Characterization of the infant gut microbiota mobilome

Karakterisering av mobile genetiske elementer i spedbarns tarmmikrobiota

Philosophiae Doctor (PhD) Thesis

Anuradha Ravi

Norwegian University of Life Sciences

Faculty of Chemistry, Biology and Food Sciences

Ås, 2017

Thesis number 2017:37

ISSN 1894-6402

ISBN 978-82-575-1438-9

Table of Contents

Acknowledgements

The work presented in this thesis has been carried out in the Microbial Diversity lab (MidDiv) at the KBM faculty in Norwegian University of Life Sciences (NMBU). Reaching to this point of my education would have been impossible without the support of several people whom only a few have been mentioned.

If there was one person I can dedicate my scientific contributions to, it will be to my supervisor Prof. Knut Rudi. I am deeply thankful for your support and enthusiasm. During the past years, you have not only supported and guided me but also given me the time to learn, to grow and be creative by myself. Your endless creativity and curiosity has inspired me to move beyond boundaries and challenge myself. Thank you very much Knut!

I would also like to thank my co-supervisor, Trine L'Abee-Lund. You introduced me to this topic and it eventually became a main part of my PhD. Thank you for your support, positive attitude and good wishes always. I would also like to thank Gaspar Perez Martinez and M. Carmen Collado for hosting me in Valencia and for all the discussion and enthusiasm while I was there. I would like to thank Steven Foley for answering even the smallest and silliest questions that I came across. I would like to thank Eva Lena, Agnes and Mari Hagbø for working with me during your master thesis. I would like to thank the administration staff especially Janne, Wenche, Heidi and Vilma for making the paper work seem very easy and helping me in difficult situations. I would like to thank Else and Rannei for helping me in ordering and delivering equipment.

To have a balance in work and life, friends are needed. Jane Ludvigsen, thank you for being in the same room with me for four years! It is hard (I know) but you have been there to give me a break and to motivate me. Katya, thank you for being there always and supporting me. You are an inspiration. Inga Leena, thank you for translating my abstract, reminding me to take breaks and caring for me. Thank you to all the people in the 2nd floor corridor for helping me and making my work days fun. A special mention to my girl gang, Maria, Bjorg-Karin, Heidrun, Guro and Elizabeth. Thank you for all your care, support and keeping me motivated. You girls are the best! I would like to also thank my Indian friends at Ås for all the dinners, fun days and laughter.

I am deeply thankful to my parents who are my guiding lights at the darkest and brightest days. I am the most luckiest to have them as parents. You have believed in my dreams and

let me fly. My biggest inspiration is my brother. You have guided, teased, protected and supported me from wherever you are. I wouldn't have come to this stage without your motivation!

Last but not the least, I would like to thank Sapna didi, Vimla ji and Suresh ji for your encouraging thoughts and prayers during the last phase of my PhD. Thank you Sandeep for joining me in the journey of life for eternity. Thank you for standing beside me and giving me all the confidence, even when I did not have confidence in myself. Your care, understanding and your jokes has kept me cheerful during the tough times. Thank you for believing in me and chasing our dreams together.

Cheers!

Anu

Summary

The human gut is densely populated with a wide diversity of bacteria. These bacteria can serve as a reservoirs for multiple Antibiotic Resistance (AR) genes that in turn are associated with wide range of Mobile Genetic Elements (MGEs). The mobilome is the collection of MGEs such as plasmids, transposons and integrons that are main contributors to Horizontal Gene Transfer (HGT). The distribution and association of the mobilome in the developing gut microbiota of infants remains largely unexplored. Therefore, the main aim of this thesis is to study the prevalence, association and characterization of plasmids and integrons that were *de novo* assembled and detected in the developing gut microbiota of full term and preterm infants.

From our study, we detected a diverse mobilome (potentially MGES) of conjugative plasmids and integrons. The MGEs that were *de novo* assembled from the shotgun metagenome data, especially conjugative plasmids harboured various AR and virulence gene factors. The integrons that are non-mobile genetic elements were closely associated with conjugative plasmids. These plasmids especially IncF and IncI conjugative plasmids were in-turn associated with the *Enterobacteriaceae* family. In addition to this, the *de novo* assembled plasmid-related contigs depicted a potential multireplicon status with shared and integrated IncF variants and shared plasmids between IncF and IncI plasmids. In total, we have *de novo* assembled 7 different IncF and IncI1 conjugative plasmids from different cohorts.

We also detected a strong correlation with the mobilome and microbiota taxonomy. We detected a significant strong association with the abundance of conjugative plasmids and different Operational Taxonomic Units (OTUs) related to *Enterobacteriaceae*. Overall, the persistence patterns of the conjugative plasmids between the different time periods of the different cohorts were surprisingly consistent.

Lastly, we isolated and characterised the functional attributes of strains carrying conjugative plasmids. We *de novo* assembled IncI and IncF plasmids and we demonstrated the mobility of these plasmids *in vitro*. We detected a mobile IncI plasmid and a non-mobile IncF plasmid, both carrying multidrug resistance genes. In addition, we also characterised a bacteriocin-producing IncFII/IncFIB conjugative plasmid from the strains.

Taken together, our results provide information on the prevalence and persistence of conjugative plasmids and integrons in three longitudinal cohorts. In addition, we

characterised the functional attributes and demonstrated transmission of the conjugative plasmids to other strains. These results reveal the innate ability of the mobilome to adapt to selective pressures in gut microbiota, in addition to spread among different bacteria.

Sammendrag

Den humane tarmen er kolonisert av et stort antall bakterier. Disse bakteriene kan tjene som et reservoar for antibiotikaresistente gener som igjen er forbundet med et stort utvalg av mobile genetiske elementer. Mobilomet består av en samling mobile genetiske elementer som plasmider, transposoner og integroner, og representerer de viktigste bidragsyterne til horisontal genoverføring. Distribusjonen og assosiasjonen av mobilomet i forbindelse med utviklingen av tarmmikrobiota hos spedbarn er fortsatt lite kjent. Hovedmålet med denne oppgaven var derfor å studere forekomsten, assosiasjonen, samt å karakterisere plasmider og integroner som ble *de novo* sammenstilt og detektert i tarm hos for tidlig og normalfødte spedbarn. I denne studien ble det påvist et mobilom av høy diversitet (potensielt mobile genetiske elementer) av konjugative plasmider og integroner. Mobile genetiske elementer som ble *de novo* sammenstilt fra shotgun metagenom data, spesielt konjugative plasmider, inneholdt flere antibiotikaresistente og virulente genfaktorer. Integronene, som er ikke-mobile genetiske elementer, var nært assosiert med konjugative plasmider. Disse plasmidene, spesielt IncF og IncI konjugative plasmider, var videre assosiert med *Enterobacteriaceae* familien. I tillegg til dette avdekket *de novo* sammenstilte plasmid-relaterte kontiger et potensielt multireplika med delte og integrerte IncF varianter samt delte plasmider mellom IncF og IncI plasmider. Totalt ble det funnet syv forskjellige IncF og IncI1 konjugerbare plasmider fra ulike kohortstudier. Det ble også oppdaget en sterk sammenheng mellom mobile genetiske elementer og taksonomi. Videre ble det oppdaget en sterk sammenheng med overvekt av konjugerbare plasmider og ulike operasjonelle taksonomiske enheter (OTUer) relatert til *Enterobacteriaceae*. Totalt viste de konjugative plasmidene overraskende konsistent persistent mønster mellom de forskjellige tidspunktene i de forskjellige kohortstudiene. Til slutt ble det isolert og karakterisert funksjonelle egenskaper av stammer som bærer konjugative plasmider. Vi *de novo* sammenstilte IncI og IncF plasmider og demonstrerte mobiliteten av disse plasmidene *in vitro*. Vi detekterte et mobilt IncI plasmid og et ikke-mobilt IncF plasmid, hvorav begge var bærere av multiresistente gener. I tillegg ble det også karakterisert et bakteriosinproduserende IncFII/IncFIB konjugativt plasmid fra stammene. Resultatene presenterer informasjon av prevalens og persistens av konjugative plasmider og integroner i tre forskjellige kohortstudier. Det ble også karakterisert funksjonelle egenskaper samt påvist overføring av konjugative plasmider til andre stammer. Disse resultatene viser en medfødt evne av

mobilomet til å tilpasse seg selektivt press i tarmen i tillegg til å spre seg mellom forskjellige bakterier.

List of papers

List of papers included in this thesis:

PAPER 1:

Ravi A, Avershina E, Foley LS, Ludvigsen J, Storrø O, Øien T, Johnsen R, McCartney LA, L'Abée-Lund MT and Rudi K (2015). The commensal infant gut meta-mobilome as a potential reservoir for persistent multidrug resistance integrons. *Scientific Reports.* Doi: 10.1038/srep15317

PAPER 2

Ravi A, Estensmo FE, L'Abée-Lund MT, Foley LS, Allgaier B, Martin RC, Claud CE and Rudi K (2017). Associations of the preterm infant gut microbiota mobilome with necrotizing enterocolitis, birthweight and hospital. Submitted to *Pediatric Research*

PAPER 3

Ravi A, Valdés-Varela L, Gueimonde M and Rudi K (2017). Transmission and persistence of IncF conjugative plasmids in the gut microbiota of full-term infants. Submitted to *FEMS Microbial Ecology*

PAPER 4

Ravi A, Hagbø M, Sunde M, Ludvigsen J, Muinck E, Diep DB, Foley LS, Collado CM, Martinez PG, L'Abée-Lund MT and Rudi K. (2017). Diversity and characterization of conjugative plasmids in the gut microbiota of a preterm twin pair. (Draft Manuscript)

Additional papers

Ravi A**,** Avershina E, Ludvigsen J, L'Abée-Lund MT and Rudi K. (2014). Integrons in the intestinal microbiota as reservoirs for transmission of antibiotic resistance genes. Pathogens 3(2), 238-248.

Avershina E, Ravi A, Storrø O, Øien R, Johnsen R and Rudi K. (2015). Potential association of vacuum cleaning frequency with an altered gut microbiota in pregnant women and their 2 year children. Microbiome 3:65:1-7. Doi: 10.1186/s40168-015-0125-2

Abbreviations

HGT: Horizontal Gene Transfer

GIT: Gastro Intestinal Tract

AR: Antibiotic Resistance

NICU: Neonatal Intensive Care Unit

MGE: Mobile Genetic Element

OTU: Operational Taxonomic Unit

SCFA: Short Chain Fatty Acids

VLBW: Very Low Birthweight

NEC: Necrotizing Enterocolitis

1. Introduction

1.1Human gut microbiota:

The microbes that live in the gastrointestinal track (GIT) of humans are termed as the gut microbiota. The gut microbiota is known for its association in human physiology and diseases, and is mainly composed of the phyla *Actinobacteria*, *Firmicutes* and *Bacteroidetes* [1]. The gut microbiota undertakes various roles in relation to protection against enteropathogens, extraction of nutrients such as Short Chain Fatty Acids (SCFA) and contribution to the immune function[2]. The microbiome is evolved within the human host from birth until death. During this time, the microbiota is constantly modifying according to the host immune system. Therefore, the proportion, diversity and composition of the gut microbiota varies throughout the different stages of human life [3]. These changes in the gut microbiota is governed by host factors such as adaptive and native immune system and external factors such as diet, illness, environmental factors and medication [4].

Full term infant gut microbiota

Initial colonization by microbes in the infant GIT has been known to influence the immune maturation and allergy development [3]. At the beginning, the infant receives a massive load of microbes through the birth process. Recent reports have suggested the presence of the microbiome in the infant meconium suggesting the presence of microbes even before birth [5, 6]. The source and composition of the microbes that colonise the infant are highly dependent on the gestational age and mode of delivery. This composition and proportion of microbes received is influenced by the mode of delivery i.e. vaginal or caesarean delivery [3]. *Firmicutes* (such as *Lactobacillaceae*) and *Proteobacteria* (such as *Enterobacteriaceae*) are initially dominated in vaginal-delivered infants[7]. On the other hand, caesarean section-delivered infants are dominated by *Firmicutes* (such as S*treptococcaceae* & *Staphylocaccaeae)*[8]. Therefore, until the gut microbiota is stabilized. it goes through major compositional changes starting with the initial colonization with aerobic bacteria[9]. Later on, when the oxygen levels are depleted, the aerobic bacteria are outcompeted by anaerobes (such as *Firmicutes* & *Actinobacteria*) [10]. With age, the development slows down and reaches the so-called adult-like state of the microbiota by the about three years[11]. Once established, the neonatal microbiome achieves a symbiotic relationship with the host and is critical for several metabolic functions.

Preterm infant gut microbiota

Preterm infants (born <37 weeks of gestation) complete their development in the extra uterine environment. The infants born <33 weeks of gestation are mostly with very low birth weight (<1500 g) and have a weak immune system. Due to this, these infants are vulnerable to many different infections [12, 13]. This can be due to the fact that they spend many months at the hospitals and are exposed to different factors such as chemicals [14], parenteral feeding [15] and exposure to neonatal intensive care unit (NICU) microbiota [14, 16]. Exposure to different postnatal therapies shape the succession of the preterm infant gut microbiome. Having said so, this early microbiome is of great importance to preterm infant health and for the development of the immune system [17]. Overall, the gestational age, birthweight and exposure factors are the most important factors that limit the support of a healthy gut microbiome in Very Low Birthweight Infants (VLBW) preterm infants. By the introduction of breast milk, the inflammatory responses have significantly reduced and has introduced a diversity of commensal bacterial species [18]. Hence, this reduces the ability of the microbes to penetrate into the host epithelium [19].

Microbiota perturbations in early life

The colonization of the GIT is perturbed by different factors such as caesarean delivery [20], birthweight, gestational age, usage of antibiotics [21] and slower GIT transit time [22].

The mode of delivery does not significantly affect the initial colonization microbe community but is highly dependent on the environment [23]. The hospital surfaces and the exposure to different microbes change the succession of these microbes colonizing the infant gut. Therefore, these factors that threaten the development of a healthy commensal microbiota result in a distinct microbiota with decreased microbial diversity and increased pathogens. This may pre-empt risk for sepsis and necrotizing enterocolitis (NEC) [15] especially in preterm infants. One of the most devastating diseases that commonly affects VLBW preterm infants is NEC [24]. Approximately 1% to 5% of VLBW preterm infants develop NEC with a mortality rate of 25% to 33% [25]. The primary risk factors for NEC are enteral feeding, abnormal gut microbiota development and prematurity of the GIT [15, 20].

Nutrition during the development of the infant play a major role in the early colonization patterns. The breast-fed infants receive a mix of nutrients, antimicrobial proteins and commensal bacterial. The antimicrobial peptides such as lactoferrin prevent the colonization of enteropathogens and stimulate growth of *Bifidobacterium* [26]. On the other hand, formula-

fed infants are exposed to a different set of nutrients and microbes. The breast-fed infants are dominated by *Bifidobacteriaceae* with decreased populations of *Enterobacteriaceae*, however the formula-fed infants are dominated by *Enterobacteriaceae* and *Clostridia[27]*. Increased number of *Firmicutes* with decreased population of *Bifidobacteria* has shown predisposition of the gut microbiota resulting in obesity[28]. Given that obesity has long-term effects on lifespan and quality of life, seeking to understand further into the metabolic actions directing towards the composition of the gut microbiota will be an important focus for research.

Antibiotic perturbations during the infant gut development disrupts the ecology of the microbiota leading to a dysbiosis [29]. Preterm infants are routinely given antibiotics including penicillin, cephalexin, gentamicin, amikacin, vancomycin, clindamycin and ampicillin. These antibiotics and combination of these antibiotics have found to increase the percentage of opportunistic bacteria while lowering the diversity of the commensal bacteria[23, 30]. The dysbiosis in the microbiota have profound effects associated with large number of health problems such as increased risk to immunological disorders such as asthma[31] and atopy[32], metabolic disruptions such as obesity[33] and developmental disorders such as autism[34]. Infants exposed to antibiotics also experience long-term disruptions with decreased abundance of *Bifidobacteriaceae* for up to 90 days after administration[35]. In addition to the disruption of the microbiota balance, antibiotic treatment equally enriches the gut-associated Antibiotic Resistance (AR) i.e. resistome development[23]. This resistome is shown to be persistent for long periods and can potentially transfer to other microbial communities and transient pathogens by horizontal gene transfer (HGT) [30, 36]. Notably, the route of antibiotic administration has played a role in the emergence of resistome population [23, 37]. Orally-given antibiotics showed greater resistome development compared to intravenous-administered antibiotics. There clearly depends on the assimilation of the antibiotics in the humans[37]. Concise information on the different exposure factors is given in Figure 1

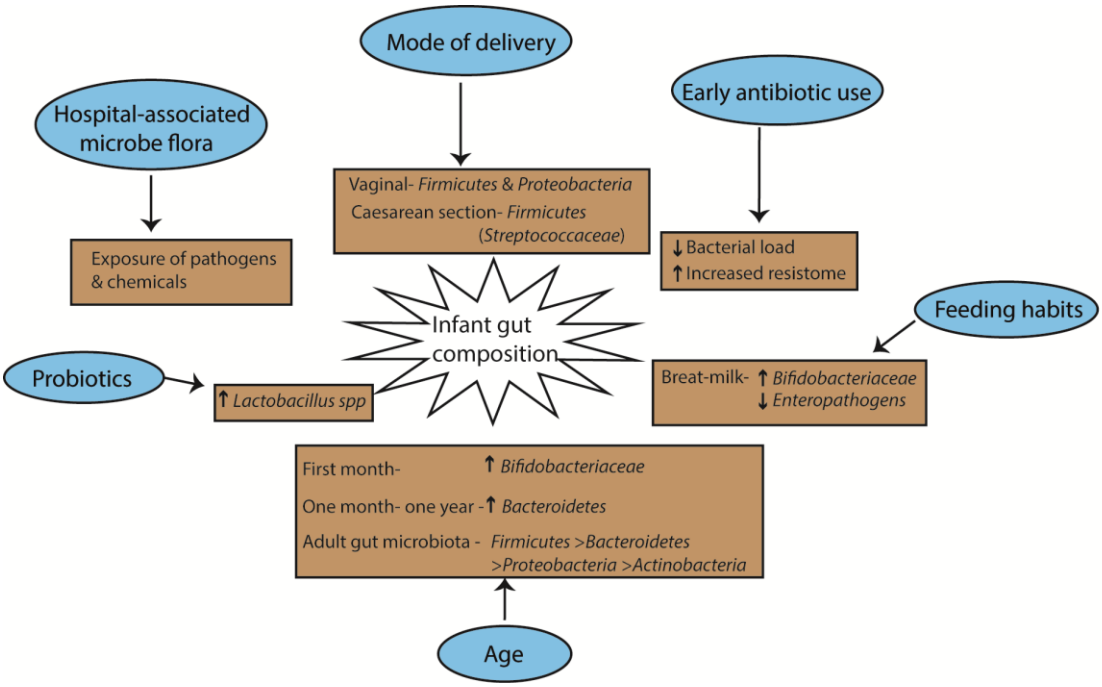


Figure 1: The influence of external factors to the infant gut microbiota composition

1.2 Gut mobilome

The microbial ecosystem in the GIT host a wide range of mobile genetic elements (MGEs) which in turn constitute the mobilome[38]. This mobilome in the complex microbial communities consists of genetic information that can be accessible by more than one bacteria[39]. HGT plays a major role in accessing this genetic information by the transfer of MGEs between different bacterial species[40]. Therefore a combination of specific bacterial phylotypes and the diverse functional attributes of MGEs can potentially alter the phenotypic properties of the bacterium.

There is limited information pertaining the role of the mobilome in the functioning and adaptability of the gut microbiota[38]. Recent studies have indicated that there is long term association between MGEs and the core gut microbes[41]. In addition to this, the role of HGT of these MGEs through the microbial communities is increasingly acknowledged [42, 43]. Therefore, understanding the role of the mobilome for the survival and persistence of several bacterial phylotypes as well as their importance in microbe-microbe interactions is yet to be explored in depth[44]. Taken together, the high level of novel gene content that is carried by the MGEs and the high diversity of the gut microbiota, makes these MGEs a potential black box relevant in understanding the functioning of the gut microbiota. Therefore understanding the distribution, diversity and persistence of these elements is of immense interest. Examples of MGEs include plasmids, transposons and integrons[40] (Figure 1).

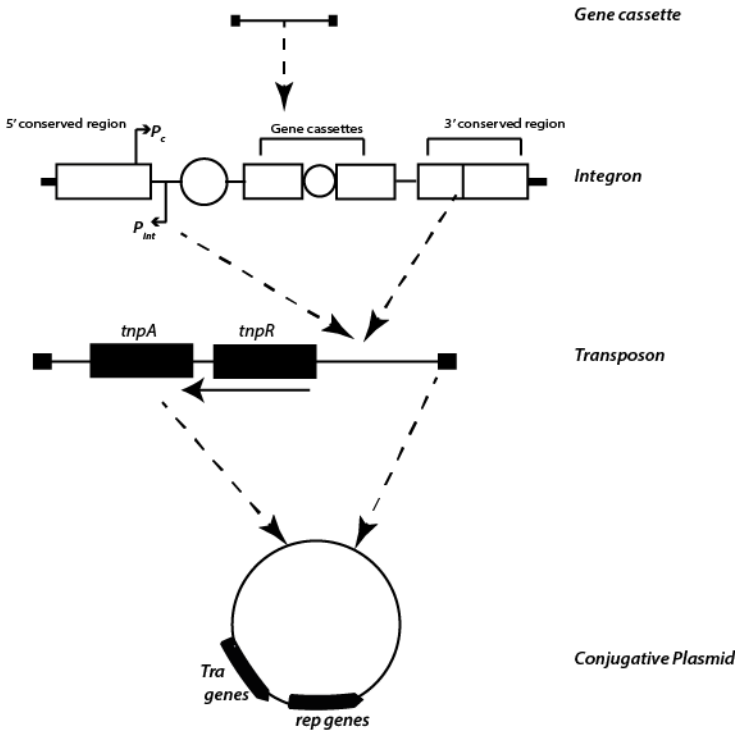


Figure 2: Schematic outline of the different MGEs that attribute the functional gene cassettes. The gene cassettes are found in integrons. Integrons are closely related to transposons which are in turn found on conjugative plasmids.

1.3 Plasmids

Plasmids are ubiquitous in any environment and are the driving force of bacterial evolution and HGT[45]. Most plasmids are capable of transferring to distinct species of bacteria and can encode a wide range of accessory elements that could be beneficial for the host. In addition, many studies have shown the carriage of multidrug resistance genes within the plasmids [46-49]. There are different types of plasmids that exists in different environments but the conjugative plasmids are of our special interest.

Conjugative plasmids

The conjugative plasmids are higher order plasmids that are large (50- 200 Kbp) in size and can replicate autonomously. They have the innate ability to spread to different sets of host population without much cost to the host bacterium. This type of transfer ensures the prevalence of the plasmid in the environment and reduces its chances of total extinction[50]. Due to this, the conjugative plasmids are termed as 'selfish' DNA elements due to their parasitic nature of transmission[50]. They also harbour several essential and non-essential genes for the host bacterium. For instance, conjugative plasmids in *Brucella meilitensis* 16 M were defined as a second chromosome due to the discovery of host-essential genes in the

plasmid[51]. However plasmids discovered in *Xanthomonas citri* are regarded as non-essential plasmids[52] since the accessory genes encoded by the plasmid belong to a virulence nature. Therefore, the non-virulent strains of *X. citri* become virulent after the introduction of plasmid pXcB.

The conjugative plasmids have two important regions that are crucial for their maintenance and stability in the host and environment [53]. These functional groups belong to the replication, stability and transfer. The genes related to this are compactly located in almost all conjugative plasmids. The *rep* (replication initiation) genes are associated with *cop* genes that aid in the maintenance of copy numbers in the host bacterium [54]. High copy number of large plasmids can cause energy loss to the host bacteria, therefore they are not preferred [55]. On the other hand, very low copy numbers can lead to plasmid cured daughter cells [55]. Therefore, stability of copy number is important for long term existence of plasmids in the host cells. Additional gene families (*par* & *mrs*) are associated to ensure copy number maintenance and vertical transmission of conjugative plasmids to the bacterial daughter cells[50]. The transfer of large conjugative plasmids are related to mobilization (*mob*) and transfer (*tra*) for consequent transmission to other bacterial cells [54]. Smaller plasmids that do not possess the conjugation machinery rely on larger transmissible plasmids by co-transfer or co-integration. Plasmid addiction systems (stb, Toxin/antitoxin systems) are modules that further ensure plasmid carrying bacterial cells. Plasmid cured cells are subsequently killed [56]. Comprehensive information on the functional groups located in conjugative plasmids is given here (Figure 3).

Figure 3: The most important functional groups of the conjugative plasmids.

The main classification of conjugative plasmids referred nowadays is through the incompatibility nature. Plasmid incompatibility is defined as the incompatibility of two plasmid groups to reside in the same bacterial cell[57]. This is due to the inability of sharing the replication apparatus between the two plasmid groups. This in turn destabilizes and degrades the inheritance of one plasmid. Up until now, 27 incompatibility groups have been recognized with variants in each group (such as IncF, IncP, IncN conjugative plasmids etc.) [58]. Even though conjugative plasmids have additional gene modules that assist in maintenance in the host cell, they have a narrow host range. Most of the conjugative plasmids have host range limiting to *Enterobacteriaceae* family [58].

IncF conjugative plasmids

IncF plasmids are low-copy number plasmids that are usually >100kbp in size and have a limited host range within the *Gammaproteobacteria* [59]. These plasmids have been associated as the primary vectors for the sudden spread and emergence of extended-spectrum β- lactamases (ESBL) [60], plasmid mediated AmpC [58], quinolone [61] and aminoglycoside resistances[62]. The plasmids are versatile in intercellular adaptation and are widely diffused to clinically relevant *Enterobacteriaceae* such as entero-pathogenic, entero-invasive and entero-haemorrhagic *E.coli*[59]. The most common variants for the IncF plasmids are IncFII, IncFIA, IncFIB and IncFIC. The IncFIC is rendered cryptic and IncFII are often associated to IncFIA or IncFIB[59]. When associated with IncFIA or IncFIB, the IncFII plasmids do not participate in the initiation of replication and are free to diverge to generate new compatible variants. Overall, the IncF plasmids contribute to the fitness of the host by providing virulence and AR determinants.

IncI conjugative plasmids

One of the first incompatibility groups defined was IncI plasmids that produces type I pili[58]. The IncI plasmids are classified into 2 variants- IncI1 & IncI2[58]. The IncI plasmids have been isolated in diverse *Enterobacteriaceae* that they are further typed through plasmid multi-locus sequence typing (pMLST) and 15 sequence types have been submitted[63]. The type IV pili are associated with IncI1 plasmids[64]. they have a complex transfer region that extends to 50 kbp with 2 types of conjugative pili: thick and thin pili[64]. The thick pili is essential for DNA transfer and conjugation in semi-solid agar and the thin pili are essential in stabilizing the conjugants and mating in liquid media.

Accessory elements

The conjugative plasmids, in addition to the replication and maintenance genes modules harbour a mosaic of different adaptive traits that are beneficial for the host. Conjugative plasmids often carry these accessory elements that benefit the host in specific environmental niches[65]. These accessory functions commonly associated are AR, rapid adaptation to specific environments and degradation of specific xenobiotics[50]. Integrons and transposons are commonly associated with conjugative plasmids[54].

The integrons are known for their carriage of multidrug resistance genes as gene cassettes[66]. They are a platform for the integration, assembly and expression of promoterless genes that code for a particular resistance [67]. The integrons are generally non-mobile but are found in close association with plasmids and transposons (Figure 1). The integron itself consists of 3 major parts, the integrase gene that helps in the integration of the gene cassettes, the attachment site is where the integrated gene cassettes are attached to and the overall common promoter for the expression of the gene cassettes[66, 68] (Figure 3). Until now 5 different classes of integrons have been characterized and the class I integrons is the most studied. The organization of the integrons in a conjugative plasmid generally increases the fitness of the bacteria[69].

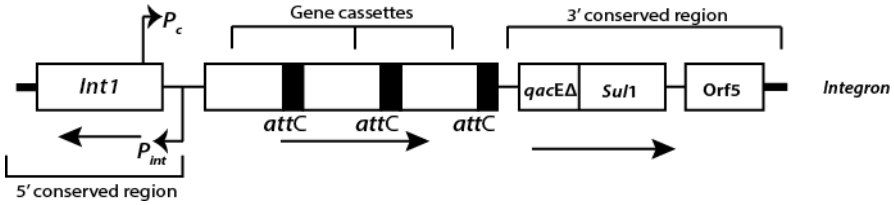


Figure 4: Representation of the integrons. The *int*1 is the integrase gene of the integron. The P_c and P_{int} are promoter genes of the integrase gene and the overall gene cassettes. *attc* are the attachment sites of the integron. *qacE*Δ and *sul*1 are resistance genes to quaternary compounds and sulphonamide respectively.

2. Aim of the thesis

Plasmid-mediated horizontal transfer of genes influences the bacterial community structure and evolution. However, the association of the microbial communities and the role of the MGEs in the gut microbiota is still elusive. The lack of existing literature and updated surveys on the extent of MGEs spread in various environments limits our knowledge on their diversity and association. Therefore the main aim of the thesis is to understand the prevalence, persistence and association of MGEs in the developing gut microbiota. The work was divided into

- Understanding the development of the gut microbiota from late pregnant mother to 2 years of the child. Identifying multidrug resistance integrons in the developing gut microbiota and their persistence and association with the microbial communities. A Norwegian mother- infant cohort (IMPACT) was used in the study [Paper 1].
- Understanding and characterizing the mobilome of the preterm infant's gut microbiota. Identifying Operational Taxonomic Units (OTUs) that are significantly related in regards to birthweight, hospital location and NEC [Paper 2].
- Characterizing and association of the IncF conjugative plasmids in the developing gut microbiota of full-term infants. [Paper 3].
- Understanding the role and functional attributes of *E.coli* strains carrying multiple Inc plasmids isolated from a preterm twin pair from Spain [Paper 4]

Information on datasets used in the thesis

Cohort	Type of Infant	Location	Collection time (days)	Condition	Number of infants	Number of samples	No of strains
IMPACT	Full term	Trondheim, Norway	Early-late pregnant mothers, 3-10, 4 months, 1 year and 2 years	Healthy	147	663	N/A
Cohort -2	Preterm	Boston, Chicago & Evanston, USA	One time & Approx. weekly	Necrotizing enterocolitis and healthy	62	160	N/A
EarlyMicroHealth	Full term	Asturias, Spain	2, 10, 30 & 90	Healthy	47	180	N/A
Cohort-4	Preterm	Valencia, Spain	One time, 20 days	Healthy	2	2	74

2.1 Longitudinal cohort information

IMPACT

The IMPACT cohort (Immunology and Microbiology in Prevention of Allergy among Children in Trondheim)[70] is a controlled non-randomized longitudinal study, which began in 2000. The study involved 720 pairs of pregnant women and their children, up to two years of age. Ninety percent of the children were vaginally delivered. The fecal samples were collected from pregnant women during the first/second trimester (7-20 weeks) and third (32-40 weeks) trimester, and from the children at 3-10 days, 4 months, 1 and 2 years of life. For the paper 1, samples from a randomly selected subgroup of 147 mother-child pairs were analysed.

Cohort 2

This study consists of premature infants with and without NEC. All the infants with NEC showed >Bell's stage 2 NEC symptoms. The infants were recruited from three different hospitals in USA- Boston, MA (n=24); Chicago, IL (n=29); and Evanston, IL (n=9). All infants recruited in the study were born with a birthweight <1500 g and a gestational age <32weeks. Even though a case control model (each NEC infant is paired with 2 non-NEC infant) was implemented in 18 of the 23 samples, the infants were treated individually in this study. This study cohort was used in paper 2.

EarlyMicroHealth

This is an unselected longitudinal cohort of 47 healthy full-term infants, born between gestational weeks 38 and 41 (average 39.1) in Asturias, Spain. The fecal samples were collected from the infant at 2 days, 10 days, 30 days and 90 days. Eighty-three percent of the infants were vaginally delivered. None of the children were given antibiotics up to end of sampling. This study cohort was used in paper 3.

Cohort 4

The preterm twin pair was a part of a prospective, single-center observational study cohort from Valencia, Spain[71]. The twins selected for this study were born preterm (gestational age 30) and weighed 1410 g and 1630 g for twin A and twin B respectively. The infants stayed at the hospital until sampling even though they did not show any complications or signs of infection. The fecal samples were collected 20 days after birth. The children were

born by emergency caesarean section and breast-fed until sampling. No antibiotics were given until sampling. The twin samples were used in paper 4.

3. Results and discussion

The major findings of this thesis relate to the association of the mobilome with the developing gut microbiota. From all the four cohorts, we detected mobilome in the gut microbiota at different ages. Most frequently, conjugative plasmids and integrons were detected in the cohorts. These plasmids and accessory elements were highly prevalent and persistent across the different longitudinal datasets. In addition, we *de novo* assembled different variants of conjugative plasmids and integrons carrying diverse AR genes.

3.1 Prevalence and persistence of MGEs

Multidrug resistance integrons

From the PAPER 1 study cohort, we observed the prevalence of integrons in the gut microbiota of developing infants. Integrons can carry multiple AR genes as gene cassettes[72]. Therefore, the spread of integrons can be related to the spread of multidrug resistance genes[73]. The *int*1, integrase gene was used as the marker for the integron[74]. The general distribution of integrons in Norwegian mother-infant cohort was 15.3% (PAPER 1). This distribution is quite low compared to the integrons detected in other studies [75, 76]. This indicates that a strict regulation on antibiotic use in humans and animals are effective methods to eradicate multidrug resistant infections. Integrons were also detected in shotgun sequenced samples (PAPER 2- PAPER 4). They were harboured within conjugative plasmids. In all the longitudinal datasets, the integrons were highly persistent between the different time periods of the infants and prevalent within conjugative plasmids. In PAPER 1, the integrons showed high persistence at 3-10 days to 4 months [$p<0.0001$], 4 months to 1 year [$p<0.0001$] and between mother and child (Late Pregnant mother to 2 year child [$p<0.05$]). The high persistence of integrons between longitudinal samples can thereby be explained by their close association with plasmids and transposons[77].

