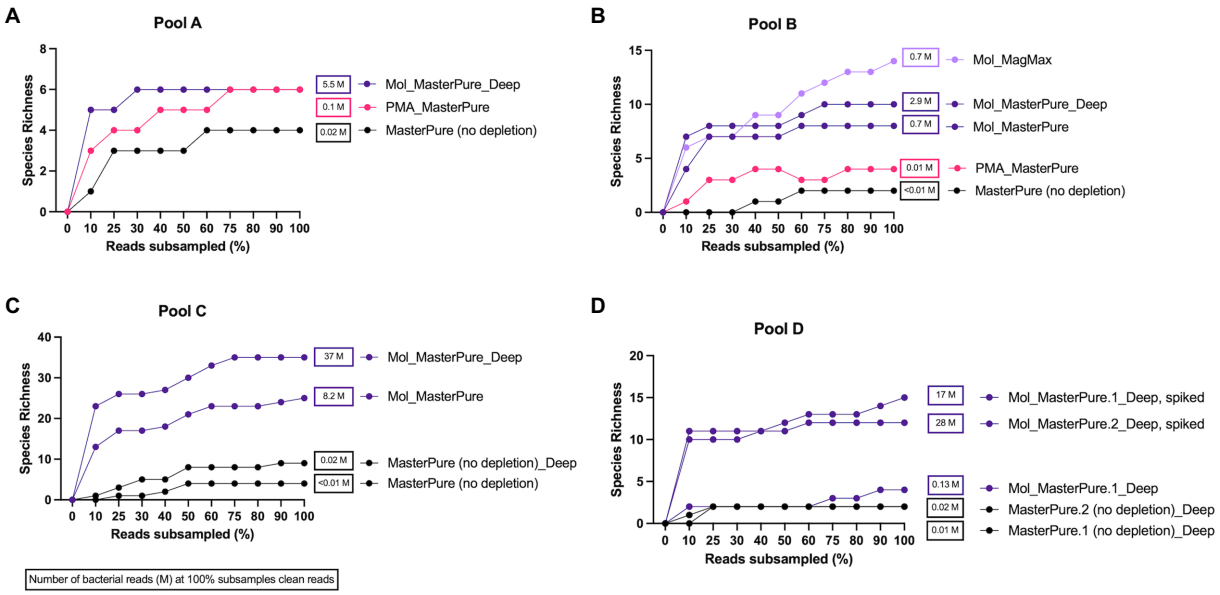

FIGURE 4
Rarefaction analysis. Rarefaction analysis. Rarefaction curves at species level for samples at various sequencing depth (% clean reads) for pools A **(A)**, B **(B)**, C **(C)**, and D **(D)**. Protocols including MolYsis for host DNA depletion retrieved highest species richness across all pools. Number of bacterial reads obtained at 100% clean reads are listed next to samples rarefaction curve.

of these 12 samples failed quality control prior to WGS. Six samples prepared using 12 PCR cycles (DNA input 6 ng) produced libraries of sufficient concentration and quality and were sequenced at an average depth of 32 M reads (details regarding library preparation, the number of initial, human and bacterial reads are listed in Table 3). Relative abundance of classified bacterial taxa, together with rarefaction analysis can be found in the online supplement (Supplementary Figure 3). Three samples were spiked with a standardized Zymo Spike-in Control II (Zymo, D6321 & D6321-10) after host DNA depletion. Their total microbial load (relative to *I. haloterans*) is shown in Figure 8A. To observe a possible correlation between host DNA content obtained with real time qPCR and WGS, four samples were also quantified using Femto quantification kit for bacterial and host DNA (Figure 8B).

Discussion

We performed a method optimization study for low microbial biomass samples with high human DNA content for the purpose of microbiome and resistome characterization with WMS using nasopharyngeal aspirates from premature infants. We found that nasopharynx aspirates of preterm infants have a high host DNA content (99%). Of the protocols tested in our study, Mol_MasterPure (composed of host DNA depletion with MolYsis™ Basic5 and DNA extraction with MasterPure™ Gram Positive DNA Purification Kit) was the most promising protocol for microbiome and resistome characterization with WMS, tested with pooled and individual patient samples.

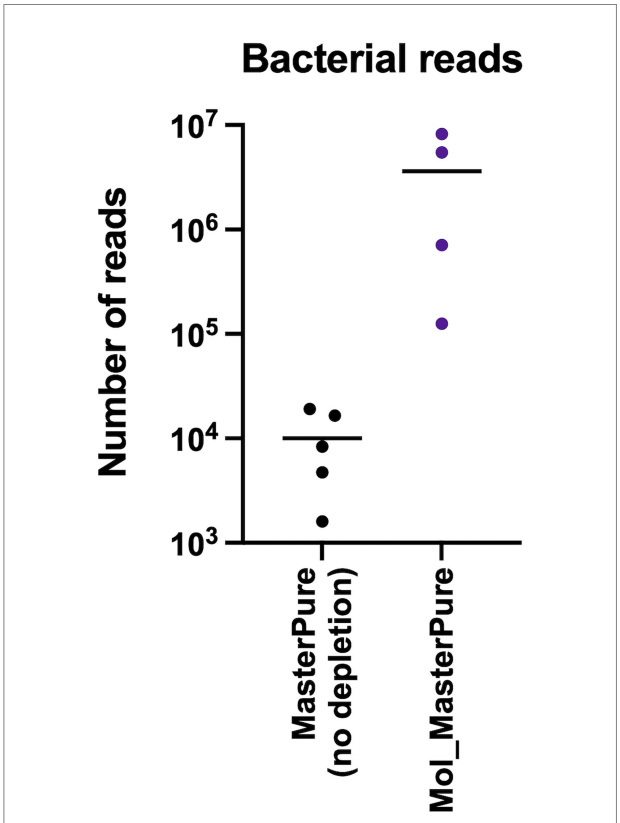

FIGURE 5
Bacterial reads after processing with MolYsis. Number of reads assigned to bacteria in non-depleted reference samples and in depleted samples using the Mol_MasterPure protocol from the four pools (A, B, C, D). The values for each individual samples are presented as dots. The horizontal lines correspond to mean values.

