UiT The Arctic University of Norway

Faculty of Health Sciences

Department of Medical Biology

Duration and dynamics of *Klebsiella pneumoniae* species complex gut carriage in a community-based adult population cohort

Wasifa Kabir

Master's thesis in Biomedicine MBI-3911 May 2023

Table of Contents

List of Tables

List of Figures

Acknowledgements

"Great things are not done by one person. They are done by a team of people."
-Steve Jobs

This thesis was carried out in the Research Group of Host-Microbe Interactions, Faculty of Medical Biology, Uit, The Arctic University of Norway. I am immensely grateful to my three musketeers, my three supervisors, Professor Arnfinn Sundsfjord, Kenneth Lindstedt, and Dorota Julia Buczek.

First and foremost, I would like to thank my main supervisor, Professor Arnfinn Sundsfjord, for giving me the opportunity to be part of his team and introducing me to the wonderful people of the KLEB-GAP project. Thanks for giving me exposure to different scientific conferences and meetings that clearly helped me to broaden my knowledge and help me realise the bigger picture of the scientific world. Thanks for your guidance, patience, feedback, and valuable time. I have learnt from you that we always have the time to say 'Hi' no matter how busy we are. You are the mentor anyone would ever have wished for.

I also would like to thank my co-supervisor, Dorota Julia Buczek, for teaching me coding, scripting, and troubleshooting with great patience, and answering all my questions, and emails even at midnight. Thanks for helping me grow as a bioinformatician, for advising me on my future career, and for making me feel the way that I never hesitated to ask any questions.

Words will not be enough to thank my co-supervisor, Kenneth Lindstedt for teaching all the labwork and bioinformatic tools and for being patient in explaining everything repeatedly. Thanks for keeping me up-to-date with the new bioinformatic tools, for thorough feedback on the thesis, and for sharing your knowledge and expertise. You always had an open door for me despite your busy schedule. This thesis would not be possible without your guidance, patience, and importantly your Spotify playlist.

I also thank all the members of the Host-Microbe Interaction (HMI) group at UiT for being so welcoming. It is the people in the group who made the working environment lively. A big thanks goes to Jonathan Hira who suggested doing my thesis with Professor Arnfinn in the first place, knowing my interest in both bioinformatics and lab work. I would like to give special thanks to Ahmed Berghet who helped me from time to time with the R script. Also, I would like to express my gratitude to Ørjan Samuelsen for his valuable feedback that improved the work.

I will take this opportunity to thank my fellow students who thoroughly encouraged me throughout the thesis. We had ups and downs, some were struggling with writing, and some had a hard time in the lab, but we supported each other throughout the period. A warm thanks goes to Shamima for providing me with company during my late-night university stay.

Finally, I would like to thank my parents who flew from Bangladesh to Tromsø just so I could focus on research while they entertained my little one, my son Rihanur Rahman. Thanks to my sunshine Rihanur, for soothing me with laughter, cuddling, and giggling after the long tiresome days and nights of writing, and for understanding that your mamma is going to the university to read ABCD. Last but not least, my dear husband Ataur Rahman, thanks for being always there when I needed it the most and even when I did not. Thanks for being the biggest critic of my work as well.

Abbreviations

AMR: Antimicrobial resistance

BSI: Blood stream infection

CAI: Community-acquired infection

CRKp: Carbapenem-resistant *Klebsiella pneumoniae*

DNA: Deoxyribonucleic acid

ESBL: Extended spectrum β-lactamases

GIT: Gastrointestinal tract

HAI: Hospital-acquired infections

hvKp: Hypervirulent *Klebsiella pneumoniae*

KpSC: *Klebsiella pneumoniae* species complex

MDR: Multidrug resistance

MLST: Multilocus sequence typing

NCBI: National Centre for Biotechnology Information

qPCR: Quantitative polymerase chain reaction

RCA: Rolling circle amplification

ROS: Reactive oxygen species

SCAI: Simmons citrate agar with 1% inositol

ST: Sequence type

StrainGE: Strain Genome Explorer

UTI: Urinary tract infections

WGS: Whole genome sequencing

WHO: World Health Organization

WMS: Whole Metagenome Sequencing

ZKIR: *zur-khe* intergenic region

Abstract

Klebsiella pneumoniae species complex (KpSC) is a group of closely related bacterial species including *Klebsiella pneumoniae sensu stricto* known as a major human opportunistic pathogen associated with healthcare. KpSC gut carriage is a major risk factor for clinical infections and elderly people (≥60y) are at the risk of KpSC gut colonisation. This master project investigated the presence, duration and dynamics of KpSC gut carriage among a community-based general adult population (n=108, 62% female; ≥47 y), recruited from the previous Tromsø7 Cross-sectional KpSC Carriage Study, during six months (September 2021 to March 2022). Monthly faecal samples were screened for KpSC by ZKIR-qPCR, and whole metagenome sequencing (WMS) was used for microbiome analysis and ST typing by StrainGE. qPCR confirmed the majority of the population being transient carriers (76.5%), a minority of the population were non-carriers (9.1%), and few were chronic carriers (14.1%). Alpha diversity analysis did not show any significant difference among the three KpSC carriage groups. Alpha diversity significantly decreased with increased KpSC relative abundance in the age group below 60y. The chronic carriers were significantly older than the transient carriers. Evaluation of StrainGE tool for analysing KpSC phylogenetic relationships from metagenome samples revealed that StrainGE accurately identified two out of the three phylogeny relations. Furthermore, ST analysis by StrainGE identified a highly diverse KpSC population with 13 STs and two phylogroups detected in 41 metagenome samples from seven participants (one transient and six chronic carriers). Three out of the six chronic carriers retained a single ST in all six months. This study indicates that there might be an association of gut colonisation with KpSC with age and gut microbiome. The accuracy of StrainGE suggests that it might be a useful tool in epidemiological surveillance and outbreak investigation using metagenomic samples. The diversity of KpSC in gut colonisation highlights the niche-specific adaptive capacity of the KpSC members and the challenges for the development of decolonisation strategies.

Keywords: *Klebsiella pneumoniae* species complex, WMS, sequence type, gut carriage, longitudinal study, antimicrobial resistance, StrainGE

1 Introduction

1.1 Antimicrobial resistance (AMR): An urgent global health threat

Antimicrobial resistance (AMR) in bacteria is one of the growing threats to public health due to the emergence and spread of drug resistance untreatable making common infections untreatable with the currently available antibiotics [1]. The World Health Organization (WHO) declared AMR a "global public health concern" in 2001 [2]. More than 33,000 deaths and 87,000 disability-adjusted-life-years (DALYs) directly attributed to AMR infections were estimated in the EU for the year 2015 [3]. Recently the global burden of AMR was assessed to 4.95 million bacterial associated deaths and 1.27 million bacterial attributable deaths in 2019 [3]. Six pathogens including *Escherichia coli, Staphylococcus aureus, Klebsiella pneumoniae, Streptococcus pneumoniae, Acinetobacter baumannii* and *Pseudomonas aeruginosa* altogether were responsible for 929,000 of the attributable deaths and 3.57 million associated deaths (Figure 1) [1].

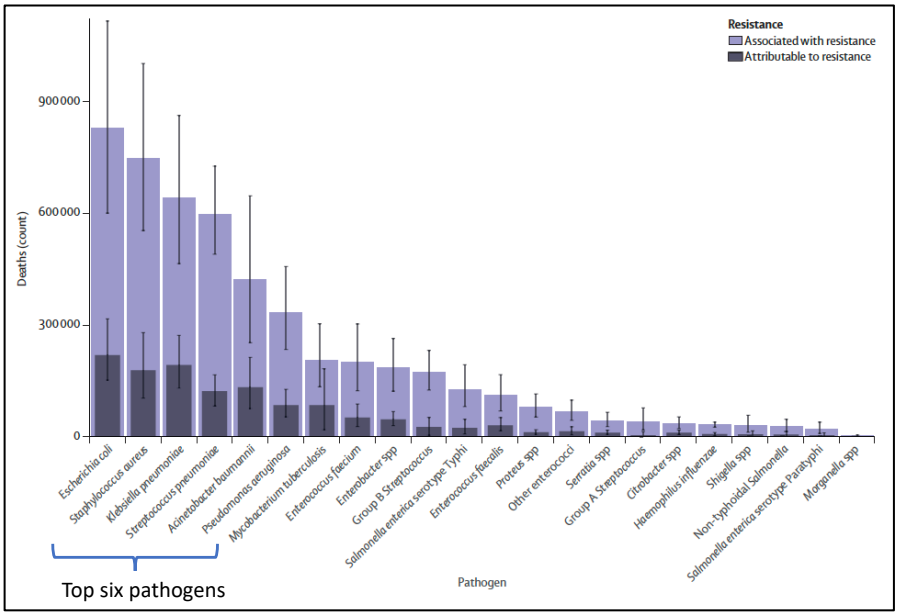


Figure 1. Global death (counts) associated and attributable to AMR by pathogen in 2019. *The blue arrow shows the top six pathogens in the list. Error bars represent 95% confidence interval. The figure is modified from the source.* [1]

AMR causes increased mortality, morbidity, prolonged treatment and increased financial burden due to the lack or unavailability of efficient antibiotics. Moreover, there is a void of discovery for new antibiotics. The combined effect threatens the practice of modern medicine like intensive care, surgery, cancer therapy and organ transplantation due to the risk of infections caused by untreatable superbugs. Global Antimicrobial Resistance and Use

Surveillance System (GLASS) 2022 has reported high level of AMR in *K. pneumoniae* among other bacteria globally (127 countries covering 72% of the global population) [4]. GLASS 2022 report also reported that 8% of the global bloodstream infections (BSI) caused by *K. pneumoniae* are resistant to carbapenem, an antibiotic class usually reserved for severe multidrug-resistant infections as the last resort β-lactams [4].