Conjugative plasmids

Two types of conjugative plasmids were detected in this thesis, IncF (PAPER 2- PAPER 4) and IncI (PAPER 1 & PAPER 4) conjugative plasmids. The IncF conjugative plasmids are well studied and represented in bacteria from several human and animal sources[38]. They play a major role in the dissemination of specific AR (such as β-lactamases[78] and plasmid-mediated quinolones[61]) and virulence genes (such as cytotoxins and adhesion factors)[79]. The highest prevalence of IncF conjugative plasmids was detected in PAPER 3 where 54%

of the full term infants were detected with the IncFIB conjugative plasmid. A small proportion of the dataset (8%) was also detected with IncFIA. However, all samples positive to IncFIA was positive to IncFIB indicating a link between the conjugative plasmids (PAPER 3). In addition to this, the IncFIB was highly persistent between 2 days to 10 days [pvalue <0.0001] and 10 days to 30 days [pvalue <0.0001] (PAPER 3). Taken together, the *Enterobacteriaceae* family is one of the dominant groups of the infant gut microbiota. The IncF plasmids have a close host range within this family, making these plasmids widely distributed within the infant gut microbiota (discussion in PAPER 3). Strains isolated from preterm infants (PAPER 4) showed 64% harboured IncFIB along with IncI1 plasmids. In these strains, the IncI1 plasmid harboured all the transfer genes but the IncFIB plasmid harboured several accessory genes but very few transfer genes (PAPER 4). The presence of such plasmids in a bacteria initiates a so-called conjugational complex that helps to initiate replication for both conjugative plasmids[64]. However, information on this mode of replication is limited in regards to literature and reproducibility.

3.2 MGEs in the longitudinal cohorts

We identified conjugative plasmids in all our longitudinal cohorts. Conjugative plasmids, in general are larger in size and carry multiple accessory genes that are essential or non-essential to the bacterial host [39]. The *de novo* assembled conjugative plasmids were identified in the preterm infant cohort and were associated with NEC-positive infants and hospital location (PAPER 2). Distinct regions of the plasmids i.e. *int*1 gene of the integron, *yih*A gene belonging to haemolysin modulating expression gene family (*Hha* family) and *rep*A gene of the IncF conjugative plasmid were detected in the *de novo* assembled IncF plasmid (PAPER 2). The *hha* family of genes regulate expression of α- hemolysin toxin and other virulence factors [80]. The α-hemolysin toxin is related to enterocolitis in humans and birds [81]. The integron within a transposon carried trimethoprim, streptomycin, β- lactam antibiotics and sulphonamides related resistance genes (PAPER 2). In addition to this, the assembled IncF conjugative plasmid contained all the genes necessary for the transfer (*tra*A- *tra*X) and replication of the IncF conjugative plasmid (PAPER 2). In the PAPER 4, we detected 2 types of conjugative plasmids of IncFIB and IncI. In most of the strains, the IncFIB was non-mobile with only TraX and FinO. However, the IncI plasmids contained all the transfer genes (TraA-TraY) and pilus genes (Pil genes). The IncI plasmids are known for their complex transfer system that extends to over 50kb with two types of conjugative pilus regions. This *de novo* assembled IncI plasmid was concordant with that [64] (PAPER 4). In relation to the detection

of plasmids with and without transfer genes, we attempted to transfer the conjugative plasmids *in vitro* to other *E.coli* strains (PAPER 4). From the transmission experiments, the IncI plasmid due to the presence of transfer genes could transfer but not the IncFIB. Therefore, the IncFIB plasmid of this strain collection were non-mobile and native plasmids for these bacterial strains (PAPER 4).

In addition to detection of plasmids in preterm infants, we *de novo* assembled conjugative plasmids with multidrug resistance genes in the healthy full term infants as well (PAPER 1 & PAPER 3). The mobilome has the potential to vary in terms of genetic diversity and functions required over the lifetime of the host and its environment. Conjugative plasmids, largely attributed to HGT is known for its role in the acquisition of multiple AR genes and novel functional genes benefitting the host[50]. Detection of MGEs in developing gut microbiota indicates the versatility of MGEs to withstand major perturbations (discussion in PAPER 1). The integrons of the IncI conjugative plasmid harboured resistance genes to aminoglycosides, sulphonamides and trimethoprim (PAPER 1). In PAPER 3, we *de novo* assembled IncF conjugative plasmids from three longitudinal datasets of the Spanish cohort. In two longitudinal datasets, conjugative plasmids exhibited multi replicon status whereby shared IncFIA/IncFIB plasmid- the transfer system is shared between IncFIA and IncFIB and integrated IncFIA/IB plasmid- the IncFIA and IncFIB are integrated into one plasmid (PAPER 3

Conjugative plasmids are ubiquitous due to their special properties in achieving persistence in complex environments[54]. Copy number control (cop genes) [82], active patitioning systems (*par*A/*par*B) [83] and post segregationally killing (TA systems such as *ccd*A/*ccd*B)[84] are such properties that help to maintain persistence. The *de novo* assembled plasmids from both the projects harbour these genes, indicating the long term persistence of these elements in the gut microbiota (PAPER 3 & PAPER 4). In addition to the persistence mechanisms, the IncFIB plasmids of the PAPER 4 include virulence genes such as *IroBCDEN* [85] and aerobactin biosynthesis gene family. The former and later gene families are associated with extraintestinal pathogenic *E. coli* (ExPEC) (discussion in PAPER 4).

Functional attributes of conjugative plasmids

Conjugative plasmids contain diverse accessory elements in addition to their transfer and replication apparatus. Therefore, the characteristics of the accessory elements define the nature of the conjugative plasmids [86]. In the gut microbiota, we identified diverse types of

conjugative plasmids assembled from different datasets (PAPER 1- PAPER 4). The detection of diverse conjugative plasmids has been related to the co-evolution of bacteria within the human host[38]. Therefore some MGEs may therefore be unique or enriched in particular datasets[38]. The plasmids detected in the preterm infant dataset of PAPER 2 harboured potential virulence genes in the IncF plasmids but the plasmids detected in the strains of the preterm twin pair of PAPER 4 harboured various AR genes and virulence factors. These plasmids portray a wide diversity and adapt to the environmental conditions. We detected a plasmid showing bacteriocin activity that inhibiting a group of commensal *Enterobacteriaceae* (PAPER 4). Bacteriocin production by conjugative plasmids, in general has been shown to augment niche competition whereas the bacteriocin producers outcompetes the non-producers[87]. They are shown to be important mediators for intra- and interspecies interactions and for maintaining the microbial diversity. The presence of plasmid survival genes in addition to AR genes harboured in the backbone of the plasmid gives the host bacterium a competitive advantage during antibiotic treatment compared to the commensal bacteria[88] (discussion in PAPER 4).

3.3 Microbiota association with MGEs

The detection of integrons in the gut microbiota could not be related to any particular bacterial phylotypes across the different individuals in our sample (PAPER 1 & PAPER 2). Hence it is unlikely that the integrons have a strict bacterial phylotype. Since integrons are immobile structures that are found in close association with plasmids and transposons, the most likely explanation for the lack of association could be due to the presence of multiple integrons in different conjugative plasmids within the samples[66]. However, when we focussed on the association of particular conjugative plasmids with the gut microbiota composition, several bacterial phylotypes showed significant correlation (PAPER2, PAPER 3).

We found a strong correlation between the abundance of OTU1 classified as *Enterobacteriaceae* and the prevalence of IncFIB conjugative plasmids over time (PAPER 3). In fact, this abundance of OTU1 in samples with and without IncFIB depict a unique development (results in PAPER 3). The significant association between IncFIB and OTU1 could be possibly due to the narrow host range of IncF plasmids and the high proportion of *Enterobacteriaceae* in the microbiota population (discussion in PAPER 3). The strongest association was observed at 2 days, 10 days, and 30 days and with vaginal delivery. The association with vaginal delivery indicates the possibility of vertical transmission of

conjugative plasmids from the mother to the child during birth [89] (discussion in PAPER 3). However, without the information of the mother's gut microbiota, we cannot be very certain.

When distinct regions of the conjugative plasmid were screened and association with the gut microbiota composition was calculated in the preterm infant cohort (PAPER 2), OTU2 classified as *Enterobacteriaceae* showed a significant positive association to the NEC and hospital location. In addition to this, the replication regulatory region (*rep*) of the IncF plasmid and *int*1 gene of the integron showed a positive association towards OTU2 (PAPER 2). Overall, the samples from Evanston had higher prevalence of the signature genes compared to the other hospitals (PAPER 2). Therefore, the potential characteristics of a particular bacterial strain could also be related to the specific genetic elements encoded extra chromosomally and not necessarily related to its phenotypic characteristics[90] (discussion in PAPER 2).

4. Conclusion

In this thesis, we have prospectively studied the mobilome of the gut microbiota and its association with the microbial community. From all the datasets, we detected a mobilome that showed diverse MGEs. We discovered a diversity of conjugative plasmids between the different datasets of full term and preterm infants. These conjugative plasmids and integrons were persistent between the longitudinal samples. In addition to the detection and persistence, these plasmids harboured different accessory elements according to the environmental exposures. This descriptive knowledge on the ecology, prevalence and persistence in longitudinal datasets has enabled us to move further into the understanding the functional attributes of MGEs. These have shown the versatility of these plasmids and their influence in adaptability and establishment of the developing gut microbiota.

5. Future perspectives

This thesis has facilitated our understanding of the general diversity of MGEs in the developing gut microbiota. However further work needs to be implemented on the role of the MGEs in the development of the gut microbiota and their involvement in community functions and interactions. Therefore, future work should be on categorizing the different MGEs in human or animal population. Overall, the gut mobilome constitutes a vast amount of genetic information that has the potential to enhance our understanding of transmission of AR genes and their functions in the microbial ecosystem. Therefore, in order to access this information, current bioinformatics tools need to be redesigned to detect MGEs and utilizing longitudinal cohorts. Multidrug resistance genes are known for their existence in MGEs, therefore targeting MGEs for the eradication of multidrug resistance rather than the strains itself could be additional barriers against multidrug resistant strains.

6. References

1. Sekirov, I., et al., *Gut Microbiota in Health and Disease.* Physiological Reviews, 2010. **90**(3): p. 859-904.
2. Putignani, L., et al., *The human gut microbiota: a dynamic interplay with the host from birth to senescence settled during childhood.* Pediatr Res, 2014. **76**.
3. Rodriguez, J.M., et al., *The composition of the gut microbiota throughout life, with an emphasis on early life.* Microb Ecol Health Dis, 2015. **26**: p. 26050.
4. Berrington, J.E., et al., *The neonatal bowel microbiome in health and infection.* Curr Opin Infect Dis, 2014. **27**.
5. Gritz, E.C. and V. Bhandari, *The Human Neonatal Gut Microbiome: A Brief Review.* Frontiers in Pediatrics, 2015. **3**: p. 17.
6. Aagaard, K., et al., *The placenta harbors a unique microbiome.* Sci Transl Med, 2014. **6**.
7. Collado, M.C., et al., *Factors influencing gastrointestinal tract and microbiota immune interaction in preterm infants.* Pediatr Res, 2015. **77**(6): p. 726-731.
8. Penders, J., et al., *Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy.* Pediatrics, 2006. **118**(2): p. 511-521.
9. Palmer, C., et al., *Development of the Human Infant Intestinal Microbiota.* PLoS Biol, 2007. **5**(7): p. e177.
10. Avershina, E., et al., *Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children.* FEMS Microbiology Ecology, 2013: p. n/a-n/a.
11. Avershina, E., et al., *Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children.* FEMS Microbiol Ecol, 2014. **87**(1): p. 280-90.
12. Morrow, A.L., et al., *Early microbial and metabolomic signatures predict later onset of necrotizing enterocolitis in preterm infants.* Microbiome, 2013. **1**(1): p. 1-16.
13. Westerbeek, E.A.M., et al., *The intestinal bacterial colonisation in preterm infants: A review of the literature.* Clinical Nutrition, 2006. **25**(3): p. 361-368.
14. Santos, J., S.E. Pearce, and A. Stroustrup, *Impact of Hospital-Based Environmental Exposures on Neurodevelopmental Outcomes of Preterm Infants.* Current opinion in pediatrics, 2015. **27**(2): p. 254-260.
15. Patel, R.M. and P.W. Denning, *Intestinal microbiota and its relationship with necrotizing enterocolitis.* Pediatr Res, 2015. **78**(3): p. 232-238.
16. Brooks, B., et al., *Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants.* Microbiome, 2014. **2**.
17. de la Cochetiere, M.-F., et al., *Early Intestinal Bacterial Colonization and Necrotizing Enterocolitis in Premature Infants: The Putative Role of Clostridium.* Pediatr Res, 2004. **56**(3): p. 366-370.
18. Updegrove, K., *Necrotizing Enterocolitis: the Evidence for Use of Human Milk in Prevention and Treatment.* Journal of Human Lactation, 2004. **20**(3): p. 335-339.
19. Meier, P.P. and L. Bode, *Health, nutrition, and cost outcomes of human milk feedings for very low birthweight infants.* Adv Nutr, 2013. **4**.
20. Warner, B.B. and P.I. Tarr, *Necrotizing enterocolitis and preterm infant gut bacteria.* Seminars in Fetal and Neonatal Medicine, 2016. **21**(6): p. 394-399.
21. Cotten, C.M., 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**.

22. Gewolb, I.H., et al., *Stool microflora in extremely low birth weight infants.* Arch Dis Child Fetal Neonatal Ed, 1999. **80**.
23. Gibson, M.K., T.S. Crofts, and G. Dantas, *Antibiotics and the developing infant gut microbiota and resistome.* Curr Opin Microbiol, 2015. **27**: p. 51-6.
24. Mai, V., et al., *Fecal Microbiota in Premature Infants Prior to Necrotizing Enterocolitis.* PLoS ONE, 2011. **6**(6): p. e20647.
25. Henry, M.C.W. and R.L. Moss, *Current issues in the management of necrotizing enterocolitis.* Seminars in Perinatology, 2004. **28**(3): p. 221-233.
26. Mastromarino, P., et al., *Correlation between lactoferrin and beneficial microbiota in breast milk and infant's feces.* BioMetals, 2014. **27**(5): p. 1077-1086.
27. Guaraldi, F. and G. Salvatori, *Effect of breast and formula feeding on gut microbiota shaping in newborns.* Front Cell Infect Microbiol, 2012. **2**.
28. Gurnani, M., C. Birken, and J. Hamilton, *Childhood obesity: causes, consequences, and management.* Pediatr Clin N Am., 2015. **62**.
29. Langdon, A., N. Crook, and G. Dantas, *The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation.* Genome Medicine, 2016. **8**(1): p. 39.
30. Moore, A.M., et al., *Gut resistome development in healthy twin pairs in the first year of life.* Microbiome, 2015. **3**(1): p. 27.
31. Pitter, G., et al., *Antibiotic exposure in the first year of life and later treated asthma, a population based birth cohort study of 143,000 children.* European Journal of Epidemiology, 2016. **31**(1): p. 85-94.
32. Kusel, M.M.H., et al., *Antibiotic use in the first year of life and risk of atopic disease in early childhood.* Clinical & Experimental Allergy, 2008. **38**(12): p. 1921-1928.
33. Turta, O. and S. Rautava, *Antibiotics, obesity and the link to microbes - what are we doing to our children?* BMC Medicine, 2016. **14**(1): p. 57.
34. Fallon, J., *Could one of the most widely prescribed antibiotics amoxicillin/clavulanate "augmentin™" be a risk factor for autism?* Medical Hypotheses, 2005. **64**(2): p. 312-315.
35. Arboleya, S., et al., *Intestinal Microbiota Development in Preterm Neonates and Effect of Perinatal Antibiotics.* J Pediatr., 2015. **166**.
36. Moore, A.M., et al., *Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes.* PLoS One, 2013. **8**(11): p. e78822.
37. Zhang, L., et al., *Antibiotic Administration Routes Significantly Influence the Levels of Antibiotic Resistance in Gut Microbiota.* Antimicrobial Agents and Chemotherapy, 2013. **57**(8): p. 3659-3666.
38. Ogilvie, L.A., S. Firouzmand, and B.V. Jones, *Evolutionary, ecological and biotechnological perspectives on plasmids resident in the human gut mobile metagenome.* Bioeng Bugs, 2012. **3**(1): p. 13-31.
39. Slater, F.R., et al., *Progress towards understanding the fate of plasmids in bacterial communities.* FEMS Microbiol Ecol, 2008. **66**(1): p. 3-13.
40. Frost, L.S., et al., *Mobile genetic elements: the agents of open source evolution.* Nat Rev Microbiol, 2005. **3**(9): p. 722-32.
41. Zaneveld, J.R., et al., *Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives.* Nucleic Acids Research, 2010. **38**(12): p. 3869-3879.
42. Smillie, C.S., et al., *Ecology drives a global network of gene exchange connecting the human microbiome.* Nature, 2011. **480**(7376): p. 241-4.
43. Huddleston, J.R., *Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes.* Infect Drug Resist, 2014. **7**: p. 167-76.

44. Jones, B.V., *The human gut mobile metagenome: a metazoan perspective.* Gut Microbes, 2010. **1**(6): p. 415-31.
45. Mazodier, P. and J. Davies, *Gene transfer between distantly related bacteria.* Annu Rev Genet, 1991. **25**: p. 147-71.
46. Kruse, H. and H. Sørum, *Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments.* Applied and Environmental Microbiology, 1994. **60**(11): p. 4015-4021.
47. Marcade, G., et al., *Replicon typing of plasmids in Escherichia coli producing extended-spectrum beta-lactamases.* J Antimicrob Chemother, 2009. **63**(1): p. 67-71.
48. Shin, S.W., et al., *Profiling of antimicrobial resistance and plasmid replicon types in β-lactamase producing Escherichia coli isolated from Korean beef cattle.* Journal of Veterinary Science, 2015. **16**(4): p. 483-489.
49. Carattoli, A., *Plasmids in Gram negatives: molecular typing of resistance plasmids.* Int J Med Microbiol, 2011. **301**(8): p. 654-8.
50. Norman, A., L.H. Hansen, and S.J. Sørensen, *Conjugative plasmids: vessels of the communal gene pool.* Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2009. **364**(1527): p. 2275-2289.
51. DelVecchio, V.G., et al., *The genome sequence of the facultative intracellular pathogen Brucella melitensis.* Proceedings of the National Academy of Sciences, 2002. **99**(1): p. 443-448.
52. El Yacoubi, B., et al., *In Planta Horizontal Transfer of a Major Pathogenicity Effector Gene.* Applied and Environmental Microbiology, 2007. **73**(5): p. 1612-1621.
53. Thomas, C.M., *Paradigms of plasmid organization.* Molecular Microbiology, 2000. **37**(3): p. 485-491.
54. Bahl, M.I., L.H. Hansen, and S.J. Sørensen, *Persistence Mechanisms of Conjugative Plasmids*, in *Horizontal Gene Transfer: Genomes in Flux*, M.B. Gogarten, J.P. Gogarten, and L.C. Olendzenski, Editors. 2009, Humana Press: Totowa, NJ. p. 73-102.
55. Watve, M.M., N. Dahanukar, and M.G. Watve, *Sociobiological Control of Plasmid Copy Number in Bacteria.* PLOS ONE, 2010. **5**(2): p. e9328.
56. Olendzenski, M.B.G.J.P.G.L., ed. *Horizontal Gene transfer, Genomes in Flux*. 1 ed. Methods in Molecular Biology, ed. J.M. Walker. 2009, Springer: Uniter Kingdom. 542.
57. Novick, R.P., *Plasmid incompatibility.* Microbiological Reviews, 1987. **51**(4): p. 381-395.
58. Carattoli, A., *Resistance plasmid families in Enterobacteriaceae.* Antimicrob Agents Chemother, 2009. **53**(6): p. 2227-38.
59. Villa, L., et al., *Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants.* J Antimicrob Chemother, 2010. **65**(12): p. 2518-29.
60. Novais, Â., et al., *Emergence and Dissemination of Enterobacteriaceae Isolates Producing CTX-M-1-Like Enzymes in Spain Are Associated with IncFII (CTX-M-15) and Broad-Host-Range (CTX-M-1, -3, and -32) Plasmids.* Antimicrobial Agents and Chemotherapy, 2007. **51**(2): p. 796-799.
61. Lascols, C., et al., *A Plasmid-Borne Shewanella algae Gene, qnrA3, and Its Possible Transfer In Vivo between Kluyvera ascorbata and Klebsiella pneumoniae.* Journal of Bacteriology, 2008. **190**(15): p. 5217-5223.
62. Galimand, M., et al., *Worldwide Disseminated armA Aminoglycoside Resistance Methylase Gene Is Borne by Composite Transposon Tn1548.* Antimicrobial Agents and Chemotherapy, 2005. **49**(7): p. 2949-2953.

63. García-Fernández, A., et al., *Multilocus sequence typing of IncI1 plasmids carrying extended-spectrum β-lactamases in Escherichia coli and Salmonella of human and animal origin.* Journal of Antimicrobial Chemotherapy, 2008. **61**(6): p. 1229-1233.
64. Thomas, C.M., *The horizontal gene transfer pool. Bacterial plasmids and gene spread*. 2005, School of Biological Sciences, University of Birmingham, UK Harwood academic publishers.
65. Ghigo, J.-M., *Natural conjugative plasmids induce bacterial biofilm development.* Nature, 2001. **412**(6845): p. 442-445.
66. Davies, J. and D. Davies, *Origins and evolution of antibiotic resistance.* Microbiol Mol Biol Rev, 2010. **74**(3): p. 417-33.
67. Su, J., et al., *Analysis of integrons in clinical isolates of Escherichia coli in China during the last six years.* FEMS Microbiol Lett, 2006. **254**(1): p. 75-80.
68. Fluit, A.C. and F.J. Schmitz, *Resistance integrons and super-integrons.* Clin Microbiol Infect, 2004. **10**(4): p. 272-88.
69. Broaders, E., C.G. Gahan, and J.R. Marchesi, *Mobile genetic elements of the human gastrointestinal tract: potential for spread of antibiotic resistance genes.* Gut Microbes, 2013. **4**(4): p. 271-80.
70. Øien, T., O. Storrø, and R. Johnsen, *Intestinal microbiota and its effect on the immune system — A nested case-cohort study on prevention of atopy among small children in Trondheim: The IMPACT study.* Contemporary Clinical Trials, 2006. **27**(4): p. 389-395.
71. Cernada, M., et al., *Genome-Wide Expression Profiles in Very Low Birth Weight Infants With Neonatal Sepsis.* Pediatrics, 2014.
72. Kovalevskaya, N.P., *Mobile Gene Cassettes and Integrons.* Molecular Biology, 2002. **36**(2): p. 196-201.
73. Mazel, D., *Integrons: agents of bacterial evolution.* Nat Rev Microbiol, 2006. **4**(8): p. 608-20.
74. L'Abee-Lund, T.M. and H. Sorum, *Class 1 integrons mediate antibiotic resistance in the fish pathogen Aeromonas salmonicida worldwide.* Microb Drug Resist, 2001. **7**(3): p. 263-72.
75. van der Veen, E.L., et al., *Effect of long-term trimethoprim/sulfamethoxazole treatment on resistance and integron prevalence in the intestinal flora: a randomized, double-blind, placebo-controlled trial in children.* Journal of Antimicrobial Chemotherapy, 2009. **63**(5): p. 1011-1016.
76. Vinué, L., et al., *Prevalence and diversity of integrons and associated resistance genes in faecal Escherichia coli isolates of healthy humans in Spain.* Journal of Antimicrobial Chemotherapy, 2008. **62**(5): p. 934-937.
77. Stalder, T., et al., *Integron Involvement in Environmental Spread of Antibiotic Resistance.* Frontiers in Microbiology, 2012. **3**: p. 119.
78. Coque, T.M., et al., *Dissemination of Clonally Related Escherichia coli Strains Expressing Extended-Spectrum β-Lactamase CTX-M-15.* Emerging Infectious Diseases, 2008. **14**(2): p. 195-200.
79. Johnson, T.J. and L.K. Nolan, *Pathogenomics of the Virulence Plasmids of Escherichia coli.* Microbiology and Molecular Biology Reviews : MMBR, 2009. **73**(4): p. 750-774.
80. Nieto, J.M., et al., *The hha gene modulates haemolysin expression in Escherichia coli.* Mol Microbiol, 1991. **5**(5): p. 1285-93.
81. Dittmar, E., et al., *Necrotizing enterocolitis of the neonate with Clostridium perfringens: diagnosis, clinical course, and role of alpha toxin.* Eur J Pediatr, 2008. **167**(8): p. 891-5.

82. Summers, D., *Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability.* Mol Microbiol, 1998. **29**(5): p. 1137-45.
83. Ebersbach, G. and K. Gerdes, *Plasmid segregation mechanisms.* Annu Rev Genet, 2005. **39**: p. 453-79.
84. Hayes, F., *Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest.* Science, 2003. **301**(5639): p. 1496-9.
85. Caza, M., et al., *Specific Roles of the iroBCDEN Genes in Virulence of an Avian Pathogenic Escherichia coli O78 Strain and in Production of Salmochelins.* Infection and Immunity, 2008. **76**(8): p. 3539-3549.
86. !!! INVALID CITATION !!! {}.
87. Kommineni, S., et al., *Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract.* Nature, 2015. **526**(7575): p. 719-722.
88. Buffie, C.G. and E.G. Pamer, *Microbiota-mediated colonization resistance against intestinal pathogens.* Nature reviews. Immunology, 2013. **13**(11): p. 790-801.
89. Mueller, N.T., et al., *The infant microbiome development: mom matters.* Trends Mol Med, 2015. **21**(2): p. 109-17.
90. Grishin, A., et al., *The role of the intestinal microbiota in the pathogenesis of necrotizing enterocolitis.* Seminars in Pediatric Surgery, 2013. **22**(2): p. 69-75.

PAPER 1

Ravi A, Avershina E, Foley LS, Ludvigsen J, Storrø O, Øien T, Johnsen R, McCartney LA, L'Abée-Lund MT and Rudi K (2015). The commensal infant gut meta-mobilome as a potential reservoir for persistent multidrug resistance integrons. *Scientific Reports.* Doi: 10.1038/srep15317.

SCIENTIFIC REPORTS

OPEN

The commensal infant gut meta-mobilome as a potential reservoir for persistent multidrug resistance integrons

Received: 03 March 2015
Accepted: 21 September 2015
Published: 28 October 2015

Anuradha Ravi[1,*], Ekaterina Avershina[1,*], Steven L. Foley[2], Jane Ludvigsen[1], Ola Storrø[3], Torbjørn Øien[3], Roar Johnsen[3], Anne L. McCartney[4], Trine M. L'Abée-Lund[5] & Knut Rudi[1]

Despite the accumulating knowledge on the development and establishment of the gut microbiota, its role as a reservoir for multidrug resistance is not well understood. This study investigated the prevalence and persistence patterns of an integrase gene (*int1*), used as a proxy for integrons (which often carry multiple antimicrobial resistance genes), in the fecal microbiota of 147 mothers and their children sampled longitudinally from birth to 2 years. The study showed the *int1* gene was detected in 15% of the study population, and apparently more persistent than the microbial community structure itself. We found *int1* to be persistent throughout the first two years of life, as well as between mothers and their 2-year-old children. Metagenome sequencing revealed integrons in the gut meta-mobilome that were associated with plasmids and multidrug resistance. In conclusion, the persistent nature of integrons in the infant gut microbiota makes it a potential reservoir of mobile multidrug resistance.

The spread of antibiotic resistance (AR) genes and development of multidrug resistance represent major threats to public health[1]. Until recently, pathogens have been the prime focus with respect to understanding the spread of multidrug resistance, with the commensal microbiota receiving much less attention. However, recent studies have shown the prevalence of AR genes in the commensal gut microbiota[2–5]. Furthermore, the gut microbiota shows a high rate of horizontal gene transfer (HGT), which was indicated to be up to 25-fold greater than that of bacteria in other environments[6]. Hence, the collective mobile genetic elements (MGEs) in the gut microbiota (i.e. the gut meta-mobilome) represent an important target for both understanding and combating the spread of multidrug resistance[5,7].

The gut microbiota forms a complex ecosystem. The gut is assumed sterile at birth[8,9] whereas just after birth, it goes through major shifts starting with facultative anaerobic bacteria (*Enterococcaceae* and *Streptococcaceae*)[10,11]. As oxygen levels deplete, strictly anaerobic bacteria (*Bifidobacteriales* and *Bacteroidetes*) take over and dominate in the gut[12]. This progression slows down as the microbiota reaches the adult-like state where an estimated 100–200 species co-exist in close proximity[13]. Although scientists have started to understand the shifts in the taxonomic composition of the developing microbiota from infancy to adulthood, the knowledge of the meta-mobilome, including the transmission and persistence of multiple antimicrobial resistance genes, is limited.

[1]Norwegian University of Life Sciences, Chemistry, Biotechnology and Food science department (IKBM), Campus Ås, Ås 1432, Norway. [2]National Center for Toxicological Research, U.S. Food and Drug Administration, Division of Microbiology, Jefferson, AR 72079. [3]Department of Public Health and General Practice, Norwegian University of Science and Technology, 9491 Trondheim, Norway. [4]Microbial Ecology & Health Group, Department of Food and Nutritional Sciences, University of Reading, Reading, UK. [5]Norwegian University of Life Sciences, Department of Food safety and Infection Biology, Campus Adamstuen, Oslo 0454, Norway. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to K.R. (email: knut.rudi@nmbu.no)

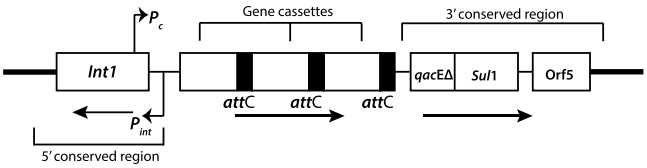


Figure 1. Structure of class I integron. A general representation of a class I integron with resistance gene cassettes at the attachment sites (*att*C) and a common promoter for the cassettes as P_c and for the integrase as P_{int}. The following cassettes are a part of the 3′ conserved region and not mobile: s*ul*1 gene encoding resistance to sulfonamides and *qac*EΔ encoding resistance to quaternary ammonium compounds.

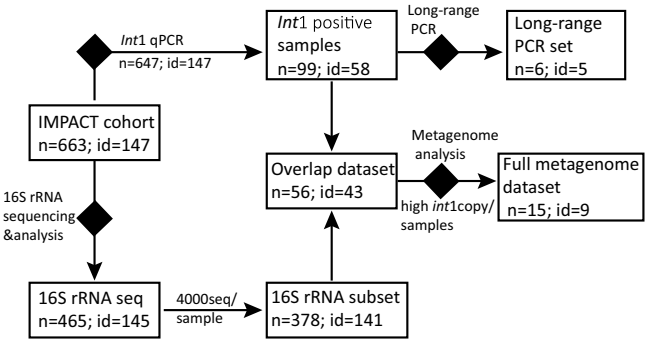


Figure 2. Workflow of experimental setup.

Antimicrobial resistance genes can be carried in integrons, which are non-mobile elements themselves, but are often found within MGEs like transposons and plasmids[14,15]. Integrons are platforms for integration, assembly and expression of specific gene cassettes within the MGEs that often encode antimicrobial resistance[16]. The individual genetic cassettes typically lack their own promoters, but are expressed by a common promoter for all the cassettes within the integron (Fig. 1). There have been 5 classes of integrons (class I–V) classified to date[5]. The class I integrons are the most widely studied and are found in a broad host range of commensal and pathogenic bacteria[17]. Class I integrons are found extensively in clinical isolates containing several different AR gene cassettes conferring resistance to antibiotics commonly used against bacterial infections[16,18]. Up to 8 gene cassettes have been found in a single class I integron[16], however hundreds of gene cassettes have been detected in so-called super-integrons[19].

The aim of the current study was to investigate the prevalence and persistence of class I integrons in a large unselected longitudinal cohort of mothers and their children. We used quantitative PCR to identify and study the persistence patterns of integrons. 16S rRNA and metagenome deep sequencing were used to analyze the phylogeny and genetic background of the integrons in the samples and to trace these elements longitudinally.

Materials and Methods

The schematic overview of the workflow is displayed in Fig. 2. The methods were performed in accordance to the approved guidelines and all experimental protocols were approved by Norwegian University of Life Sciences.

Cohort description. IMPACT (Immunology and Microbiology in Prevention of Allergy among Children in Trondheim) study is a controlled non-randomized longitudinal study, which began in 2000. The regional committee for Medical Research Ethics for Central Norway has approved the IMPACT study (ref. 120–2000). This study was granted a license by the Norwegian Data Inspectorate to process personal health data and one of the parents of each child signed a written informed consent form (r. 2003/953-3 KBE/-). Current controlled trials registration number: ISRCTN28090297.

The study involved 720 pairs of pregnant women and their children (up to two years of age). Ninety percent of the children were vaginally delivered and at term. Ninety-seven percent of the infants were breast-fed exclusively for the first six weeks of life. Fecal samples were collected from the pregnant women during the first/second (7–20 weeks) trimester and the third (32–40 weeks) trimester, and from the children at 3–10 days, 4 months, 1 and 2 years of age. In the current study, samples from a randomly selected subgroup of 147 mother-child pairs from the IMPACT cohort were analyzed. Information on allergy related hereditary diseases, atopy and antibiotic usage; health and exposure factors for the parent and child is summarized in Supplementary Table S1.

Sample collection. Fecal samples from the subjects of the IMPACT cohort were collected in Cary-Blaire transport and holding medium (BD Diagnostics, Sparks, MD). The samples were frozen at −20 °C within 2 h from collection. The samples were then stored at −80 °C within one month for children and mothers.

DNA purification. Fecal DNA was purified with an automated protocol using DNA extraction kit based on paramagnetic particles (LGC Genomics, UK). In brief, the samples were subjected to mechanical lysis using glass beads and the DNA was purified by eluting from the paramagnetic particles by downstream processes as described by manufacturer. The DNA was stored at −40 °C.

Gene quantification. The abundance of integrons (using the integrase (*int1*) gene[20] as a proxy) in the samples was calculated relative to the 16S rRNA gene by quantitative PCR. Each PCR reaction (25 μl) contained 1× HOT FIREPol PCR mix (Solis BioDyne, Estonia); 200 nM forward and reverse primers; one μl of sample DNA and water. The reaction mix was run on LightCycler 480 (Roche, Germany). Following the thermal cycling the raw fluorescence data was exported into LinRegPCR program. The software performed baseline corrections and calculated the mean PCR efficiency. For the *int1* amplicon, we also used High Resolution Melting (HRM) curve analysis, in addition to Sanger sequencing using the BigDye Terminator v.1.1 chemistry (Applied Biosystems) for verification.

The thermal cycling for the 16S rRNA primer-pair (5′-TCCTACGGGAGGCAGCAGT-3′; 5′-GGACTACCAGGGTATCTAATCCTGTT-3′) was an initial denaturation of 95 °C for 15 min followed by 40 cycles of 95 °C for 30 sec and 60 °C for 30 sec. This primer-pair targets conserved regions of the 16S rRNA gene[21]. The primers flanking the *int1* gene (5′-ACGAGCGCAAGGTTTCGGT-3′; 5′-GAAAGGTCTGGTCATACATG-3′) from Sørum *et al.*[11] were used with thermal cycling conditions 95 °C for 15 min and 40 cycles of 95 °C for 30 sec, 53 °C for 30 sec and 72 °C for 30 sec.