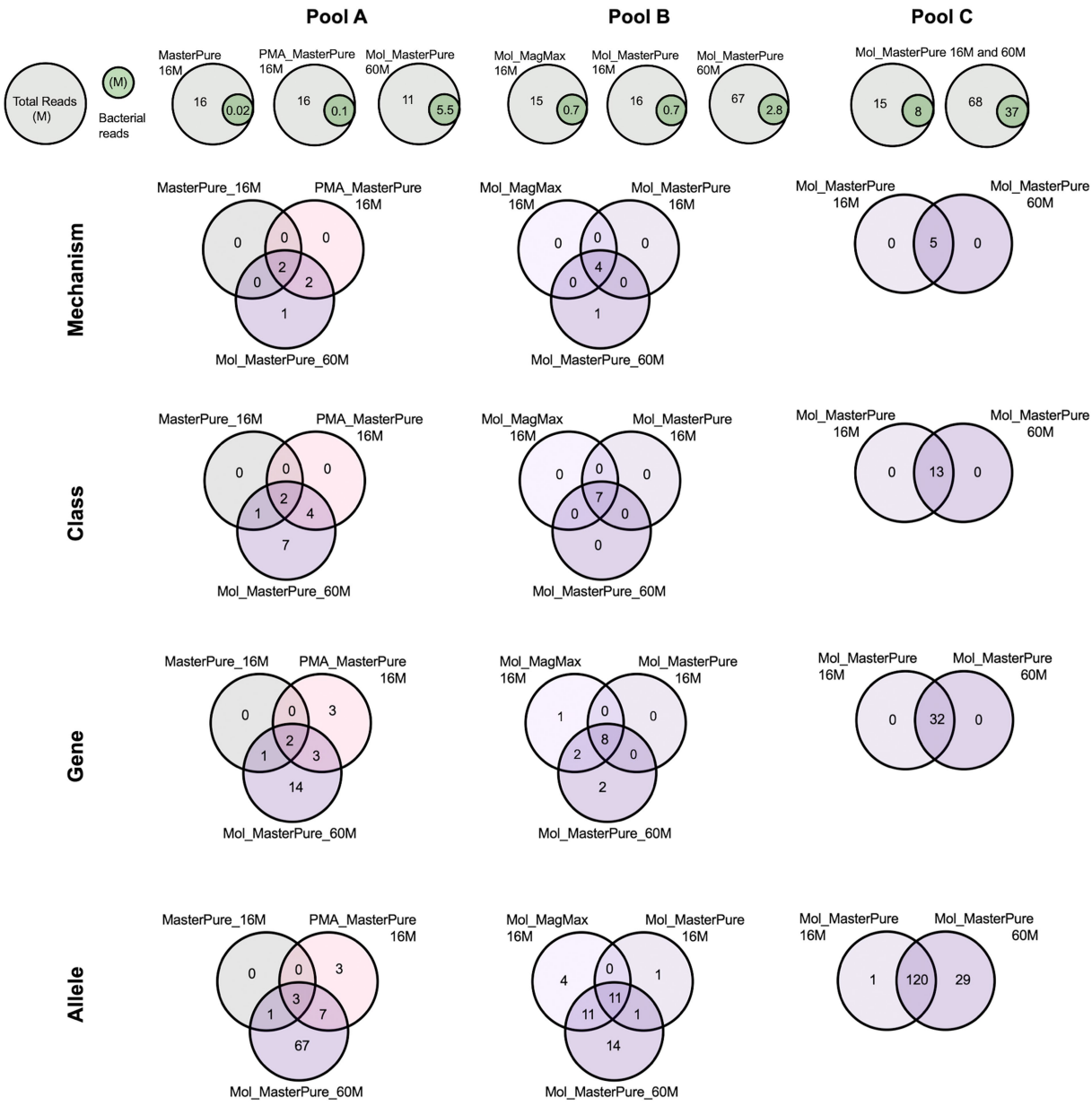

FIGURE 6
Venn diagrams representing intersection of different resistome annotation levels (Mechanism, Class, Gene, Allele) for patient samples from pools A, B, and C, processed according to different protocols and sequenced at different sequencing depths.

Microbial DNA may be partially lost during both host DNA depletion and DNA extraction procedures. This is particularly critical in samples with low microbial biomass. We explored this using a defined microbial mock community. Of the three DNA extraction kits only MasterPure retrieved the expected DNA yield, thus becoming a reference. Host DNA depletion processes remove all extracellular DNA. Since the D6300 mock is stored in RNA/DNA shield, some of the bacterial cells will be lysed prior to processing. Hence, the retrieved DNA yield or the retrieved microbial composition could not be compared with the non-depleted reference.