1.2 *K. pneumoniae*: A key global source and vehicle of AMR

K. pneumoniae is an opportunistic pathogen that was first identified by Carl Friendländer in 1882 [5]. The most common infections caused by *K. pneumoniae* are pneumonia, urinary tract infection (UTI), wound infection, bloodstream infection and neonatal sepsis [5-7].

K. pneumoniae as one of the ESKAPE pathogens (Gram positives *Enterococcus faecium, Staphylococcus aureus;* and Gram-negatives *K. pneumoniae, A. baumannii, P. aeruginosa, Enterobacter* species) is throwing challenges towards twenty-first century´s treatment for hospital-acquired infections (HAI) due to the ability to acquire AMR and develop multidrug resistance (MDR) [8]. WHO has categorized *K. pneumoniae* as one of the critical priority pathogens in 2017 due to its ability to acquire extended-spectrum β-lactamases (ESBL) and develop carbapenem-resistant strains (CRKp) [9]. ESBL-producing and CRKp strains are alone responsible for more than 90,000 infections and 7,000 deaths annually in Europe [3]. There was no report of CRKp before 2005 in Europe but as of 2015, CRKp has emerged as a major nosocomial pathogen in several countries including Romania, Italy and Greece [6]. Globally more than 250,000 deaths are associated to *K. pneumoniae* infections and at least 100,000 deaths are due to CRKp and third-generation cephalosporin-resistant *K. pneumoniae* strains [1]. According to the European Antimicrobial Resistance Surveillance Network database (http://atlas.ecdc.europa.eu/public/index.aspx?Instance = GeneralAtlas), resistance in *K. pneumoniae* has steadily increased over the years against four major antibiotics classes, including third generation cephalosporins, aminoglycosides, fluoroquinolones and carbapenems. Non-susceptible rates against all classes of antibiotics including carbapenem are higher in *K. pneumoniae* than *E. coli* [6]. The higher resistance rates in *K. pneumoniae* than *E. coli* indicates the importance of the former pathogen in AMR.

1.3 Characteristics of *K. pneumoniae*

K. pneumoniae (Kp) is a gram-negative, facultative anaerobic, encapsulated, non-motile, non-flagellated, non-spore forming, rod-shaped, lactose fermenting and oxidase negative opportunistic bacterium [10-13]. Kp rods are arranged as single, paired in short chains, or occasionally in clusters (Figure 2) [14].

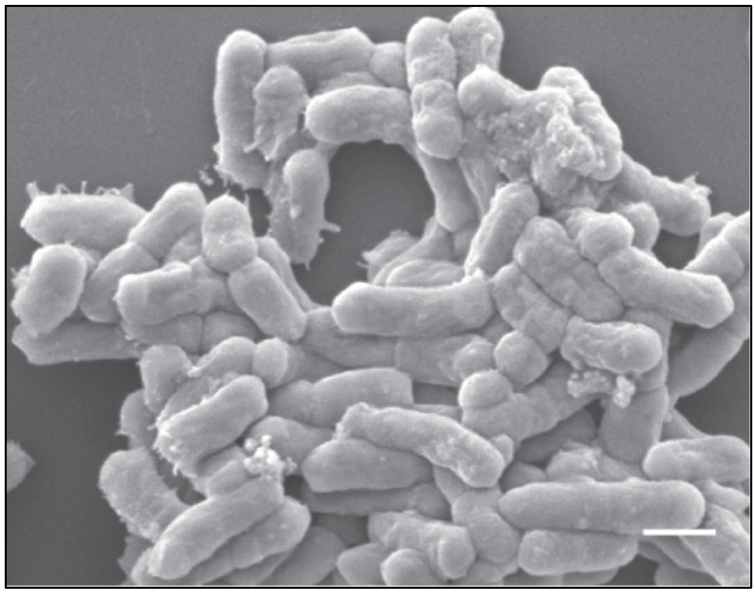


Figure 2. Scanning electron microscopy picture of biofilm producing Kp. *Magnifications: 200,000x and Scale bar: 1 µm.* [15]

1.3.1 Taxonomic composition and species designations

K. pneumoniae is classified under the order of *Enterobacterales* and the family of *Enterobacteriaceae* like many other clinically relevant bacteria (*E. coli* and the notorious human pathogens *Salmonella, Yersinia, Serratia, Enterobacter, Citrobacter, Kluyvera, Leclercia, Raoultella, Cronobacter,* etc) [5]. The genus *Klebsiella* was named by Trevisan (1885) honouring the German microbiologist Edwin Klebs (1834-1913) [13]. The diverse species belonging to the *Klebsiella* genus include the members of *K. pneumoniae* species complex (KpSC) and others (*Klebsiella indica, Klebsiella terrigena, Klebsiella spallanzanii, Klebsiella huaxiensis, Klebsiella oxytoca, Klebsiella grimontii, Klebsiella pasteurii and Klebsiella michiganensis*) [16].

KpSC shares only 90% of the average nucleotide identity with the other *Klebsiella* species (Figure 3) [5]. Members of the KpSC share 95%-96% average nucleotide identity (ANI) with *K. pneumoniae sensu stricto* and were originally designated as the phylogroups of Kp [17-19]. Later confirmed as different species according to the designation of new species (≥3% genome-wide ANI) of by the whole genome sequencing (WGS) [20, 21]. At the time of this writing, there were seven members consisting of five species classified as KpSC including *K. pneumoniae sensu stricto* (Kp1), *K. quasipneumoniae subsp. quasipneumoniae* (Kp2) and *K.*

variicola subsp. variicola (Kp3), *K. quasipneumoniae subsp. similipneumoniae* (Kp4) and *K. variicola subsp. tropica* (Kp5), *K. quasivariicola* (Kp6) and *K. africana* (Kp7) [22]. Kp1 is the clinically most important one as it constitutes more than 85% of the clinical KpSC isolates [5]. Henceforth, "Kp" will be used for Kp1 and KpSC will refer to all the members of the KpSC.

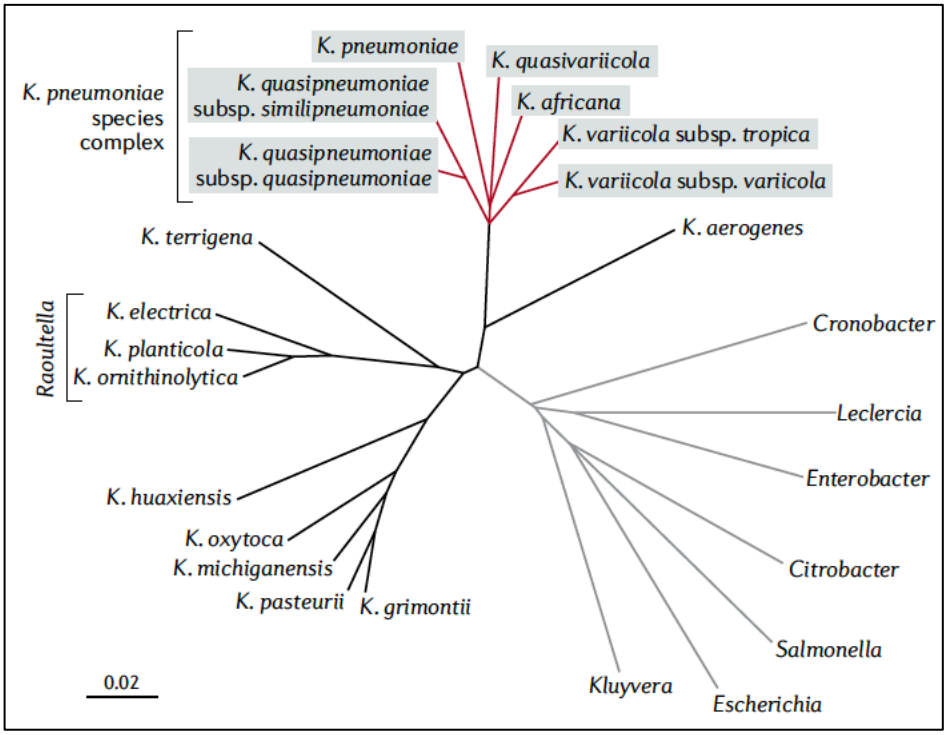


Figure 3. Phylogenetic relationship between Kp, members of KpSC and other Klebsiella genus. *In the figure, Klebsiella pneumoniae represents Kp.* [5]

It has been a challenge to distinguish between different members of KpSC in the clinical laboratory due to the lack of suitability of the existing microbiology methods [19]. Even modern identification techniques like WGS and mass spectrometry (MALDI-TOF) platforms can fail the proper identification of different KpSC species due to the lack of an updated reference database [5]. The Kp multilocus sequence typing (MLST) scheme based on allelic differences in seven housekeeping genes can be used to properly identify different KpSC members as the database contains the allele sequences and profile definition of the whole species complex [23].

1.3.2 Genomic diversity and global problem clones of Kp

Kp has large and diverse genomes of ~ 5-6 Mbp in size, including ~ 5000-6000 genes [5]. The number of core genes which are conserved in all the members of the species is ~1700 [20]. The pangenome (core genes and accessory genes together found in different Kp strains) is very diverse and likely contains more than 100,000 protein-coding sequences [5]. The large

pangenome provides broad metabolic and other capabilities to adapt to a wide range of niches [12]. Kp has an open pangenome, meaning more genes will be detected as more genome is sequenced [20]. The diverse pangenome indicates a high capacity of horizontal gene transfer (HGT), including plasmid-mediated conjugation, phage-mediated transduction and integrative conjugative elements [5].