Microbial community analyses. Microbial communities were assessed using Illumina sequencing of 16S rRNA gene amplicons (n = 465), with subsets subjected to full metagenome (n = 15) and long-range PCR amplicon (n = 6) analyses. For full metagenomics analysis, samples were selected based on the high relative quantities of *int1* gene in the samples. For a long-range PCR, six *int1*-positive samples were randomly chosen for amplification.

Long-range primers were used to amplify the sequence flanking the region from *attI* to the 3′ consensus region including the gene cassettes (5′-GGCATCCAAGCAGCAAG-3′; 5′-AAGCAGACTTGA CCTGA-3′)[11] with the TaKaRa LA PCR kit Ver.2.1. The thermal cycling conditions of 94 °C for 5 min followed by 35 cycles of 98 °C for 10 min, 54 °C for 30 sec and 72 °C for 1 min, with the final extension step at 72 °C for 5 min. The resultant PCR products were analyzed by agarose gel electrophoresis and Illumina sequencing.

For full metagenome, long-range amplicon and metagenome analyses, gDNA was randomly fragmented, tagged, amplified and prepared for sequencing using Nextera XT kit (Illumina, USA).

Portions of the 16S rRNA genes were amplified using PRK341F/PRK806R primers targeting V3-V4 regions[22], modified by addition of Illumina-specific adapters. Each PCR reaction (25 μl) contained 1× HOT FIREPol PCR mix (Solis BioDyne, Estonia); 200 nM uniquely tagged forward and reverse primers; 1 μl of sample DNA and water. The thermal cycling conditions were 95 °C for 15 min and 30 cycles of 95 °C for 30 sec, 50 °C for 1 min and 72 °C for 45 sec. PCR products were then pooled, based on their concentrations measured using Quant-iT™ PicoGreen® dsDNA assay kit (Life Technologies, USA), column-purified using E.Z.N.A.® Cycle Pure kit (Omega Bio-tek, USA) and submitted for sequencing.

Sequencing was performed on MiSeq platform (Illumina, USA) using V3 sequencing chemistry with 300 bp paired-end reads. 16S rRNA gene amplicon samples were processed at Norwegian Sequencing Centre (Oslo, Norway), whereas full metagenome samples were sequenced in-house.

Bacterial culturing. For isolation of *Bifidobacterium* species, 10-fold dilutions of fecal samples in 1% peptone water were anaerobically cultured on Beerens agar at 37 °C. Isolated colonies were then subcultured to purity using the same conditions. DNA was extracted for sequencing of 16S rRNA gene as described above to confirm isolates belonging to *Bifidobacterium* genus.

Three-fold serial dilutions of fecal samples from the cohort were prepared in distilled water, cultured on lactose agar and in tryptic soya broth with 5% horse blood, incubated at 37 °C for 24 h. The broth was supplemented with 0.1% of both Tween 80 and magnesium chloride to recover damaged *Enterobacteriaceae* cells.

Data analyses. 16S rRNA gene amplicon data were analyzed using QIIME pipeline[23]. Sequences were first quality filtered (*split_libraries.py*; sequence length between 200 bp and 1000 bp; minimum average quality score 25; not more than 6 ambiguous bases; and no primer mismatch allowed) and then clustered at 99% homology level using closed-reference *uclust* search against Greengenes database[24] (*pick_closed_reference_otus.py*). Persistence of operational taxonomic units (OTUs) over time in individuals was assessed using multi-way decomposition PARAFAC analysis of mean-centered abundance data[25]. This analysis allows detection of the OTUs that bring most of the variation into the system, simultaneously with detecting the time points at which these OTUs are most pronounced. Simpson's reciprocal

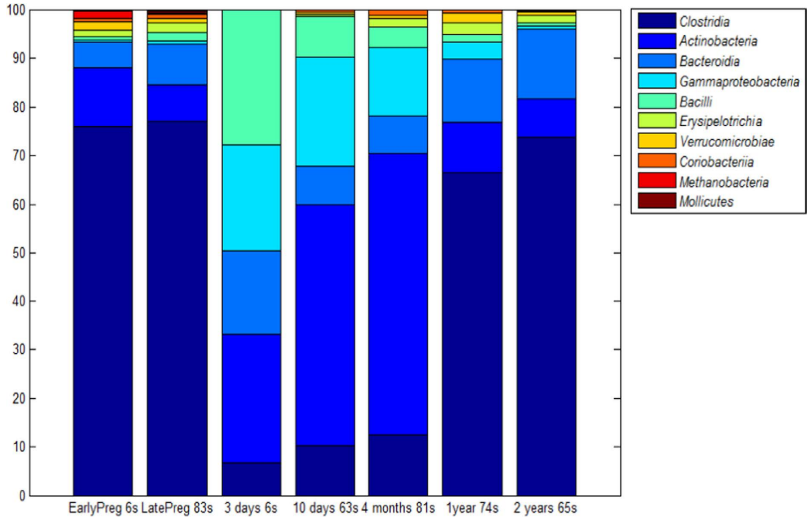


Figure 3. Bacterial class composition of stool samples of infants (from 3 days to 2 years of age) and their mothers during early (1/2 trimester; EarlyPreg) and late (final trimester; LatePreg) pregnancy based on the deep sequencing of 16S rRNA gene amplicons. s, Number of samples at each time period.

diversity index and Bray-Curtis dissimilarity index were used for alpha- and beta-diversity assessment, respectively.

Metagenome data mapping and assembly was performed using Geneious pipeline following authors' recommendations[26]. MG-RAST metagenome analyzer was used to analyze the taxonomy and functional classification of the samples[27]. PATRIC database[28] in MG-RAST was used to check the integron abundance in the samples. E-value $< 10^{-5}$ was used as the cut-off to select integron hits.

Int1 gene persistence was calculated as the ratio of the number of mother-child pairs in who *int1* was detected at both time points to the total number of mother-child pairs for who information for both time points was available. The odds ratio for *int1* gene detection was calculated by the ratio of *int1* persistence to the prevalence of *int1* at a later time point.

Fisher exact test, Pearson correlation coefficient and Spearman correlation coefficient were used for pairwise comparisons of *int1* and 16S rRNA data (including diversity, OTU abundance and bacterial class abundance data). The significance of the change over time was tested with Friedman's test - a non-parametric version of ANOVA test which takes into account repeated measurements. The change in *int1* gene relative abundance was also compared to the change in log-transformed OTU relative abundances over time in an attempt to identify OTUs that correlated to *int1*. Regression and classification decision trees were also built in an attempt to identify bacterial classes that correlated to *int1*. Data analyses were performed using MATLAB® R2014a software (The MathWorks Inc., Natick MA, USA).

Results

Microbiota composition and development. The phylogenetic composition of the microbiota was assessed using deep 16S rRNA gene sequencing. All samples that were amplified with 16S rRNA gene-targeting primers and further amplified with Illumina-adapted primer set were included in the analysis. In total, sequencing data were available for 451 samples. In addition, seven of the samples were analyzed in triplicate to determine technical variation, which was found to be low (Supplementary Fig. S1). The average quality score for the sequence range of 250–299 bp was 25.

On average, 21,277 sequences per sample were generated after quality filtering and assembly. To ensure even amount of sequencing information, 4,000 reads per sample were randomly picked from the full dataset based on the recommendations by Sørensen *et al.*[29]. The final dataset after quality filtering and unification of the sequencing information per sample comprised 378 samples, with a total of 8,288 OTUs belonging to 27 classes. The 10 most abundant classes comprised nearly 100% of the microbiota at all ages (Fig. 3). Stool samples from newborns and 4-month-old infants were significantly lower in alpha-diversity and significantly higher in beta-diversity than stool samples from 2-year-olds and their mothers (Fig. 4). At 1 year of age, both alpha- and beta-diversity estimates were significantly higher than that of 4 month-olds. There was a high dominance of *Clostridia* in stool samples from mothers, as well as from 1- and 2-year-olds. Five bacterial classes were relatively equal in abundance in neonatal stool samples collected soon after birth (3 days), whereas *Actinobacteria* became dominant thereafter (4–10 days) and remained so through at least first 4 months of age. By 1 year of age, the average profile of stool samples from children had started converging towards the adult profile. However, pronounced differences in the abundance of *Actinobacteria* and *Bacteroidia* were seen between adults and 2-year-old children, suggesting climax adult community was still not reached by 2 years of age.

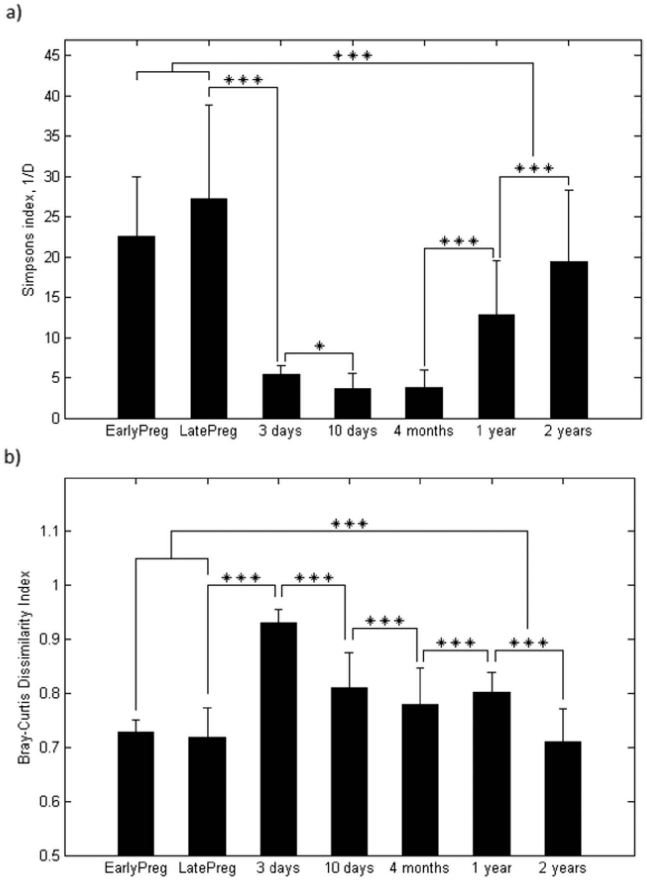


Figure 4. Diversity characteristics of stool samples of infants (from 3 days to 2 years of age) and their mothers during early (1/2 trimester; EarlyPreg) and late (final trimester; LatePreg) pregnancy based on the deep sequencing of 16S rRNA gene amplicons. (**a**) Simpson's reciprocal index of alpha-diversity. (**b**) Bray-Curtis dissimilarity index of beta-diversity. *p value < 0.05; and ***p value < 0.001.

Microbiota persistence and stability. The persistence of 599 most abundant OTUs in the dataset (with an abundance level $\geq$0.5% in at least one sample) were analyzed using PARAFAC. No significant associations of OTUs to age were identified when only considering the detected/non-detected information. When abundance levels were considered, two OTUs belonging to *Bifidobacterium* species (*B. longum* OTU594044 and *B. breve* OTU484303), and one assigned to *Enterobacteriaceae* family (OTU1109087), showed highest stability over time in the cohort (Fig. 5). Spearman correlation test identified the persistence of the *B. longum*-assigned OTU, which had a highest loading in PARAFAC, from 3–10 days to 4 months of age (correlation coefficient = 0.49; p = 0.007). The two other OTUs, however, did not show any significant correlations between the age groups.

Integron distribution and persistence. The distribution of integrons was analyzed by quantitative PCR of the *int1* gene. All samples were included and amplification was controlled by 16S rRNA gene amplification. Out of initial 663 IMPACT samples, 16 failed to amplify PCR products using 16S rRNA gene-targeting primers and thus were excluded from the analysis. In total, 99 of the 647 samples analyzed showed the presence of integrons. The prevalence of the integron-positive samples was highest from 4-month-old children compared to any other age (Fig. 6a). The highest persistence patterns for integrons were seen in children between 3–10 days and 4 months, and 4 months to 1 year (Fig. 6b). Persistence between some mother-child pairs was also detected. The *int1* gene copy numbers of the positive samples, corrected for the estimated genome equivalents[30], were significantly higher in samples from infants (3–10 days and 4 months) compared to both pregnant mothers and 2-year-old children (Fig. 6c).

For the children with persistent *int1* genes, 17% (1 of 6 children with antibiotic usage information) received antibiotics during the first year of life. In addition, 31% (46 out of 147) of the children in the whole cohort had antibiotic usage information documented.

Correlation of *int1* gene to 16S rRNA gene. Detection of *int1* gene did not correlate to alpha-diversity (Simpson's reciprocal index 1/D = 12.3 ± 1.74 [mean ± SEM] and 1/D = 13.7 ± 0.66 for *int1*-positive and *int1*-negative subgroups, respectively) or to beta-diversity (Bray-Curtis Dissimilarity

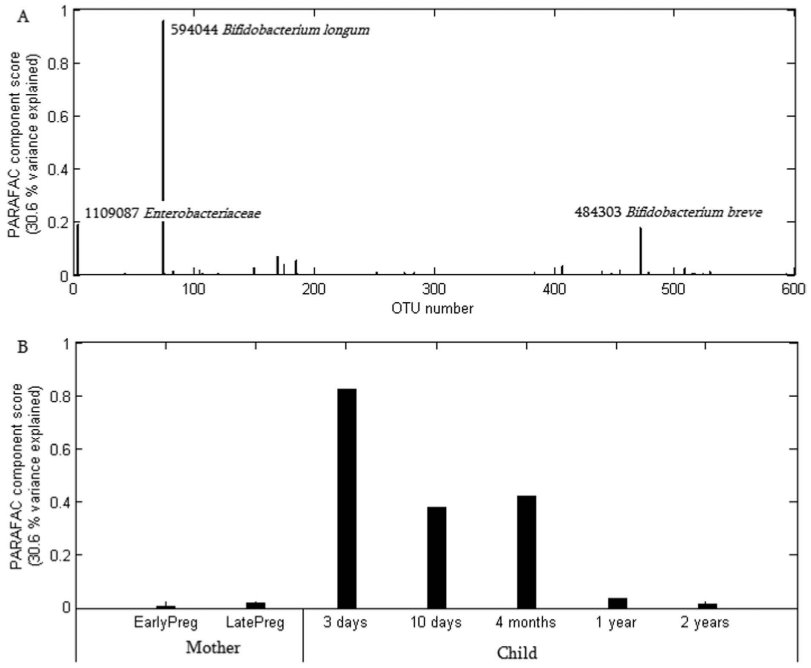


Figure 5. Summary of PARAFAC analysis (multi-way decomposition which defines most influential OTUs over time) on the relative abundances of 599 most abundant bacterial OTUs. (**A**) Most influential OTUs. (**B**) Time points associated with most influential OTUs.

index BC = 0.85 ± 0.03 and BC = 0.86 ± 0.04 for *int1*-positive and *int1*-negative subgroups, respectively). There was also no significant correlation detected between alpha-diversity and *int1* gene relative abundance (correlation coefficient = −0.389, p = 0.45).

With respect to OTU quantity, the most persistent OTU (*B. longum* OTU594044) showed a positive correlation with the *int1* gene (p = 0.03) at 3–10 days. No other significant correlations, however, were found (Supplementary Fig. S2). Additionally, it was investigated whether a change in OTU relative abundance could be associated to the change in *int1* gene relative abundance over time, but there was not an OTU identified that was significantly associated to the *int1* gene. Finally, the analyses concentrated on the OTUs that were detected in all samples for which *int1* were detected (Supplementary Table S2); however, these OTUs did not show any quantitative correlations with *int1* either.

There were additional attempts to find bacterial classes that might correlate to *int1* detection or *int1* gene abundance; however, no significant pairwise correlations between bacterial classes and *int1* gene abundance were detected (Supplementary Fig. 3). Regression and classification decision trees were then built to test for the cumulative effects of bacterial classes, but these analyses also suggested weak correlations between 16S rRNA gene and *int1* gene data (Supplementary Fig. S4 and Supplementary Fig. S5 for regression and classification, respectively).

Search for *int1* gene in bacterial isolates. The detection of the *int1* gene in the genomes of sequenced representatives of persistent/stable OTUs identified by PARAFAC was carried out by BLAST searching whole-genome sequencing data from 16 *B. longum*, 2 *B. breve* and 10 *E. coli* strains, isolated from previously published subset of the IMPACT dataset[31,32]. Despite high numbers of hits to 16S rRNA gene per isolate (451.5 ± 37.8), which was used as a proxy for the genome coverage, the analyses failed to identify reads that showed homology to the *int1* gene sequence identified in our work.

Long-range PCR and amplicon sequencing. Six integron-positive samples were randomly selected for long-range PCR and sequencing. The reads were assembled into contigs and two contigs of lengths 1541 bp and 1019 bp showed BLAST hits to *E. coli* strain DK510 (GQ906578.1) containing dihydrofolate reductase (*dfrA17*) and aminoglycoside adenylyltransferase (*aaDA5*) genes (E-value = 0) with 100% identity (Fig. 7a) and *E. coli* strain A30 (KF921570.1) containing dihydrofolate reductase (*dfrA12*), hypothetical protein (*orfF*) and aminoglycoside adenylyltransferase (*aadA2*) gene cassettes (E-value = 0) with 100% identity (Fig. 7b), respectively.

Integron presence in shotgun metagenome data. Fifteen samples (late pregnancy, n = 1; 3–10 days, n = 6; 4 months, n = 5; and 2 years, n = 3) having microbiota profile information and highest relative abundance of *int1* were selected for shotgun metagenome sequencing. On average, 837,048 reads with a size range from 35 bp to 301 bp were obtained for each sample. By NCBI BLAST searches 699

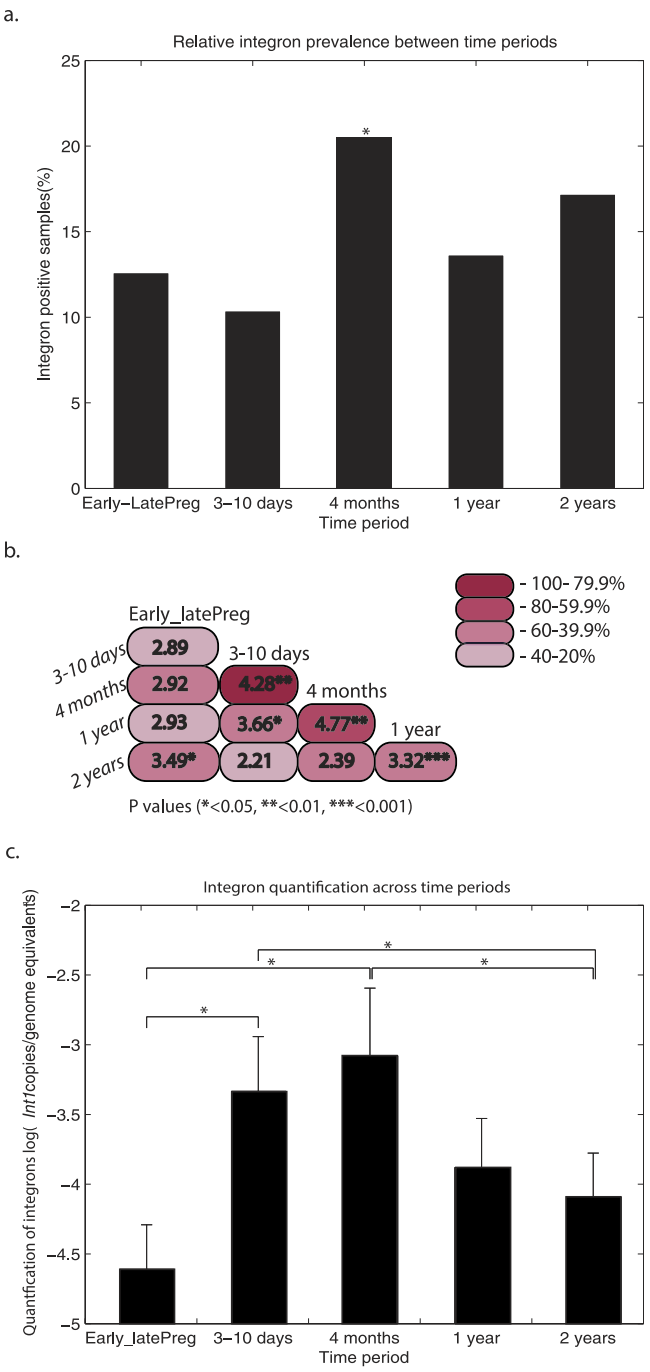


Figure 6. Prevalence, persistence of integron-positive samples and relative quantity of integrons in the positive samples between time points. (**a**) Relative prevalence of integron-positive samples in the dataset. *Binomial testing between the highest abundance (4 months) and the rest (p value = 0.005). (**b**) Persistence of integrons at each time point. The numbers represent the odds-ratio; the color gradient represents the percentage of persistence between time points. Significant p values by Fisher exact test are also indicated (*p value < 0.05; **p value < 0.01; ***p value < 0.001). (**c**) Relative integron quantification at each time point (log (*int*1 copies/genome equivalent[1]) for integron-positive samples. Error bars represent standard error of the mean (SEM). The significant difference between sample groups was calculated by Kruskal-Wallis test; p value < 0.05 is indicated by bracketing. Early_latePreg, samples collected from mothers during early (7–20 weeks) and late (32–40 weeks) pregnancy; 3–10 days, samples from 3- to 10-day-old infants; 4 months, 1 year and 2 years; samples from 4-month-old infants, 1-year-old and 2-year-old children, respectively. [1]16S rRNA copies of all samples from different age groups were normalized to reflect genome equivalents taking into account copy number information given by Vetrovsky *et al.*[30].

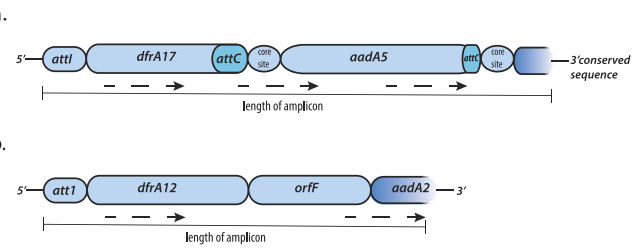


Figure 7. Integrons detected by long-range PCR. (**a**) 1.5 kb partially sequenced integron by long-range PCR product. (**b**) 1.1 kb partially sequenced integron by long-range PCR. Cylindrical boxes show individual genes that are size dependent, i.e. larger box is longer gene; dotted arrows indicate the direction of transcription; and gradient blue color the end of the acquired sequence. Gene and structural features: *attI*, primary recombination site; *dfrA17* and *dfrA12*, dihydrofolate reductase; *attC*, recombination site; *aadA5* and *aadA2*, aminoglycoside adenylyltransferase; and *orfF*, hypothetical protein.

shotgun metagenomic reads from 12 samples were identified that showed high homology to the *int1* gene (E-value $< 10^{-5}$; average identity [range] 97.5% [85.1%; 100%]; average query coverage 99.7% [98.4%; 100.0%]).

Using the MG-RAST metagenome analyzer[27], it was found that all the samples showed the presence of integrons and integron-related genes. The identity of the integron hits of the samples were obtained from PATRIC database (Supplementary Table S3).

Metagenome assembly and identification of complete integrons. The reads were extracted that showed *int1* homology in only one direction of the paired-end reads (n = 71) to investigate the genetic background of their paired mates. By BLAST searching of these sequences against NCBI database, candidate plasmid pSH1148_107 (GenBank JN983049) was identified that was most prevalent among the hits (Supplementary Table S4). The metagenomic reads were then mapped onto the complete plasmid sequence and approximately 60% of the plasmid was encompassed by the metagenomic reads. Seventeen of the 25 conjugation proteins of the plasmid mapped to our reads, including the *Inc1* conjugative transfer proteins, DNA primase and pilus biogene (Supplementary Fig. S6). The reads partially covered the origin of replication. There was one child who showed high prevalence of a plasmid related to pSH1148_107 (more than 1% of all reads) in stool samples from both 3–10 days and 4 months (20× and 34× mean coverage for 3–10 days and 4 months, respectively). The 3–10 days and 4 months reads mapped similarly to the plasmid. The *de novo* assembly of the reads mapped to a transposon containing integron with the *sul*1 gene and *aad*A gene cassette, which was similar to the resistance genes in pSH1148_107, and an additional *dfrA17* gene cassette (Fig. 8). The gene cassettes encode resistance to sulphonamides, spectinomycin and streptomycin, and trimethoprim respectively.

The long-range PCR amplicon contigs were also mapped to the integron assembled from our metagenome. The 1541 bp-long contig showed 97% coverage, suggesting both assemblies came from the same integron. The other contig of 1019 bp length had different gene cassettes and thus showed only partial coverage.

Taxonomic range of the integrons identified by long-range PCR. BLAST searching of the NCBI database with the *int1*-containing contigs identified by long-range PCR revealed high homology (100% pairwise identity with 100% query coverage) towards plasmids isolated from *E. coli*, *Kluyvera georgiana*, *Salmonella enterica* and *Shigella flexneri* (Supplementary Table S5), all belonging to *Enterobacteriaceae* family.

Search for integrons in other metagenomes. To search for the same integron in other publically available metagenomes, data was extracted from 60 metagenome samples from the cohort provided by Yatsunenko *et al.*[33]. The available cohort contained fecal samples from healthy children and adults in Malawi, United States and Venezuela; and 20 metagenomes from each of the respective countries was analyzed. Eleven (18.3%) of the metagenomes showed the presence of *int1* gene. Seven of the *int1*-positive metagenome samples also contained reads mapping to the transposon flanking regions. However, the integron-associated gene cassettes were not similar to those detected in our dataset (Supplementary Table S6).

Discussion

Several studies have shown a high prevalence of AR genes in infants with the absence of antibiotic treatments[3,34,35] which is in line with our findings. However, to our knowledge, this study is the first one to observe high *int1* gene prevalence and persistence. A high prevalence of integrons was found at 3–10 days and 4 months of age. In the early periods of life the resistance against colonization by exogenous bacteria is low[35], therefore opening for the possibility of establishment of bacteria from the environment.

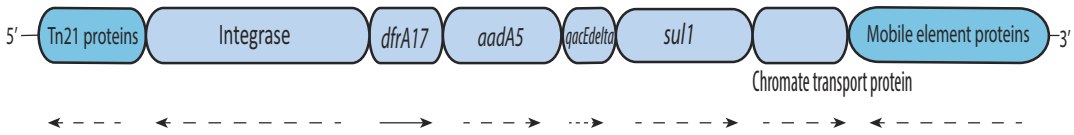


Figure 8. Graphical representation of a transposon-containing integron by *de novo* assembly. The boxes illustrate the coding region of genes, dark blue represents genes of the transposon and the light blue indicates genes of the integron. Genetic features: *dfrA17,* trimethoprim resistance protein; *aad*A5, streptomycin and spectinomycin resistance protein; *qacEdelta*, quaternary ammonium compound resistance protein; *sul1*, sulphonamide resistance protein.

A plausible explanation for the high integron prevalence at early age could be the hospital environment, since children are first exposed to this atmosphere[36]. There was also persistence of *int1* gene throughout the first two years of life and between mothers and their 2-year-old children, pointing towards maternal source as another potential route for transmission. Similar patterns have also been detected in transposon-associated genes in mother-infant pairs[3,31]. An alternative explanation, though, could be the colonization by various integrons at different ages. However, taking into account the increased likelihood of *int*1 detection at one time period given it was detected previously, the more probable explanation would be the persistence of the same integron rather than the detection of independent multiple colonization events.

The persistence of integrons in the gut microbiota indicates the versatility of MGEs to endure the drastic changes that occur during first years of life[10,33]. However, it is unlikely that antibiotic treatment influences the presence of multidrug resistance integrons since we did not find any alteration of persistence patterns in our dataset associated with antibiotic usage.

Diversity estimates of the cohort corresponded well with previously published observations of increase in alpha- and decrease in beta-diversity with age[11,33]. Interestingly, when *int1* gene abundance was highest at early days of life, the microbial diversity was lowest, suggesting that *int1* gene should be associated to those few bacteria that are established by then. However, there was no correlation between *int1* gene and diversity estimates or bacterial classes. Moreover, despite numerous attempts, we could not associate *int1* gene to any particular phylotype across individuals within our cohort. Hence, it is unlikely that the integrons have a strict phylotype association. In addition, when we tried to search for *int1* gene in *Bifidobacterium* isolates that represent the most abundant bacterial group in infancy and that was the only bacterial genus correlating to *int1* gene abundance at early infancy, we failed to find any indication of integron presence in its genomes. Lack of association between integrons and phylotypes across large phylogenetic distances has previously been observed[37]. Statistical inconsistencies have been reported when phylogenetic trees were obtained for *int1* gene and molecular marker for phylogeny such as RNA polymerase subunit B (*rpo*B)[37]. Therefore, given the broad host range for integrons[38], the most plausible explanation for the lack of phylotype association is high rates of HGT. In concordance with potentially high HGT rates, a possible transposon carrying an integron was identified in our samples, suggesting MGEs as the likely vehicle for mobility of integrons. The mobile nature of integrons-associated MGEs has been previously observed in pathogenic bacteria[39,40], environmental samples[41] and in hospital environments[42]. We also observed the persistence of a transposon-containing integron on a potential conjugative plasmid in one infant at two time periods. This integron contained genes associated with aminoglycosides and sulfonamide resistance similar to the conjugative plasmid pSH1148_107, along with and additional trimethoprim resistance gene.

We expanded our search for integrons from different samples in our dataset by involving long-range PCRs that could amplify the whole integron. Two class I integrons were identified with potential association to a mobile element having resistance genes to trimethoprim, streptomycin and spectinomycin. Interestingly, a study by Shahcheraghi *et al.* also found a similar integron containing resistance genes in enteropathogenic *E. coli* strains (JX442969.1) isolated from fecal samples of children less than 5 years of age[43]. These evidences give further support that integrons can be reservoirs for AR genes in infants, with the potential for transmission to pathogens[4,35]. Additionally, we also detected integrons with different gene cassettes in publicly available metagenomes, suggesting the diversity of integrons in global human populations.

Our observation of integron-containing elements regardless of antibiotics intake suggests that they can persist without outer selection pressure. A recent study on the gut microbiota of an isolated group of Yanomani Amerindian tribe showed a similar pattern of the carriage of a pool of mobilizable next-generation antibiotic resistance genes without any prior antibiotic pressure[44]. Moreover, Stern and colleagues found over 10,000 contigs containing potential mobile elements in the MetaHIT dataset[45], which were likely to be quite common constituents of the gut microbiota since all were identified as targets for CRISPR elements. Interestingly, only around 10% of these contigs were of viral nature, leaving the rest to plasmids and MGEs, suggesting that the host actually counter selects these mobile

elements. This finding supports the selfish parasitic-like spread of conjugative plasmids associated integrons in the gut.

The overall results of the study provide evidence for high prevalence of integrons in the fecal microbiota at early stages of life and further suggest that the commensal gut microbiota can serve as a reservoir for multidrug resistance, potentially contributing to its rapid spread.

References

1. McKenna, M. Antibiotic resistance: the last resort. *Nature* **499,** 394–396, doi: 10.1038/499394a (2013).
2. Hu, Y. *et al.* Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature communications* **4,** 2151, doi: 10.1038/ncomms3151 (2013).
3. de Vries, L. E. *et al.* The Gut as Reservoir of Antibiotic Resistance: Microbial Diversity of Tetracycline Resistance in Mother and Infant. *PloS one* **6,** e21644, doi: 10.1371/journal.pone.0021644 (2011).
4. Sommer, M. O. A., Dantas, G. & Church, G. M. Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science* **325,** 1128–1131, doi: 10.1126/science.1176950 (2009).
5. Ravi, A. *et al.* Integrons in the Intestinal Microbiota as Reservoirs for Transmission of Antibiotic Resistance Genes. *Pathogens* **3,** 238–248 (2014).
6. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480,** 241–244, doi: 10.1038/nature10571 (2011).
7. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499,** 219–222, doi: 10.1038/nature12212
8. Sekirov, I., Russell, S. L., Antunes, L. C. M. & Finlay, B. B. Gut Microbiota in Health and Disease. *Physiol. Rev.* **90,** 859–904, doi: 10.1152/physrev.00045.2009 (2010).
9. Jiménez, E. *et al.* Isolation of Commensal Bacteria from Umbilical Cord Blood of Healthy Neonates Born by Cesarean Section. *Curr. Microbiol.* **51,** 270–274, doi: 10.1007/s00284-005-0020-3 (2005).
10. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the Human Infant Intestinal Microbiota. *PLoS Biol.* **5,** e177, doi: 10.1371/journal.pbio.0050177 (2007).
11. Avershina, E. *et al.* Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *Fems Microbiol. Lett.* n/a-n/a, doi: 10.1111/1574-6941.12223 (2013).
12. Avershina, E. & Rudi, K. Is it who you are or what you do that is important in the human gut? *Benef. Microbes*. in press (2013).
13. Faith, J. J. *et al.* The Long-Term Stability of the Human Gut Microbiota. *Science* **341,** doi: 10.1126/science.1237439 (2013).
14. Kovalevskaya, N. P. Mobile Gene Cassettes and Integrons. *Mol. Biol.* **36,** 196–201, doi: 10.1023/A:1015361704475 (2002).
15. Fluit, A. C. & Schmitz, F. J. Resistance integrons and super-integrons. *Clin. Microbiol. Infect.* **10,** 272–288, doi: 10.1111/j.1198-743X.2004.00858.x (2004).
16. Mazel, D. Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.* **4,** 608–620, doi: 10.1038/nrmicro1462 (2006).
17. Gillings, M. R. *et al.* Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J.* doi: 10.1038/ismej.2014.226 (2014).
18. Gillings, M. R., Xuejun, D., Hardwick, S. A., Holley, M. P. & Stokes, H. W. Gene cassettes encoding resistance to quaternary ammonium compounds: a role in the origin of clinical class 1 integrons? *ISME J.* **3,** 209–215, doi: 10.1038/ismej.2008.98 (2009).
19. Rowe-Magnus, D. A. *et al.* The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc. Natl. Acad. Sci. USA* **98,** 652–657, doi: 10.1073/pnas.98.2.652 (2001).
20. Sorum, H., L'Abee-Lund, T. M., Solberg, A. & Wold, A. Integron-containing IncU R plasmids pRAS1 and pAr-32 from the fish pathogen Aeromonas salmonicida. *Antimicrob. Agents Chemother.* **47,** 1285–1290 (2003).
21. Nadkarni, M. A., Martin, F. E., Jacques, N. A. & Hunter, N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* **148,** 257–266 (2002).
22. Yu, Y., Lee, C., Kim, J. & Hwang, S. Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol. Bioeng.* **89,** 670–679, doi: 10.1002/bit.20347 (2005).
23. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7,** 335–336, doi: 10.1038/nmeth.f.303 (2010).
24. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–5072, doi: 10.1128/AEM.03006-05 (2006).
25. Bro, R. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems* **38,** 149–171, doi: 10.1016/s0169-7439(97)00032-4 (1997).
26. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649, doi: 10.1093/bioinformatics/bts199 (2012).
27. Meyer, F. *et al.* The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9,** 386, doi: 10.1186/1471-2105-9-386 (2008).
28. Wattam, A. R. *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42,** D581–591, doi: 10.1093/nar/gkt1099 (2014).
29. Sørensen, L. L., Coddington, J. A. & Scharff, N. Inventorying and Estimating Subcanopy Spider Diversity Using Semiquantitative Sampling Methods in an Afromontane Forest. *Environ. Entomol.* **31,** 319–330, doi: 10.1603/0046-225X-31.2.319 (2002).
30. Vetrovsky, T. & Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PloS one* **8,** e57923, doi: 10.1371/journal.pone.0057923 (2013).
31. Avershina, E. *et al.* Bifidobacterial succession and correlation networks in a large unselected cohort of mothers and their children. *Appl. Environ. Microbiol.* **79,** 497–507, doi: 10.1128/AEM.02359-12 (2013).
32. de Muinck, E. J. *et al.* Diversity, transmission and persistence of Escherichia coli in a cohort of mothers and their infants. *Environ. Microbiol. Rep.* **3,** 352–359, doi: 10.1111/j.1758-2229.2010.00231.x (2011).
33. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–227, doi: 10.1038/nature11053 (2012).
34. Moore, A. M. *et al.* Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. *PloS one* **8,** e78822, doi: 10.1371/journal.pone.0078822 (2013).
35. Sepp, E. *et al.* The occurrence of antimicrobial resistance and class 1 integrons among commensal Escherichia coli isolates from infants and elderly persons. *Ann. Clin. Microbiol. Antimicrob*. **8,** 34 (2009).
36. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74,** 417–433, doi: 10.1128/MMBR.00016-10 (2010).
37. Boucher, Y., Labbate, M., Koenig, J. E. & Stokes, H. W. Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol.* **15,** 301-309, doi: 10.1016/j.tim.2007.05.004 (2007).