Three pre-extraction host DNA depletion methods with selective lysis of human cells and extracellular DNA degradation (Figure 9) were compared in five protocols with pooled patient samples. These depletion methods were chosen as they have been shown to be superior to other pre- and post-extraction host DNA removal methods (e.g., filtration and selective removal of CpG-methylated host DNA) in studies analysing samples with high host DNA content (Thoendel et al., 2016; Marotz et al., 2018; Heravi et al., 2020; Rubiola et al., 2020).

Previously, the lyPMA protocol was found to be more effective than both MolYsis and QIAamp protocols in host DNA removal

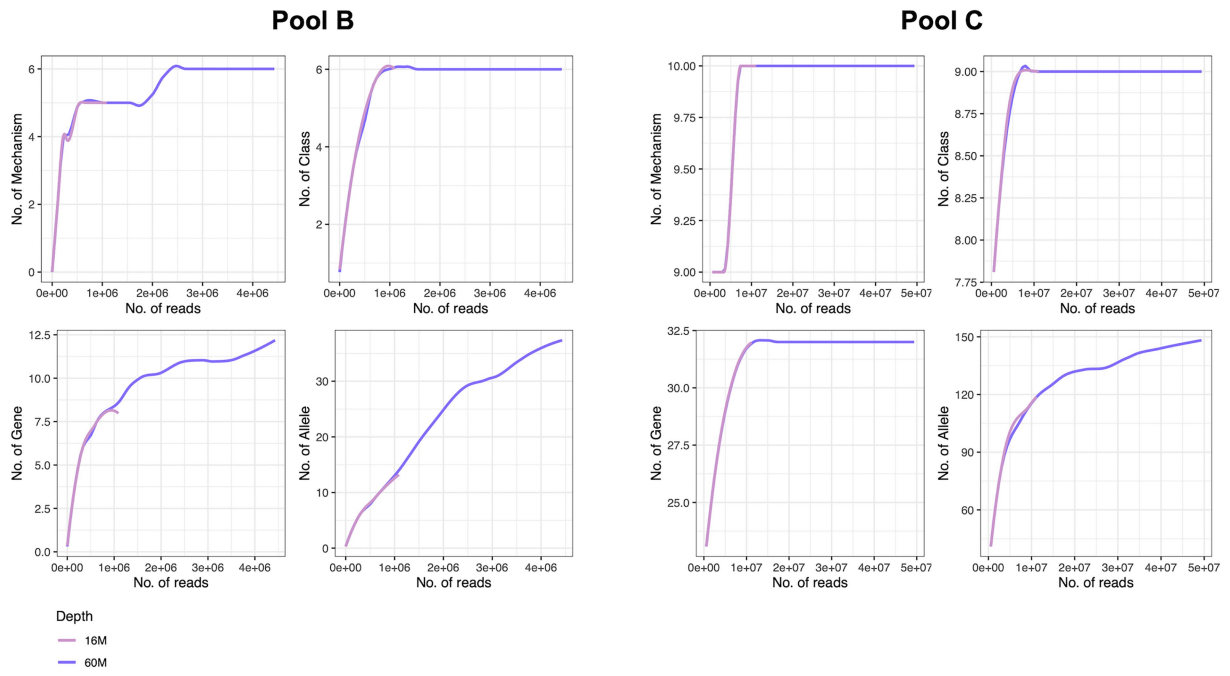

FIGURE 7
Rarefaction analysis of two samples with different host DNA content at two sequencing depths at mechanism, class, gene and allele levels, performed on remaining reads (after removal of host reads). Both samples were processed with Mol_MasterPure protocol. Host DNA content was 92% for the sample originating from pool B, and 15% for the sample originating from pool C. Number of total reads for sample from pool B processed with Mol_MasterPure protocol was 18M (pink line) and 68M (purple line). Number of reads assigned to bacteria was 0.7M and 2.9M, respectively. Number of total reads for sample from pool C processed with Mol_MasterPure was 15M (pink line) and 68M (purple line). Number of reads assigned to bacteria was 8.2M and 27M, respectively.

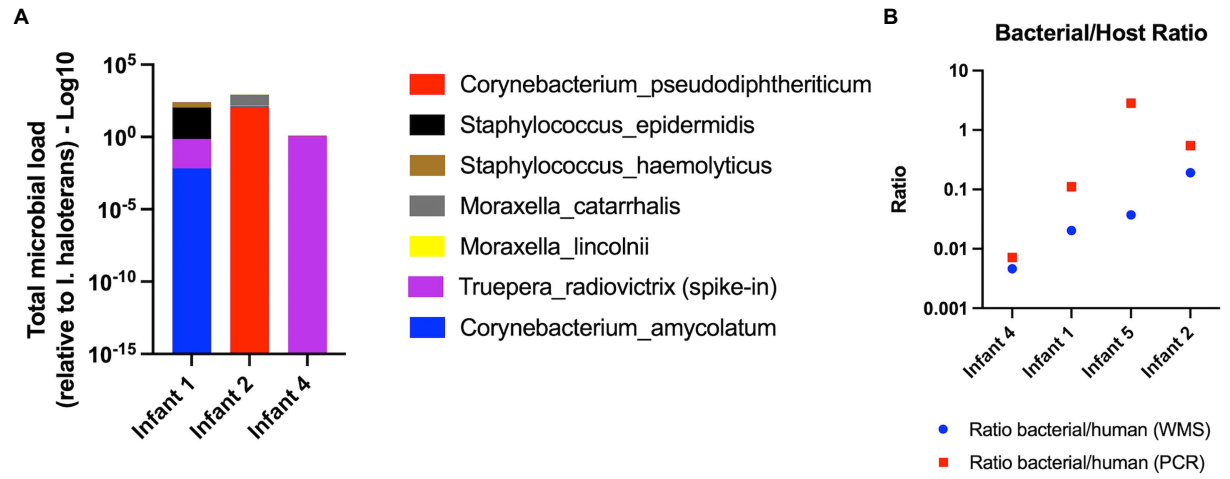

FIGURE 8
Individual patient samples. **(A)**. Total microbial load relative to the abundance of I. haloterans (log scale). **(B)** Bacterial/Host DNA ratio for individual patient samples evaluated with real time qPCR (red) and WGS (blue).