Core genome-based phylogenetic analysis revealed hundreds of deep-branching lineages termed as "clones" with only ~0.5% nucleotide divergence [5]. Clonal groups (CGs) or clones are closely related Kp strains originating from a recent common ancestor. CGs can be identified either by allelic differences in seven core genes by multi-locus sequence typing (MLST) scheme or by core genome multi-locus sequence typing (cgMLST) schemes. cgMLST schemes provide higher discriminatory power between CGs as it contains 694 genes (Institut Pasteur's BIGSdb-Kp) or 2,365 genes (SeqSphere+, Ridom GmbH, Muenster, Germany) for comparison [21]. Different clones can also be separated from each other based on accessory genes as clone-specific niche adaptation occurs through HGT [24].

The diverse sequence types (STs) of Kp, which are identified by allelic differences in the housekeeping genes, differ in prevalence of pathogenicity and virulence factors between clones [25]. Some clones are locally distributed with limited spread in single hospitals or health care communities, but some of the clones are referred as 'global problem clones' due to the high burden of global disease and their association with virulence and ability to develop MDR [5]. Some clones belong to the classical Kp (cKp) pathotype which is prone to develop multidrug resistance and has non-overlapping geographical distribution with clones of hypervirulent Kp strains [26, 27].

Multidrug-resistant (MDR) Kp clones are defined by clones which are resistant against three or more antimicrobial classes in addition to intrinsic resistance to ampicillin [5]. MDR clones are mostly associated with HAI/outbreaks such as causing UTIs and bacteremia in immunocompromised patients (Figure 4) [28]. For example, local outbreaks in Kilifi, Kenya and Melbourne, Australia were caused by MDR clones ST70 and ST323, respectively [29, 30]. The most common MDR-Kp clones include ST11, ST258 and ST437 [5].

Hypervirulent Kp (hvKp) clones are defined as virulent pathogens causing infections in healthy persons of any age at any site of the body and often multiple sites of the body are affected (Figure 4) [31]. hvKp was first reported in 1980 and by 2004 it was the cause of 80% of pyogenic liver abscess (PLA) cases in Taiwan [31]. hvKp infections are caused by a small subset of lineages and are highly prevalent in the Asia Pacific rim, although the global occurrence is also increasing [5]. The most common hvCGs are CG23, CG65 (ST67 and ST 375) [5]. hvKp is associated with community-acquired invasive infections, including PLA in the absence of biliary tract disease, non-hepatic abscesses, pneumonia, necrotizing fasciitis, endophthalmitis and meningitis [31]. hvKp clones rarely possess acquired AMR genes in contrast to MDR clones [24]. However, they harbour key virulence loci for increased pathogenicity [24].

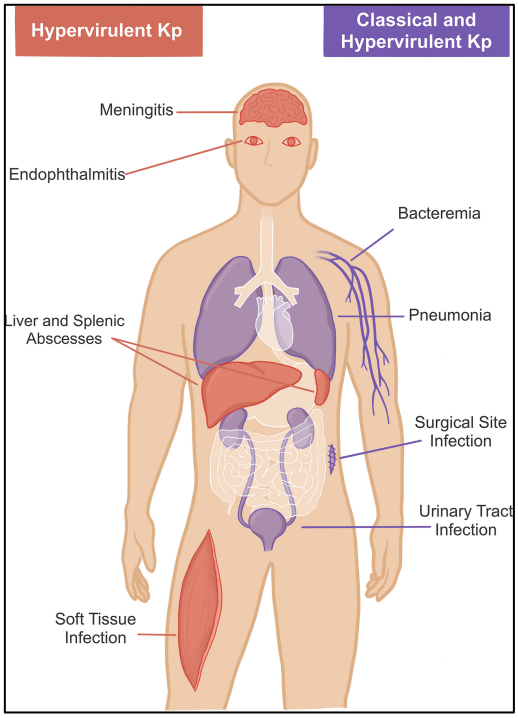


Figure 4. Kp infection sites in human body. [27]

The emergence of superbug: convergence of MDR Kp with hvKp

The presence of multiple acquired resistance genes or distinct virulence loci separates MDR Kp from hvKp creating separate lineages (Figure 5) [5]. These two non-overlapping lineages are increasingly converging in the last few years and developing clinically more difficult-to-treat clones with the combination of MDR and hypervirulence (MDR-hvKp) [16]. Since the first report of MDR-hvKp in 2015, MDR-hvKp clones have been isolated from Asia (China (China mainland, Hong Kong and Taiwan), India, Iran), Europe (France, Norway, the United Kingdom, Russia), Africa (Egypt), South (Brazil) and North America [32].

The mechanisms involving the convergence of MDR and hypervirulence may include the acquisition of AMR genes by hvKp clones, the transfer of virulence plasmids into MDR clones, the transmission of mobile genetic elements (MGE) to create hybrid vectors (AMR genes into virulence plasmids or recombination of AMR genes with virulence plasmid backbones or transfer of AMR genes into MDR plasmids) [5, 33].

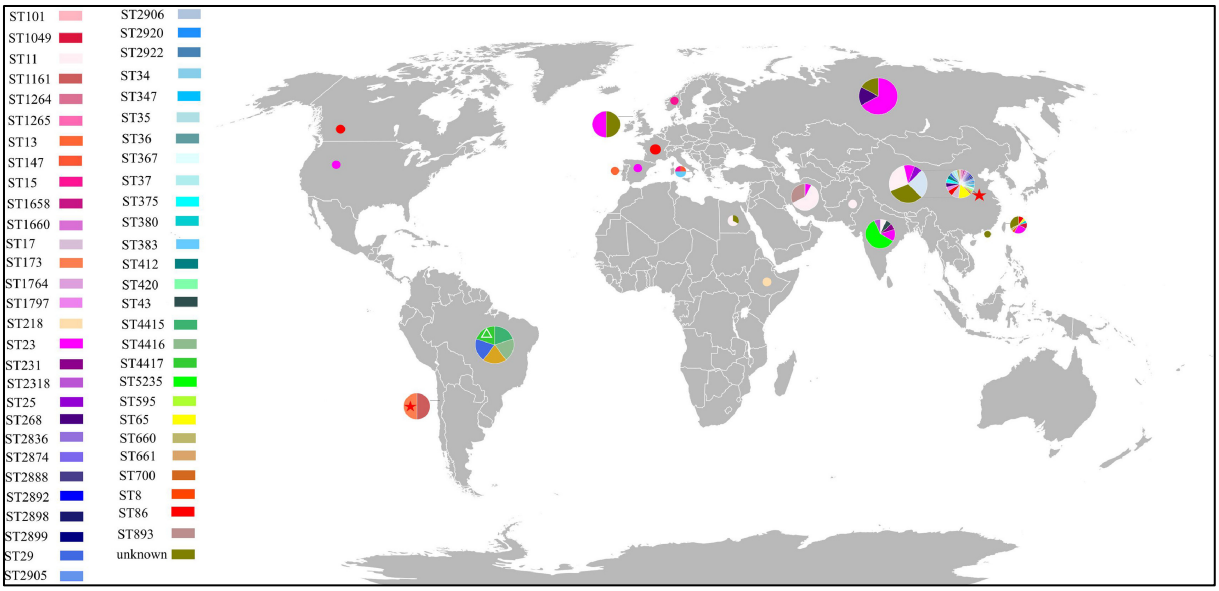


Figure 5. Global distribution of MDR-hvKp. *The triangle and asterisk represent K. quasipneumoniae subsp. similipneumoniae and K. variicola subsp. Variicola, respectively, the rest of them represent the prevalence of MDR-hvKp.* [16]

Some of the MDR clones and hvKp clones have become global problems termed 'global problem clones' or 'high-risk clones'. Global MDR clones are highly resistant lineages (resistant against third-generation cephalosporins and/or carbapenems) including CG258, CG15, CG20, CG29, CG37, CG101, CG147 and CG307. Global hvKp clones include CG23, CG25, CG65, CG66, CG86 and CG380 [5, 34]. The more recent ST307 and ST147 emerged in Europe during the 1990s and spread globally in a very short period causing worldwide nosocomial outbreaks [35]. Both clones acquired plasmids with ESBL-and/or carbapenemase-encoding genes aiding their global spread [35]. Carbapenemase-producing ST307 has been reported to be the cause of longer hospital stays and over 50% mortality of the patients with Kp-BSI in tertiary care centres compared to other strains in a two-year surveillance study in Medellín, Colombia [34].

1.3.3 AMR in Kp

Kp is well known for harbouring AMR genes and spreading AMR through HGT within and between species [5]. Both intrinsic and acquired AMR genes in Kp are responsible for the development of MDR [5]. Kp is intrinsically resistant to penicillins, including ampicillin and piperacillin, due to the production of class A β-lactamase enzyme SHV by core gene *bla*$_{SHV}$ in Kp, *bla*$_{LEN}$ in *K. variicola*, *bla*$_{OKP}$ in *K. quasipneumoniae* and other species [20]. Moreover, mutations and mobilization of numerous core genes on plasmids and transposons (mobile forms) have been associated with clinical resistance (Table 1) [5].