38. Nandi, S., Maurer, J. J., Hofacre, C. & Summers, A. O. Gram-positive bacteria are a major reservoir of Class 1 antibiotic resistance integrons in poultry litter. *Proc. Natl. Acad. Sci. USA* **101,** 7118–7122, doi: 10.1073/pnas.0306466101 (2004).
39. Meng, H. *et al.* Characterization and horizontal transfer of class 1 integrons in Salmonella strains isolated from food products of animal origin. *Int. J. Food Microbiol.* **149,** 274–277, doi: 10.1016/j.ijfoodmicro.2011.07.006 (2011).
40. Dalsgaard, A., Forslund, A., Serichantalergs, O. & Sandvang, D. Distribution and content of class 1 integrons in different Vibrio cholerae O-serotype strains isolated in Thailand. *Antimicrob. Agents Chemother* **44,** 1315–1321 (2000).
41. Vaisvila, R., Morgan, R. D., Posfai, J. & Raleigh, E. A. Discovery and distribution of super-integrons among pseudomonads. *Mol. Microbiol.* **42,** 587–601 (2001).
42. Gonzalez, G. *et al.* Presence of integrons in isolates of different biotypes of Acinetobacter baumannii from Chilean hospitals. *FEMS Microbiol. Lett.* **161,** 125–128 (1998).
43. Shahcheraghi, F. *et al.* Identification and characterization of class 1 integrons among atypical enteropathogenic Escherichia coli isolated from children under 5 years of age. *Iran. J. Microbiol.* **6,** 156–162 (2014).
44. Yeagle, P. Microbiome of uncontacted Amerindians. *Science* **348,** 298, doi: 10.1126/science.348.6232.298-a (2015).
45. Stern, A., Mick, E., Tirosh, I., Sagy, O. & Sorek, R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* **22,** 1985–1994, doi: 10.1101/gr.138297.112 (2012).

Acknowledgements

This work was supported by Quota scholarship and funding from Norwegian University of Life Sciences, Ås, Norway and Norwegian University of Science and Technology, Trondheim, Norway.

Author Contributions

A.R.- main author of the article. E.A.-main author of the article. S.F.- Provided plasmids and integron-related experiments. J.L.- Culturing of fecal samples. O.S.- Sample collection for IMPACT study. T.Ø.- Sample collection for IMPACT study. R.J.- Sample collection for IMPACT study. A.L.M.- Isolation of *Bifidobacterium*. T.M.L.- review of scientific content in the article. KR- main lead of the article

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ravi, A. *et al.* The commensal infant gut meta-mobilome as a potential reservoir for persistent multidrug resistance integrons. *Sci. Rep.* **5**, 15317; doi: 10.1038/srep15317 (2015).

PAPER 2

Ravi A, Estensmo FE, L'Abée-Lund MT, Foley LS, Allgaier B, Martin RC, Claud CE and Rudi K (2017). Associations of the preterm infant gut microbiota mobilome with necrotizing enterocolitis, birthweight and hospital. Submitted to *Pediatric Research*

Associations of the preterm infant gut microbiota mobilome with necrotizing enterocolitis, birthweight and hospital

Running title: Preterm infant gut mobilome

Anuradha Ravi[1], Eva Lena F. Estensmo[1], Trine M. L'Abée-Lund[2], Steven L. Foley[3], Bernhard Allgaier[4], Camilia R. Martin[5], Erika C. Claud[6], Knut Rudi[1]*

[1]Norwegian University of Life Sciences, Chemistry, Biotechnology and Food science department (IKBM), Campus Ås, Ås, Norway; [2]Norwegian University of Life Sciences, Department of Food safety and Infection biology, Campus Adamstuen, Oslo, Norway; [3]U.S. Food and Drug Administration, National Center for Toxicology Research, Division of Microbiology, Jefferson, Arkansas; [4]NorthShore University HealthCenter, Department of Pediatrics, Evanston, Illininois; [5]Department of Neonatology and Division of Translational Research, Beth Israel Deaconess Medical Center, Boston, Massachusetts; [6]The University of Chicago, Biological Sciences Division, Departments of Pediatrics and Medicine, Chicago, Illinois.

#Address correspondence to Knut Rudi, Department of Chemistry, Biotechnology and Food Science (IKBM), Norwegian University of Life Sciences. Christian M Falsensvei 1, N-1432 Ås,Norway. Work Phone +47 67232546. Email for correspondence: knut.rudi@nmbu.no.

Financial support: The work was supported by Quota scholarship and funded by Norwegian University of Life Sciences, Ås campus, Norway.

Category of study: Basic science

Word count of abstract: 200

Word count of manuscript: 4000

Disclosure- The views presented in this manuscript do not necessarily reflect those of the U.S. Food and Drug Administration. There are no conflicts of interests.

Abstract

Background: Preterm infants are a vulnerable group at for risk for necrotizing enterocolitis(NEC). Although the preterm microbiota has been extensively studied, the mobilome i.e. mobile genetic elements (MGEs) in the gut microbiota has not been considered. Therefore, the aim of this study was to investigate the association of the mobilome with occurrence of NEC, hospital and birthweight in preterm infants microbiota.

Methods: The dataset consists of fecal samples from 62 preterm infants with and without NEC from three different hospitals. We analysed the gut microbiome by using 16S rRNA amplicon sequencing, shot-gun metagenome sequencing and quantitative PCR. Predictive models and other data analyses was performed using MATLAB and QIIME.

Results: The microbiota composition was significantly different between NEC positive and negative infants and significantly different between hospitals. An OTU showed strong positive and negative correlation to NEC and birthweight respectively, while none showed significance for mode of delivery. Metagenome analyses revealed high levels of conjugative plasmids with MGEs and virulence genes. By quantitative PCR, plasmid signature genes were signifinatly higher in NEC positive infants, in addition to being different between hospitals.

Conclusion: Our results point towards association of the mobilome in preterm infants with respect to both hospital and NEC.

Introduction

Preterm infants are a unique patient population completing development in an extrauterine environment influenced by a concommitantly developing microbiome. Furthermore, this patient group largely acquires its initial microbiota within the hospital environment. This early microbiome is of great importance to preterm infant health and is potentially modifiable by alterations to the hospital environment. Multiple preterm infant outcomes have been linked to the microbiome including risk for sepsis and in particular neonatal necrotizing enterocolitis (NEC) (1,2).

NEC is a devastating disease which most commonly affects very low birthweight premature infants (1). Feeding intolerance, abdominal distension and bloody stools are some of the major symptoms for NEC (3). Approximately 1% to 5% of very low birthweight preterm infants (<1500 g; < 37 weeks gestation) (4) develop NEC with a mortality rate of 25%- 33% (5) .

The primary risk factors for NEC are prematurity and bacterial colonization, however studies have failed to identify a specific pathogen. Recent studies have related microbial communities with a bloom of Gammaproteobacteria such as *Escherichia coli* and *Klebsiella pneumoniae* (1,2) to NEC. Additionally, reports have shown gram-positive (*Enterococcus faecalis*) (6) and anearobic bacteria (*Clostridium spp*) (7) contributing to NEC. Interestingly, NEC has not been observed in germ-free animals (1) indicating the role of the gut microbiota as a major contributing factor (8). There is evidence for changes in the microbiota prior to the onset (2,9), with high level of antibiotic usage being a potential contributing factor (10).

While differences in the bacterial taxa contributing to NEC have been extensively studied, genetic factors such as linkage between antibiotic resistance (AR) and virulence genes have not been investigated with respect to preterm infants or NEC. Therefore, the aim of this work was to investigate the mobilome or a collection of mobile genetic elements of the preterm infant gut microbiota and its potential association with NEC, birthweight and hospital location.

The mobilome of the infant gut microbiota includes transposons, plasmids and bacteriophages (11). Some of the major constituents of the gut mobilome are conjugative plasmids. Conjugative plasmids are self replicating genetic elements that propagate in an infectious manner (12). They can harbor several accessory functional elements that help to maintain longterm stability in a microbial population (13). In addition, two different conjugative plasmids having identical replication machinery are incompatible in the same bacterial cell, hence plasmids are identified by incompatibility groups (13). 27 incompatibility groups defined to to date. However, the most prominent is the incompatibility group F (IncF) plasmids commonly found in *Enterobacteriaceae* (14). These plasmids have been detected in bacteria from several human and animal sources. IncF conjugative plasmids contain an assortment of other MGEs and virulence genes (15). Plasmid mediated antibiotic resistance and virulence to hospital-acquired infections has been previously reported explaining the influence of conjugaal transfer of virulence factors and inducement of bacterial biofilms (16,17). Virulence traits associated with MGEs includes bacterial toxins (18), secretion systems (19) and haemolysins. These properties can transform the characteristics of the host cell.

Integrons are accessory components of conjugative plasmids (20). They are genetic elements that are capable of integration and expression of genetic cassettes by an overall common

promoter (20,21). The integron consists of 3 main parts, an integrase (*int*1) gene that helps in the integration of specific gene cassettes; an attachment (*att*1) site into which the gene cassettes are integrated; and a common promoter (P_c) for expression of the gene cassettes. We investigated the intestinal microbiome of preterm infants from 3 neonatal intensive care units. Amplicon sequencing was used to identify compositional signatures in the microbiota taxonomic composition. Then, full metagenome deep sequencing was used for analysing the phylogeny and genetic background of the MGEs in selected samples. Finally, quantitative PCR was used to study the prevalence and quantity of plasmid signature sequences and thier association with MGEs.

Material and Methods

Workflow:

A workflow of the experimental design and total number of samples used is shown in Figure 1.

Cohort description:

A summary regarding the cohort features and description is given in Table 1

The study consists of premature infants with and without NEC. All the infants with NEC showed ≥ Bell's stage 2 NEC symptoms with mild to moderate systemic illness and pneumatosis intestinalis. The infants were recruited from three different hospitals in USA- Beth Israel Hospital in Boston, MA (n=24); Comer Children's Hospital at The University of Chicago in Chicago, IL (n=29); and NorthShore University HealthSystem Hospital in Evanston, IL (n=9). Sixty-nine percent of the infants were born through caesarean section. These infants resided in the neonatal intensive care unit of the respective hospitals. All infants recruited in the study were born with a birthweight <1500 g and a gestational age < 32weeks. Each NEC positive infant was matched with two NEC negative infants of the same gestational age and day of life. This case control model was implemented in 18 of the 26 NEC positive infants. In total the study consists of 42% (n=23) of NEC positive infants and 58% (n=39) of NEC negative infants. All the infants including NEC positive and NEC negative patients were given antibiotics based on their respective clinical criteria. The faecal samples were collected weekly after spontaneous defecation. In total, the NEC positive infants have 63 samples where 51 samples are from longitudinal datasets. The NEC negative infants have 97 samples, where 73 samples are collected from longitudinal datasets. Informed consent was obtained from the preterm

infants' parents for faecal samples collection and storage. The samples were collected directly from the diaper and into the collection tube using the wooden end of a sterile cotton swab. The samples were immediately frozen at -80°C until processed. The samples were sent to Genetic Analysis, Ås, Norway for long time storage and DNA extraction.

DNA extraction:

DNA was isolated from 160 samples using an automated protocol of MagNA Pure Compact System (Roche Applied Science, Basel, Switzerland). DNA from a subset of the samples in the dataset was also manually extracted by QIAamp DNA Stool mini kit (Qiagen, Venlo, Netherlands). These were termed as duplicates. Fifty mg of the frozen fecal sample was dissolved in 1 ml extraction buffer [50 mM Tris (pH 7.4), 100 mM EDTA (pH 8.0), 400 mM NaCl, 0.5% SDS] containing 20 µL proteinase K (20 mg/ml) and 500 µL of 0.1-mm-diameter zirconia/silica beads (BioSpec Products, Bartlesville, OK, USA) were added into the extraction tubes and a Mini-Beadbeater-16 (BioSpec Products) was used to lyse the microbial cells. The lysed cells were centrifuged and 50 µL of the supernatant was taken for DNA isolation. For the MagNA Pure Compact System, the supernatant was mixed with paramagnetic beads and was eluted using a 96 super Magnet plate (Alpaqua,Beverly, MA, USA). For the QIAamp DNA stool mini kit, purified DNA was extracted using QIAamp mini Spin columns according to the manufacturer's protocol.

DNA concentration and quality were determined by fluorometry (using a Qubit system (Invitrogen) and stored at -40°C until further use.

Polymerase chain reaction and gene quantification

The primers used in the study are shown in Table 2. Each 25µl PCR reaction contained 1X HOT FIREPol PCR mix (Solis BioDyne, Tartu, Estonia); 200nM forward and reverse primers; 1µl of sample DNA and sterile deionized water. The reaction mix was amplified using LightCycler 480 (Roche) and resultant flourescence data was uploaded into the LinRegPCR program (22) to perform baseline correction and calculate mean PCR efficiency. High resolution melting (HRM) curve analysis and DNA sequencing using BigDye Terminator v1.1 chemistry (Thermo Fisher Scientific, Waltham, MA, USA) was used to verify the identity of the PCR products. The thermal cycling conditions for the 16S rRNA primer pair targetting the conserved regions of the 16SrRNA gene were 95°C initial denaturation for 15 mins followed by 40 cycles of 95°C for 30 sec and 60°C for 30 sec (11). Primers flanking the *int*1 gene of the integron (23), *rep*A gene of the conjugative plasmid and *yigB* gene of the haemolysin expression modulating protein (hha) gene family were used with thermal cycling conditions of 95°C for 15 min and 40 cycles of 95°C for 30 sec, specified annealing temperatures for the genes (Table 2) and 72°C for 30 sec.

Microbial community analysis

Microbial community structure of the samples was assessed using Illumina amplicon sequencing of 16S rRNA gene. The 16S rRNA genes were amplified using PRK341F/PRK806R primers that target the V3-V4 hypervariable regions and were modified to contain illumina specific adapters. Each PCR reaction contained HOT FIREPol PCR mix (Solis Biodyne); 200 nM illumina-adapter attached forward and reverse primer; 1µl of sample DNA and water. The thermal cycling conditions were 95°C for 15 min and 30 cycles of 95°C for 30 sec, 50°C for 1 min and 72°C for 45 sec. The PCR amplicons were pooled and concentration was measured using the PerfeCta NGS quantification kit (Quanta Biosciences, Beverly, MA, USA) and purified

using Agencourt AMPure XP-PCR Purification kit (Beckman Coulter, Brea, CA, USA). The purified products were sequenced with the Miseq platform (Illumina, San Diego, CA, USA) using V3 chemistry with 300bp paired-end reads.

Sequences from the 16S rRNA amplicon data were analysed using the QIIME pipeline (24). Sequences were quality-filtered (*split_libraries.py*; sequence length 200-600bp; minimum average quality score 25; no more than 6 ambiguous bases, but with no primer mismatches) and then clustered at 97% homology level using Usearch version 8 against the Greengenes database (25).

Shotgun metagenome sequencing and analysis

The metagenome was fragmented, tagged and quantified according to the Nextera XT Sample preparation guide (Illumina). Concentration of the pooled library was normalised using the PerfeCta NGS quantification kit (Quanta Biosciences). Sequencing was done in-house on a MiSeq platform using V3 chemistry and 300bp paired end reads.

Metagenome data mapping and assembly was performed on Geneious (26) following the recommended criteria. De novo assembling of the reads was performed by Geneious Read Mapper (Geneious, Biomatters, New Zealand). MG-RAST metagenome analyzer (27) (Argonne National Laboratory, Lemont, IL, USA) was used to analyze the functional classification in the samples using the SEED (subsystem) database that houses collections of functionally related protein families (28). The ResFinder program (DTU, Copenhagen, Denmark), an online tool was used to find antimicrobial resistance genes in the sequences based on the NCBI database (29). The RAST (Rapid Annotation using Subsystem Technology) server using SEED-based annotation

was used to identify genes within the contigs built by Geneious (30). Reference genomes for assembly and annotation were downloaded from the NCBI database.

Validation and statistical analyses

Technical variation was detetermined by Pearson regression analyses between the technical duplicates. To account for the uneven sampling and presence of duplicates across the individuals, we used the average microbiota and average quantification of genes across all sampling points for each individual in the comparative statistical analyses.

Fisher Exact test, Pearson correlation and binomial testing were used for pairwise comparisons of relative abundances of *rep*A, *int*1 and *yig*B genes within the 16S rRNA amplicon analyses and between the relative abundances of the individual genes across different hospitals.Correction for multiple testing was done using Benjamini and Hochberg false discovery rate (BHFDR) test.Predictive models using OTUs in the study were made using Partial Least Squares (PLS) discriminant analysis (DA), (Eigenvector Research, Manson, WA,, USA). The models were calibrated using a subset of the dataset and cross-validated using Venetian Blinds procedure where the data is split into subsets and each subset is validated to fit the model. Cross-validated models with an accuracy of classification >0.5 indicate significance. Predictive models were made for predicting hospital location, detection of NEC and association of NEC with plasmid signature genes. Correlations with birth weight were identified using PLS regression. Variables important in the models were identified by the VIP score, with scores >1 indicating importance to the model. All data analyses was performed using MATLAB R2014a software (The MathWorks, Natick, MA,USA).

Results

Microbiota composition

On average, 44,194 sequences per sample were generated by Illumina V3-V4 16S rRNA amplicon sequencing after quality filtering and chimera removal. To ensure even amounts of sequence information and to gather information on the most abundant operational taxanomic units (OTUs) from all the samples, 6000 sequences/sample were randomly picked from the whole dataset. The final dataset after quality filtering and integration of the sample information contained 192 samples, of those 58 were technical duplicates. The technical duplicates showed a mean squared regression coefficient of 0.75 and a standard deviation of 0.33 for pairwise OTU level comparisons, while comparison of different samples gave squared regression coefficients <0.3. In total, the sequences in the dataset belonged to 299 OTUs of 13 bacterial classes. Overall, the gut microbiota composition was mainly composed of *Proteobacteria* with lower levels of *Firmicutes*.

Microbiota associations to metadata

We found no major differences in the α- diversity between the NEC positive and negative infants (Supplementary figure 1a) but when calculated between the different hospitals, infants from Evanston displayed higher diversity than those from Boston and Chicago (p= 0.003, Boston-Evanston; p=0.003, Chicago and Evanston, Kruskal Wallis test) (Supplementary figure 1b). The β-diversity estimates from principle coordinates (PC) 1, on the other hand showed significant differences in NEC positive (median= 0.16) and negative samples (median= -0.01)

(p=0.00001, Kruskal Wallis test), but no differences among hospitals (median= 0.12 [Chicago]; median= 0.07 [Boston]; median= 0.16 [Evanston]) (p= 0.35, Kruskal Wallis test).

The proportion of *Enterobacteriaceae* was significantly more abundant, on average in NEC positive (59%) to NEC negative infants (44%, p=0.001, Kruskal Wallis test). An OTU classified as *Enterobacteriaceae* (referred as OTU2) revealed the strongest association to NEC with a VIP score of 40 in a PLS-DA predictive model (classification accuracy of 0.80 for the calibrated- and 0.65 for the cross-validated model). OTU2 also showed a direct significant correlation to NEC (p= 0.04, Kruskal Wallis test) (OTU2 abundance, median =25 [NEC]; median =5 [No NEC]). There were, however, no OTUs that were signifcantly related to mode of delivery (BHFDR corrected Kruskal Wallis test).

In regard to the association of microbiota composition to hospital location, predictive models using PLS- DA showed an accuracy of classification of location based on the microbiota (calibrated/cross-validated) for Boston 0.78/0.63, Chicago 0.67/0.56 and Evanston 0.74/0.64, indicating predictive information in the microbiota for all locations. Specifically, an OTU classified as *Enterobacteriaceae* (referred to as OTU9) showed pronounced association with the hospital location, having a median of 5.0% for Boston, 0.7% for Chicago and 0.3% for Evanston ($p<0.0005$. Kruskal Wallis test). OTU2 also showed significant associations with Evanston (median= 11.4%) as opposed to 0.1% and 0.2% in Chicago and Boston respectively (p=0.05, Kruskal Wallis test). A predictive model for the association of microbiota composition and birth weight by PLS-DA showed an accuracy of classification as 0.76/0.59 (calibrated and cross-validated) in the median binarized dataset. OTU2 and birth weight when directly

267 correlated showed the strongest negative correlation (Spearman rho=0.45;p=0.005) whereas
268 OTU9 showed strongest positive correlation (Spearman rho=0.45;p=0.004).

269 **Shotgun metagenome analyses**

270 Since OTU2 was positively associated with the detection of NEC and negatively associated with
271 birth weight, we selected longitudinal samples from 3 patients having high abundance of
272 OTU2: patient 17 from Chicago and patient 49 from Evanston positive for NEC and patient 89
273 from Boston negative for NEC. In addition to this, longitudinal samples of patient 86 from
274 Boston and patient 22 from Chicago having low abundance of OTU2 and positive for NEC were
275 also selected (Supplementary table 1). On average, 691,759 sequences were generated per
276 sample with a size range of 35 bp to 301 bp. The unassembled reads were uploaded into MG
277 RAST metagenome analyser. Functional abundance of genes related to conjugative plasmids,
278 MGE and virulence were analysed in the metagenomes by SEED Subsystem Annotation
279 database (mininum identity 90%; minimum alignment length 50 bp). However, there were no
280 clear differences in the gene distribution between NEC positive and negative samples (Figure
281 2).

282 Given that all the infants received antibiotics at least one time point, we looked into the
283 presence of AR genes in the shotgun metagenomes. The unassembled raw reads were
284 uploaded to ResFinder to locate AR genes in the samples. Genes associated with resistance to
285 β-lactams, macrolides and aminoglycosides were found in almost all samples (threshold
286 pairwise identity 99%) (Table 3). Longitudinal carriage of particular resistance genes was
287 observed in all the infants, however no clear association was identified for AR genes and NEC.

Metagenome assembly

The reads were trimmed (error probability 0.05) and paired using Geneious. The paired reads were then built into contigs by Geneious Read Mapper. On average, 1,800 contigs greater than 1,000 bp in length with at least 96 contigs greater than the N50 length were assembled per sample by the assembler. The contigs from all the samples were evaluated for the presence of an OTU2 representative sequence (Supplementary table 2). The contigs with OTU2 representative sequence of each sample showed highest identity to HG428755, an enteropathogenic *E coli* (EPEC) (E value=0;identity >96%; query coverage >80%) that was used as a model to study host-pathogen interactions. In order to understand the coverage of this genome by our metagenomic reads, the samples were mapped directly towards this genome and its corresponding plasmids (CBTO010000001 and CBTO010000002) (Supplementary table 3).

To identify potential complete conjugative plasmids assembled from our dataset, the denovo assembled contigs from each sample having plasmid related genes were annotated using the RAST annotation server. All the identified contigs were 97% identical with 98% pairwise identity with each other. A representative contig of 61058 bp in length annotated by RAST was found to belong to a conjugative plasmid homologue of IncF group of plasmids (Figure 3). This annotated plasmid contained genes for transfer (*tra*A-*tra*X); replication (*rep*A); and resistance genes for trimethoprim, streptomycin and sulfonamides carried in an integron. In addition there were genes for haemolysin expression modulating (hha) family (*yih*A, *yig*B and *fin*O) that regulate production of α-haemolysin toxin and several invasin genes (31). NCBI-

BLAST analysis of this contig revealed similar IncF conjugative plasmids in *E. coli* (E value 0; identity 100%; average query coverage 58% range [35%-78%]).

To determine the presence of other conjugative plasmids in our dataset, the metagenomic reads from all the samples were mapped towards the de novo assembled conjugative plasmid. Seven of the 15 samples covered >80% of the assembled conjugative plasmid with 98% pairwise identity (sampling day 4, 9^1 of Patient 89; sampling day 12 of Patient 86; sampling day 46 of Patient 17; sampling day 11 of Patient 22 and sampling day 46 of Patient 49). The seven samples with >80% coverage also covered the integron with the gene cassettes and the replication and transfer genes of the IncF plasmid family (coverage >80%; pairwise identity >97%). Eight samples including the 7 samples and day 11 sample of Patient 86 covered >90% of plasmid sequences mapped to hha gene family (coverage >80%; pairwise identity >97%) (Supplementary table 4).

Quantification of signature sequences of conjugative plasmids

Distinct regions of the de novo assembled conjugative plasmid were selected as signature sequences. Replication machinery (replication regulatory gene-*rep*A), virulence (hha gene family- *yig*B) and carrier of multidrug resistance genes (Class I integron integrase gene- *int*1) were targeted and screened in our dataset using quantitative PCR. In total, 23% of the samples from the dataset contained at least one of these genes. Interestingly, the relative gene abundance of *rep*A strongly correlated with *yig*B ($p<0.0001$, Pearson correlation; $r^2=0.8$) indicating the replication genes and virulence genes are likely in the same genetic element

(Figure 4). No significant correlations were found between *int*1 with *rep*A or *int*1 with *yig*B (r^2<0.5)

With respect to the association of OTUs and signature genes, the genes showed a significant microbiota association with an accuracy of classification (calibrated/validated) of 0.80/0.67 for *int*1 gene, 0.85/0.74 for *rep*A gene and 0.8/0.66 for *yig*B gene using PLS-DA. OTU2 showed significant association with *rep*A, hha genes and *int*1 showing a median of 5.6% for *int*1 positive samples and 0.0006% for *int*1 negative samples (p=0.015. Kruskal Wallis test) and 8.9% for *rep*A positive sample and 0.0006% for *rep*A negative samples (p<0.0005. Kruskal Wallist test). However, there was no significant association between OTU2 and *yig*B (p=0.13. Kruskal Wallis test), with a median of 5.3% for *yig*B positive samples and 0.0006% for negative samples. Samples from Evanston showed higher prevalence of the signature genes compared to the other hospitals (Figure 5). There were no direct significant correlations between the signature genes with NEC, nor mode of delivery.

Longitudinal associations of OTU2 and signature sequences

Samples were plotted on a longitudinal time scale from time of birth to end of sampling in order to detect temporal acquisition of plasmid related signature genes and co-occurence of OTU2 (Table 4). The diagnosis of NEC was significantly associated with high levels of OTU2 (>25%) (p=0.01, Fisher Exact test)

PLS-DA revealed that NEC is associated with signature sequences with an accuracy of classification of 0.79/0.56 (calibrated/validated). r*ep*A and *int*1 showed the highest VIP score

(>1) associated with NEC. All NEC positive infants showed an increase in the levels of *rep*A and *yig*B at the time of NEC diagnosis.

Discussion

While many studies have attempted to characterize the microbiome of preterm infants, this work, to our knowledge is the first to investigate the mobilome as possible means of genetic transfer as microbial functional cassettes. Typically an infectious disease is associated with a particular pathogen, but the virulence potential within a bacterial species may vary and be attributed to specific genetic elements encoded by specific strains (9,32).

We identified an association of NEC with a conjugative plasmid containing virulence genes and multiple drug resistance genes. These findings may potentially help to explain why specific pathogens attributed to NEC have not yet been identified (5,33). It is possible that a combination of specific bacterial phylotypes along with such conjugative plasmids could promote pathogenicity. In support of this hypothesis, it has also been recognized that the fecal resistome could serve as a pool of genes to facilitate genetic transfers due to their immense ability to disseminate among pathogenic bacteria (34).

The de novo assembled conjugative plasmid contained the genes necessary for conjugal transfer, virulence genes and AR genes. We believe, the virulence factors are within the conjugative plasmid, as we found a significant correlation of *rep*A, a replication regulatory gene and *yig*B, a gene from the hha family. The hha family of genes plays a role in regulating the expression of virulence genes and the α-haemolysin gene family in response to virulence factor expression (31,35). The α-haemolysin toxin has been previously shown to have a role in

development of enterocolitis in humans and animals (36). In addition, a correlation between the hha gene family and other conjugative plasmids has been previously reported in other studies (37).

An integron that contained trimethoprim and streptomycin resistance gene cassettes was also assembled within the conjugative plasmid. The integron is a genetic element most commonly found within transposons that carry multiple resistance genes (20). We detected a high prevalence of different AR genes and found AR genes linked to virulence genes in the assembled conjugative plasmid. The use of antibiotics can drive the selection pressure to antibiotic resistant bacteria in the gut. Increased use of antibiotics in preterm very low birthweight infants is shown to be associated with increased risk of NEC (38).

Interestingly, there were clear differences in the distribution of the plasmids and OTUs among the three hospitals investigated. The hospitals in Evanston and Chicago, which are in the same metropolitan area had significant differences in the microbial populations and plasmid content. Previous studies have shown a high prevalence of multidrug resistance genes at early age in full term and preterm infants (39) indicating that the hospital environment is an important reservoir for both bacteria and plasmids (40).

In summary, even though this dataset has limitations of small size and irregular sampling times the study data suggest that the preterm infant gut microbiota indeed can contain a mobilome with antibiotic resistance and virulence genes that may be transmitted within individual nurseries and between different host microbes. As preterm infants spend many months in the hospital environment, understanding the transmission of mobile genetic

391 elements in addition to the transmission of microbes will be critical for optimizing the health
392 of these vulnerable infants.

393 **References**

394 1. Mai V, Young CM, Ukhanova M, et al. Fecal Microbiota in Premature Infants Prior to
395 Necrotizing Enterocolitis. PloS one 2011;6:e20647.

396 2. Claud EC, Keegan KP, Brulc JM, et al. Bacterial community structure and functional
397 contributions to emergence of health or necrotizing enterocolitis in preterm infants.
398 Microbiome 2013;1:1-11.

399 3. Neu J, Walker WA. Necrotizing Enterocolitis. N Engl J Med 2011;364:255-64.

400 4. Lin PW, Stoll BJ. Necrotising enterocolitis. The Lancet;368:1271-83.

401 5. Henry MCW, Moss RL. Current issues in the management of necrotizing enterocolitis.
402 Semin Perinatol 2004;28:221-33.

403 6. Hällström M, Eerola E, Vuento R, Janas M, Tammela O. Effects of mode of delivery and
404 necrotising enterocolitis on the intestinal microflora in preterm infants. Eur J Clin Microbiol
405 Infect Dis 2004;23:463-70.

406 7. de la Cochetiere M-F, Piloquet H, des Robert C, Darmaun D, Galmiche J-P, Roze J-C.
407 Early Intestinal Bacterial Colonization and Necrotizing Enterocolitis in Premature Infants: The
408 Putative Role of Clostridium. Pediatr Res 2004;56:366-70.

409 8. Lee JS, Polin RA. Treatment and prevention of necrotizing enterocolitis. Semin
410 neonatol:SN 2003;8:449-59.

411 9. Zhou Y, Shan G, Sodergren E, Weinstock G, Walker WA, Gregory KE. Longitudinal
412 Analysis of the Premature Infant Intestinal Microbiome Prior to Necrotizing Enterocolitis: A
413 Case-Control Study. PloS one 2015;10:e0118632.

414 10. Wang Y, Hoenig JD, Malin KJ, et al. 16S rRNA gene-based analysis of fecal microbiota
415 from preterm infants with and without necrotizing enterocolitis. Isme J 2009;3:944-54.

416 11. Ravi A, Avershina E, Foley SL, et al. The commensal infant gut meta-mobilome as a
417 potential reservoir for persistent multidrug resistance integrons. Sci Rep 2015;5:15317.

418 12. Norman A, Hansen LH, Sørensen SJ. Conjugative plasmids: vessels of the communal
419 gene pool. Philos Trans R Soc London B: Biol Sci 2009;364:2275-89.

420 13. Bahl MI, Hansen LH, Sørensen SJ. Persistence Mechanisms of Conjugative Plasmids. In:
421 Gogarten MB, Gogarten JP, Olendzenski LC, eds. Horizontal Gene Transfer: Genomes in Flux.
422 Totowa, NJ: Humana Press; 2009:73-102.

423 14. Carattoli A. Plasmids in Gram negatives: molecular typing of resistance plasmids.
424 International journal of medical microbiology : IJMM 2011;301:654-8.

425 15. Ogilvie LA, Firouzmand S, Jones BV. Evolutionary, ecological and biotechnological
426 perspectives on plasmids resident in the human gut mobile metagenome. Bioeng bugs
427 2012;3:13-31.

428 16. Ghigo J-M. Natural conjugative plasmids induce bacterial biofilm development. Nature
429 2001;412:442-5.

17. Ramirez MS, Traglia GM, Lin DL, Tran T, Tolmasky ME. Plasmid-Mediated Antibiotic Resistance and Virulence in Gram-negatives: the Klebsiella pneumoniae Paradigm. Microbiol spectr 2014;2:1-15.

18. Foster TJ. Plasmid-determined resistance to antimicrobial drugs and toxic metal ions in bacteria. Microbiol Rev 1983;47:361-409.

19. Wallden K, Rivera-Calzada A, Waksman G. Type IV secretion systems: versatility and diversity in function. Cell Microbiol 2010;12:1203-12.