from saliva samples (Marotz et al., 2018). LyPMA has the advantage of lower costs and short handling time. However, the combination of osmotic lysis and DNA fragmentation with photolysis did not work as efficiently in our experiments. Host DNA content remained as high as in the reference samples indicating that method performance could depend on the sample type. A study using bovine milk samples (also low biomass, with high host DNA content) reported similar results to ours after lyPMA treatment of

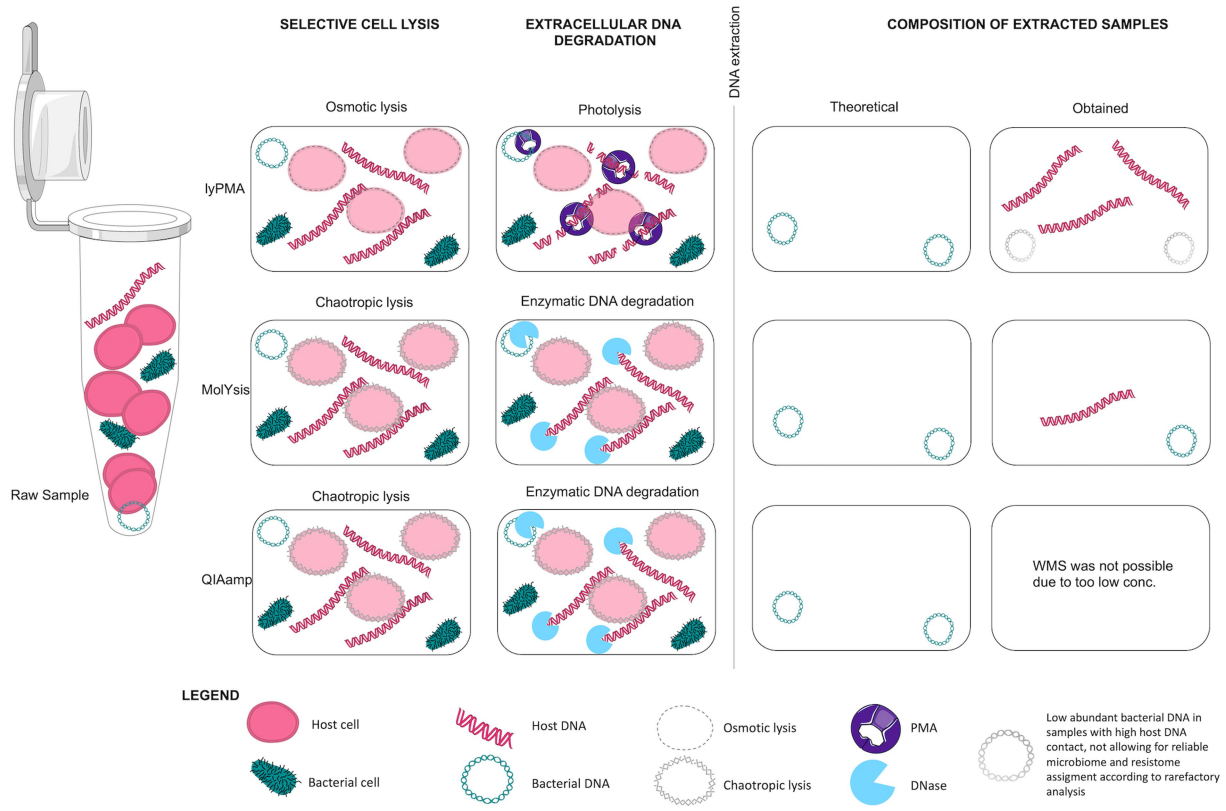

FIGURE 9
Mechanism of Host DNA depletion protocols. Graphical illustration of different host DNA depletion protocols according to their mechanism of action. LyPMA protocol is composed of osmotic lysis and DNA fragmentation with photolysis. MolYsis and QIAamp use chemical lysis of host cells followed by enzymatic degradation of extracellular DNA. In theory, all methods result in selective lysis of human cells, followed by removal of extracellular DNA, human and bacterial. In our study, only samples depleted with MolYsis™ showed a reduction in host DNA content. Parts of the figure were drawn by using pictures from Servier Medical Art by Servier (smart.servier.com).

the samples, even after optimising lyPMA concentration from 10 to 20 µM (Ganda et al., 2021). QIAamp depletion protocol works through a similar mechanism as MolYsis (chemical lysis of host cells and enzymatic degradation of extracellular DNA) and has previously shown to outperform MolYsis in host DNA removal for some samples (Marotz et al., 2018; Heravi et al., 2020). However, in our study, further analysis of the samples processed with QIAmp protocol failed due to too low DNA yield. A threshold of 1 pg DNA/µl for microbiota detection has previously been proposed (Biesbroek et al., 2012). We therefore decided to exclude samples with <1 pg. Although MolYsis was initially developed to selectively isolate and purify bacterial DNA from whole blood samples in aid of sepsis diagnosis (Gebert et al., 2008; Horz et al., 2008), its efficiency in host DNA removal has broadened its use to samples of different origin for microbiome and resistome studies (Hasan et al., 2016; Rubiola et al., 2020; Yap et al., 2020).