Table 1. Core genes of Kp involved in AMR

Core genes	Functions	Alterations	Resistance
***bla*SHV**	β-lactamase for hydrolysis of amide bonds	None	Penicillin including ampicillin and piperacillin
		Point mutations	ESBL activity (resistance to cephalosporins)
gyrA, gyrB, ***parC, parE***	DNA replication	Mutations in quinolone resistance-determining region (QRDR)	Quinolones
fosA	Glutathione-S-transferase for hydrolysis of drugs	Higher expression due to mobile forms	Fosfomycin
oqxAB /acrAB	Efflux pump	Higher expression due to mobile forms or modification of regulatory molecules	Ciprofloxacin, fluoroquinolones, nitrofurantoin, tigecycline, chloramphenicol and carbapenems
ompK36	Porin	Point mutations and deletions	Carbapenem
***mgr*B**	Lipid A production	Inactivation	Colistin
phoPQ, pmrAB or crrAB		Point mutations	

Most of the resistance in Kp is conferred by the acquired accessory AMR genes. Kp has twice the amount of reported acquired AMR genes (≥ 400) than *E. coli* [12]. The acquired AMR genes are mostly plasmid-associated but chromosomal integration has also been reported [5]. The distribution of acquired AMR genes follows a bimodal distribution separating MDR clones with high load of acquired AMR genes from hvKp clones with rare occurrences of acquired AMR genes [5]. Possible factors influencing this difference in distribution have been proposed including the difference in recombination rates, distribution of AMR genes on the same mobile genetic elements (MGE) and fitness cost due to maintenance of the plasmid [5]. One of the major threats to modern antibiotic treatment is the spread of ESBL-producing Kp. Thus, the distribution of AMR genes among ESBL-producing Kp will be discussed further.

ESBL-producing Kp are resistant to β-lactams, including third-generation cephalosporins (cefotaxime, ceftazidime and ceftriaxone), fourth-generation cephalosporin (cefepime) and monobactams (aztreonam) [36]. According to the Giske *et al.* classification, ESBLs are divided

into three main groups consisting of $ESBL_A$, $ESBL_M$ and $ESBL_{CARBA}$ [37]. The $ESBL_A$ enzymes include CTX-M (cefotaxime-hydrolyzing β-lactamase) and variants of SHV (sulphydryl variable) and TEM (Temoneira) [37]. $ESBL_A$ hydrolyses penicillin, monobactams and cephalosporins but cannot hydrolyse cephamycin, carbapenems and penicillin + β-lactam inhibitors leaving them as useful antimicrobial agents [37]. $ESBL_M$ contains a diverse group of β-lactamases including plasmid-mediated AmpC and OXA-ESBLs which hydrolyses most cephalosporins, all monobactams and penicillins but is unable to hydrolyse carbapenems and 4th generation cephalosporins [37, 38]. $ESBL_{CARBA}$ has hydrolytic activity against carbapenems, most often hydrolyses all β -lactams and includes Ambler class B metallo β-lactamases (MBL) such as Verona integrated-encoded MBL (VIM) and New Delhi MBL (NDM), class A and class D carbapenemases such as *Klebsiella pneumoniae* carbapenemase (KPC) and OXA-48 family, respectively [37-39]. Most of the ESBL genes are present on plasmids and disseminated by HGT.

1.3.4 Pathogenicity and virulence factors of Kp

Pathogenicity is the ability of a pathogen to cause disease and virulence factors enhance pathogenicity [28]. Kp has **intrinsic virulence factors** encoded chromosomally and present among all the members, which are basic requirements for causing opportunistic infections including capsular polysaccharide (K antigen), lipopolysaccharide LPS (O antigen), fimbriae (type 1 and 3) and siderophores (Figure 6) [5].

Capsular polysaccharides (K antigen) play an important role in gut colonisation, protect against host-mediated clearance and contribute to interactions with mucus [40]. Only two of the serotypes K1 and K2 among 77 serological serotypes are associated with invasive infections outside the hospital setting [41]. The distribution of capsule types is conserved in hypervirulent Kp, where K1 is present in CG23 and the rest have K2. K3 and K5 are associated with rhinoscleromatis (a rare chronic granulomatous infection mainly occurring in the upper respiratory tract, including the nasal cavity and nasopharynx) and liver abscess, respectively [42, 43]. MDR clones have diverse K locus with only the exception of CG307, which carries K-loci 102 (KL102) [44].

LPS (O antigen) is an integral component of the outer membrane in Gram-negative bacteria and has numerous functions like resisting host antibodies and clearance of microbes, including evasion of antibody-dependent complement-mediated lysis due to modification [40, 45]. Clinical Kp mostly has serotypes O1 and O2, which might provide enhanced defence against

phagocytosis [41]. An ex-vivo experiment showed that O2afg (a subtype of O2) suppresses the production of pro-inflammatory cytokines and chemokines, assisting the host immune evasion [46].

Fimbriae are filamentous organelles expressed on the surface of the bacteria and fimbrial adhesins play important role in adhesion to mucosal membrane [47]. Both type 1 (*fim*) and 3 (*mrk*) fimbriae help Kp in the colonisation and persistence [48]. They are critical for biofilm formation on biotic and abiotic surfaces like respiratory devices or recently implanted devices [48]. Type 1 and 3 fimbriae have been associated with adhesion to urogenital epithelial cells and bladder endothelial cells, respectively [48]. However, type 1 fimbriae are dispensable for the Kp gut colonisation [49].

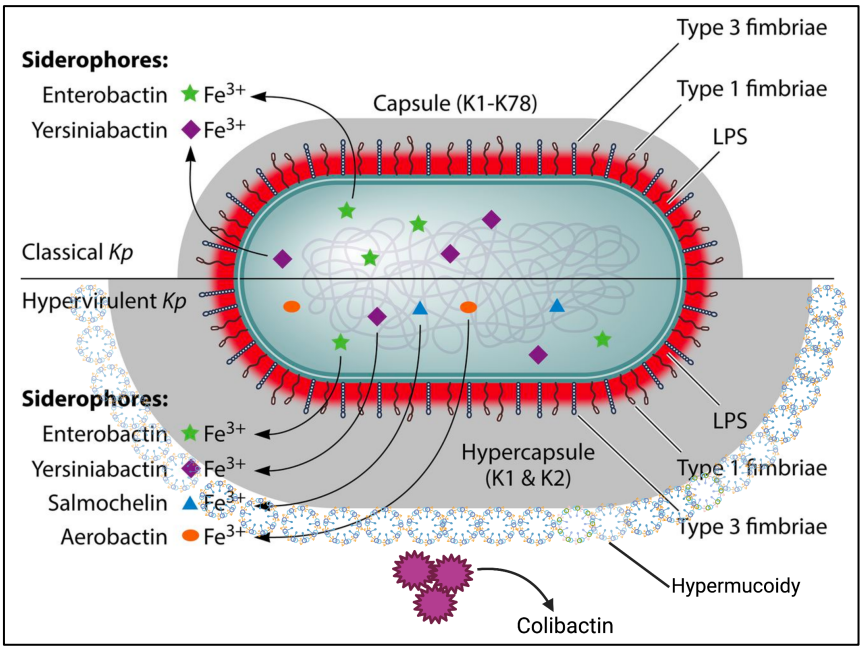


Figure 6. Virulence factors in Kp. *The figure is modified from reference.* [50]

Some of the virulence factors are only present in few Kp encoded by the accessory loci enhancing pathogenicity [5]. These virulence factors are known as **acquired virulence factors**, usually plasmid-encoded and give Kp the ability to transform from an opportunistic pathogen to a true pathogen such as hvKp causing CAIs [5].

Siderophores are low molecular weight molecules with high binding affinity for iron and are essential for acquiring iron from extracellular environment to ensure growth in most niches [51]. The core siderophore in Kp, enterobactin (Ent), is encoded by the *ent locus* and acquired siderophores, yersiniabactin, aerobactin and salmochelin, are encoded by the *ybt*, *iuc* and *iro*

loci, respectively [5]. All the accessory siderophore systems increase virulence by providing an advantage over the host's iron-withholding devices [5]. For example, glycosylated Ent known as salmochelin, produced by *iro* locus can evade host lipocalin 2 (Lcn2) secreted by neutrophils and continue iron scavenging, whereas core Ent-based iron scavenging activity is disrupted due to the binding with Lcn2 followed by host inflammatory response [52, 53].

Hypermucoidy and colibactins are also acquired virulence factors, among which hypermucoidy is the most well-known in Kp [5]. Hypermucoidy can be described by, but not limited to, the overproduction of the capsule due to the presence of accessory regulatory genes, regulator of the mucoid phenotype *rmpA1* or *rmpA2* and confirmed by the phenotypic string test [54]. However, hypermucoidy is not a core feature of all hypervirulent clones and its role in disease progression is not well understood [27]. Colibactin, a genotoxin that induces DNA damage in eukaryotic cells, is present in ~10% of Kp [46]. It has been associated with liver abscess and invasion from the intestine to the vascular system [5].

1.3.5 Plasmid diversity in Kp

Plasmid-mediated gene transfer is the main mechanism in the spread of AMR genes spread in Kp [5]. The presence of multiple plasmids with multiple AMR genes in Kp comes with a fitness cost for the maintenance of the plasmids [5]. The majority of the acquired accessory genes are on large plasmids capable of self-transmission via conjugation [5]. Self-transmissible plasmids in Kp belong to a small number of incompatibility groups (IncFII, IncN, IncR and IncX3)[5]. However, AMR genes in Kp are also found on mobilizable but not self-transmissible plasmids such as dominant virulence plasmids KpVP-1 and KpVP-2 [55]. Although plasmid variation among single clones has been found, there is also evidence of specific plasmid maintenance among some problem clones [56]. For example, the FIB_K blaKPC plasmid pKpQIL47, FII_K blaCTX-M-15 plasmid and FIB_K Kp virulence plasmid (KpVP-1) have been maintained in ST258, ST307 and ST23 respectively [44, 56]. Tracking of plasmid transmission through genomics is now used as infection control and strategy to lessen disease burden during outbreaks in hospitals [5]. There is a knowledge gap regarding Kp-plasmid diversity in transient carriers (colonised with Kp from time to time) and persistent carriers (colonised with Kp continuously for longer period), which might help to better understand the colonisation dynamics of Kp.