20. Ravi A, Avershina E, Ludvigsen J, #039, Abée-Lund T, Rudi K. Integrons in the Intestinal Microbiota as Reservoirs for Transmission of Antibiotic Resistance Genes. Pathogens 2014;3:238-48.

21. Mazel D. Integrons: agents of bacterial evolution. Nat Rev Microbiol 2006;4:608-20.

22. Ruijter JM, Pfaffl MW, Zhao S, et al. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications. Methods 2013;59.

23. L'Abee-Lund TM, Sorum H. Class 1 integrons mediate antibiotic resistance in the fish pathogen Aeromonas salmonicida worldwide. Microb Drug Resist 2001;7:263-72.

24. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335-6.

25. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006;72:5069-72.

449 26. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable
450 desktop software platform for the organization and analysis of sequence data. Bioinformatics
451 2012;28:1647-9.

452 27. Meyer F, Paarmann D, D'Souza M, et al. The metagenomics RAST server - a public
453 resource for the automatic phylogenetic and functional analysis of metagenomes. BMC
454 Bioinformatics 2008;9:386.

455 28. Overbeek R, Begley T, Butler RM, et al. The Subsystems Approach to Genome
456 Annotation and its Use in the Project to Annotate 1000 Genomes. Nucleic Acids Res
457 2005;33:5691-702.

458 29. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial
459 resistance genes. J Antimicrob Chemother 2012;67:2640-4.

460 30. Aziz RK, Bartels D, Best AA, et al. The RAST Server: Rapid Annotations using Subsystems
461 Technology. BMC Genomics 2008;9:1-15.

462 31. Nieto JM, Carmona M, Bolland S, Jubete Y, de la Cruz F, Juarez A. The hha gene
463 modulates haemolysin expression in Escherichia coli. Mol Microbiol 1991;5:1285-93.

464 32. Grishin A, Papillon S, Bell B, Wang J, Ford HR. The role of the intestinal microbiota in
465 the pathogenesis of necrotizing enterocolitis. Semin Pediatr Surg 2013;22:69-75.

466 33. Gibbs K, Lin J, Holzman I. Necrotising enterocolitis: The state of the science. Indian J
467 Pediatr 2007;74:67-72.

34. Moore AM, Patel S, Forsberg KJ, et al. Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. PloS one 2013;8:e78822.

35. Madrid C, Nieto JM, Juárez A. Role of the Hha/YmoA family of proteins in the thermoregulation of the expression of virulence factors. Int J Med Microbiol 2001;291:425-32.

36. Dittmar E, Beyer P, Fischer D, et al. Necrotizing enterocolitis of the neonate with Clostridium perfringens: diagnosis, clinical course, and role of alpha toxin. Eur J Pediatr 2008;167:891-5.

37. Paytubi S, Dietrich M, Queiroz MH, Juárez A. Role of plasmid- and chromosomally encoded Hha proteins in modulation of gene expression in E. coli O157:H7. Plasmid 2013;70:52-60.

38. Cotten CM, Taylor S, Stoll B, Goldberg RN, Hansen NI. 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.

39. Schwiertz A, Gruhl B, Lobnitz M, Michel P, Radke M, Blaut M. Development of the Intestinal Bacterial Composition in Hospitalized Preterm Infants in Comparison with Breast-Fed, Full-Term Infants. Pediatr Res 2003;54:393-9.

40. In: Knobler SL, Lemon SM, Najafi M, Burroughs T, eds. The Resistance Phenomenon in Microbes and Infectious Disease Vectors: Implications for Human Health and Strategies for Containment: Workshop Summary. Washington (DC)2003.

Figure 1

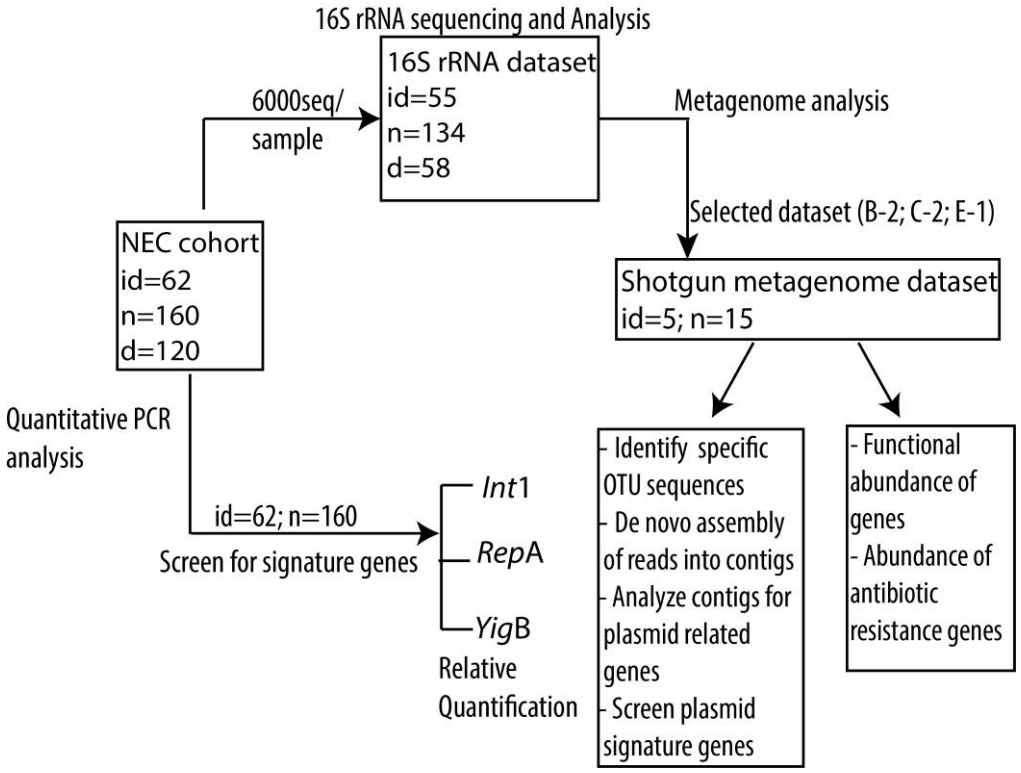
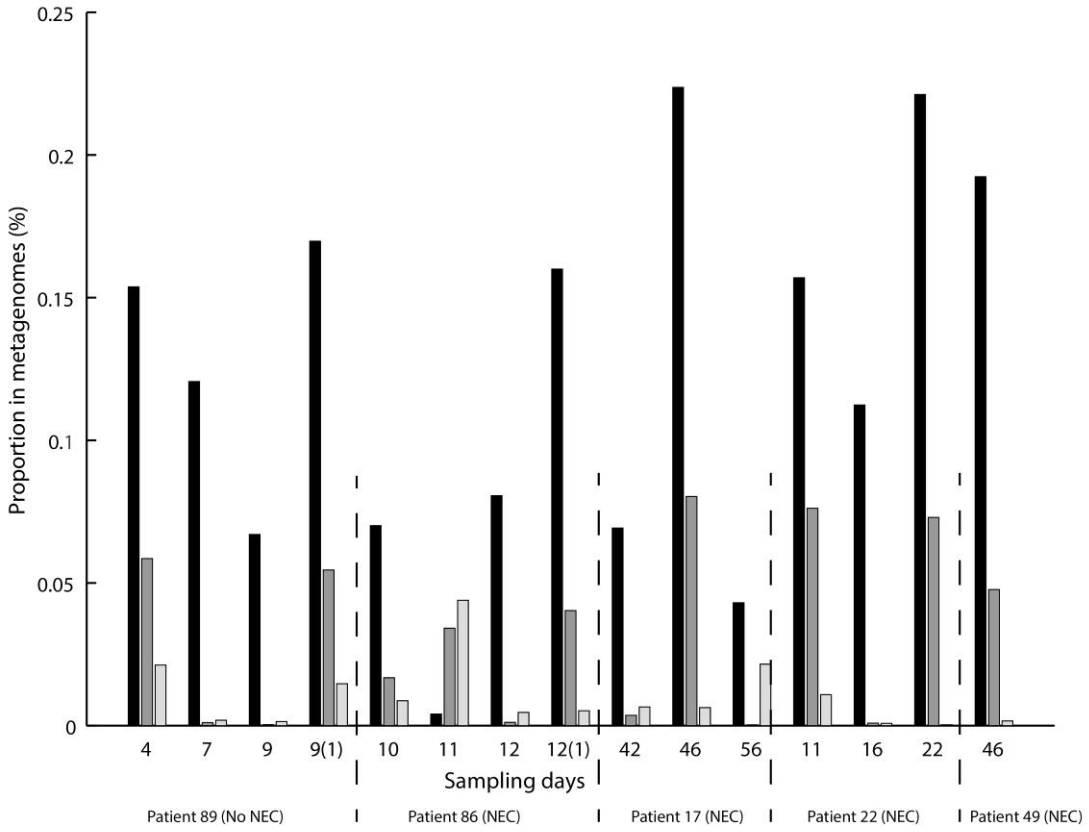


Figure 1: Workflow of the experimental setup. n=number of samples included; d=duplicate samples; id= number of patients; B=Boston;C=Chicago; E=Evanston

501 **Figure 2**



502

503 **Figure 2**: Abundance of functional genes for conjugative plasmid, MGE and virulence genes.

504 Maximum e-value 1e-5; minimum identity 90%; minimum alignment length 50bp was

505 regarded as hit. Black= hits to conjugative plasmid; dark grey= hits to mobile genetic elements;

506 light grey= hits to virulence and invasin genes; (1)= extra sample with same sampling day.

507

508

509

510

Figure 3

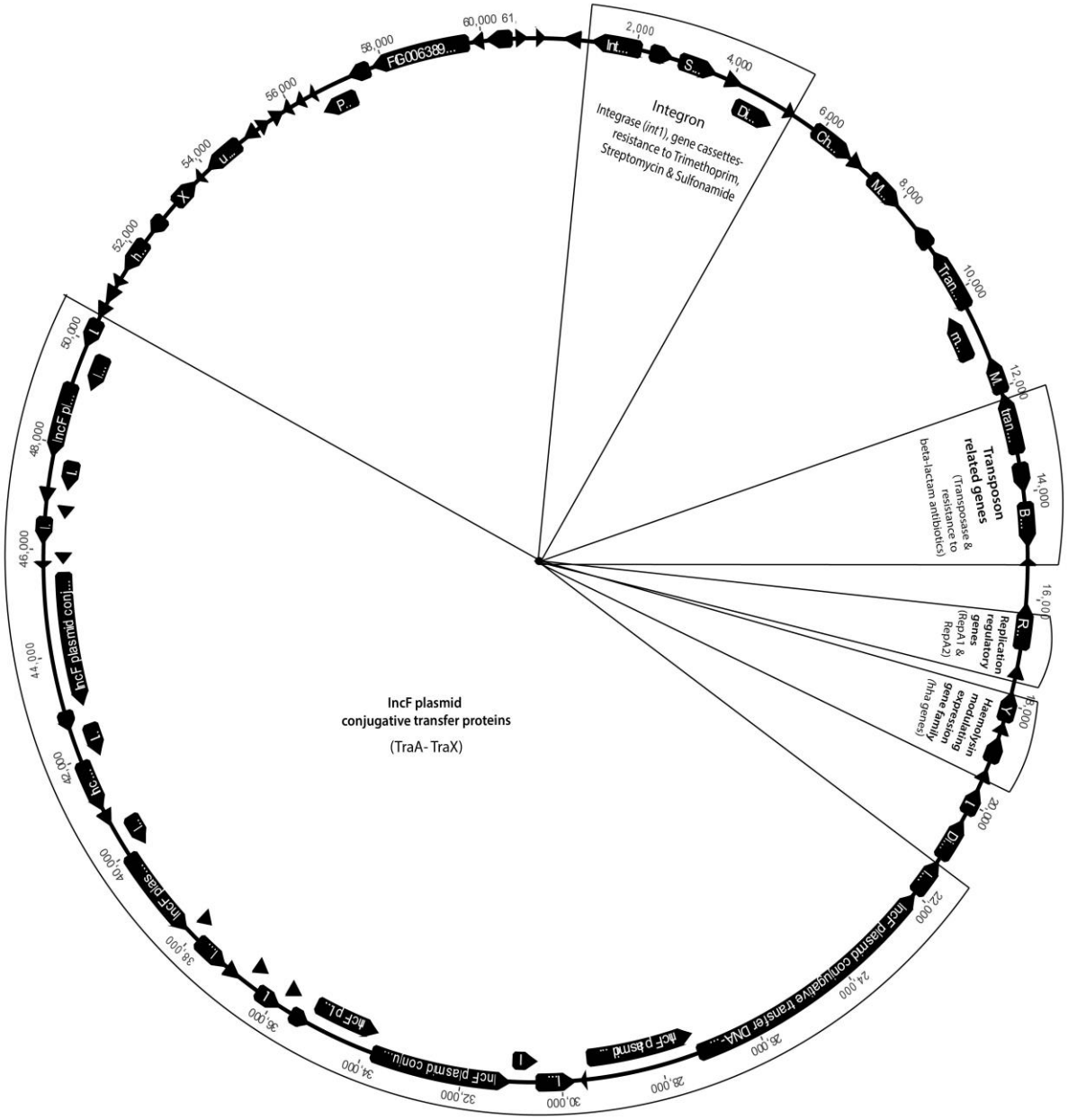


Figure 3: de novo assembled conjugative plasmid. A conjugative plasmid of 61058bp was assembled by de novo assembling of metagenomic reads and annotated by RAST using SEED subsystem database.

Figure 4

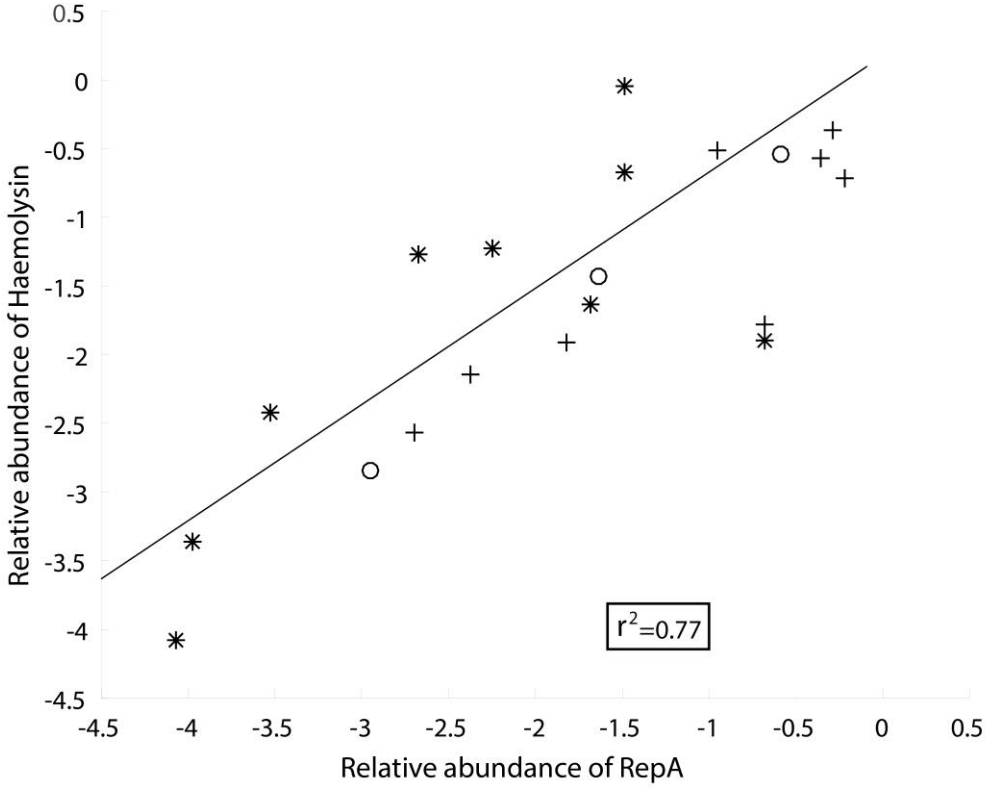
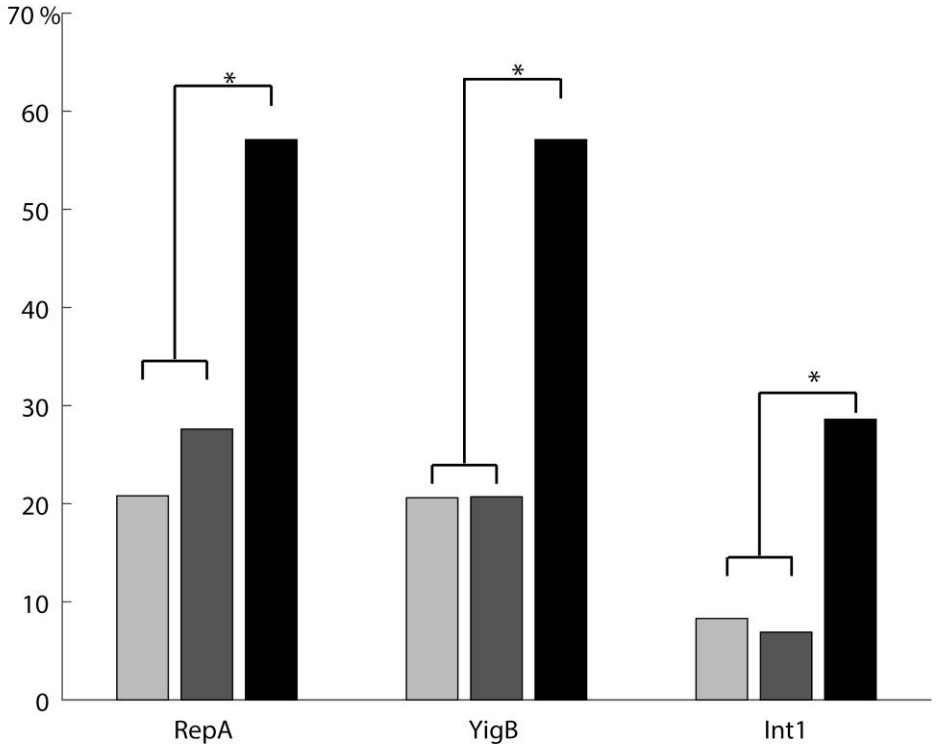


Figure 4: Correlation analysis of *rep*A and *yig*B gene abundances. Pearson correlation of *rep*A and *yig*B genes in samples from Boston, Chicago and Evanston. *=Samples from Boston ; += samples from Chicago; o= samples from Evanston

527 **Figure 5**



528

529 **Figure 5:** Geographical distribution of plasmid signature genes. Relative proportion of samples

530 positive to *rep*A, *yig*B and *int*1 genes in Boston, Chicago and Evanston. Black= samples from

531 Evanston; Dark grey= samples from Chicago; Light grey= samples from Boston;.* P

532 value>0.0001 (binomial testing).

533

534

535

536

537

538

Tables

Table 1

Description of cohort[a]

	NEC positive			NEC negative		
	Boston	Chicago	Evanston	Boston	Chicago	Evanston
Number of infants	8	12	3	16	17	6
Infants with longitudinal data[b]	7/ 5 ± 2	5/ 4 ± 1	0	14/ 4 ± 1	3/ 6 ± 1	0
Gestational age (week)	29.1 ± 2.6	26.2 ± 2.9	25.8 ± 3.6	27.9 ± 1.2	25.9 ± 2.4	22.7 ± 2.9
Birthweight (g)	1169 ± 382	903 ± 329	905 ± 441	1271 ± 502	886 ± 299	886 ± 303
Day of life when NEC was diagnosed	31.5 ± 1.9	31 ± 2.8	31 ± 0.95	n/a	n/a	n/a
Total number of days antibiotics was given	9 ± 8.2	ND	11.3 ± 10.6	5.2 ± 5.6	ND	6.6 ± 6.3
% of caesarean-born infants	75	42	33	75	76	100

[a] Erros are given by standard deviations

[b]The representation is given by: number of infants/number of samples per infant.

545 **Table 2**

546 Primers used in the study

Gene		Sequence	Annealing temperature(°C)
*Int*1	F	ACGAGCGCAAGGTTTCGGT	
	R	GAAAGGTCTGGTCATACATG	53
*Rep*A	F	GGGCACGTTTTATTCCACGG	
	R	GTTTGCTGCCCTTGATGTGT	59
*Yig*B	F	TGACTGATGAACATCGCCGT	
	R	GTGCTGCCTGTCCTTCAGAA	59.6
16S rRNA	F	TCCTACGGGAGGCAGCAGT	
	R	GGACTACCAGGGTATCTAATCCTGTT	60

547

548

549 **Table 3**

550 Antibiotic resistance genes found in longitudinal samples of the patients taken from

Patient	Institution	N[a]	β-lactam	Macrolide	Sulphonamide	Aminoglycoside	Trimethop-rim	Fosfom-cin	Tetracycl-ine	Phenic-ol
89	Boston	4	+++	+++	+++	+++	+++	-	-	-
86	Boston	4	++	+	-	-	-	+++	-	-
17	Chicago	3	++	+	-	+	-	-	+	-
22	Chicago	3	+++	+	+++	+++	++	+++	-	-
49	Evanston	1	+++	-	+++	+++	+++	-	-	+++

551 ResFinder.

552 [a]'N' number of samples. '+++' located in all the samples; '++' in more than one sample; '+' in
553 one sample only;'-' absent in all samples

554

555 **Table 4**

556 Abundance levels of OTU2 and signature genes in longitudinal datasets of NEC positive and

557 negative infants**.**

Patient	NEC	Sampling day	NEC Sample[b]	OTU2	*Rep*A[a]	*Yig*B[a]	*Int*1[a]
37	Yes	12	t-15	+	+	+	-
		17	t-10	+	-	++	-
		21	t-6	+	+++	+++	-
		24	t-3	++	++	++	-
		27	t-0	++	++	+++	-
86	Yes	10	t-2	+	-	++	++
		11	t-1	+	-	++	++
		12	t-0	+	+++	-	+++
15	No	38	No	+	-	-	-
		42	No	+	-	-	-
		48	No	+	-	-	-
		52	No	+	-	-	-
		56	No	+	-	-	-
		67	No	++	+++	+++	-
17	Yes	11	t-45	+	++	++	-
		42	t-14	++	++	++	++
		46	t-10	+++	+++	+++	++
		56	t-0	+++	+++	+++	+
89	No	4	No	+++	++	-	+++
		7	No	+++	+++	+++	+++
		9	No	+++	+++	+++	+
41	No	23	No	+	++	-	-
		26	No	+	++	-	-
		27	No	+	++	-	-
94	No	7	No	+	-	-	-
		10	No	+	-	-	-
		11	No	+	-	+	-
		12	No	+	-	-	-
25	No	7	No	+	-	-	-
		10	No	+	-	-	-
		11	No	+	+	+	-
		12	No	+	-	-	-
	No	6	No	+	-	-	-
		8	No	+	-	-	-

116	9	No	+	-	-	-
	10	No	+	-	-	-
	11	No	+	-	++	++

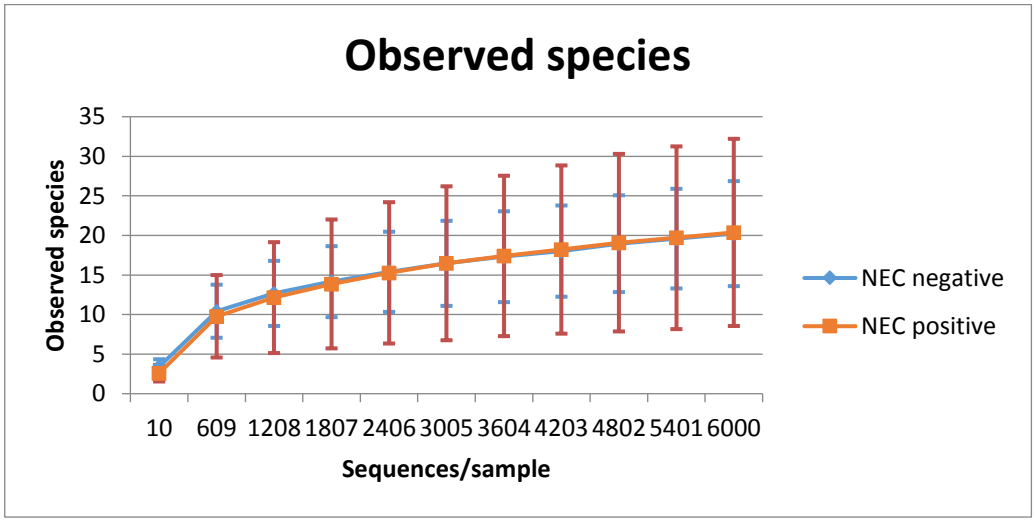
[a] The gene abundances relative to 16S rRNA gene was calculted for *Rep*A, *Yig*B and *Int*1 column, '+' -4 to -5; '++' -2 to -3.99; '+++' 0 to -1.99 are relative gene abundance values. The number of 16S rRNA sequences for OTU2 column, '+' <8% (500 seqs); '++' >8% to <42% (2500 seqs); '+++' >42%;

[b] t= day of life; t=0 is the day of NEC diagnosis, while the other numbers indicate days prior to NEC diagnosis. NA do not have the taxonomy information.

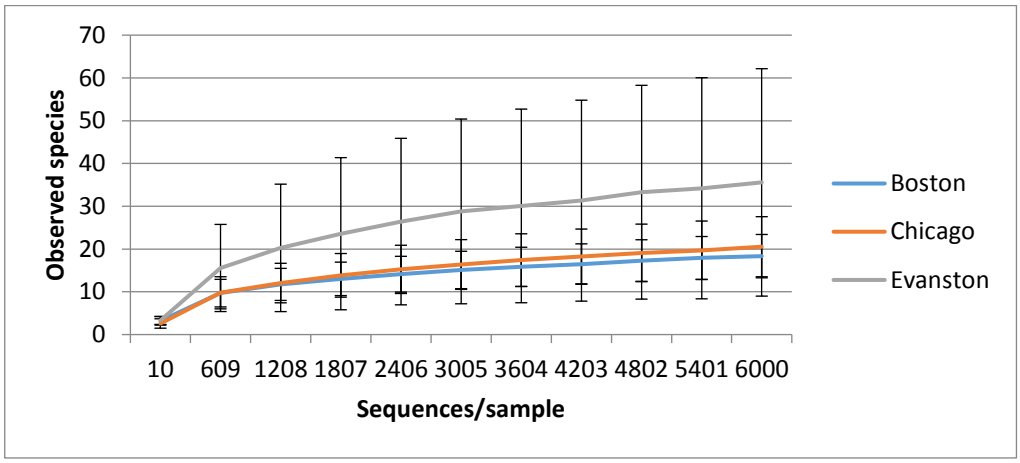
Supplementary data

Supplementary Figure 1

a)



b)



Supplementary Figure 1: Rarefraction curves of observed species in the number of sequences per sample(Average±SEM) a) in NEC positive and negative b) in hospitals of Boston, Chicago and Evanston

Supplementary Table 1: Metadata for the samples chosen for shotgun metagenome analyses.

Patient	Sampling day	Hospital	NEC (Y/N)[a]	Age of NEC diagnosis	Abundance of OTU2
	4				5991
89	7	Boston	N	N/A[b]	3288
	9				2971
	9[1]				3495
	10				5
86	11	Boston	Y	12	2
	12				1
	12[1]				2
	42				437
17	46	Chicago	Y		3710
	56				3765
	11				6
22	16	Chicago	Y	25	0
	22				0
49	46	Evanston	Y	46	4878

[a]N- No; Y- Yes; [b]N/A not applicable; [1]extra sample of the same sampling time point

583 **Supplementary Table 2:** Contigs having OTU2 representative sequence (Pairwise identity >97%).

Patient	Sampling day	Contigs with OTU2 representative sequence	Length of the contig	Pairwise identity (%)
89	4	22	93,813	97,7
	7	89	986	98,9
		97	3073	99,5
	9	79	2058	99,5
	9[1]	222	4036	99
86	11	91	2905	99,1
	12	40	125961	97
17	42	121	1959	99.3
	46	139	2940	99,7
	56	12	111647	97
		127	19052	99
49	46	6	24080	97

584 [1] extra sample of the same sampling time point

585

586 **Supplementary Table 3:** Metagenome reads mapped to reference genome (HG428755).

Patient	Sampling day	% of mapped reads	Pairwise identity (%)	Reference genome covered (%)	Reference plasmid (CBTO010000001) covered (%)	Reference plasmid (CBTO010000002) covered (%)
89	4	74	99	87.1	74	59.5
	7	38	95.6	60	37.7	31.6
	9	58	98.8	89.3	65	46,4
	9[1]	36	73.4	56	28	13.8
	10	36.2	92.5	48.3	21.3	9
86	11	74	92.5	48.3	47	32.7
	12	61	99.2	89.6	66.2	49
	12[1]	34.5	92.9	70	43	31.8
	42	38	98.3	56	25	10.3
17	46	63.5	98.6	94.2	83.4	62
	56	20	92	56.1	37.5	27.4
49	46	83.5	99	88.3	63.7	43

587 [1] extra sample of the same sampling time point

588

589

590

591

592

Supplementary Table 4: Coverage of the assembled conjugative plasmid by the metagenomic reads from every sample.

Patient	Sample number	Pairwise identity (%)	% of mapped reads	Coverage of plasmid (%)	Coverage of integron	Pairwise identity	Coverage of hha gene	Pairwise identity	Coverage of replication and transfer genes	Pairwise identity
	4	99	0.017	91.4	84,8	98,5	100	92,5	99	99,5
89	7	97.8	0.01	41	81,7	97	0	0	40	98,3
	9[1]	99	0.017	100%	100	99,5	100	99,9	100	99,1
	9	98	0.01	36	30,8	98,6	0	0	31,3	98,3
	10	98.6	0.01	30.8	61	99,4	0	0	37,3	98,8
86	11	69.4	0.01	64	69	99	91,2	99,8	69,6	99
	12	99.4	0.023	100	100	99,5	100	99,9	1010	99
	12[1]	97	0.02	46	81,4	99,8	1,4	70,6	42,2	97,2
	42	99.2	0.002	32.3	55,4	99,4	0	0	39,3	99,2
17	46	99	0.004	84.3	16,6	100	100	100	100	99,6
	56	98.3	0.015	43.1	78,7	99,7	4,2	64,5	42,8	98
	11	96.3	0.014	100	100	99,5	100	99	100	99
22	16	98.8	0.03	30.6	1,5	90	10,8	98	38,9	97
	22	98	0.001	81.4	0	0	100	99,5	98,9	99
49	46	81.4	0.05	100	100	99,8	100	99,5	100	99,5

[1] extra sample of the same sampling time point

PAPER 3

Ravi A, Valdés-Varela L, Gueimonde M and Rudi K (2017). Transmission and persistence of IncF conjugative plasmids in the gut microbiota of full-term infants. Submitted to *FEMS Microbial Ecology*

Transmission and persistence of IncF conjugative plasmids in the gut microbiota of full-term infants

Anuradha Ravi[1], Lorena Valdés-Varela[2], Miguel Gueimonde[2] and Knut Rudi[1*].

[1]Norwegian University of Life Sciences, Chemistry, Biotechnology and Food science department (KBM), Campus Ås, Ås, Norway;

[2] Department of Microbiology and Biochemistry of Dairy Products, IPLA-CSIC, Villaviciosa, Asturias, Spain.

Keywords: Conjugative plasmids, gut microbiota, full-term infants, Multireplicon, Incompatibility, mobile genetic elements

Address of correspondence to Knut Rudi, Department of Chemistry, Biotechnology and Food Science (IKBM), Norwegian University of Life Sciences. Christian M Falsensvei 1, N-1432 Ås, Norway. Work Phone +47 67232546. Email for correspondence: knut.rudi@nmbu.no.

Abstract

Conjugative plasmids represent major reservoirs for horizontal transmission of antibiotic resistance (AR) and virulence genes. Our knowledge about the ecology and persistence of these plasmids in the gut microbiota remains limited. The IncF plasmids are the most widespread in clinical samples and in healthy humans. Therefore, the main aim is to study their ecology and association with the developing gut microbiota. Using a longitudinal (2, 10, 30 and 90 days) cohort of full-term infants, we investigated the transmission and persistence of IncFIA and IncFIB plasmids. The IncFIB plasmids had higher representation than IncFIA in the cohort, while IncFIA always co-occurred with IncFIB. However, the relative gene abundance of IncFIA was significantly higher than IncFIB for all time periods, indicating that IncFIA may be a higher copy-number plasmid. Through the LEFse and OTU-level associations, we observed major differences in the abundance of *Enterobacteriaceae* in samples positive and negative to IncFIB. This association was significant at 2, 10 and 30 days and showed an association with vaginal delivery. From the shot-gun analyses, we *de novo* assembled multi-replicon shared (IncFIA/IncFIB) and integrated (IncFIA/IB) plasmids that were persistent through the dataset. Overall, the study demonstrates the nature of IncF plasmids in complex microbial communities.

Introduction

The human gut microbiota harbors a complex microbial ecosystem that goes through major compositional and functional changes from mass colonization at birth to the adult-like state (Avershina et al. 2013; Palmer et al. 2007). Although our knowledge about the shifts in composition during the early-life microbiota development are relatively well established, we know very little about the infant intestinal mobilome, i.e. the collection of mobile genetic elements (MGEs) of the gut microbiota (Palmer et al. 2007; Yatsunenko et al. 2012). We have recently shown that MGEs containing antibiotic resistance genes (AR) are prevalent and surprisingly persistent in the infant gut even across mother to child up to 2 years of age (Ravi et al. 2015) . Furthermore, we found these elements linked to conjugative plasmids, suggesting an important role in the reservoir for AR transmission.

Conjugative plasmids are autonomous and often encode multiple accessory elements and addiction systems to ensure their maintenance and stability in the host cell (Norman et al. 2009). Due to this, they are often regarded as parasites of the bacterial cell (Olendzenski 2009). Conjugative plasmids are classified based on their genetic traits and compatibility. Two different plasmid groups with the same replication machinery are incompatible in the same bacterial cell (Norman et al. 2009; Villa et al. 2010). Using this feature, 27 incompatibility (Inc) groups have been recognized, with variants within each group (Carattoli 2009). Incompatibility F group (IncF) plasmids have a narrow host range and have been reported in different *Enterobacteriaceae,* including *Escherichia coli* (Hopkins et al. 2006; Karisik et al. 2006; Marcade et al. 2009), *Salmonella enterica* (Hopkins et al. 2006; Park et al. 2009) and *Enterobacter aerogenes* (Park et al. 2009). They are usually low copy-number plasmids. The most common variants of the IncF conjugative plasmids are IncFII, IncFIA &IncFIB (Kline 1985; Villa et al. 2010). The IncFII plasmid do not participate in the initiation of replication and are often found in association with IncFIA and IncFIB plasmids (Osborn et al. 2000; Toukdarian 2004).