Patient samples were initially pooled rather than processed individually, as we expected a large variation between individual samples, potentially preventing us from comparing different processing methods. Variation in species and ARG richness was seen also between the pools. As the samples were blindly pooled, no metadata was collected to supplement the interpretation of our results. We found that nasopharyngeal aspirate samples from premature infants contained a high content of host DNA and removing host DNA with MolYsis prior to DNA extraction was the only successful method for enriching microbial fraction sufficiently for both microbiome and resistome analysis of WMS data. Due to the variation between patient pools and the small number of samples processed with each protocol sent to WMS, no meaningful statistical tests could be implemented to compare host DNA depletion efficiency of different protocols.

High host DNA content interferes with the sensitivity of WMS for taxonomic profiling, even at greater sequencing depths (Pereira-Marques et al., 2019). In addition, methods that amplify specific sequences, such as qPCR and 16S rRNA sequencing, have been demanding to implement for low biomass samples with high host DNA content. A previous study by Gallacher et al. (2020) in a cohort of premature infants showed that only 6.7% of nasopharyngeal aspirate samples obtained in the first 3 days after birth had a high enough bacterial load for 16S amplicon sequencing. Similar low bacterial load was found in some of our samples, reflecting the very low biomass in samples obtained soon after birth.

10.3389/fmicb.2022.1038120

TABLE 3 Detailed information on individual patient samples.

Sample name	DNA quantity (ng)	Library DNA input (ng)	Library conc. (ng/μl)	Molarity pool sent for seq. (nM)	Total reads	Preprocessed reads	Human reads		Remaining reads		Bacterial reads	Bacterial reads (% of Remaining reads)	ARG richness (No of ARG)	ARG associated reads	ARG associated reads (% of Remaining reads)
Infant 1*	195.6	6.5	28.2	25	35.23 M	35.23 M	33.40 M	95%	1.83 M	5.2%	0.68 M	37.0%	11	8,624	0.47%
Infant 2*	714	6	11.2	25	22.98 M	22.98 M	15.89 M	69%	7.09 M	30.9%	3.04 M	42.9%	5	10,980	0.15%
Infant 3	6.8	6.6	15.7	25	46.24 M	46.24 M	18.69 M	40%	27.55 M	59.6%	12.33 M	44.8%	5	33,812	0.12%
Infant 4*	44.1	6.3	6.6	25	26.70 M	26.70 M	26.04 M	98%	0.66 M	2.5%	0.12 M	18.1%	0	0	0%
Infant 5	199.2	6	17.6	25	25.38 M	25.38 M	23.66 M	93%	1.72 M	6.8%	0.88 M	51.2%	12	131,665	7.65%
Infant 6	6.8	6.6	10.5	25	37.63 M	37.63 M	36.62 M	97%	1.01 M	2.7%	0.02 M	1.5%	2	90	0.01%

Six individual patient samples were processed with Mol_MasterPure protocol. Samples from Infant 1, Infant 4, Infant 5, were from infants between 0 and 7 days old, and samples from Infant 2 and Infant 3 between 56 days and 6 months old. *Samples were spiked with 20 μl (0.4 ng) of Zymo Spike-in Control II for Low Microbial Load samples (Zymo, D6321, and D6321-10) after host DNA depletion.

The most promising protocol Mol_MasterPure was further applied also to individual patient samples. Despite variations in composition between individual samples, all samples met the yield and quality parameters recommended for WMS. Individual patient samples analysed in our study showed variation in obtained DNA yield (Table 3) and host DNA content (from 40 to 98%; details in Table 3). Samples collected in the first week after birth had a higher host DNA content despite host DNA depletion processing, while samples obtained later in life showed greater enrichment. This is expected due to rapid microbial colonization and increase in microbial density of nasopharyngeal microbiome from birth on (Theodosiou et al., 2020).

To evaluate microbial enrichment of our samples prior to WMS and to assist in the estimation of required sequencing depth, we performed real-time qPCR using primers targeting the 16S gene. The microbial enrichment seen with qPCR was indicative of the extent of host DNA removal seen with WMS (Figures 3C, 5). Relative qPCR could serve as a time and cost-efficient triage prior to WMS. However, this would require having a reference sample (without depletion) for every patient sample, which might not be feasible in practice. Alternatively, targeted qPCR absolute quantification methods for both bacterial and host DNA can be used to predict library composition for WMS and help determine needed sequencing depth (Cho et al., 2021). In our individual patient samples, 2 out of 6 samples did not yield sufficient material for both WMS library preparation and two qPCR reactions. The ratio of bacterial / host WMS reads and bacterial / host DNA quantity (qPCR) for the remaining four samples showed some correlation (Figure 8B), but more samples would be required to suggest a possible prediction model.

Equimolar library pooling is necessary to obtain comparable number of total reads for all submitted samples and is preferred over equal-volume pooling for use in patient derived samples with multiple bacterial species (Muller et al., 2019). We initially set the threshold to 10 nM for samples sent to WMS and successfully adjusted it to 6 nM (Pool D; Table 2) due to lower yield from samples obtained within 24 h after birth.

One of the limitations of WMS is providing only relative information on microbial composition. Spiking samples with known absolute abundance serves as a positive control, and additionally enables quantification of microbiome composition (Figure 8A) and contributes towards more unbiased interpretation of dynamics and interactions in the microbiome (Wang et al., 2021). Therefore, spike with known absolute abundance can be helpful in microbiome studies where determining total microbial load is relevant to the aims of the study (Stämmler et al., 2016).