1.3.6 Phages of Kp

Bacteriophages (phages) are viruses infecting bacteria and utilizing bacterial replication machinery for their replication, producing progeny and releasing the progeny in the lytic cycle through host cell lysis [57]. Prophages are dormant phages integrated into the host genome during the lysogenic cycle [57]. Prophages remain dormant for an unspecified amount of time and replicate with the bacterial genome without harming the host [57]. Prophages can become activated under certain stimulation and initiate full lytic cycle. Prophages in Kp are very diverse and accessory genomes of Kp contain many prophages [58]. Prophages give several benefits to the host including defending against adverse environmental stress or attack from similar phages [59]. Global problem clone CG258 has its conserved prophages participating in the evolution and fitness benefit [58]. Genetic changes during the lysogenic cycle may cause degradation of the phage genome or transfer of phage genes into the host that leads to toxin production and gain of antibiotic resistance capability by the host [60]. Furthermore Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas system in Kp can prevent the expansion of the prophages [58]. However, most of the studies on prophages in Kp are based on bioinformatics based and lack the experimental evidence. The diversity and role of prophages in long-term persistence of Kp is another unexplored field.

1.3.7 Ecological niches and cross-niche transmission of KpSC

KpSC has several ecological niches apart from the human body, including soil, water, plants, insects, birds, reptiles and other mammals [12]. A cross-sectional estimation with culture-based techniques indicated a gut prevalence of 39% and 44% Kp (and/or the broader KpSC as the study did not distinguish between phylogroups) in the intestinal tract of dogs and dairy cattle, respectively [61, 62]. KpSC can both be free living and host-associated [5]. The bacteria can co-exist with the host as a commensal organism and cause opportunistic infections [63].

KpSC members, *K. quasipneumoniae* and *K. variicola* are mostly associated with healthy human gut and agricultural niches, respectively [19, 64]. Previously, they were thought to represent the human gastrointestinal tract (GIT) colonisation instead of infection [20]. The Tromsø 7 cross-sectional KpSC carriage study (T7-study) based on culture-based WGS reported the gut carriage of 6.4% and 27.7% of Kp2 and Kp3 in community-based adults in Norway, respectively [65]. Later, they were also reported as the cause of minority *Klebsiella* infections in patients at health care system [5, 66]. WGS of non-ESBL isolates from the nationwide Norwegian surveillance programme on resistant microbes in humans (NORM)

reported the presence of Kp (69.1%), Kp2 (3.1%), Kp3 (24.5%) and Kp4 (3.3%) in KpSC bacteraemias and gut colonisation is the prerequisite for the infection [67]. One out of every four cases of KpSC bacteraemia in Norway is caused by susceptible Kp3 [67]. ESBL-producing *K. quasipneumoniae* and Kp3 strains were found as virulent as Kp, causing fatal infections with similar mortality rates [19].

In recent years, the environment has become a potential reservoir for the spread of Kp and antibiotic resistance in humans (Figure 7) [68]. A recent study on the Norwegian marine environment reported a close genetic relationship between Kp isolate (2016-1200 ST25) from blue mussels (*Mytilus edulis*) with a clinical isolate from Germany [68]. This finding suggests the possible transmission of Kp from the marine environment to humans via seafood [68]. In another longitudinal study on healthy adults in Austria, it was also hypothesized that the high diversity of the Kp isolates found in the study was due to the consumption of contaminated foods [69]. The study revealed that individuals who shared meals had identical strains several times over the one-year study period indicating food as a source of Kp transmission [69].

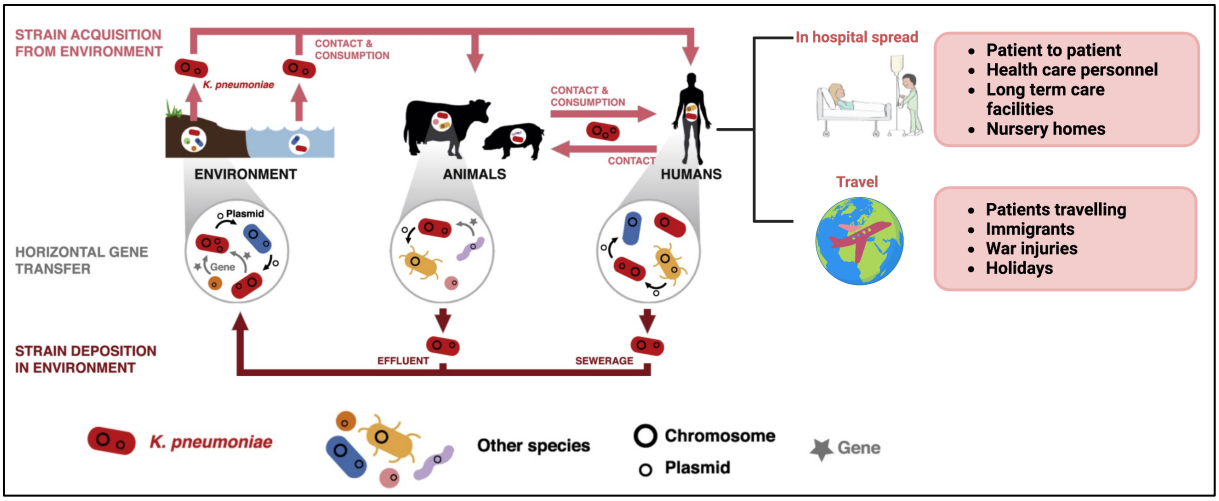


Figure 7. Routes of transmission of Kp. Kp strains can disseminate from one niche to another along with their plasmids and AMR genes. *The direction of transmission can either be from the environment to the human or animal as host via contaminated water, food and contact or from the host back to the environment through sewage or effluent. Transmission between humans can also occur through direct contact from person to person and global dissemination can occur through travelling. The figure is modified from the references.* [6, 12]

Cross-niche transmissions, such as from human and companion animals and vice versa, have also been reported in previous studies. One study examining the Kp gut colonisation in household members and companion dogs found that they shared the same Kp STs, indicating that dogs may act as a reservoir for Kp in humans and vice versa. [61]. Limited zoonotic transmission of Kp was also reported in another study based on pig farms in Thailand, where

sharing of the same clones of some of the STs (ST4788, ST3541, ST661 and ST29) were confirmed between humans and pigs [70]. One study in Norwegian healthy poultry also revealed the presence of overlapping Kp STs (ST13, ST17, ST35, ST37 and ST290) between humans and poultry animals, indicating transmission among humans and animals [71]. However, a recent large-scale One Health study based on WGS of *Klebsiella* species isolates from a wide range of clinical, community, animal and environmental samples in Pavia, Italy, revealed that the most common way of *Klebsiella* transmission is person-to-person contact rather from animals or the environment [72]. Although, the study had limitations concerning food-borne and wastewater-mediated transmission, which might have a crucial impact on the transmission cycle among humans, animals and the environment [72].

The lack of large-scale systematic sampling has created a void of knowledge regarding the ecology and distribution of KpSC in different niches and their main reservoirs. Hence, we lack information about to what extent the cross-niche transmission is happening, the strain diversity and relative abundance of KpSC over the time in a particular niche and the factors influencing the carriage. The current study focuses to address the duration of KpSC gastrointestinal carriage, as well as strain diversity and dynamics, over the six-months study period in community-based adult population.

1.3.8 Human gut colonisation of Kp

Following acquisition, Kp mostly colonises the mucosal surfaces in humans, including the nasopharynx and the GIT [7]. Kp can also be found transiently on the skin [73]. Colonisation is defined by the presence of bacteria on the body surface but not causing any symptoms or disease [74]. Thus, colonised individuals can act as silent carriers [75]. The silent carriage can turn into a clinical Kp infection when the host becomes immunocompromised or acquire other risk factors for infection [76]. The rate of Kp-colonisation depends on the type of human body site and the type of acquisition, whether hospital-acquired (HA) or community-acquired (CA) [73]. Studies from Brazil, Indonesia and Vietnam showed that the rate of colonisation in healthy children and adults varied a lot between 3-15% for CA nasopharyngeal colonisation and up to 19% for HA nasopharyngeal colonisation [73]. However, studies based on intensive-care unit (ICU) and cancer patients in USA and Australia indicated gastrointestinal Kp colonisation rates of up to 20% in hospitalized patients [73]. The prevalence of Kp gut colonisation in the community (USA and Australia) has been reported lower (~4-6%) than in recent health care exposed individuals (~25%) [77-80]. Culture based detection of KpSC in the T7 study on

community-based adults (≥ 40 years) in Norway reported 16.3% gut carriage of KpSC [65]. Additionally, culture-based detection of Kp also confirmed a higher percentage (18-87%) of Kp gut carriage in healthy adults in many other countries such as Korea and Chinese healthy adults in Japan, Singapore, Taiwan and Malaysia [81, 82].

Intestinal carriage has been established as a key risk factor for Kp infection [77, 79]. A direct comparison of carriage and disease-causing strains of Kp revealed a four times elevated risk of infection in colonised patients (cancer and ICU) compared to non-colonised patients [77, 79]. Combined analysis from two different studies (one in Australia and another in the USA) based on rectal swabs from ICU patients confirmed that 80% of the HA Kp infections were due to the patient's own carriage strains [77, 83]. Furthermore, the six most prevalent STs in the T7-study were also listed as the frequently observed bacteraemia STs in the NORM study [65, 67]. One of the main concerns is the colonisation of the hypervirulent strains, which are prevalent in Asia. These strains possess the threat of metastatic spread from GIT to other tissues causing life-threatening abscesses and sepsis [84]. Furthermore, the increased relative abundance of Kp within the gut has also been associated with the risk of bloodstream infection in long-term intensive-care hospital patients [85].