The *Enterobacteriaceae* are one of the most dominant groups in the neonatal microbiota (Arboleya et al. 2012). The IncF conjugative plasmids in *Enterobacteriaceae* are of particular interest since they contribute to the carriage and spread of AR and virulence genes (Carattoli 2011). These plasmids have been associated with the unexpected emergence of plasmid-

mediated extended-spectrum β-lactamases (ESBLs) (Coque et al. 2008; Novais et al. 2007), quinolone (Lascols et al. 2008) and aminoglycoside resistances (Carattoli 2009). They also carry specific virulence traits such as cytotoxins and adhesion factors as accessory genes (Timothy J. Johnson and Nolan 2009). The IncF variants are one of the most represented plasmid types in clinical samples and are also represented in healthy humans (T. J. Johnson et al. 2007).

Due to the role of conjugative plasmids in the persistence of AR, the aim of the current study was to investigate the prevalence, persistence and association of conjugative plasmids in the gut microbiota of 47 healthy, full-term infants. The fecal samples were collected from 2 days to 90 days post-delivery. We used quantitative PCR to determine the prevalence, abundance and stability of conjugative plasmids. Amplicon and shotgun metagenome sequencing was used to categorize the microbiota taxonomy, analyze association with the IncF variants, characterize the conjugative plasmid and trace these plasmids longitudinally.

Materials and methods

Cohort description & sample collection:

The study consists of an unselected longitudinal cohort of 47 healthy, full-term infants born between gestational weeks 38 and 41 (average 39.1) after an uncomplicated pregnancy at the Central University Hospital of Asturias in northern Spain. The infants' birth weight ranged between 3050 and 4120 g (average 3370), and all infants remained healthy during the length of the study. Eighty-three percent of the infants were vaginally delivered, and until the end of the study, 66% were exclusively breastfed. None of the children was given antibiotics up to the end of sampling, and all were discharged from the hospital on their second or third day of life.

DNA extraction

Fecal samples were collected at two, 10, 30 and 90 days of age in a sterile container and immediately frozen at -20ºC. Samples were sent within one week to the laboratory where they were stored at -80ºC until analysis. For DNA extraction the samples were thawed, weighed (1 g) and diluted ten times in sterile PBS solution for homogenization in stomacher at full-speed for five minutes (LabBlender, Sussex, UK). DNA was then extracted from 1 mL of homogenate

by using the QIAamp DNA stool kit (Qiagen GmbH, Hilden, Germany) as previously described (Arboleya et al. 2012). Extracted DNA was kept frozen at -80ºC until analysis.

Gene quantification:

The abundance of the IncF variants in the samples was calculated relative to the 16S rRNA gene by quantitative PCR. For the identification of the IncF variants by quantitative PCR, the replication regulatory region (*rep*A) (Carattoli et al. 2005) and iteron region (Carattoli et al. 2005) were targeted as marker genes for IncFIA and IncFIB respectively. Each PCR reaction of 20µl contained 1X HOT FIREPol EvaGreen qPCR mix (Solis BioDyne, Estonia); 200nM of forward and reverse primers and one µl of DNA. The reaction mix was run in a LightCycler 480 apparatus (Roche, Germany). Thermal cycling condition for the 16S rRNA V3-V4 region(Nadkarni et al. 2002) (5'-TCCTACGGGAGGCAGCAGT-3'; 5'-GGACTACCAGGGTATCTAATCCTGTT-3') consisted on an initial denaturation of 95°C for 15 mins followed by 95°C for 30s and 60°C for 30s for 40 cycles. The thermal cycling condition for iteron region of IncFIA (5'-CCATGCTGGTTCTAGAGAAGGTG-3'; 5'-GTATATCCTTACTGGCTTCCGCAG-3') and *rep*A of IncFIB (5'-GGAGTTCTGACACACGATTTTCTG-3'; 5'- CTCCCGTCGCTTCAGGGCATT-3') was initial denaturation at 95°C for 15 mins followed by 95°C for 30 s, 60°C for 30 s and 72°C for 1 min for 40 cycles. After the thermal cycling, the raw CT values were exported into LinRegPCR program (Ruijter et al. 2013) for baseline correction and average PCR efficiency. High resolution melting (HRM) curve analysis and targeted restriction digestion was used to verify the PCR amplicons.

16S rRNA profiling analyses

Illumina sequencing was used to analyze the microbial communities through 16S rRNA gene amplicon (n=180) sequencing. For PCR amplification, the 16S rRNA primers PRK341F and PRK806R(Yu et al. 2005), targeting the V3-V4 hypervariable region, were used under the following conditions: 95°C for 15 min followed by 95°C for 30 s, 50°C for 1 min and 72°C for 45 s. These primers were modified to contain Illumina-specific adapters. Each PCR reaction contained 1X HOT FIREPol DNA polymerase (Solis BioDyne, Estonia); 200 nM of uniquely tagged forward and reverse primers; 1µl of DNA in a total reaction volume of 25 µl. The PCR products were purified using Agencourt AMPure XP-PCR Purification kit (Beckman Coulter,

Indianapolis, IN, USA) and pooled based on their concentration measured by Qubit 1.0 fluorometer (ThermoFisher Scientific, Waltham, MA, USA). The pooled products were again purified by Agencourt AMPure XP-PCR Purification kit (Beckman Coulter). Then, the concentration was measured using QX200 droplet digital PCR system (Biorad, Oslo, Norway) using Illumina adapter specific primers and the normalized amplicon pool was sequenced on Miseq platform (Illumina, San Diego, CA, USA) using V3 chemistry with 300 bp paired-end reads.

Sequences were analyzed using the QIIME pipeline (Caporaso et al. 2010). Sequences were quality-filtered (*split_libraries.py*; minimum sequence length 350bp; minimum average quality score 25; average error estimation as parameter) and then clustered at 97% homology level using Usearch version 8 using Greengenes database (DeSantis et al. 2006).

Shotgun metagenome analyses

The shotgun metagenome sequencing of selected samples was carried out by Illumina sequencing. The metagenomes of the samples were fragmented, tagged, quantified and normalized according to the Nextera XT protocol using manufacturer's recommendations. Sequencing was done in-house using the same platform as the 16S rRNA amplicon sequencing.

Data analysis and assembly of the metagenome reads were performed by Geneious R10 (Kearse et al. 2012) following the recommended guidelines. Geneious Read Mapper (Geneious, USA) was used for *de novo* assembly of the reads into contigs. For the metagenome assembly, the reads were trimmed (error probability 0.05) and merged (overlap length 10bp). PlasmidFinder program, an online tool for identifying plasmid related genes (Carattoli et al. 2014), was used to identify plasmid related contigs. The contigs from the *de novo* assembler were annotated using RAST (Rapid Annotation using Subsystem Technology) annotation server. This server uses SEED based annotation to identify functional genes (Aziz et al. 2008). ProgressiveMauve multiple genome alignment (Darling et al. 2010) was used to compare the different contigs between the samples.

Validation and statistical analyses:

Fisher exact test, Spearman correlation and Kruskal Wallis test were used to test the pairwise comparisons of the relative gene abundances of the IncF variants and the Operational

Taxonomic Units (OTU) and between the individual gene abundance within different sampling times. Correction of multiple testing was carried out using Benjamini and Hochberg false discovery test (BHFDR). The error bars were calculated using standard error of mean (SEM). ANOVA-simultaneous component analysis (ASCA) (Smilde et al. 2005) method was used to determine OTU-level associations between different time periods within the IncF variants (Eigenvector Research Incorporated, USA). To do this, the samples with binarized data of with and without IncF variants along with the corresponding OTU abundances was used. The data analyses was performed using MATLAB® R2016a software (The MathWorks Inc., USA).

The persistence of IncFIB across the time periods was calculated as the ratio of the number of IncFIB positive samples between the two time periods to the total number of IncFIB positive samples for which information on both time periods was available.

To categorize the bacterial species that are more or less in the samples with and without IncFIA and IncFIB and to account for significant changes of the microbial diversity, LEFse (Linear Discriminant Analysis (LDA) effect size) analysis was used (Segata et al. 2011). This performs a nonparametric Wilcoxon sum-rank test followed by a LDA analysis to measure the effect size of each taxon. Binarized data on the presence/absence of IncF variants along with the taxonomy information with the OTU abundances/sample was submitted for the analysis.

Results

Microbiota composition

On average, 35,317 sequences/samples were generated from the V3-V4 region of 16S rRNA gene after quality filtering and chimera removal. The final dataset comprised of 391 OTUs belonging to 13 bacterial classes. The final dataset, after quality filtering and rarefying at 5000 sequences/sample contained 167 samples belonging to 47 full-term infants.

The ten most abundant taxonomy classes consisted of nearly 99.95% of the microbial composition (Figure 1). In general, at 2 days, the population of *Gammaproteobacteria* (60%) in particular *Enterobacteriaceae* (48.4%) was the highest with lower levels of *Actinobacteria* (6% of the population). The population of *Gammaproteobacteria* dropped to 43% by 30 days and increased to 49% at 90 days, whereas the population of *Actinobacteria* increased with age, reaching 25% by the age of 3 months. For infants born through vaginal delivery (n=31), *Gammaproteobacteria* (64%) was higher with lower levels of *Bacilli* (15.5%) in particular *Streptococcaceae* (4.6%) compared to infants born via C- section (n=6; 15% [*Gammaproteobacteria*]; 68.3 [*Bacilli*]; 39% [*Streptococcaceae*]). In addition, there were major differences in the population of *Bacteroidia* in infants receiving breast milk (n=23; 12%) compared to formula-fed infants (n=14; 1%).

Distribution of IncF variants in the cohort

In total, 97 of 180 samples in the cohort showed the presence of IncF variants. The prevalence of IncFIB positive samples was higher (54% of all samples) compared to that of IncFIA (8%). At two days of age, the prevalence was 50% and 5% for IncFIB and IncFIA, respectively. The IncFIB indicated the highest prevalence at 10 days and 30 days of age (58%, of the cohort) and for IncFIA at 90 days (10%). Samples positive for IncFIA were also positive for IncFIB.
In the samples positive to both IncFIA and IncFIB, the relative gene abundance of IncFIA was two times higher on average than that of IncFIB. Between the time periods, the relative gene abundances of IncFIA and IncFIB varied throughout the study, being higher at two days and 90 days of age and lower at 10 and 30 days (Figure 2). In addition to this, the relative gene abundance of IncFIA and IncFIB from all time periods displayed a significant correlation

(n=14, p=0.01, ρ=0.6 [IncFIA- IncFIB], Spearman correlation), indicating a positive interaction between the IncF variants.

We then investigated whether the most prevalent of the IncF variant, i.e. IncFIB, was stable across the time periods. We found that IncFIB exhibited highest persistence patterns between two days and 10 days (p=0.0001, Fisher exact test) and 10 days and 30 days (p=0.0001, Fisher exact test) (Figure 3). Significant persistence across 10 days and 30 days and two days and 30 days was also detected (p=0.001, Fisher exact test).

Microbiota association with IncFIA and IncFIB

The differences in the microbiota taxonomic composition by LEFse analysis showed significant taxon associations in the samples with IncF variants compared to samples without (Figure 4). The samples with IncFIB had a significantly higher population of *Gammaproteobacteria* at all the sampling times ($p<0.05$, LEFse analysis), particularly within the *Enterobacteriaceae* family. On the contrary, the IncFIB-positive samples exhibited a negative association with the population of *Actinobacteria* and *Bacillales*, particularly with the families *Bifidobacteriaceae* and *Enterococcaceae*, respectively. Diverse bacterial classes (*Verrucomicrobiae*, *Alphaproteobacteria* and *Gammaproteobacteria*) showed a positive association to samples with IncFIA, whereas *Actinobacteria*, in particular *Bifidobacteriaceae*, showed a negative association at 30 days and 90 days of age.

ASCA-ANOVA analyses showed significant associations in samples with and without IncFIB at two, 10 and 30 days of age (n=21, p=0.0001; n=24, p=0.0065; n=25, p=0.001, respectively). However, this association was not significant at 90 days of age. The model also showed significant interaction between IncFIB and mode of delivery, in particular vaginal delivery (n=82, p=0.0003). Table 1 illustrates OTUs involved in the significant interactions with IncF variants at two, 10 and 30 days. OTU1 classified as *Enterobacteriaceae* showed the most substantial changes in the microbiota in samples positive to IncFIB and vaginal delivery. Samples positive to IncFIB exhibited higher abundances of OTU1 compared to IncFIB negative samples.

Since OTU1 indicated strong associations with IncFIB, we looked into whether these associations were constant over time (Figure 5). Major differences in the OTU1 abundance between IncFIB positive and negative samples were observed. The IncFIB negative samples

displayed an increase in OTU1 abundance over time, whereas the IncFIB positive samples showed a decrease over time. There was a significant change in OTU1 abundance in the IncFIB positive samples at two days with 30 days and 90 days (p=1.03e-04 [2 days- 30 days]; p=3.32e-04 [2 days- 90 days], Kruskal Wallis test, BHFDR tested).

In addition to this, we investigated whether the abundance of some OTUs varied according to the changes in relative abundance of the IncF variants over time. The IncFIA indicated OTU116 as *Veillonaceae* at two days, OTU379 as *Paraprevotella* at 10 days and OTU117 as *Lactococcaceae* at 90 days as significant OTUs (pvalue <0.001; Kruskal wallis test, BHFDR tested). The IncFIB showed OTU1 and OTU15 as *Streptococcaceae* at two days as significant OTUs and OTU1 at 10 days and 30 days (pvalue<0.001, Kruskal wallis test, BHFDR tested).

Shotgun metagenome assembly

A subset of ten samples with high and low levels of IncFIB relative gene abundance is highlighted in the Supplementary Table 1, with microbiota profile information. These samples were chosen for shotgun metagenome sequencing. The samples are named along with their metadata and their corresponding time periods. For instance, PA-V-B-2 represents infant A, vaginally delivered, breastmilk- fed and 2 days old, and PB-V-F-10 represents infant B, vaginally delivered, formula-fed and 10 days old. On average, the samples had 1,393,880 reads with read lengths from 35bp to 301bp. The filtered and merged reads were built into contigs, obtaining an average of 1,623 contigs with over 1,000bp in length. The average N50 length was 145,467 bp with at least 143 contigs >=N50 length per sample.

Given that the samples had high and low quantities of the IncF conjugative plasmids, the contigs were submitted to PlasmidFinder for identifying plasmid related genes. Seven out of the eight samples with high levels of IncF gene abundances contained IncFIA and IncFIB related contigs. NCBI-BLAST analysis of these contigs depicted different BLAST hits for IncFIA and IncFIB, except for PC-V-B-10 in which the IncF genes were present on the same contig (E value =0; average identity 99.2% [range 98% -100%]; average query coverage 87.8 [range 54%-100%]) (Supplementary Table 2).

The filtered and merged reads of the samples were mapped towards its corresponding BLAST hit that had the highest query coverage. In all samples, the reads covered 63% on average of the

IncFIA reference genome and 77.6% on average for the IncFIB reference genome (average pairwise identity 97% [Range 95%- 100%]) (Supplementary Table 3). We then investigated whether the detected IncFIA and IncFIB conjugative plasmids were linked or independent. This was done by mapping the reads/sample towards the corresponding IncF reference genomes to identify unique and shared reads. These reads were mapped to their contigs in corresponding samples. More information is given in the supplementary text. Using this approach, we determined unique reads for each corresponding assembled conjugative plasmid and shared reads that matched both the assembled IncF conjugative plasmids. On average, 28,760 reads were unique for IncFIA, and 12,249 reads were unique for IncFIB (average pairwise identity 96%) (Supplementary Table 4). In PA-V-B and PB-V-B, the annotation of the unique reads identified replication genes along with integrons, toxin/antitoxin system and plasmid partitioning gene family. The presence of these genes in the unique reads potentially indicates separate conjugative plasmids. However, the annotation of the shared reads identified transfer genes (tra & trb) of conjugative plasmids, indicating the potential sharing of the transfer genes between the IncF plasmids.

In order to look further into the corresponding contigs carrying the shared and unique reads, the reads were mapped back to the built contigs. The contigs that mapped ≥1000 of the unique or shared reads and covered ≥10% of the contig, was considered as plasmid related contigs. In PA-V-B, the coverage of the contigs with unique reads mapping to IncFIA (avg. coverage 64.7) was significantly higher than for reads mapping to IncFIB ([avg. coverage, 23] [p=0.04, Kruskal Wallis test]) (Table 2). The contigs with shared reads in PA-V-B showed the same coverage (avg. coverage [IncFIA] 40.1; [IncFIB] 41). On the other hand in PC-V-B-30, contigs with the shared reads showed the same coverage (coverage 20.6), and no contigs with >1000 unique reads were found. Therefore, a multireplicon model of an assembled shared (IncFIA/ IncFIB) conjugative plasmid in PA-V-B and integrated IncFIA/IB conjugative plasmid in PC-V-B seems to be present in this cohort (Figure 6).

Finally, we investigated whether the IncF related contigs were persistent between the different time periods for each infant. Infants with the IncFIA and IncFIB related contigs in their longitudinal samples were selected, i.e. PA-V-B and PC-V-B. The contigs with the shared reads across the different time periods were multiple aligned using Mauve. In PC-V-B-10 and 30, up

to 50% of the assembled contigs between the two time periods show 93% identity and the rest show 99% identity. The drop in identity initially could be due to chimeric regions between the time periods. In PA-V-B, the contigs shared between IncFIA and IncFIB related reads (2 days, 10 days and 90 days of age) show 99.1% identity (Supplementary Figure 1) indicating the persistence of conjugative plasmids across the time period analyses.

Discussion

In this study, we have used the 16S rRNA information to study the interactions of microbial community profiles with the abundance of IncF variants throughout the first months of life. The gut microbiota succession indicates a higher proportion of *Enterobacteriaceae* with increasing abundance of *Actinobacteria* as previously reported for healthy, full-term infants (Avershina et al. 2013; Ravi et al. 2015). We observed a high occurrence of IncF variants in the fecal samples in our cohort of infants during the first three months of life. In general, IncF variants have a limited host range in the *Gammaproteobacteria* and show wide association within the *Enterobacteriaceae* family (Lyimo et al. 2016). Therefore, the high proportion of *Enterobacteriaceae* in the microbiota population and the limited host range of the IncF variants are the likely explanation for the observed strong association between IncFIB and the developing gut microbiota. During vaginal delivery, the vaginal and fecal microbiota are the major contributors to the initial colonization of the infants (Mueller et al. 2015). Therefore, the possibility of a vertical transmission of conjugative plasmids from the mother to the infant during delivery seems very plausible. However, this observed association of IncFIB and developing gut microbiota was lost by 90 days, as the abundance of OTU1 in IncFIB positive samples gradually decreased. Though not statistically significant, we observed a higher tendency of plasmid loss at 90 days, whereas the overall level of OTU1 abundance in IncFIB negative samples gradually increased. This could potentially lead to a weaker association of OTU1 and IncFIB, resulting in horizontal gene transfer of IncFIB plasmids to other bacterial species.

While we observed a high prevalence of IncFIB in our cohort, the prevalence of IncFIA was much lower. Nevertheless, IncFIA portrayed higher relative gene abundance compared to IncFIB, indicating that IncFIA as a higher copy number plasmid. Even though high abundant plasmids are favored in intra-host selection, i.e. over-replication of plasmid copies to increase

chances on fixating to a cell, they are outcompeted by low copy-number plasmid-borne cells or by plasmid-free cells (Watve et al. 2010). Therefore, for achieving persistence of large plasmids in complex environments as in the gut microbiota, IncF variants are often associated with copy number control (Summers 1998), active partitioning systems (Ebersbach and Gerdes 2005) and post-segregationally killing (Hayes 2003). The toxin/antitoxin systems are important for the plasmid stabilization and to effectively kill plasmid free cells (Unterholzner et al. 2013), while the plasmid partitioning system ensures proper partitioning of the plasmid to the corresponding daughter cells (Bignell and Thomas 2001). This in turn ensures copy number maintenance and inheritance of plasmid-borne daughter cells (Slater et al. 2008). In relation to this, the observed high persistence of IncFIB could potentially be due to the relative low gene abundance of IncFIB along with the presence of a plasmid partitioning system (*ParA & ParB*) and toxin/antitoxin systems (*ccd*A & *ccd*B) gathered from our metagenome analyses.

By combining the information on the relative gene abundance of the IncF variants and the coverage of the contigs with plasmid related genes, we assembled a multireplicon-based IncF plasmid from two longitudinal datasets. In the shared IncFIA/IncFIB conjugative plasmid, IncFIA showed two times higher gene abundance compared to IncFIB, which was further confirmed by our assembled data where the coverage of the unique contigs of the corresponding IncF related genes was also two times higher in IncFIA. However, the coverage of the shared contigs between the IncF variants did not show any differences, indicating shared trait between the conjugative plasmids. On the other hand, the integrated multireplicon plasmid displayed no difference in the relative gene abundance between the IncF variants and displayed no unique contigs. Commonly, variants of the Inc groups share homologous genes such as transfer genes (tra & trb), suggesting their relatedness (Suzuki et al. 2010). In addition to this, the prevalence of IncFIA plasmid always along with IncFIB plasmid suggests similar integrated or shared conjugative plasmids in our dataset. Therefore, we believe these complex conjugative plasmids have major factors that ensure their widespread and versatile rapid adaptation to drastic changes in the developing gut microbiota (Carattoli et al. 2005; Ravi et al. 2015; Villa et al. 2010).

Overall, our study shows the persistence of IncF conjugative plasmids in the developing gut microbiota especially within the *Enterobacteriaceae* family. This strongly suggests the

presence of an active mobilome containing multidrug resistance genes in the developing infant gut microbiota.

Funding

This work was supported by a Quota scholarship and funding from Norwegian University of Life Sciences, Ås, Norway, and the JPI HDHL initiative and Spanish MINECO (Project EarlyMicroHealth, ref. PCIN-2015-233)

Acknowledgements

We would like to thank Trine L'Abée-Lund for her suggestions on the preparation of this manuscript and Paulo Jorge de Almeida Borges for his help with the bioinformatics analyses.

References

Arboleya, S., et al. (2012), 'Establishment and development of intestinal microbiota in preterm neonates', *FEMS Microbiol Ecol*, 79 (3), 763-72.

Avershina, Ekaterina, et al. (2013), 'Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children', *FEMS Microbiology Ecology*, n/a-n/a.

Aziz, Ramy K., et al. (2008), 'The RAST Server: Rapid Annotations using Subsystems Technology', *BMC Genomics*, 9 (1), 1-15.

Bignell, Colin and Thomas, Christopher M. (2001), 'The bacterial ParA-ParB partitioning proteins', *Journal of Biotechnology*, 91 (1), 1-34.

Caporaso, J. G., et al. (2010), 'QIIME allows analysis of high-throughput community sequencing data', *Nat Methods*, 7 (5), 335-6.

Carattoli, A. (2009), 'Resistance plasmid families in Enterobacteriaceae', *Antimicrob Agents Chemother*, 53 (6), 2227-38.

Carattoli, A. (2011), 'Plasmids in Gram negatives: molecular typing of resistance plasmids', *Int J Med Microbiol*, 301 (8), 654-8.

Carattoli, A., et al. (2005), 'Identification of plasmids by PCR-based replicon typing', *J Microbiol Methods*, 63 (3), 219-28.

Carattoli, A., et al. (2014), 'In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing', *Antimicrob Agents Chemother*, 58 (7), 3895-903.

Coque, Teresa M., et al. (2008), 'Dissemination of Clonally Related Escherichia coli Strains Expressing Extended-Spectrum β-Lactamase CTX-M-15', *Emerging Infectious Diseases*, 14 (2), 195-200.

Darling, Aaron E., Mau, Bob, and Perna, Nicole T. (2010), 'progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement', *PLOS ONE*, 5 (6), e11147.

DeSantis, T. Z., et al. (2006), 'Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB', *Appl Environ Microbiol*, 72 (7), 5069-72.

Ebersbach, G. and Gerdes, K. (2005), 'Plasmid segregation mechanisms', *Annu Rev Genet*, 39, 453-79.

Hayes, F. (2003), 'Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest', *Science*, 301 (5639), 1496-9.

Hopkins, K. L., et al. (2006), 'Replicon typing of plasmids carrying CTX-M or CMY beta-lactamases circulating among Salmonella and Escherichia coli isolates', *Antimicrob Agents Chemother*, 50 (9), 3203-6.

Johnson, T. J., et al. (2007), 'Plasmid replicon typing of commensal and pathogenic Escherichia coli isolates', *Appl Environ Microbiol*, 73 (6), 1976-83.

Johnson, Timothy J. and Nolan, Lisa K. (2009), 'Pathogenomics of the Virulence Plasmids of Escherichia coli', *Microbiology and Molecular Biology Reviews : MMBR*, 73 (4), 750-74.

Karisik, E., et al. (2006), 'Molecular characterization of plasmids encoding CTX-M-15 beta-lactamases from Escherichia coli strains in the United Kingdom', *J Antimicrob Chemother*, 58 (3), 665-8.

Kearse, M., et al. (2012), 'Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, 28 (12), 1647-9.

Kline, Bruce C. (1985), 'A review of mini-F plasmid maintenance', *Plasmid,* 14 (1), 1-16.
Lascols, Christine, et al. (2008), 'A Plasmid-Borne Shewanella algae Gene, qnrA3, and Its Possible Transfer In Vivo between Kluyvera ascorbata and Klebsiella pneumoniae', *Journal of Bacteriology,* 190 (15), 5217-23.
Lyimo, Beatus, et al. (2016), 'IncF Plasmids Are Commonly Carried by Antibiotic Resistant Escherichia coli Isolated from Drinking Water Sources in Northern Tanzania', *International Journal of Microbiology,* 2016, 3103672.
Marcade, G., et al. (2009), 'Replicon typing of plasmids in Escherichia coli producing extended-spectrum beta-lactamases', *J Antimicrob Chemother,* 63 (1), 67-71.
Mueller, N. T., et al. (2015), 'The infant microbiome development: mom matters', *Trends Mol Med,* 21 (2), 109-17.
Nadkarni, M. A., et al. (2002), 'Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set', *Microbiology,* 148 (Pt 1), 257-66.
Norman, Anders, Hansen, Lars H., and Sørensen, Søren J. (2009), 'Conjugative plasmids: vessels of the communal gene pool', *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 364 (1527), 2275-89.
Novais, Ângela, et al. (2007), 'Emergence and Dissemination of Enterobacteriaceae Isolates Producing CTX-M-1-Like Enzymes in Spain Are Associated with IncFII (CTX-M-15) and Broad-Host-Range (CTX-M-1, -3, and -32) Plasmids', *Antimicrobial Agents and Chemotherapy,* 51 (2), 796-99.
Olendzenski, Maria Boekels Gogarten; Johann Peter Gogarten; Lorraine (ed.), (2009), *Horizontal Gene transfer, Genomes in Flux*, ed. John M. Walker (1 edn., Methods in Molecular Biology, Uniter Kingdom: Springer) 542.
Osborn, A. Mark, et al. (2000), 'Mosaic plasmids and mosaic replicons: evolutionary lessons from the analysis of genetic diversity in IncFII-related replicons', *Microbiology,* 146 (9), 2267-75.
Palmer, Chana, et al. (2007), 'Development of the Human Infant Intestinal Microbiota', *PLoS Biol,* 5 (7), e177.
Park, Y. J., et al. (2009), 'Accumulation of plasmid-mediated fluoroquinolone resistance genes, qepA and qnrS1, in Enterobacter aerogenes co-producing RmtB and class A beta-lactamase LAP-1', *Ann Clin Lab Sci,* 39 (1), 55-9.
Ravi, Anuradha, et al. (2015), 'The commensal infant gut meta-mobilome as a potential reservoir for persistent multidrug resistance integrons', *Scientific Reports,* 5, 15317.
Ruijter, J. M., et al. (2013), 'Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications', *Methods,* 59.
Segata, N., et al. (2011), 'Metagenomic biomarker discovery and explanation', *Genome Biol,* 12.
Slater, F. R., et al. (2008), 'Progress towards understanding the fate of plasmids in bacterial communities', *FEMS Microbiol Ecol,* 66 (1), 3-13.
Smilde, A. K., et al. (2005), 'ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data', *Bioinformatics (Oxford, England),* 21.
Summers, D. (1998), 'Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability', *Mol Microbiol,* 29 (5), 1137-45.
Suzuki, Haruo, et al. (2010), 'Predicting Plasmid Promiscuity Based on Genomic Signature', *Journal of Bacteriology,* 192 (22), 6045-55.
Toukdarian, Aresa (2004), 'Plasmid Strategies for Broad-Host-Range Replication in Gram-Negative Bacteria', *Plasmid Biology* (American Society of Microbiology).

Unterholzner, Simon J., Poppenberger, Brigitte, and Rozhon, Wilfried (2013), 'Toxin–antitoxin systems: Biology, identification, and application', *Mobile Genetic Elements*, 3 (5), e26219.

Villa, L., et al. (2010), 'Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants', *J Antimicrob Chemother*, 65 (12), 2518-29.

Watve, Mukta M., Dahanukar, Neelesh, and Watve, Milind G. (2010), 'Sociobiological Control of Plasmid Copy Number in Bacteria', *PLOS ONE*, 5 (2), e9328.

Yatsunenko, T., et al. (2012), 'Human gut microbiome viewed across age and geography', *Nature*, 486 (7402), 222-7.

Yu, Y., et al. (2005), 'Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction', *Biotechnol Bioeng*, 89 (6), 670-9.

Figures

Figure 1

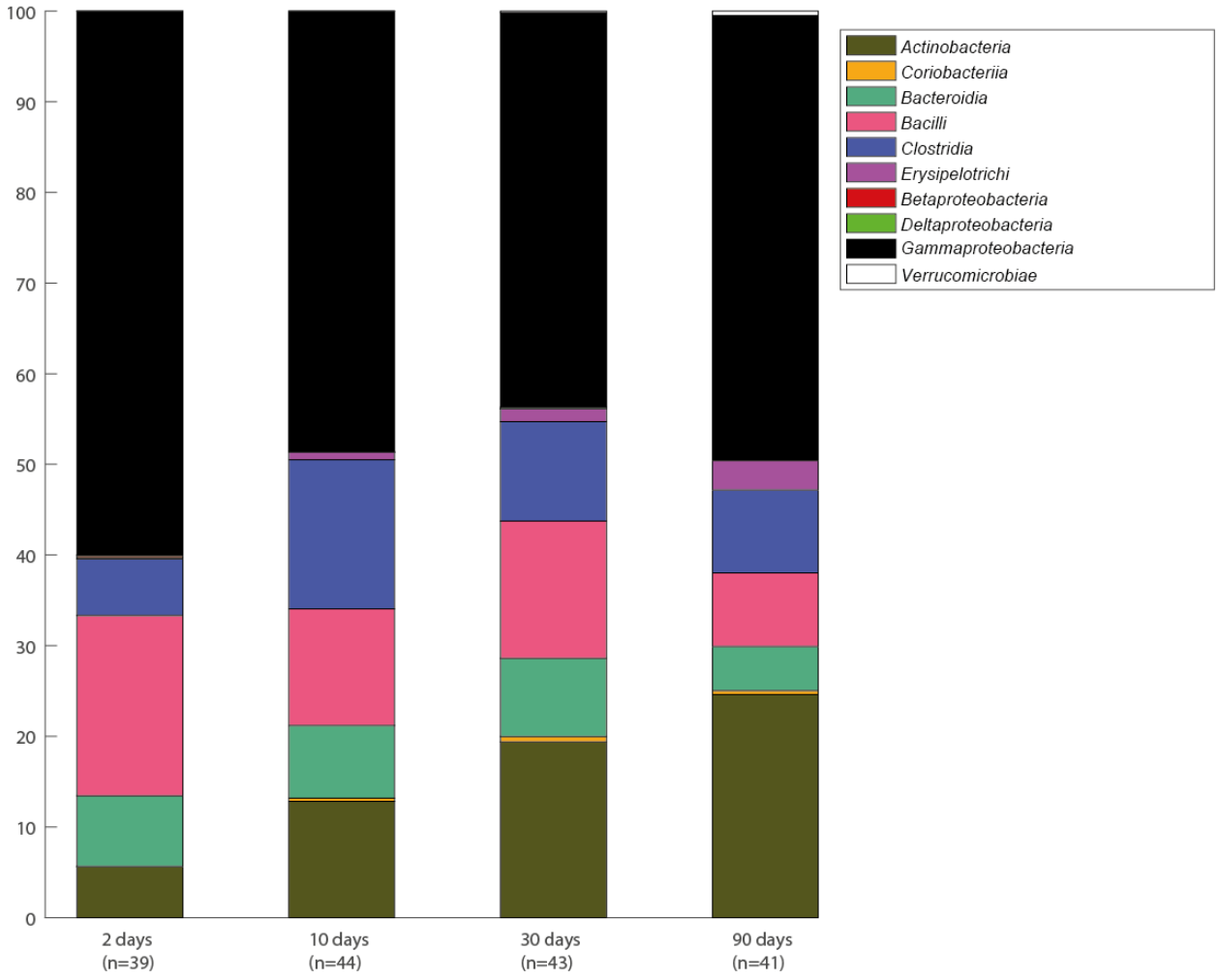
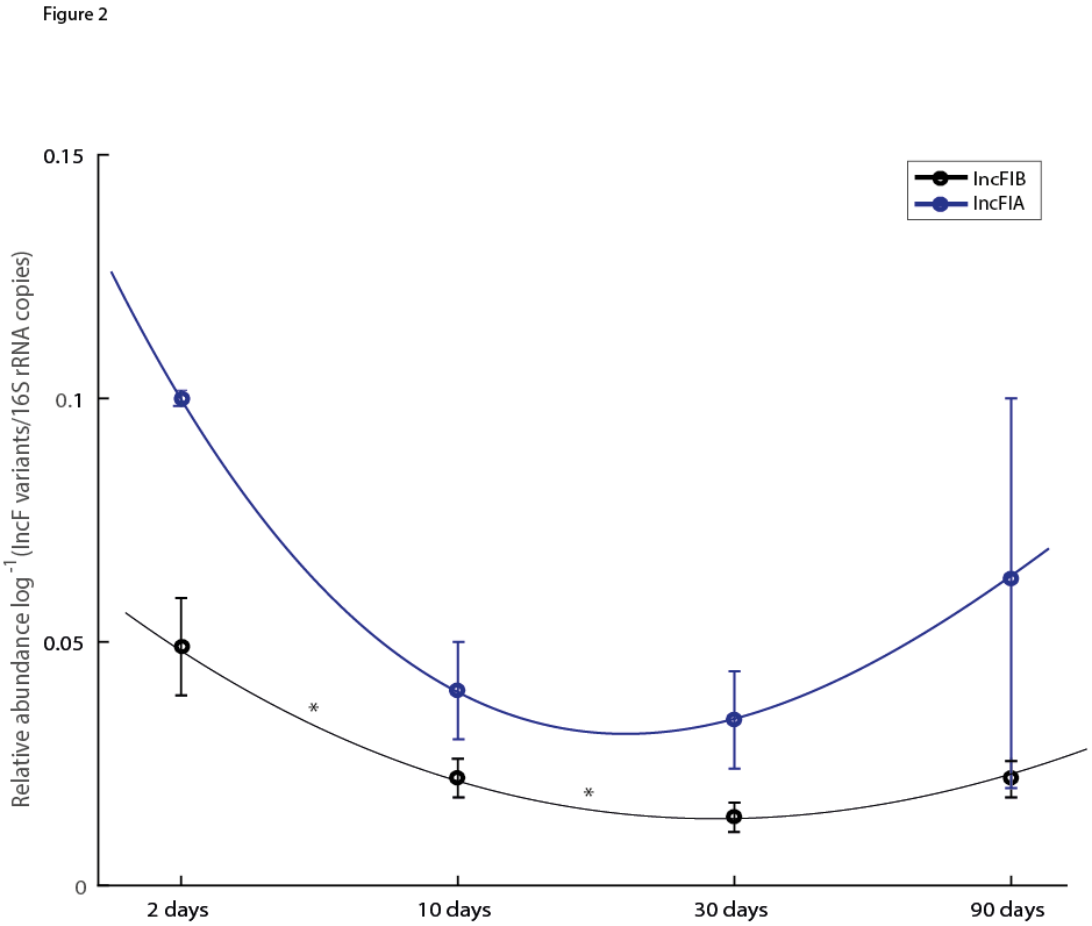


Figure 1: Bacterial class composition of the study cohort based on the 16S rRNA gene amplicons. n represents the total number of samples per time period.