Methods without extracellular DNA removal might overestimate the bacterial composition in analysed patient samples (DNA from viable and non-viable bacteria), restricting interpretation of a possible taxonomic bias between depleted and reference samples. It is however a concern that besides removing extracellular DNA from unviable cells, DNA from bacteria with a thin or missing cell wall could also be lost during host DNA depletion steps (Horz et al., 2010). This could introduce a taxonomic bias especially as a loss of Gram-negative

species (Horz et al., 2010; Heravi et al., 2020; Rubiola et al., 2020), as we observed in our samples from Pool D that were spiked with mock community (Zymo D6300) (Supplementary Figure 1). Even though the respiratory tract is mainly colonised by Gram-positive bacteria, some Gram-negative bacteria are also relevant, including for instance *Moraxella* (Gram-negative) (Toivonen et al., 2021). We were not able to obtain enough bacterial reads for microbiome and resistome classification from any of the reference (non-depleted) samples to be able to compare them with their depleted parallels. Our study was, however, not designed for bias analysis. Further, our study's limitations were using a mock (Zymo, D6300) stored in RNA/DNA shield that could cause cell lysis prior to processing, and not using the producer's recommended amount of mock community since we aimed to have a better representation of the low biomass of our samples of interest. To address this problem, a mock with viable bacterial cells from cultures of species commonly found in the respiratory microbiome should be created, for a detailed investigation of the possible bias introduced with host DNA depletion protocols. This was not feasible in our study.

In this study we describe how different protocols for host DNA depletion and DNA extraction performed on mock community standards, pooled and individual patient samples from nasopharyngeal aspirates of premature infants. Microbiome and resistome composition from low biomass samples with high host DNA content was best characterized applying a protocol combining depletion with MolYsis™ and extraction with MasterPure™. Analysis of samples obtained immediately after birth remains challenging, and our protocol should be further tested and optimized in settings of a larger study. Our findings may contribute to broadening and improving use of WMS in respiratory and other low biomass microbiota studies.

Data availability statement

The datasets presented in this study can be found at: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA876384.

Ethics statement

The studies involving human participants were reviewed and approved by Hospital's Data Protection Officer and the Regional Committee for Medical and Health Research Ethics–South East, Norway (2018/1381 REKD), and by the Danish National Committee for Health Research Ethics (H-180512193). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

All authors contributed to the design of the study. PR, UL-T, and KH collected the samples. PR, GS, and HÅ carried out the laboratory experiments. PR and AD carried out the data analysis. PR wrote the manuscript. KH and FP supervised the overall study. All authors discussed the results, critically revised the manuscript, and agreed to the published version of the manuscript.

Funding

This work was supported by the Norwegian Research Council (NFR) project number 273833, by Olav Thon Foundation, by the Faculty of Dentistry at the University of Oslo and by Oslo University Hospital.

Acknowledgments

We would like to thank the parents of participating infants, and the clinical staff at the Neonatal Intensive Care units at Ullevaal, Oslo University Hospital, Oslo, Norway and Rigshopitalet, Copenhagen, Denmark, for their assistance with sample collection. Further, we would like to thank the Olav Thon Foundation for awarding support to the study. The sequencing service for this work was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities. The computations were performed on resources provided by Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway. Figures were created using Prism9 (GraphPad Software, LLC, San Diego, CA, United States).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1038120/full#supplementary-material

References

Andrews, S. (2010). Fast QC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Accessed July 1, 2022).

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bio bakery 3. *elife* 10:e65088. doi: 10.7554/eLife.65088

Biesbroek, G., Sanders, E. A., Roeselers, G., Wang, X., Caspers, M. P., Trzciński, K., et al. (2012). Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. *PloS One* 7:e32942. doi: 10.1371/journal.pone.0032942

Biesbroek, G., Tsivtsivadze, E., Sanders, E. A. M., Montijn, R., Veenhoven, R. H., Keijser, B. J. F., et al. (2014). Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am. J. Respir. Crit. Care Med.* 190, 1283–1292. doi: 10.1164/rccm.201407-1240OC

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Cho, M. Y., Oliva, M., Spreafico, A., Chen, B., Wei, X., Choi, Y., et al. (2021). Two-target quantitative PCR to predict library composition for shallow shotgun sequencing. *mSystems* 6, e00521–e00552. doi: 10.1128/mSystems.00552-21

Doster, E., Lakin, S. M., Dean, C. J., Wolfe, C., Young, J. G., Boucher, C., et al. (2020). MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res.* 48, D561–D569. doi: 10.1093/nar/gkz1010

Douglas, C. A., Ivey, K. L., Papanicolas, L. E., Best, K. P., Muhlhausler, B. S., and Rogers, G. B. (2020). DNA extraction approaches substantially influence the assessment of the human breast milk microbiome. *Sci. Rep.* 10:123. doi: 10.1038/s41598-019-55568-y

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., and Weyrich, L. S. (2019). Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol.* 27, 105–117. doi: 10.1016/j.tim.2018.11.003

Florensa, A. F., Kaas, R. S., Clausen, P., Aytan-Aktug, D., and Aarestrup, F. M. (2022). Res finder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb Genom.* 8:000748. doi: 10.1099/mgen.0.000748