Risk factors for KpSC gut colonisation in a general adult community-based population have recently been examined the T7-study (Figure 8) [65]. The T7-study has shown that increased KpSC colonisation is positively correlated with age ≥60y [65]. It is shown that alcohol consumption increases the prevalence of oropharyngeal Kp colonisation, but a larger cohort is needed to find its association with gut colonisation [65]. Antibiotic treatment also predisposes individuals to KpSC gut colonisation which is likely due to the disruption of the gut microbiome [65, 86]. This suggests the role of gut microbiota in maintaining the barrier protecting from KpSC colonization. Non-antibiotic drugs such as proton pump inhibitors (PPIs) or non-steroidal anti-inflammatory drugs (NSAIDs) were also positively associated with KpSC gut carriage supporting the influence of non-antibiotic drugs on gut microbiota selecting specific bacterial species [87][65].

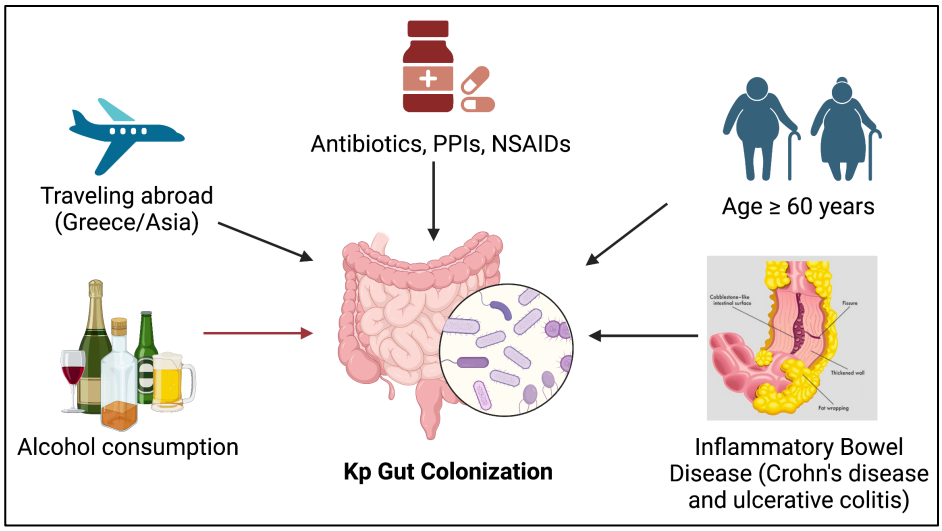


Figure 8. Risk factors associated with KpSC gut carriage. Travelling abroad, medications (antibiotics, PPIs, NSAIDs), age and inflammatory bowel disease are key risk factors positively associated with Kp carriage. *The role of alcohol consumption needs to be further investigated. The black arrow represents established risk factors and the red arrow represent possible risk factor. Figure was created by Biorender.*

Crohn´s disease and ulcerative colitis have also been associated with an increased prevalence of Kp gut carriage [65]. One of the influencing factors might be the disease-specific alteration of the gut microbiota in Crohn´s disease or ulcerative colitis. There is a significant increase in the abundance of Enterobacterales with a decrease in gut microbial diversity in patients with Crohn´s disease [88]. Travelling abroad to Asia/Greece was also a risk factor for Kp gut carriage, which could be explained by the high Kp carriage among the local population [65, 82, 89]. Personnel hygiene, consumption of specific food (raw meat or vegetables), hospitalization of household members and animal contacts have also been associated as risk factors for Kp carriage among pregnant women living in low-income countries [89].

The dynamics and duration of human gut Kp-carriage, however, are still not well studied. A previous study based on ESBL- producing Kp outbreak in neonatal intensive care units (NICUs) at Stavanger University Hospital in Norway between November 2008 and April 2009 reported that the duration of gut colonisation by the same Kp-clone (ST17) in children can be more than twelve months [90, 91]. However, another longitudinal study in Austria with only six healthy adult study subjects reported the presence of specific STs only for two consecutive calendar weeks [69]. The two above-mentioned studies had different study subjects and contexts, which might affect the different observations. Data of strain diversity in colonisation and infection is important for vaccine research and a better understanding of the cross-niche transmission from

reservoirs facilitating human Kp colonisation [12, 92]. Moreover, we still lack knowledge of potential role of gut microbiota composition on Kp carriage.

Human gut microbiota

Microbes are the leading population both in diversity and numbers, dominating life on Earth [93]. Humans have developed a close relationship with microbes homing trillions of these tiny friends and foes [94]. According to the latest studies, humans have approximately 30 trillion cells while harbouring ten to one hundred trillion microbial cells spread throughout the skin, inside the gut and almost everywhere on the human body surface [94]. Microbial genes outnumber ours by 150 times and are often referred by scientists as our 'second genome' [95]. The term 'microbiome' was coined by Nobel Laureate Joshua Lederberg in 2001, which includes the microbiota (living microorganisms of different kingdoms including Bacteria, Archaea, Protozoa and Fungi and algae), structural elements (microbial structures, mobile genetic elements, etc.), microbial metabolites and the environmental conditions of the habitat known as the 'biome' (pH, temperature, oxygen level, etc.) [96, 97]. The human microbiome is dynamic, differing from one site of the body to another and between individuals and consists of bacteria, archaea, protozoa, fungi, algae and viruses [98, 99]. The human gut microbiome is the most diverse among other sites of the body with thousands of species and has been associated with human health and disease (Figure 9) [100, 101].

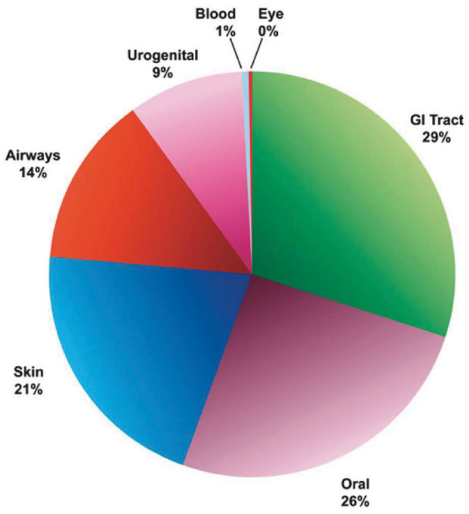


Figure 9. Distribution of bacteria in human body sites. *It shows gut is the major reservoir.* [94]

Human gut microbiota has coevolved with the host into a complex relationship shaping the immune system, providing protection against colonisation of pathobionts, assisting digestion of indigestible food, production of essential vitamins and homeostasis maintenance [102]. Human gut microbiota composition is influenced by host genetics, age, diet, lifestyle, ethnicity and living environment [102]. In children, gut microbiota composition is also influenced by the mode of infant delivery (vaginal or c-section) and feeding (breastfed or formula-fed) [103]. Dysbiosis in the gut, which is the alteration of the microbiota composition from normal state, including change in relative abundance, membership, or localization of microorganisms, has

been associated with various immune, metabolic and neurological disorders in both intestinal and extra-intestinal sites, including inflammatory bowel disease, irritable bowel syndrome, diabetes mellitus, obesity and colorectal cancer [104-107].

We need to know the composition of **healthy gut microbiota** to understand dysbiosis. There is no clear definition of a healthy gut microbiota [108]. Most of the studies investigating healthy gut microbiota are from socioeconomically developed societies in Europe and North America, which might not reflect the microbiome of individuals from developing, non-industrialized countries [108]. However, current studies and data from Human Microbiome Project (HMP) show that most of the healthy adult human gut microbiota is dominated by Firmicutes and Bacteroidetes, followed by Proteobacteria and Actinobacteria [93, 98, 109]. The relative abundance (presented as mean ± standard deviation) of the dominant Phyla in healthy adult human faecal samples from two cross-sectional studies, Segata *et al.* 2012 and King *et al.* 2019 is presented in Table 2 [93, 109].

Table 2. Relative abundance of dominant phyla in healthy human faecal samples

Phyla	Segata *et al.* (2012)	King *et al.* (2019)
Bacteroidetes	65.20 ± 20.50	73.13 ± 22.16
Firmicutes	29.64 ± 19.26	22.20 ± 18.66
Proteobacteria	2.91 ± 4.74	2.15 ± 10.39
Actinobacteria	0.51 ± 2.82	1.82 ± 3.00
Verrucomicrobia	0.44 ± 1.74	0.70 ± 1.68
Fusobacteria	0.07 ± 0.50	Not detected

There is also a difference in the relative proportions of bacterial families along the whole gut due to the physio-chemical differences [110]. For example, lower pH, higher amount of antimicrobial peptides, presence of bile salts and higher oxygen tension in the small intestine limit the bacterial load compared to the colon [110]. The small intestine is dominated by *Lactobacillaceae* and *Enterobacteriaceae*, whereas *Bacteroidaceae*, *Prevotellaceae*, *Rikenellaceae*, *Lachnospiraceae* and *Ruminococcaceae* are common in the colon [110]. A healthy individual starts shaping gut microbiota very early in life and one study has shown that intestinal microbiota is not fully developed until after the first five years of life [111, 112]. As we grow older, we achieve relative stability in gut microbiota composition, including developed enterotypes [113].