481 Figure 2



482

483 **Figure 2:** Relative gene abundance of the samples positive to IncF variants between time
484 periods. Gene abundance is calculated relative to the 16S rRNA copies for that sample. The
485 error bars represent the standard error of mean. *p<0.05, Kruskal Wallis test. Correction for
486 multiple testing done by BHFDR testing. Errors bars are represented by SEM

487

Figure 3

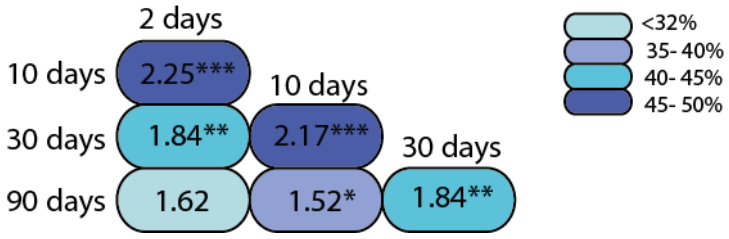
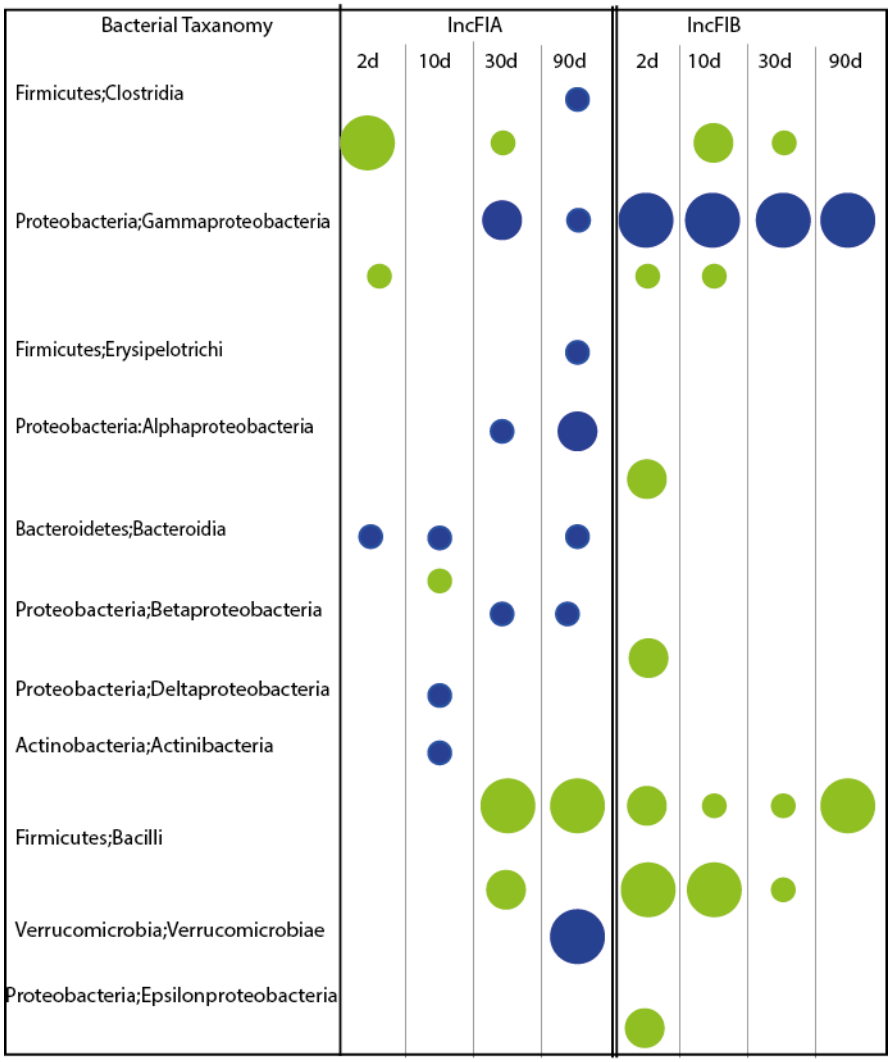


Figure 3: Persistence of IncFIB positive samples at each time point. The numbers represent the ratio of the number of IncFIB positive samples between the two time periods to the total number of IncFIB positive samples on both time periods. *p <0.05; ** <0.001, *** <0.0001, Fisher exact test. The color gradient represents the percentage of samples persistent between one time periods to the corresponding time period.

508 Figure 4

Figure 4

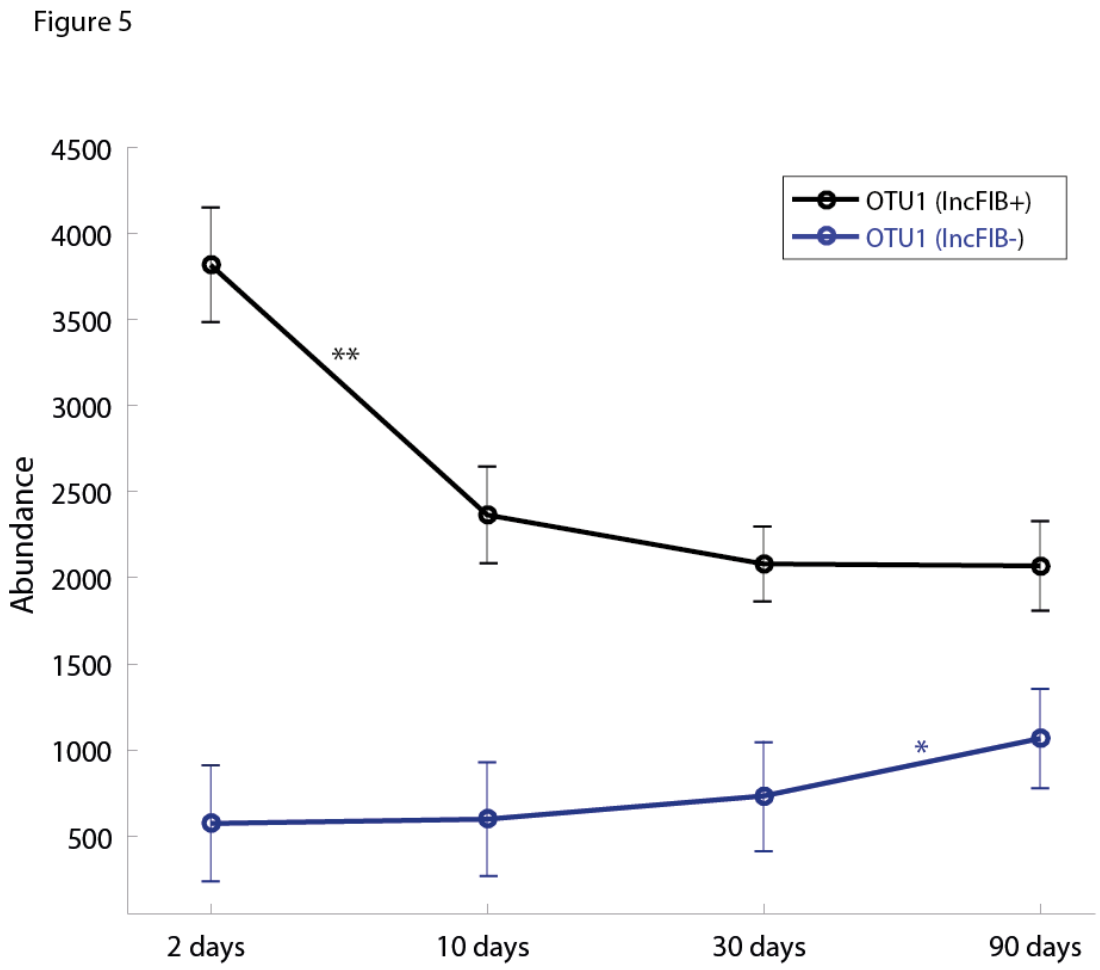


509

510 **Figure 4:** Association of bacterial taxonomy to IncF groups ($p<0.05$, LEFse analysis). Blue
511 represents bacterial groups associated with samples positive to IncF variants; green represents
512 bacterial groups negatively associated with samples positive to IncF variants. Big size circle
513 represents taxonomy association at the class level; medium size circle represent association to
514 family level and; small size circle represents association to species level.

515

516 Figure 5



517

518 **Figure 5:** OTU1 abundance in IncFIB positive and IncFIB negative samples over time. *p

519 <0.05; ** <0.001, Kruskal Wallis test. Error bars are represented by SEM.

520

521

522

523

524

525

Figure 6

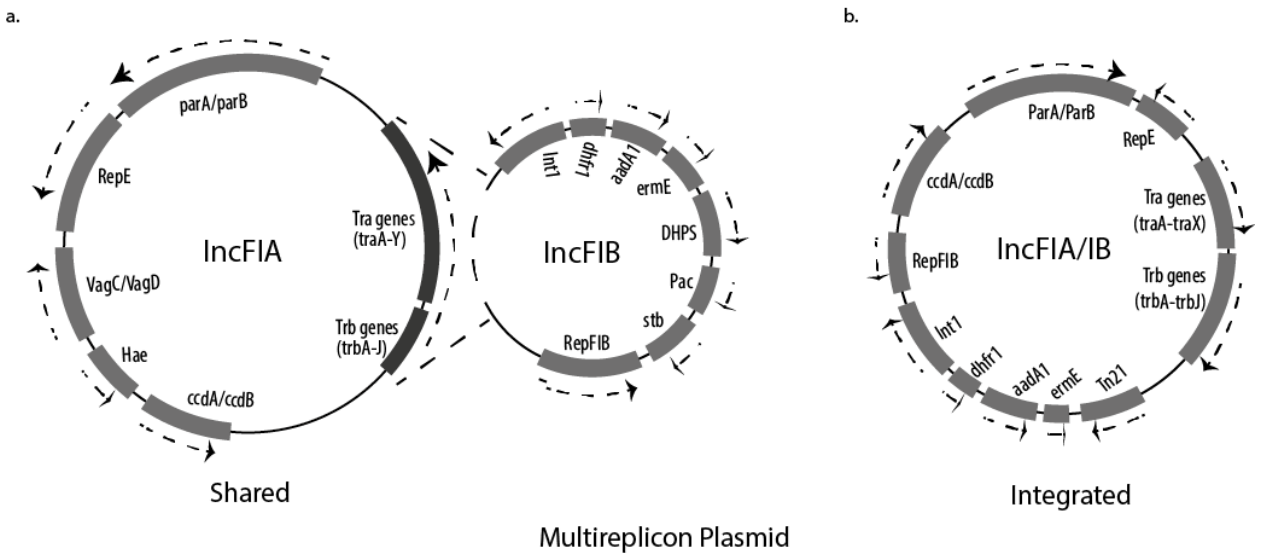


Figure 6: Representation of shared and integrated conjugative plasmid. De novo assembly of IncFIA and IncFIB conjugative plasmid. a) shared multireplicon plasmid; b) Integrated multireplicon plasmid. parA/parB, plasmid partitioning genes; RepE, replication gene; VagC/VagD, virulence associated protein; hae, haemolysin gene; ccdA/ccdB, toxin/antitoxin system, int1, class I integron integrase; dhfr1, dihydrofolate reductase; aadA1, Streptomycin O-adenyltransferase; ermE, ethidium bromide-methyl viologen resistance; DHPS, dihydropterate synthase; Pac, puromycin N-acetyltransferase; stb, plasmid stability genes; RepFIB, replication gene.

Tables

Table 1: OTU associations at different time periods

Time periods	Variable	OTU	Classification	P value[1]	Median[2]
2 days	IncFIB	1	*Enterobacteriaceae*	1.36E-05	4222 (+) 37 (-)
	IncFIB	3	*Enterococcaceae*	8.00E-03	74 (+) 10 (-)
	Delivery	1	*Enterobacteriaceae*	0.026	3801(v) 36.5 (c)
	Delivery	2	*Streptococcus*	7.08E-04	2(v) 936.5 (c
10 days	IncFIB	1	*Enterobacteriaceae*	5.99E-05	1952(+) 11.5(-)
30 days	IncFIB	1	*Enterobacteriaceae*	1.30E-04	2183(+) 9(-)

[1]P values calculated by Kruskalwallis test. [2] '+' samples with IncFIB; '-' samples without IncFIB, 'v' vaginal delivered, 'c' caesarean section.

Table 2: Contigs mapping to unique and shared reads

			IncFIA		IncFIB	
Infant	Sample (Days)	Reads	No. of contigs	Avg. coverage	No. of contigs	Avg. Coverage
PA-V-B	2	Shared	4	46,6	4	46
	10		5	31.8	5	47.06
	90		5	42	5	40.3
PB-V-F	10		4	19.4	3	18.6
PC-V-B	30		2	20.6	2	20.6
PA-V-B	2	Unique	4	84,9		26.2
	10		3	38,6	3	16.5
	90		5	70.6		26.2
PB-V-F	10		3	19.6	3	12.5
PC-V-B	30		0	0	0	0

Supplementary data

Supplementary Table 1

Metadata of the samples chosen for shotgun metagenome analyses

		Relative gene abundance	
Infant	Sampling day	IncFIA	IncFIB
PA-V-B	2	0.098	0.089
	10	0.069	0.04
	30[a]	0.028	0.0056[b]
	90	0.18	0.046
PB-V-F	2	0.10	0.001[b]
	10	0.023	0.01
	30	0.007[b]	0.002[b]
	90	0.043	0.026
PC-V-B	10	0.027	0.025
	30	0.057	0.043

[a]No sequences were produced and was discarded from further analyses [b] samples with low relative gene abundances.

559 **Supplementary Table 2**

560 BLAST hit for the contigs with plasmid related genes

Infant	Days	Gene	BLAST hit	Sequence length (bp)	Query coverage (%)	Identity (%)
PA-V-B	2	IncFIA	KP453775	43,002	91,07	100
	10			52,415	98,1	100
	90			37,140	99,86	99,9
PB-V-F	2		CP015912	191,970	88,6	99,6
	10		CP015239	11,935	84,8	95,5
PC-V-B	10		LO017738	8,965	90,93	100
	30	IncFIA/IncFIB	CO010172	97,965	84,75	97
PA-V-B	2	IncFIB	EU935739	42,933	54,35	99,8
	10			42,901	88,54	100
	90			294,819	85,33	100
PB-V-F	10		KP398867	4638	93,10	100
	90		AM886293	6,077	84,3	98,4
PC-V-B	30		CP015131	5,511	98.31	99.5

561

562 **Supplementary Table 3**

563 Mapping reads towards the genome with the highest alignment score

Infant	Sample (Days)	Gene	BLAST hit	Pairwise identity (%)	Coverage of reference sequence (%)	Average Coverage
PA-V-B	2	IncFIA	KP453775	97,7	53,7	65
	10			97,9	63,8	178
	90			97,3	53,6	60
PB-V-F	10		CP015239	95	64,4	74 30,6
PC-V-B	10		LO017738	95	77	20
	30	IncFIA/IncFIB	CO010372	97	84,5	22
PA-V-B	2	IncFIB	EU935739	97,1	76	103
	10			98,6	81	73
	90			96,7	75,7	91
PB-V-F	10		KP398867	96,7	78,5	176
	90		AM886293	95,7	70	170
PC-V-B	30		CP015131	95.2	84.7	24

564 **Supplementary Table 4**

565 Unique and shared reads of IncFIA and IncFIB conjugative plasmids

Infant	Sample (Days)	Used reads of	Unique reads	Shared reads	Pairwise identity (%)	Coverage of reference sequences (%)
PA-V-B	2	IncFIA (KP453775)	49,621	51,321	96,4	61,7
	10		35,806	37,347	96,5	61
	90		43,741	43,978	95,8	61.7
PB-V-F	10	IncFIA (CP015239)	9316	12,177	95,6	40
PC-V-B	30	IncFIA (LO017738)	5,320	10,991	95,4	50
PA-V-B	2	IncFIB (EU935739)	15,616	49,067	96,2	30
	10		11,388	37,055	96,5	30,2
	90		12,696	43,510	96,9	29,4
PB-V-F	10	IncFIB (KP398867)	13,582	10,899	95	38,5
PC-V-B	30	IncFIB (CP015131)	7,966	11,400	95.2	52.2

566

PAPER 4

Ravi A, Hagbø M, Sunde M, Ludvigsen J, Muinck E, Diep DB, Foley LS, Collado CM, Martinez PG, L'Abée-Lund MT and Rudi K. (2017). Diversity and characterization of conjugative plasmids in the gut microbiota of a preterm twin pair. (Draft Manuscript)

Diversity and characterization of conjugative plasmids in the gut microbiota of a preterm twin pair

Anuradha Ravi[1], Mari Hagbø[1], Marianne Sunde[2], Jane Ludvigsen[1], Eric de Muinck[3], Dzung B. Diep[1], Steven L. Foley[4], Maria C. Collado[5], Gaspar P. Martinez[5], Trine M. L'Abée-Lund[6], Knut Rudi[1]

[1] Norwegian University of Life Sciences, Chemistry, Biotechnology and Food science department(IKBM), Campus Ås, Ås, Norway;

[2] Norwegian Veterinary Institute, Section for food hygiene, zoonotic diseases and antimicrobial resistance, Oslo, Norway.

[3] Centre for Ecological and Evolutionary Synthesis, Dept. of Biosciences, University of Oslo, Oslo, Norway

[4] National Center for Toxicological Research, U.S. Food and Drug Administration, Division of Microbiology, Jefferson, AR 72079

[5] Institute of Agrochemistry and Food Technology; Spanish National Research Council (IATA-CSIC); Department of Biotechnology; Unit of Lactic Acid Bacteria and Probiotics; Valencia, Spain

[6] Norwegian University of Life Sciences, Department of Food safety and Infection Biology, Campus Adamstuen, Oslo 0454, Norway.

#Address correspondence to Knut Rudi, Department of Chemistry, Biotechnology and Food Science, KBM), Norwegian University of Life Sciences. Christian M Falsensvei 1, N-1432 Ås, Norway. Work Phone +47 67232546. Email for correspondence: knut.rudi@nmbu.no.

Abstract:

Conjugative plasmids represent major reservoirs for multidrug resistance genes. Preterm infants, due to their extended stays at the hospital and enteral feeding harbour a less diverse microbiome with dominating populations of opportunistic pathogens such as *Enterobacteriaceae*. This bacterial family are known to be carriers of mobile genetic elements especially conjugative plasmids and integrons. Therefore, the main aim is to study the diversity of conjugative plasmids through whole genome sequencing and characterise the functional attributes carried by the plasmid. We isolated 74 strains from fecal samples of a preterm twin pair. The strains indicated resistance to wide range of antibiotics and with possible AmpC/ Extended Spectrum β-lactamase production. The *de novo* assembly for 11 of the 17 strains harboured an IncI1 plasmid with all transfer genes and an IncFIB plasmid with potential virulence genes and minimal transfer genes. In addition to this, integrons with multidrug resistance genes were detected. This plasmid and integron was detected in both the twins. Another strain harboured an IncFII/IncFIB conjugative plasmid that contained bacteroicin-production related genes. The strain showed no resistance to antibiotics but produced bacteroicin against a few *E.coli*- related strains. Through qPCR screening, we detected most of the *E.coli* strains consisted of IncI1 and IncFIB plasmid. The rest of the strains were *Enterococcus spp.* Transmission experiments demonstrated the transmission of IncI plasmid and no transmission of IncFIB indicating this plasmid as native for the bacterial strains. The integrons also transferred but by only broth mating indicating presence of smaller plasmids. Overall, the study demonstrated the vast diversity of conjugative plasmids and integrons in a relatively small population of *Enterobacteriaceae*

Introduction

The complexity of the gut microbiota in preterm infants is perturbed by number of biotic and abiotic factors[1]. The microbial composition is less diverse than for infants delivered at term, with higher abundance of *Enterobacteriaceae* (such as *Escherichia coli* & *Klebsiella spp*), *Enterococcaceae* (such as *Enterococcus spp*) & *Clostridia*[2]. The preterm gut microbiota has been linked to the presence of Mobile Genetic Elements (MGEs) within these bacterial families especially within the *Enterobacteriaceae*[3]. However, there is a lack of knowledge on the transmission of these MGEs. Therefore, the main aim of our study is to identify the diversity and functional traits of MGEs residing within the isolates of *Enterobacteriaceae* family isolated from the gut microbiota of a preterm twin pair. The conjugative plasmids are of special interest since these elements are known for their carriage of multi-drug resistance genes and virulence factors[4].

Conjugative plasmids harbours 3 functional groups relating to replication, transfer and stability. Due to the presence of their complex transfer system, conjugative plasmids can replicate and transfer autonomously[5]. Plasmid addiction systems harboured by many plasmids, such as plasmid partitioning, toxin/antitoxin and stability genes ensures the stability of these plasmids within the microbial populations[5]. They are often regarded as parasites of the bacterial cell. In addition to these groups, the plasmids also harbour accessory elements such as integrons that are generally non-mobile but are found harbouring several antibiotic resistance (AR) genes as gene cassettes[6]. They are often found in association with transposons and plasmids.

The classification of conjugative plasmids is through their incompatibility nature[7]. Plasmid incompatibility is the inability of sharing the replication apparatus between two plasmid groups[4,7]. This, in turn destabilizes and degrades the inheritance of one plasmid. Until now, 27 incompatibility groups have been discovered and most of these conjugative plasmids have a narrow host range within the *Gammaproteobacteria*[8].

To study the various types of conjugative plasmids residing the preterm fecal sample, we isolated strains harbouring multidrug resistance properties and studied the different strains and conjugative plasmid variants by whole genome sequencing. We then explored the functional attributes of conjugative plasmids harboured within the strains using culturing and quantitative PCRs.

Materials and methods

Sample description

Faecal samples were collected from a preterm twin pair that was a part of a prospective, single-centre, observational study. The infants were admitted to the University and Polytechnic Hospital La Fe in Valencia, Spain. The twins selected for this study (preterm twin I & twin II) were born preterm (gestational age 30) and weighed 1410 g and 1630 g respectively. The infants stayed at the hospital until sampling even though they did not show any complications or signs of infection. The fecal samples were collected 20 days after birth. The children were born by emergency caesarean section and breast-fed. No antibiotics were given until sampling. The collected fecal samples were frozen and kept at -80°C for later analysis.

Bacterial strain isolation from fecal samples

Mueller Hington (MH) agar (Sigma Aldrich, Madrid, Spain) was used to plate 0.2 g of faecal sample diluted to up to 10^{-4} dilution from the corresponding twins. The plates were incubated at 37°C overnight. Random individual colonies from all species were picked out at 10^{-3} and 10^{-4} dilution plates and streaked onto fresh MH agar plates to get pure cultures. The isolated pure cultures were then stored with 35% glycerol at -80°C until further analysis.

Antibiotic susceptibility test for the isolated strains

The antibiotic susceptibility tests was determined using the standard Kirby-Bauer disk diffusion method[9]. The susceptibilities of the isolates were tested for 6 different antibiotics groups belonging to penicillin (amoxicillin-clavulanic acid 30 µg/disc); cephalosporin (cefpodoxime 10 µg/disc); fluoroquinolones (Ciprofloxacin 5 µg/disk); aminoglycosides (gentamicin 5 µg/disk); trimethoprim 5 µg/disk and sulphomethoxasole 25 µg/disc. The antibiotic susceptibility cartridges were obtained from Oxoid, ThermoFisher Scientific, and Waltham, MA, USA.

All the strains from the corresponding samples were included for the antibiotic susceptibility testing. The bacterial suspensions were adjusted to a turbidity of 0.5 McFarland standard and streaked onto MH agar plates. The antimicrobial discs were placed on the surface of the agar plate and was incubated at 37 °C overnight. The diameter of the inhibition zones surrounding the antimicrobial discs were interpreted according to the EUCAST guidelines[10].

DNA extraction

The strains were revived from the glycerol stock by inoculating 100 µl of the glycerol stock in 5 ml of MH broth (Sigma-Aldrich, Oslo, Norway) and incubated at 37°C overnight. For the DNA isolation, 200 µl of the overnight-incubated broth was mixed with 200 µl of S.T.A.R buffer (Roche, Oslo, Norway). In addition to this, 0.25g of acid-washed glass beads <106 µm (Sigma-Aldrich) was added and the cells were lysed in FastPrep96 (MP Biomedicals, France) at 1800 rpm for 40 seconds for 3 rounds. The lysed cells were centrifuged at 13000 rpm for 5 mins and 50 µl of the supernatant was used for the DNA isolation. An automated protocol based on paramagnetic particles (LGC Genomics, UK) was used for the DNA isolation. In brief, lysed cells were mixed with paramagnetic beads and eluted using a 96-super magnet plate (Alpaqua, Beverly, MA, USA). The concertation of the eluted DNA (1.5 – 30.6 ng/ µl) was determined by fluorometer using a Qubit system (Invitrogen). The DNA was then stored at -40°C until further use.

Gene Quantification

For the presence and quantification of conjugative plasmids, accessory elements and taxonomy of the strains, quantitative PCR was used. Each PCR reaction of 25 µl contained 1X HOT FIREPol PCR mix (Solis Biodyne, Tartu, Estonia); 200 nM of forward and reverse primers: 1 µl of DNA sample and sterile deionised water. The reaction mix was then amplified using a Lightcycler 480 (Roche). The fluorescence data was then uploaded to LinRegPCR program[11] to perform baseline correction and calculate mean PCR efficiency. High Resolution Melting (HRM) curve analysis was used to verify the identity of the PCR products. The primers used in the study is given in Table 1. The thermal cycling conditions for the *E.coli*23S857[12] specific to *E coli* strains, IncFIA & IncFIB[7] variants of the IncF conjugative plasmid, IncI conjugative plasmid[7] and *int*1 gene[13] of the class I integron was an initial denaturation of 95°C for 15 mins and 40 cycles of 95°C for 30 sec, 60°C for 30 sec, as for the *int*1 gene the annealing temperature was 53°C. The PCR products were verified by Sanger sequencing (GATC biotech, Solna, Sweden).

To test the association between the plasmids and accessory elements, spearman correlation was used. The data analyses was performed using MATLAB® R2016a software (The MathWorks Inc., USA)

Whole genome sequencing and analysis

The whole genome sequencing of the strains was carried out using the Illumina sequencing. The metagenome was fragmented, tagged and quantified using the protocol recommended by the Nextera XT sample preparation protocol (Illumina). Sequencing was done in-house using V3 chemistry and 300 bp paired end reads.

The analysis of the whole genome reads was performed using Geneious R10 following the recommended guidelines. The raw reads from the sequencer were merged using an in-house designed merge application and de novo assembling of the reads was performed using Geneious Read mapper[14] (Geneious, Biomatters, New Zealand). The ResFinder[15], PlasmidFinder[16], pMLST[17] & MLST[18] (DTU, Copenhagen, Denmark), are online applications used for the detection of the antimicrobial genes, plasmids, plasmid MultiLocus Sequence Typing (pMLST) & MultiLocus Sequence Typing (MLST) of the strains in the *de novo* assembled contigs based on the NCBI database. RAST (Rapid Annotation using Subsystem Technology) server based on the SEED system was used to identify gene within the *de novo* assembled contigs[19]. BLAST Ring Image Generator (BRIG) was used to display the regions of the plasmids that were similar to each other across the strains[20]. Bactibase, a repository for bacteriocin related genes in both gram-positive and gram-negative bacteria was used to check for potential bacteriocin-related genes in the strains[21].

Bacteriocin production assay

To screen for bacteriocin producers in our strains, the *E.coli* MH1[22,23] was initially used as the indicator. The first round of screening was performed using soft agar overlay assay[22]. Briefly, the overnight grown indicator strains in Luria Bertani (LB) agar (ThermoFisher Scientific) were 100-fold diluted in 5 ml of LB soft agar (0.8%) that was plated out as lawn cultures. The strains were spotted on the indicator lawn cultures and incubated at 37 °C for 18-24 h. The inhibition was defined as clear zones around the producing bacterial colony.

For the strains that showed bacteriocin production activity, the crude protein precipitate of the strains were used to test the activity. The protein precipitate was extracted by adding 50% saturation of ammonium sulphate in a 50 ml LB broth with overnight cultures. The flasks were incubated without shaking at 4°C for 24 h before spinning down the supernatant at 25000 rpm

for 45 mins. The protein precipitate containing the crude bacteriocin was then re-suspended in 500 µl of Phosphate Buffered Saline (PBS). In case of any re-suspended cells in the pellet, the protein pellet was heat-treated to 70 °C for 5 mins. This protein precipitate was used to test the bacteriocin activity similar to the agar diffusion method.

Sensitivity of protease was tested by adding 2 µl of proteinase K (Sigma-Aldrich) at 20 µg/ml near the spotted cells. Sensitivity was seen when the indicator was not affected in the regions close to the proteinase K application.

Conjugation experiment

Selected strains were used for the conjugation assay. The recipient strains used in this study was an E.coli DH5α rif^{+}, which was resistant to 32 rifampicin. Solid agar mating and liquid mating were performed for this experiment. For the liquid mating, 500µl of the recipient and 10 µl of the donor were mixed in 4 ml of LB broth and incubated at 37 °C for 4 h and 24 h. A 10^{-1} dilution of the mix was then streaked out on MH agar plates containing 32 mg/ml of rifampicin. For the solid agar mating, 1 µl loop of donor and recipient colonies are mixed together. The mixed colonies are diluted up to 10^{-2} using NaCl and streaked together on a MH agar plate with 32 mg/ml of rifampicin. Disks with antimicrobial agents corresponding to the resistance profiles of the donor strains were placed onto the surface of the agar plates, followed by incubation at 37 °C for 4 h and 24 h.

Presumptive transconjugants growing within the inhibition zones on the rifampicin containing plates were subcultured and further subjected to DNA isolation and gene quantification.

Results and Discussion

Bacterial isolates and antimicrobial susceptible tests

In total, 22 strains originating from preterm twin I and 52 isolates from preterm twin II were isolated. In total, from the qPCR screening, 44 (Twin I- 9; II- 36) of the 74 strains were *E.coli* positive, the rest of the strains belonged to *Enterococcus spp*. The antibiotic susceptibility testing was carried out for all the isolates where 71 strains showed resistance to at least one antimicrobial agent (Figure 1). Resistance to cephalosporin where most prevalent as almost 93% of the isolates were resistant to cefpodoxime. In addition, 17 isolates were resistant to gentamicin, Furthermore, all the isolates resistant to amoxicillin-clavulanic acid (n=19) was

also resistant to cefpodoxime. The resistance patterns suggest extended spectrum β-lactamase (ESBL) production. The ESBLs are enzymes that can hydrolyse most of the penicillins such as amoxicillin and cephalosporins such as third-generation cefpodoxime[24,25] . The ESBL enzymes encoding genes are usually harboured within plasmids. Clavulanic acid is commonly used as an inhibitor for β-lactamases and is generally combined with penicillin for targeting ESBL-borne infections[26]. However, the resistance to both clavulanic acid and cefpodoxime indicates an AmpC-type cephalosporinases where the AmpC-type enzymes are poorly inhibited by the classical ESBL-inhibitors such as clavulanic acid[27]. However, for the confirmation of AmpC-ESBL production, a screening breakpoint of >1mg/L is recommended for cefotaxime, ceftriaxone and ceftazidime[28], in addition to testing against other substances.. Since cefpodoxime is the most sensitive individual indicator for detection of ESBL production, further confirmatory testing with the above-mentioned compounds needs to be done[24,25].

In addition to this, gentamicin is commonly given in combination with β-lactam antibiotics. This combination is commonly administered for infants since it provides a synergistic effect against the most commonly encountered pathogens[29]. From a survey on antimicrobial resistance in 265 Spanish hospitals and 19,081 *E.coli* isolates from 2001- 2010, it was demonstrated that a significant strong correlation towards the rate of usage of fluoroquinolones, and third-generation cephalosporin in *E.coli* strains to the resistance against both the antibiotics[30]. Therefore the presence of such spectrum of resistance genes to several antibiotics could potentially be due to the increased prevalence of MGEs. The resistance to sulphamethoxazole could be due to the presence of integrons since integrons consists of a 3' conserved segment that confers resistance to sulphonamides (*sul*1) and ammonium quaternary compounds (*qac*EΔ1)[31].

Whole genome sequencing

Strains showing diverse AR profiles that are from *E.coli* and *Enterococcus spp* were chosen for whole genome sequencing. Seventeen strains (Twin I-6; Twin II-11) were selected from the strain collection (Supplementary Table 1). The strains are marked alongside their twin pair. For instance, A-II means strain A of twin II and B-I means strains B of twin I. On average, 827,634 reads were generated per genome with read length from 35 bp- 301 bp. The reads were merged

and built using Geneious Read mapper, obtaining an average of 73 contigs over 1,000 bp in length. The average N50 length was 210,522 bp with at least 10 contigs >=N50 length/ sample.