Gallacher, D., Mitchell, E., Alber, D., Wach, R., Klein, N., Marchesi, J. R., et al. (2020). Dissimilarity of the gut-lung axis and dysbiosis of the lower airways in ventilated preterm infants. *Eur. Respir. J.* 55:1901909. doi: 10.1183/13993003.01909-2019

Ganda, E., Beck, K. L., Haiminen, N., Silverman, J. D., Kawas, B., Cronk, B. D., et al. (2021). DNA extraction and host depletion methods significantly impact and potentially bias bacterial detection in a biological fluid. *mSystems* 6:e0061921. doi: 10.1128/mSystems.00619-21

Gasparrini, A. J., Wang, B., Sun, X., Kennedy, E. A., Hernandez-Leyva, A., Ndao, I. M., et al. (2019). Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. *Nat. Microbiol.* 4, 2285–2297. doi: 10.1038/s41564-019-0550-2

Gebert, S., Siegel, D., and Wellinghausen, N. (2008). Rapid detection of pathogens in blood culture bottles by real-time PCR in conjunction with the pre-analytic tool Mol Ysis. *J. Infect.* 57, 307–316. doi: 10.1016/j.jinf.2008.07.013

Greathouse, K. L., Sinha, R., and Vogtmann, E. (2019). DNA extraction for human microbiome studies: the issue of standardization. *Genome Biol.* 20:212. doi: 10.1186/s13059-019-1843-8

Hasan, M. R., Rawat, A., Tang, P., Jithesh, P. V., Thomas, E., Tan, R., et al. (2016). Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *J. Clin. Microbiol.* 54, 919–927. doi: 10.1128/JCM.03050-15

Hasrat, R., Kool, J., de Steenhuijsen Piters, W. A. A., Chu, M. L. J. N., Kuiling, S., Groot, J. A., et al. (2021). Benchmarking laboratory processes to characterise low-biomass respiratory microbiota. *Sci. Rep.* 11:17148. doi: 10.1038/s41598-021-96556-5

Heravi, F. S., Zakrzewski, M., Vickery, K., and Hu, H. (2020). Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *J. Microbiol. Methods* 170:105856. doi: 10.1016/j.mimet.2020.105856

Horz, H.-P., Scheer, S., Huenger, F., Vianna, M. E., and Conrads, G. (2008). Selective isolation of bacterial DNA from human clinical specimens. *J. Microbiol. Methods* 72, 98–102. doi: 10.1016/j.mimet.2007.10.007

Horz, H. P., Scheer, S., Vianna, M. E., and Conrads, G. (2010). New methods for selective isolation of bacterial DNA from human clinical specimens. *Anaerobe* 16, 47–53. doi: 10.1016/j.anaerobe.2009.04.009

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, H. (2012). Seqtk toolkit for processing sequences in FASTA/Q formats. *GitHub* 767:69.

Manenzhe, R. I., Dube, F. S., Wright, M., Lennard, K., Zar, H. J., Mounaud, S., et al. (2020). Longitudinal changes in the nasopharyngeal resistome of south African infants using shotgun metagenomic sequencing. *PloS One* 15:e0231887. doi: 10.1371/journal.pone.0231887

Marotz, C. A., Sanders, J. G., Zuniga, C., Zaramela, L. S., Knight, R., and Zengler, K. (2018). Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome.* 6:42. doi: 10.1186/s40168-018-0426-3

Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., et al. (2017). The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiol. Mol. Biol. Rev.* 81, 3–13. doi: 10.1128/MMBR.00036-17

Mizgerd, J. P. (2014). The infant nose. Introducing the respiratory tract to the world. *Am. J. Respir. Crit. Care Med.* 190, 1206–1207. doi: 10.1164/rccm.201410-1919ED

Morley, V. J., Woods, R. J., and Read, A. F. (2019). Bystander selection for antimicrobial resistance: implications for patient health. *Trends Microbiol.* 27, 864–877. doi: 10.1016/j.tim.2019.06.004

Muller, B. H., Mollon, P., Santiago-Allexant, E., Javerliat, F., and Kaneko, G. (2019). In-depth comparison of library pooling strategies for multiplexing bacterial species in NGS. *Diagn. Microbiol. Infect. Dis.* 95, 28–33. doi: 10.1016/j.diagmicrobio.2019.04.014

Nelson, M. T., Pope, C. E., Marsh, R. L., Wolter, D. J., Weiss, E. J., Hager, K. R., et al. (2019). Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles. *Cell Rep.* 26, 2227–40.e5. doi: 10.1016/j.celrep.2019.01.091

Pereira-Marques, J., Hout, A., Ferreira, R. M., Weber, M., Pinto-Ribeiro, I., van Doorn, L.-J., et al. (2019). Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Front. Microbiol.* 10:1277. doi: 10.3389/fmicb.2019.01277

Rautava, S., Luoto, R., Salminen, S., and Isolauri, E. (2012). Microbial contact during pregnancy, intestinal colonization and human disease. *Nat. Rev. Gastroenterol. Hepatol.* 9, 565–576. doi: 10.1038/nrgastro.2012.144

Rubiola, S., Chiesa, F., Dalmasso, A., Di Ciccio, P., and Civera, T. (2020). Detection of antimicrobial resistance genes in the Milk production environment: impact of host DNA and sequencing depth. *Front. Microbiol.* 11:1983. doi: 10.3389/fmicb.2020.01983

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12:87. doi: 10.1186/s12915-014-0087-z

Shi, Y., Wang, G., Lau, H. C.-H., and Yu, J. (2022). Metagenomic sequencing for microbial DNA in human samples: emerging technological advances. *Int. J. Mol. Sci.* 23:2181. doi: 10.3390/ijms23042181