Enterotypes of human gut microbiota are defined as clusters of bacteria contributing collectively to the preferred community compositions, which remain stable from adulthood and can be restored upon modification [113]. Three enterotypes have been identified in one of the previous studies that differ in species and functional composition [113]. Enterotypes are identified by the level of one of the three genera dominating the microbiota- *Bacteroides* (enterotype 1), *Prevotella* (enterotype 2) and *Ruminococcus* (enterotype 3) [113]. Bacterial species in each enterotype have distinctive routes to generate energy, for example, bacteria clusters of enterotype 1 utilize mostly carbohydrates, using principally glycolysis and pentose phosphate pathways, while bacteria clusters of enterotypes 2 and 3 degrade mucins resulting in the uptake of simple sugars from gut mucosal layer [113]. How gut microbiota resists Kp colonisation and the mechanisms involved in overcoming colonisation resistance by Kp are questions yet to be fully answered.

Colonisation resistance is the phenomenon when the native symbiotic microbiota limits the entry and colonisation of pathogens under undisturbed conditions [114]. Gut microbiota can also limit the expansion of indigenous potentially threatful 'pathobionts' along with resisting colonisation of pathogens or innocuous species like probiotics [115]. Colonisation resistance can be achieved by direct mechanisms or indirect mechanisms [116].

Direct colonisation resistance includes **exploitative competition**, such as spatial or nutrient competition, where bacteria compete for niche availability, metabolites and essential nutrients (Figure 10) [117-120]. A recent study in mice has revealed that *Klebsiella oxytoca* strains outcompete MDR Kp with the help of other commensals like *Blautia coccoides, Enterococcus faecalis and Enterocloster clostridioformis* in gut colonisation by specific carbohydrate competition involving the utilization of specific beta-glucosides [121]. Microbiota can also resist colonisation directly by producing inhibitory molecules, toxic compounds such as short-chain fatty acids (SCFA) and type VI secretion system **(T6SS)** mediated killing [116]. While a plasmid-encoded tellurium resistance operon called *ter* has been shown to improve Kp fitness in gut colonisation in a microbiome-dependent manner, SCFA administered exogenously has been shown to reduce the fitness when *ter* operon is inactivated [122]. The role of SCFA in colonisation resistance also aligned with the protective role of *Bifidobacterium longum* 5^{1A} in the pulmonary Kp infection [123]. *B. longum* 5^{1A}, which belongs to the genus *Bifidobacterium,* one of the largest producers of SCFA, protects against Kp infection by inducing the production of reactive oxygen species (ROS) and killing Kp by alveolar macrophages, suppressing

proinflammatory mediators and finally participating in the resolution phase of inflammation by IL10 [123]. Furthermore, SCFA limits the expansion of CRKp in the gut by intracellular acidification, countering the competitive advantage in respiration to utilize O_2 and NO_3 by Kp [124]. The above-mentioned studies support the direct colonisation resistance by microbiota enriched with SCFA producers against Kp. In another animal study, reduction in *K. quasipneumoniae* gut colonisation was achieved by using *Klebsiella* species-specific colicin-type bacteriocins [125]. T6SS mediated colonisation and killing of *K. oxytoca* by *Salmonella enterica* subsp. *enterica* serovar Typhimurium has been reported [126]. However, there is a clear knowledge gap regarding T6SS-mediated colonisation resistance against Kp.

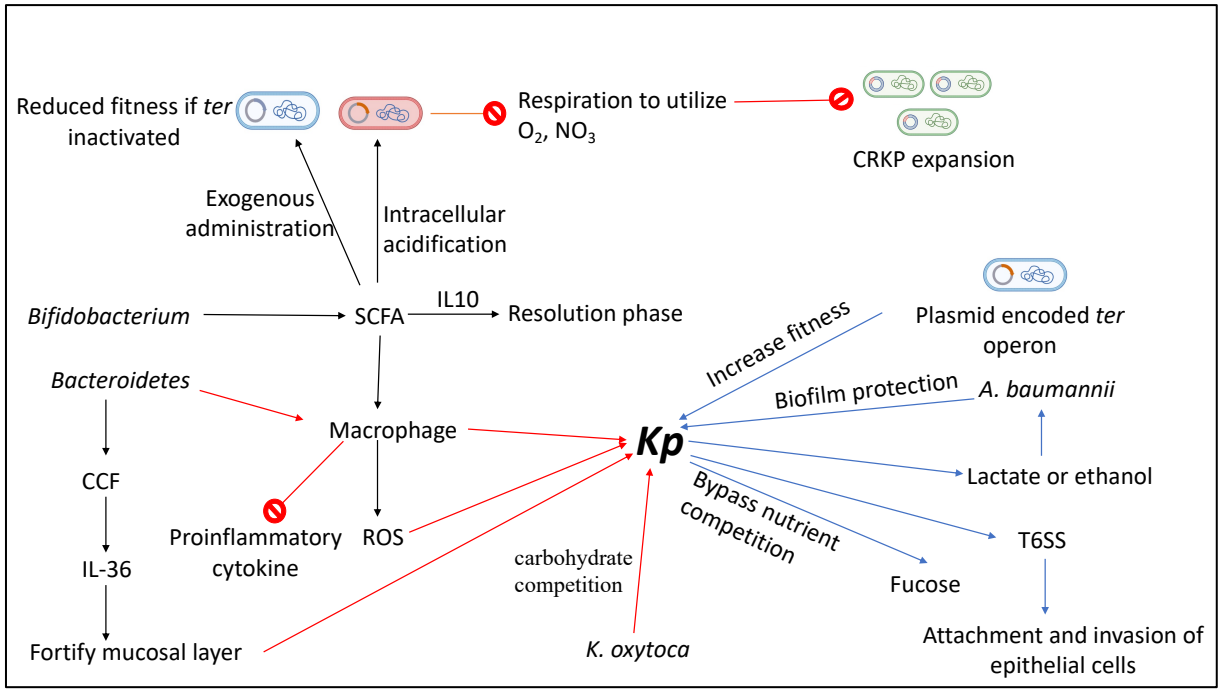


Figure 10. Mechanisms of colonisation resistance to Kp. *The red arrow indicates a negative relationship and the blue arrow indicates a positive association.*

Indirect colonisation resistance mechanisms include the mucus layer, oxygen gradients, antimicrobial proteins and peptides produced by host cells and host innate and adaptive immune responses, including the production of certain cytokines modulated by the gut microbiota (Figure 10) [116]. Bacteroidetes provide intestinal colonisation resistance to Kp by fortifying the mucosal barrier via IL-36 signalling and macrophages but depend on the polysaccharide utilization locus known as conserved commensal colonisation factor (CCF) [127].

Gut microbiota mediated Kp gut colonisation

Along with protection against colonisation, gut microbiota can also help in colonisation as well (Figure 10). *In vitro* analysis has shown that cross-feeding between Kp and *Acinetobacter baumannii* promotes their co-existence [128]. Lactate or ethanol produced by Kp is utilized by *A. baumannii*, while biofilm produced by *A. baumannii* protects Kp from antibiotics such as cefotaxime and gentamicin [128]. Kp also utilizes alternative carbon sources, such as underutilized fucose liberated from mucin by gut commensals to overcome colonisation resistance by nutrient competition [129]. T6SS in Kp promotes colonisation by attaching and invading epithelial cells along with increased bacterial dissemination due to enhanced virulence [130]. T6SS in Kp also decreases intestinal microbiota richness and colonisation resistance [131]. Members of the Oscillospiraceae family have been suggested to be specific competitors for long-term gut colonisation by Kp [131]. However, which mechanisms are involved in Kp non-carriers to achieve the colonisation resistance or to overcome the colonisation resistance for long-term Kp carriers needs further investigation. We need longitudinal human colonisation studies to approach these questions.

1.4 Whole metagenome sequencing (WMS): Emerging diagnostic and surveillance tool in infectious diseases

WMS is the sequencing of total genomic DNA isolated from a given sample [132]. WMS is a powerful culture-independent tool to analyse complex microbial communities and provides information at the species or even strain level (Figure 11) [133]. Culture independence of WMS ensures an advantage over conventional culture-dependent diagnostic tools like whole genome sequencing (WGS) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) where difficult-to-culture or non-cultivable microorganisms are present in the sample [134].

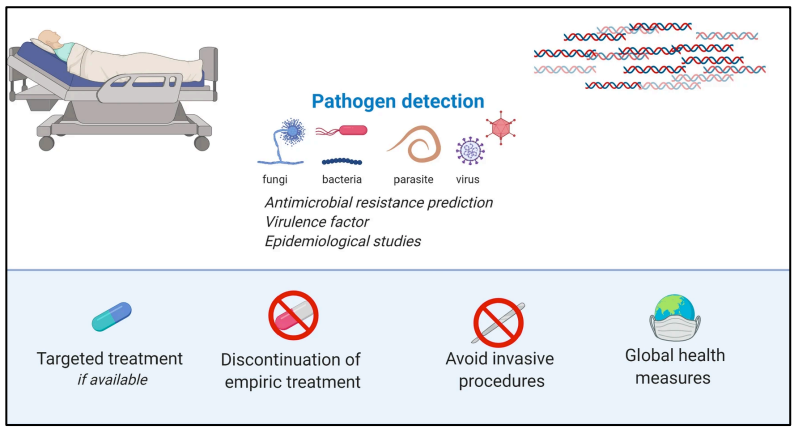


Figure 11. Impact of WMS in the clinical setting. *The figure is edited from reference.* [135]

The next-generation sequencing (NGS) has given WMS a speed, which is unmatched by traditional culture-based detection techniques. [134]. A standard Illumina instrument takes more than 18 hours of sequencing run providing information regarding a broad range of pathogens (viruses, bacteria, fungi and/or parasites), which is almost half of the time required by the conventional culture-based techniques [134]. The development of bioinformatic tools, which do not require the assembly of the sequence reads, has also made the bioinformatic analysis to identify species and strains faster than before [136]. Some of the limitations, including updated reference database, cost and clinical utility, are currently preventing the routine implementation of WMS in patient care settings [134]. However, the development of rapid diagnostic platforms, such as nanopore sequencing and surveillance networks, may facilitate the use of WMS to detect and identify specific pathogens for therapeutic and infection control purposes [134].