The contigs from all the strains were submitted to PlasmidFinder for the detection of conjugative plasmids in the genome sequences. Twelve of the 17 strains showed the presence of IncF and IncI conjugative plasmids (Twin I- 3; Twin II- 9). The strain J-I harboured IncFII & IncFIB conjugative plasmids. The rest of the strains contained IncFIB and IncI conjugative plasmid. Even though the strain J did not show resistance to any antibiotics through disk-diffusion, ResFinder indicated resistance genes to aminoglycosides from the genome data. As for the other 11 strains that harboured conjugative plasmids, ResFinder indicated resistance genes to aminoglycosides, tetracycline, sulphonamides, trimethoprim, and β-lactam and macrolide resistance. In addition to this, the MLST indicated that the strains harbouring the conjugative plasmids belonged to 3 different *E.coli* sequence types (ST) where strain B-II and strain E-I belonged to ST345, Strain J-II belonged to ST34 and the rest of the strains belong to ST636. All these sequence types are known ESBL strains with varied AR genes and MGEs[32-34].

Assembly of conjugative plasmids

IncFIB conjugative plasmid

In the 11 strains that showed multidrug resistance properties, we detected contigs-related to IncFIB. The *de novo* assembled IncFIB plasmid harboured only TraX, TraI and FinO as transfer genes. None of the other IncF transfer genes were detected. The absence of transfer genes suggests IncF plasmid could be a non-mobile plasmid. However, the plasmid harboured special properties for achieving persistence in complex environments[35]. Copy number control (cop genes)[36], active partitioning systems (parA/parB)[37] and post segregationally killing (TA systems such as ccdA/ccdB)[38]. These are the main gene families relating to persistence in microbial communities. The presence of such genes in a non-mobile plasmid indicates long term persistence of these elements in the gut microbiota. In addition to the presence of addiction systems, the assembled IncF plasmids harboured potential virulence genes such as *IroBCDEN*[39] and aerobactin biosynthesis gene family[40]. The former and later gene families are associated with extraintestinal pathogenic *E. coli* (ExPEC). Both the gene families are associated with the ability of microbes to grow in iron-limiting conditions that could potentially promote systemic infections[41,42]. The former and latter families are commonly associated with plasmid genes and

are a part of the mobile virulence gene family[43,44]. In addition to this, the strains also harboured streptomycin resistance genes (*str*A&B) and in accordance to that, the strains were resistant to streptomycin by disk diffusion method.

In order to look for similar IncF plasmids across the different strains, the assembled plasmid from one of the samples (strain K-II) was compared together with the other *de novo* assembled IncF plasmids (Figure 2). This particular IncFIB assembled plasmid was detected in 11 of the 12 strains.

On the other hand, the strain J-II that did not harbour the same IncFIB conjugative plasmid harboured an IncFII/IncFIB conjugative plasmid. This particular plasmid contained all the transfer and replication regulatory genes needed to be mobile. In addition to this, the plasmid harboured bacteriocin producing genes (colicin I), potential virulence factors (aerobactin siderophore) & haemolysin expressing genes (*hae/Yih*), copy number control genes (*cop*B) and TA systems (*ccd*A&B) (Figure 3). This particular plasmid was native to preterm twin B and was not found in any other strains.

IncI conjugative plasmid

For the strains that we detected a potential non-mobile IncFIB plasmid, we also detected IncI plasmids in the same bacterial strain (n=11). The contigs related to the IncI conjugative plasmid, on the other hand harboured all the transfer (TraA- TraY) and the pilus genes (Pil genes). This *de novo* assembled IncI plasmid consisted of a complex transfer system extending to over 50kb with two types of conjugative pilus regions. In general, these plasmids are known for their extensive transfer and pilus associated genes[45]. Many diverse IncI plasmids have been detected and sequence-typed[17], where strains from this dataset belong to IncI1 conjugative plasmid. In addition to the presence of the transfer system, the IncI plasmids also harboured a plasmid SOS system (psiA-psiB family) and the replication initiation genes. The assembled IncI plasmid from one strain (strain K-II) was compared to the assembled plasmids from the other strains and we observed the same type of assembled plasmid across the 11 of the 12 strains, the exception being Strain J-II (Figure 4).

Integrons

In 11 of the 12 strains, integrons were detected in the metagenome sequences. The consensus integron showed 98% similarity between the integrons in the other samples. The integrons harboured resistance gene cassettes to Streptomycin (*aadA*1), spectinomycin (*spc*) and chloramphenicol resistance (*CmR*) along with resistance to sulphonamides (DHPS) (Figure 5).

Overall, in 11 of the 12 strains, the conjugative plasmids with IncI/IncFIB was detected in both the preterm twins. Except the strain J-II, the conjugative plasmids indicated sharing of the mobilome. With regards to coverage of the conjugative plasmids, the IncI plasmids indicated twice as much coverage as for the IncFIB indicating IncI plasmids with 2 copies higher than IncFIB. The hospital environment are major breeding grounds for the adaptation and prevalence of multidrug resistance genes[46,47]. The IncF and IncI plasmids are also known for the carriage of several types of β-lactamases and other resistance genes[8,46]. Therefore, the hospital-acquired mobilome is a major concern especially in the neonatal intensive care unit of the hospital. In most of the strains, we detected conjugational complex involving IncI and IncFIB conjugative plasmid. This conjugational complex is created when some of the genes related to replication such as mob and FinO genes are shared between the plamsids[48]. However, information on this mode of replication is limited in regards to literature and reproducibility.

Prevalence of IncI and IncF conjugative plasmids

In order to detect similar conjugative plasmids in our strain collection, distinct regions of the IncI and IncFIB conjugative plasmids, integrase gene of the integrons were targeted and screened. The *E.coli* strains were positive to IncFIB. In addition, 3 *Enterococcus spp* were positive to IncFIB. For IncI and integron, 44 of the 74 strains were positive including one *Enterococcus spp* strain. The distribution of conjugative plasmids and integrons across the strain collection indicates the presence of similar conjugative plasmids across the gram-positive and gram-negative strains. The plasmids were also shared between the preterm infants. In general, most of the conjugative plasmids that have been identified in gram positive bacteria are from *Streptococci* and *Enterococci spp.*[49,50]. The IncFIB and IncI were prevalent together in 44 of 74 strains and in that 43 strains were *E.coli*-positive.

Bacteriocin activity assay

The bacteriocin genes harboured in the IncFII/IncFIB conjugative plasmid was unique to strain J-II plasmid. The crude precipitate showed bacteriocin activity against *E.coli* MH1 and three of the 10 commensal *E.coli* strains. This bacteriocin activity was inhibited by the presence of proteinase K. Bacteriocins are antimicrobial proteins that have bactericidal properties[51]. They have a narrow host range and normally target species that are closely related to the producer. They are shown to be important mediators for intra- and interspecies interactions and for maintaining the microbial diversity. Therefore, the role of these bacteriocin-producing strains in the development of the gut microbiota lack detailed assessment. However, the bacteriocin production by conjugative plasmids, in general has shown to augment niche competition whereas the bacteriocin producers outcompetes the non producers[52]. We believe, our strains produce a variant of colicins since we detect the genes related for colicin production and transportation. These bacteriocins are known to inhibit the growth of related *E.coli* strains[53]. The *TonB* gene that is detected in our metagenome, is a machinery involved by colicin to enter into a cell. However, further work on the characterization of colicin produced by the strains needs to be completed.

Transmission assay

The major mechanism for the transfer of mobilizable genes between different bacteria is through conjugation. Therefore to check the transferability of the conjugative plasmids in our collection, 5 representative strains were chosen (strains C-II, D-II, G-II, I-I, & L-II) for conjugation experiment. After conjugation, the DNA extracted from the transconjugants were screened for IncF, IncI and integron (Table 2). All the transconjugants showed the transfer of IncI plasmids whether through agar or broth mating. On the other hand, six of the eight transconjugants by broth mating showed the transfer of the integron indicating that the integron is not found within the IncI and IncFIB conjugative plasmid. The co-transfer of the integrons only by broth and not my liquid media could be due to the presence of smaller plasmids that carries the integron. In addition to this, the IncFIB was not mobilizable due to the lack of transfer genes. The co-transfer of Inc plasmids has been demonstrated in other studies, however, in our study the IncF conjugative plasmids are native to the bacterial strains.

Conclusion

The genetic characterization of a relatively small population of *E.coli* from fecal samples of preterm infants displays a vast diversity of conjugative plasmids and AR genes. Such studies on

plasmid diversity are impeded due to the fundamental knowledge on the diversity of plasmids and MGEs in the gut microbiota. The presence of such conjugative with wide functional attributes in the gut microbiota is alarming and is partly due to the exposure factors. The biotic factors that can be related to maternal transmission and the abiotic factors relate to the hospital environment. In our dataset, we detected two different conjugative plasmids of particular interest. The IncI/IncFIB plasmids have an evolutionary relationship where they share similarities between the transfer genes and both have a narrow host range compared to the other Inc plasmds[54]. However, in one other strain, we detected a completely different plasmid with no genotypic sharing of the genes and showed bacteriocin production. The bacteriocins are known to mediate the survival of pathogenic bacteria and are observed to eliminate multidrug resistance, therefore this could be a part of a safe-keeping strain group useful for protecting the microbial communities. However, further work on the activity of these strains against other gut microbes should be done. Therefore, conjugative plasmids are diverse elements that can potentially change the phenotypic attributes of a bacterial community. Therefore understanding their mobility and diversity in complex environments can help to gain better overview on spread of multidrug resistance genes.

Acknowledgements

Einar Nilssen (Helse Møre og Romsdal/University Hospital of North Norway, Tromsø) is acknowledged for donation of the *E. coli* recipient strain used in the conjugation experiments.

Future work

We observed resistance to several different antibiotics, however confirmatory test for the production of different β-lactamases and the type of ESBL production needs to be done. We assembled two different plasmids (IncI and IncF) that showed varied functional characteristics. We also show the presence of two different strains carrying the same conjugative plasmid. However, this was done only for the whole genome sequences samples. Diagnostic MLST screening for different strains of *E.coli* in our strain collection dataset should be done. Not many studies in the late 2000s have focussed on the biology of conjugative plasmids and the co-replication of mobile and non-mobile plasmids. Therefore, the co-replication of the two different plasmids in our bacterial strains will be interesting. Even though we demonstrated the conjugation of the IncI plasmids, the transmission efficiency of the conjugative plasmids need to be calculated. Further to this, the transmission was shown only *in vitro*, with laboratory controlled conditions and between the same species. Therefore, future work will be on transmission of the plasmids between different species and using mice models to demonstrate the persistence and *in vivo* transmission of these plasmids. Finally, we believe integrons are harboured in small plasmids which were not detected by the whole genome sequencing, therefore sequencing of the transconjugants will be vital to identify the genetic background of the integrons.

Figures

Figure 1

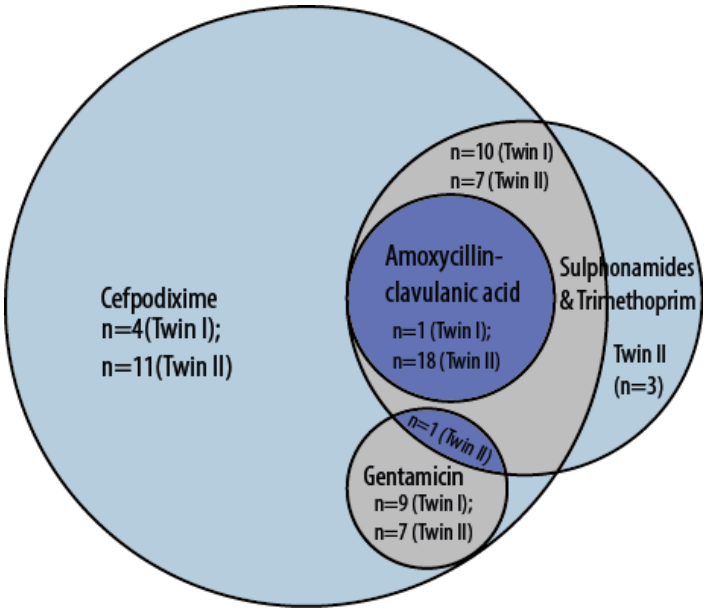
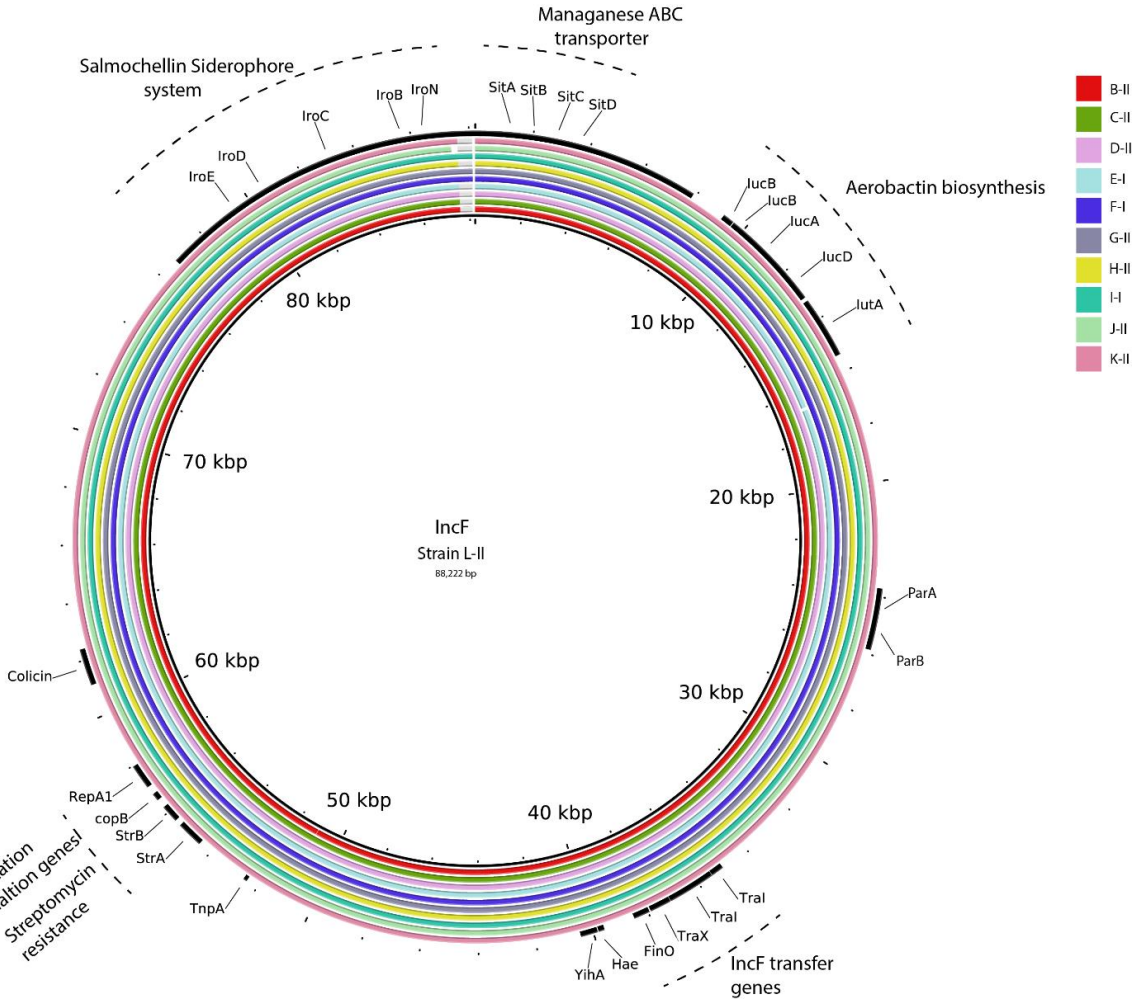


Figure 1: Strains isolated from the preterm twins showing resistance towards the different antibiotic groups.

Figure 2



393

394 **Figure 2:** Blast Ring Image Generator (BRIG) alignment of de novo assembled IncFIB

395 plasmid. *Par*A/*Par*B- plasmid portioning system; *Tnp*A- transposase; *hae*/*yig* gene, haemolysin

396 expressing genes.

397

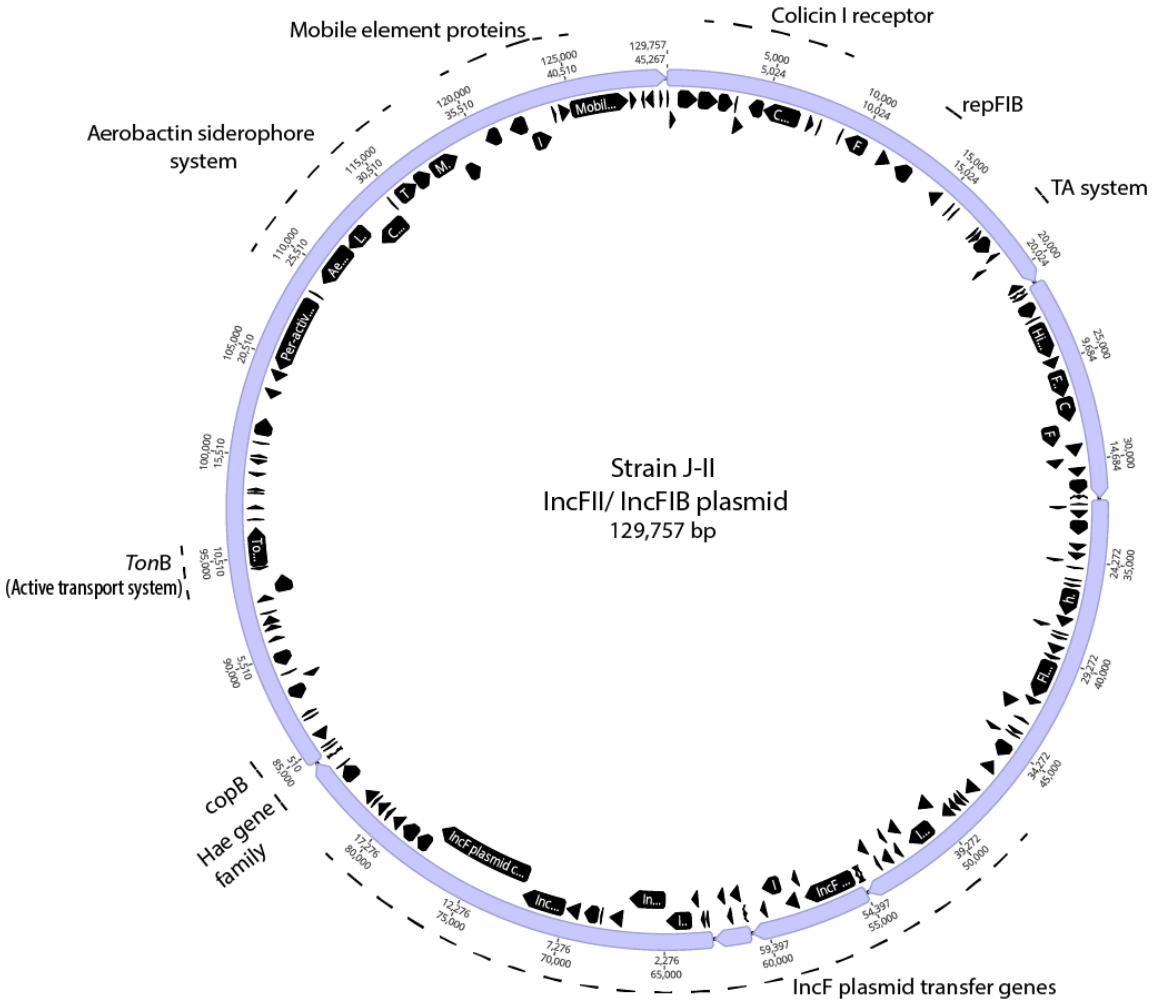
398

399

400

401

402 **Figure 3**



403

404 **Figure 3:** De novo assembly of IncFII/IncFIB plasmid of strain J-II. *Cop*B- copy number
405 control genes, TA system- Toxin/antitoxin system (ccdA&ccdB); repFIB- replication
406 regulatory gene of IncFIB.

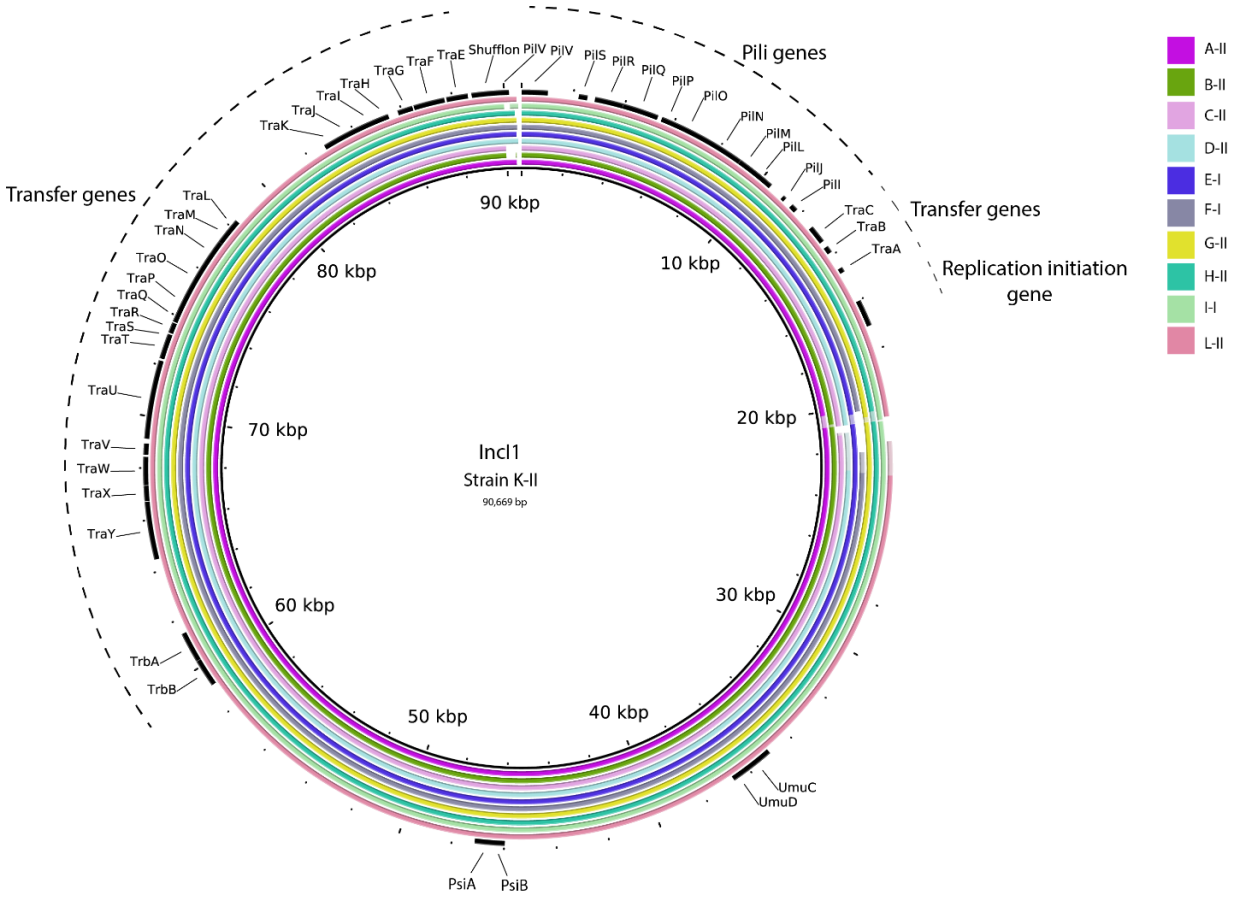
407

408

409

410

411 **Figure 4**



412

413 **Figure 4:** BRIG alignment of de novo assembled IncI plasmid. *Psi*A/*Psi*B genes- Plasmid SOS
414 inhibition system; *Umu*C/*Umu*D- UV mutagenesis and repair system.

415

416

417

418

419

420

421

422

423

Figure 5

Figure 5: *de novo* assembled integron from 11 strains. *Int*1- integrase gene; *aadA*1-streptomycin 3"-O-adenylyltranferase; CmR- chloramphenicol resistance; *spc*- Spectinomycin 9- adenylytransferase; DHPS- dihydropteroate synthase

Tables

Table 1

Primers used in the study

Gene		Sequence
EC23S857	F	GGTAGAGCACTGTTTtGGCA
	R	TGTCTCCCGTGATAACtTTCTC
IncFIA	F	CCATGCTGGTTCTAGAGAAGGTG
	R	GTATATCCTTACTGGCTTCCGCAG
IncFIB	F	GGAGTTCTGACACACGATTTTCTG
	R	CTCCCGTCGCTTCAGGGCATT
IncI	F	CGAAAGCCGGACGGCAGAA
	R	TCGTCGTTCCGCCAAGTTCGT
*Int*1	F	ACGAGCGCAAGGTTTCGGT
	R	GAAAGGTCTGGTCATACATG

436 **Table 2: Screening of the transconjugants**

Strain	Mating	IncI1	Integron
C-II	Agar	x	
D-II		x	
I-I		x	
L-II		x	
C-II colony 1	Broth	x	x
C-II colony 2		x	
D-II colony 1		x	x
D-II colony 2		x	
I-I colony 1		x	x
I-I colony 2		x	x
L-II		x	x
G-II		x	x

437

438

Supplementary table 1: Strains selected for whole genome sequencing

		Antibiotics susceptibility				
Sample name	Taxonomy	Amoxicillin-Clavulanic acid	Cefodoxime	Gentamicin	Trimethoprim	Sulfonamides
A-II	*E.coli*	x	x		x	x
B-II	*E.coli*		x	x	x	x
C-II	*E.coli*	x	x		x	x
D-II	*E.coli*		x		x	x
E-I	*E.coli*		x	x		
F-I	*E.coli*	x	x		x	x
G-II	*E.coli*				x	x
H-II	*E.coli*		x			
I-I	*E.coli*		x		x	x
J-II	*E.coli*					
K-II	*E.coli*	x	x		x	x
L-II	*E.coli*		x		x	x
M-I	*Enterococcus spp*		x	x		
N-I	*Enterococcus spp*		x			
O-II	*Enterococcus spp*		x			
P-II	*Enterococcus spp*		x	x		
Q-I	*Enterococcus spp*		x	x		

References

1. Warner BB, Tarr PI. Necrotizing enterocolitis and preterm infant gut bacteria. Seminars in Fetal and Neonatal Medicine 2016;21:394-9.
2. Unger S, Stintzi A, Shah P, Mack D, O'Connor DL. Gut microbiota of the very-low-birth-weight infant. Pediatr Res 2015;77:205-13.
3. Morowitz MJ, Denef VJ, Costello EK, Thomas BC, Poroyko V. Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. Proc Natl Acad Sci USA 2011;108.
4. Carattoli A. Evolution of Plasmids and Evolution of Virulence and Antibiotic-Resistance Plasmids. Evolutionary Biology of Bacterial and Fungal Pathogens: American Society of Microbiology; 2008.
5. Thomas CM. Paradigms of plasmid organization. Molecular microbiology 2000;37:485-91.
6. Kovalevskaya NP. Mobile Gene Cassettes and Integrons. Molecular Biology 2002;36:196-201.
7. Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ. Identification of plasmids by PCR-based replicon typing. Journal of microbiological methods 2005;63:219-28.
8. Carattoli A. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother 2009;53:2227-38.
9. Bauer AW, Kirby WM, Sherris JC, Turck M. Antibiotic susceptibility testing by a standardized single disk method. Am J Clin Pathol 1966;45:493-6.
10. Matuschek E, Brown DFJ, Kahlmeter G. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clinical Microbiology and Infection 2014;20:O255-O66.
11. Ruijter JM, Pfaffl MW, Zhao S, et al. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications. Methods 2013;59.
12. Chern EC, Siefring S, Paar J, Doolittle M, Haugland RA. Comparison of quantitative PCR assays for Escherichia coli targeting ribosomal RNA and single copy genes. Lett Appl Microbiol 2011;52:298-306.
13. L'Abee-Lund TM, Sorum H. Class 1 integrons mediate antibiotic resistance in the fish pathogen Aeromonas salmonicida worldwide. Microb Drug Resist 2001;7:263-72.

14. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 2012;28:1647-9.
15. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy 2012;67:2640-4.
16. Carattoli A, Zankari E, García-Fernández A, et al. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. Antimicrobial Agents and Chemotherapy 2014;58:3895-903.
17. Carattoli A, Zankari E, Garcia-Fernandez A, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 2014;58:3895-903.
18. Larsen MV, Cosentino S, Rasmussen S, et al. Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. Journal of Clinical Microbiology 2012;50:1355-61.
19. Aziz RK, Bartels D, Best AA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 2008;9:1-15.
20. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 2011;12:402-.
21. Hammami R, Zouhir A, Ben Hamida J, Fliss I. BACTIBASE: a new web-accessible database for bacteriocin characterization. BMC Microbiology 2007;7:89.
22. Håvarstein LS, Holo H, Nes IF. The leader peptide of colicin V shares consensus sequences with leader peptides that are common among peptide bacteriocins produced by Gram-positive bacteria. Microbiology 1994;140:2383-9.
23. Casadaban MJ, Cohen SN. Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. Journal of Molecular Biology 1980;138:179-207.
24. Oliver A, Weigel LM, Rasheed JK, McGowan Jr JE, Jr., Raney P, Tenover FC. Mechanisms of decreased susceptibility to cefpodoxime in Escherichia coli. Antimicrob Agents Chemother 2002;46:3829-36.
25. Hope R, Potz NA, Warner M, Fagan EJ, Arnold E, Livermore DM. Efficacy of practised screening methods for detection of cephalosporin-resistant Enterobacteriaceae. J Antimicrob Chemother 2007;59:110-3.

26. Soares GMS, Figueiredo LC, Faveri M, Cortelli SC, Duarte PM, Feres M. Mechanisms of action of systemic antibiotics used in periodontal treatment and mechanisms of bacterial resistance to these drugs. Journal of Applied Oral Science 2012;20:295-304.
27. Jacoby GA. AmpC β-Lactamases. Clinical Microbiology Reviews 2009;22:161-82.
28. Testing ECoAS. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. . 7.0 ed2017.
29. Chattopadhyay B. Newborns and gentamicin—how much and how often? Journal of Antimicrobial Chemotherapy 2002;49:13-6.
30. Asensio A, Alvarez-Espejo T, Fernandez-Crehuet J, et al. Trends in yearly prevalence of third-generation cephalosporin and fluoroquinolone resistant Enterobacteriaceae infections and antimicrobial use in Spanish hospitals, Spain, 1999 to 2010. Euro Surveill 2011;16.
31. Ravi A, Avershina E, Ludvigsen J, #039, Abée-Lund T, Rudi K. Integrons in the Intestinal Microbiota as Reservoirs for Transmission of Antibiotic Resistance Genes. Pathogens 2014;3:238-48.
32. Zankari E, Hasman H, Kaas RS, et al. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. Journal of Antimicrobial Chemotherapy 2013;68:771-7.
33. Ma L, Siu LK, Lin J-C, et al. Updated molecular epidemiology of carbapenem-non-susceptible Escherichia coli in Taiwan: first identification of KPC-2 or NDM-1-producing E. coliin Taiwan. BMC Infectious Diseases 2013;13:599.
34. Karah N, Dwibedi CK, Sjöström K, et al. Novel Aminoglycoside Resistance Transposons and Transposon-Derived Circular Forms Detected in Carbapenem-Resistant Acinetobacter baumannii Clinical Isolates. Antimicrobial Agents and Chemotherapy 2016;60:1801-18.
35. Bahl MI, Hansen LH, Sørensen SJ. Persistence Mechanisms of Conjugative Plasmids. In: Gogarten MB, Gogarten JP, Olendzenski LC, eds. Horizontal Gene Transfer: Genomes in Flux. Totowa, NJ: Humana Press; 2009:73-102.
36. Summers D. Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. Molecular microbiology 1998;29:1137-45.
37. Ebersbach G, Gerdes K. Plasmid segregation mechanisms. Annu Rev Genet 2005;39:453-79.

38. Hayes F. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science 2003;301:1496-9.
39. Caza M, Lépine F, Milot S, Dozois CM. Specific Roles of the iroBCDEN Genes in Virulence of an Avian Pathogenic Escherichia coli O78 Strain and in Production of Salmochelins. Infection and Immunity 2008;76:3539-49.
40. Ling J, Pan H, Gao Q, et al. Aerobactin Synthesis Genes iucA and iucC Contribute to the Pathogenicity of Avian Pathogenic Escherichia coli O2 Strain E058. PloS one 2013;8:e57794.
41. Bullen JJ. The Significance of Iron in Infection. Reviews of Infectious Diseases 1981;3:1127-38.
42. Payne SM, Finkelstein RA. The critical role of iron in host-bacterial interactions. Journal of Clinical Investigation 1978;61:1428-40.
43. Sorsa LJ, Dufke S, Heesemann J, Schubert S. Characterization of an iroBCDEN Gene Cluster on a Transmissible Plasmid of Uropathogenic Escherichia coli: Evidence for Horizontal Transfer of a Chromosomal Virulence Factor. Infection and Immunity 2003;71:3285-93.
45. Thomas CM. The horizontal gene transfer pool. Bacterial plasmids and gene spread. School of Biological Sciences, University of Birmingham, UK Harwood academic publishers; 2005.
46. Naas T, Cuzon G, Robinson AL, et al. Neonatal infections with multidrug-resistant ESBL-producing E. cloacae and K. pneumoniae in Neonatal Units of two different Hospitals in Antananarivo, Madagascar. BMC Infectious Diseases 2016;16:275.
47. Novais Â, Cantón R, Moreira R, Peixe L, Baquero F, Coque TM. Emergence and Dissemination of Enterobacteriaceae Isolates Producing CTX-M-1-Like Enzymes in Spain Are Associated with IncFII (CTX-M-15) and Broad-Host-Range (CTX-M-1, -3, and -32) Plasmids. Antimicrobial Agents and Chemotherapy 2007;51:796-9.
48. Morzejko E, Panek E, Cebrat S. Conjugational cotransfer of IncFI and IncI conjugative plasmids forming aggregate in a pathogenic Citrobacter freundii strain. FEMS Microbiology Letters 1990;68:7-11.
49. Schaberg DR, Zervos MJ. Intergeneric and interspecies gene exchange in gram-positive cocci. Antimicrobial Agents and Chemotherapy 1986;30:817-22.

50. Clewell DB. Movable genetic elements and antibiotic resistance in enterococci. Eur J Clin Microbiol Infect Dis 1990;9:90-102.
51. Cascales E, Buchanan SK, Duché D, et al. Colicin Biology. Microbiology and Molecular Biology Reviews 2007;71:158-229.
52. Kommineni S, Bretl DJ, Lam V, et al. Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. Nature 2015;526:719-22.
53. Budič M, Rijavec M, Petkovšek Ž, Žgur-Bertok D. Escherichia coli Bacteriocins: Antimicrobial Efficacy and Prevalence among Isolates from Patients with Bacteraemia. PloS one 2011;6:e28769.
54. Suzuki H, Yano H, Brown CJ, Top EM. Predicting Plasmid Promiscuity Based on Genomic Signature. Journal of Bacteriology 2010;192:6045-55.