Stämmler, F., Gläsner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P. J., et al. (2016). Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome* 4:28. doi: 10.1186/s40168-016-0175-0

Teo, S. M., Tang, H. H. F., Mok, D., Judd, L. M., Watts, S. C., Pham, K., et al. (2018). Airway microbiota dynamics uncover a critical window for interplay of pathogenic bacteria and allergy in childhood respiratory disease. *Cell Host Microbe* 24, 341–52.e5. doi: 10.1016/j.chom.2018.08.005

Theodosiou, A. A., Dorey, R. B., Read, R. C., and Jones, C. E. (2020). The infant pharyngeal microbiomes: origin, impact and manipulation. *Curr. Opin. Infect. Dis.* 33, 548–555 . doi: 10.1097/QCO.0000000000000691

Thoendel, M., Jeraldo, P. R., Greenwood-Quaintance, K. E., Yao, J. Z., Chia, N., Hanssen, A. D., et al. (2016). Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *J. Microbiol. Methods* 127, 141–145. doi: 10.1016/j.mimet.2016.05.022

Toivonen, L., Schuez-Havupalo, L., Karppinen, S., Waris, M., Hoffman, K. L., Camargo, C. A., et al. (2021). Antibiotic treatments during infancy, changes in nasal microbiota, and asthma development: population-based cohort study. *Clin. Infect. Dis.* 72, 1546–1554. doi: 10.1093/cid/ciaa262

Wang, X., Howe, S., Deng, F., and Zhao, J. (2021). Current applications of absolute bacterial quantification in microbiome studies and decision-making regarding different biological questions. *Microorganisms* 9:1797. doi: 10.3390/microorganisms9091797

Wickham, H. (2016). *Data Analysis in ggplot 2.* Cham: Springer.

World Health Organization (2014). Antimicrobial resistance: global report on surveillance. World Health Organization. Available at: https://apps.who.int/iris/handle/10665/112642 (Accessed October 15, 2021).

Yap, M., Feehily, C., Walsh, C. J., Fenelon, M., Murphy, E. F., McAuliffe, F. M., et al. (2020). Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome. *Sci. Rep.* 10:21665. doi: 10.1038/s41598-020-78773-6

Yigit, E., Feehery, G. R., Langhorst, B. W., Stewart, F. J., Dimalanta, E. T., Pradhan, S., et al. (2016). A microbiome DNA enrichment method for next-generation sequencing sample preparation. *Curr. Protoc. Mol. Biol.* 115, 7.26.1–7. 14. doi: 10.1002/cpmb.12

II

II

Impact of antibiotics and hospitalization on the nasopharyngeal microbiome in very preterm infants: Observational study using whole metagenomic sequencing

Polona Rajar[1,2]†, Achal Dhariwal[2]†, Gabriela Salvadori[2], Heidi Aarø Åmdal[2], Dag Berild[3,4], Ulf R. Dahle[5], Drude Fugelseth[1,4], Gorm Greisen[6], Ulrik Lausten-Thomsen[6], Ola Didrik Saugstad[7], Fernanda Cristina Petersen[2], Kirsti Haaland[1]*

[1] Department of Neonatal Intensive Care, Division of Paediatric and Adolescent Medicine, Oslo University Hospital, Oslo, Norway.

[2] Institute of Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway.

[3] Department of Infectious Diseases, Oslo University Hospital, Oslo, Norway.

[4] Institute of Clinical Medicine, Faculty of Medicine, Oslo University, Oslo, Norway.

[5] Centre for Antimicrobial Resistance, Norwegian Institute of Public Health, Oslo, Norway

[6] Department of Neonatology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark.

[7] Department of Paediatric Research, University of Oslo, Oslo, Norway.

† These authors contributed equally: Polona Rajar, Achal Dhariwal.

* Correspondence:

Kirsti Haaland

Department of Neonatal Intensive Care Unit Ullevål, Division of Paediatric and Adolescent Medicine, Oslo University Hospital,

Kirkeveien 166, 0450 Oslo, Norway

Tel: 0047 22117622

uxkila@ous-hf.no

III

III

The landscape of antibiotic resistance genes in the nasopharynx of preterm infants: Prolonged signature of hospitalization and effects by antibiotics

Achal Dhariwal[1]†, Polona Rajar[1,2]†, Gabriela Salvadori[1], Heidi Aarø Åmdal[1], Dag Berild[3,4], Ola Didrik Saugstad[5], Drude Fugelseth[2,4], Gorm Greisen[6], Ulf Dahle[7], Kirsti Haaland[2], Fernanda Cristina Petersen[1]*

[1] Institute of Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway.
[2] Department of Neonatal Intensive Care, Division of Pediatric and Adolescent Medicine, Oslo University Hospital Ullevål, Oslo, Norway.
[3] Department of Infectious Diseases, Oslo University Hospital, Oslo, Norway.
[4] Institute of Clinical Medicine, Faculty of Medicine, Oslo University, Oslo, Norway.
[5] Department of Pediatric Research, Oslo University Hospital, University of Oslo, Oslo, Norway.
[6] Department of Neonatology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark.
[7] Centre for Antimicrobial Resistance, Norwegian Institute of Public Health, Oslo, Norway

†Shared First authorship
*Correspondence: Fernanda Cristina Petersen
Institute of Oral Biology
Faculty of Dentistry,
University of Oslo, Oslo, Norway.
Tel: +47-22840312
E-mail: f.c.petersen@odont.uio.no