2 Aim of the study

This master project is part of a larger project with the overarching aim of investigating adult human gut carriage of KpSC in a longitudinal community-based context: chronic, transient or non-carriage, abundance and associated factors, including microbiota composition.

The **primary aim** of the master project was to investigate the following:

"The duration and dynamics of KpSC human gut carriage in adult community-based population"

The following secondary questions were addressed to investigate the primary aim -

1) Are there any differences in gut microbiota diversity between KpSC carriers of different carriage duration and KpSC non-carriers?
2) Are there any associations between gut microbiota diversity and KpSC abundance?
3) Is KpSC carriage duration related to age?
4) Is it possible to rebuild accurate phylogenetic ST relationships from the metagenomic samples?
5) Is chronic carriage of KpSC associated with a specific ST or multiple STs
6) What is the ST dynamic in chronic carriers? Whether chronic carriers regularly turned over strain or maintained the same throughout the study.

3 Materials and Methods

3.1 Participants and faecal samples

For the longitudinal KpSC gut carriage study, participants were recruited from the previous T7-study where risk factors for Kp-colonisation and Kp-genomic diversity were analysed in a general adult population (Figure 12) [65]. WGS of Single isolates from the previous T7-study and metagenome sequences of sweep samples from the comparative study of KpSC detection methods were used to evaluate phylogenetic relationship from metagenome sample by Strain Genome Explorer (StrainGE) toolkit (Figure 12) [65, 137].

3.1.1 The longitudinal KpSC gut carriage study

The longitudinal KpSC gut carriage study was designed to examine KpSC carriage duration and dynamics in participants recruited from the previous T7-study. Briefly, gender and age-matched community-based participants (n=240) from the previous T7 study (n = 2975) were selected for invitation, among which 50% had been KpSC culture positive and 50% culture-negative when examined in 2015-16. Both groups were matched for age and sex. A total 116 accepted the invitation. Two different faecal sampling kits, ESwab 490CE.A (Copan, Brescia, Italy; for viable bacteria) and Norgen Nucleic Acid Preservation system (Norgen Biotek, cat# 53700); for DNA preservation, extraction and WMS, were sent to the participants each month for six months between September 2021 and March 2022. Additionally, questionaries regarding age, sex, cohabitation, work in health sector, travel, health conditions (diabetes, Inflammatory Bowel Disease), hospitalization and medications (antibiotics, NSAIDs, PPIs) (last 4 months) were also provided (Figure S1). These risk factors were selected since they have previously been associated with Kp gut carriage in the T7-study. ESwab samples were stored as described below (section 3.2.1) and Norgen samples were directly stored at -80°C and subjected to DNA extraction (section 3.3). Previously, it has been shown that ESwab sample collection can have a bias effect of enriching Enterobacterales in stool samples [137]. Therefore, ESwab samples are not reliable for analysis of the overall microbiome [137]. However, Norgen Nucleic Acid Preservation system has been demonstrated to preserve the composition of the stool sample, which makes it suitable for microbiome analysis [137]. ESwab samples were therefore used in this study to collect culture viable samples for further enrichment of KpSC by Simmons citrate agar (SCA) with 1% inositol (SCAI) media as ESwab preserves the viability of bacteria [137].

3.1.2 KpSC single isolates for the evaluation of phylogeny relation by StrainGE

Kp single isolate whole genome sequence (WGS) data used in this master's project for the evaluation of phylogeny relation by StrainGE were selected from the previous T7-study [65]. In the T7-study, community-based adult participants (age ≥ 40 years) were recruited as part of the seventh survey of the Tromsø Study (The Tromsø Study: Tromsø7) during 2015-2016 and a collection of 2975 faecal samples was analysed for cross-sectional KpSC carriage [65]. Briefly, samples were self-collected by the participants with nylon-flocked ESwab 490CE (Copan, Brescia, Italy) stool collection kit. Samples were stored at -80°C with 200 µL added glycerol (85%) and cultured on SCAI medium enabling single colony picking. KpSC colonies were picked during this study based on their morphology of large, yellow, glossy and dome shape. Species identity confirmed by MALDI-TOF from subcultures of the suspected colonies. Upon confirmation by MALDI-TOF, DNA was isolated followed by WGS (Illumina MiSeq platform) and MLST typing by Kleborate v2.0.0 [22, 65].

3.1.3 SCAI sweep samples for the evaluation of phylogeny relation by StrainGE

Culture-enriched metagenome sequencing data of SCAI sweep samples used in this master project for the evaluation of phylogeny relation by StrainGE were selected from the previous comparative study of Kp detection methods [136, 137]. In the comparative study of KpSC detection methods, a subset of 103 samples from the T7-study were selected for WMS based on Kp culture positive (n=52) and culture negative (n=51) and the time between initial collection to freezing the samples (less than 2 days) [137]. Samples were grown on SCAI media and sweep (section 3.2) samples were collected for WMS during this previous study [137].

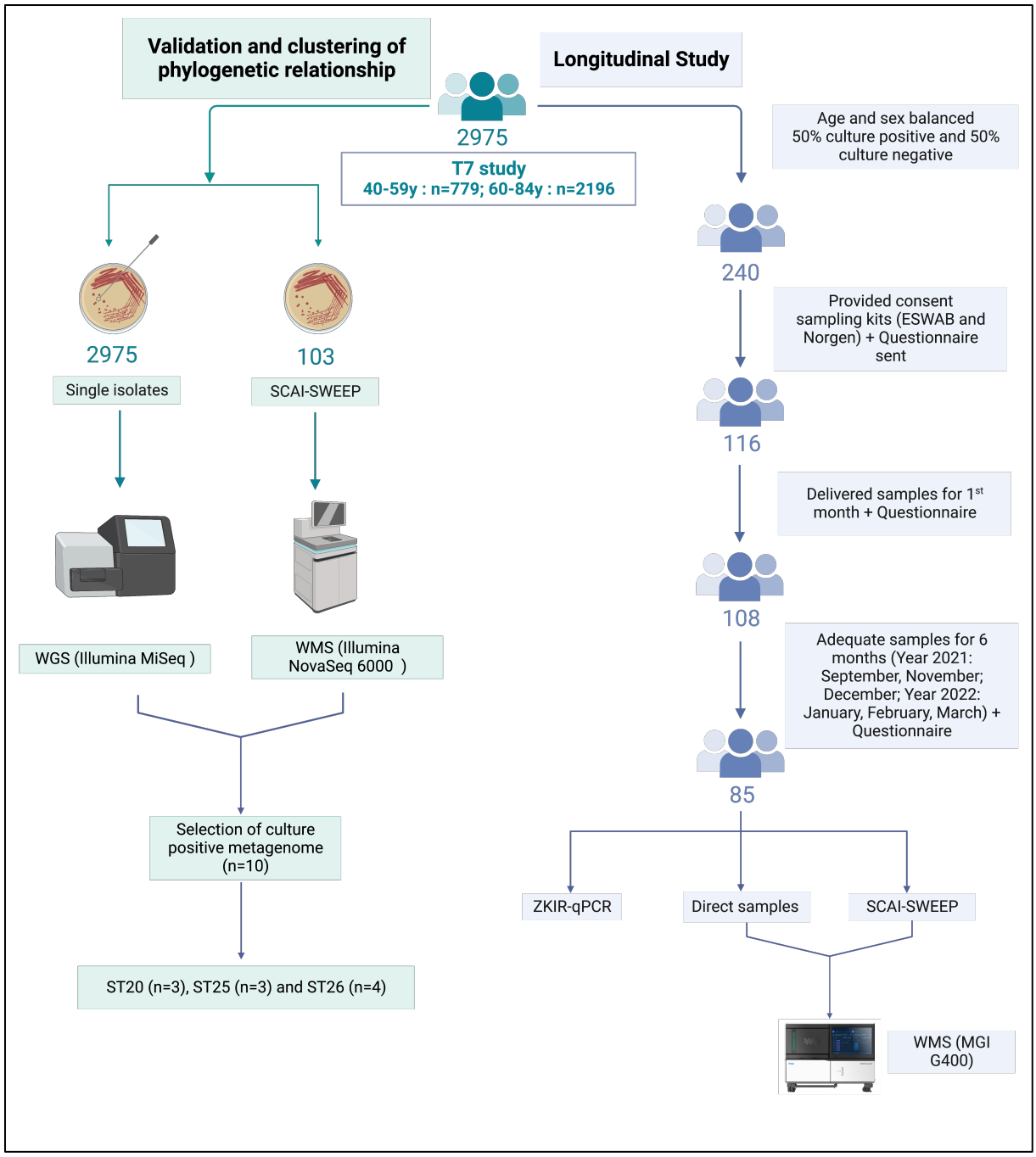


Figure 12. Study populations and experimental outline of the master project. *Created with BioRender.com.*

3.2 Bacterial culture on SCAI media

All KpSC single isolates and sweep samples used in this master's project were cultured from human faecal samples collected in ESwab collection kit using SCAI media. SCAI media was selected due to its selectivity for KpSC [138]. All Kp phylogroups can grow on SCAI, but *E. coli* (a major competitor in faecal cultures)has shown limited growth on this media, which gives a competitive growth advantage for lower abundance KpSC species [138]. This is due to SCAI

containing citrate and inositol as the sole carbon sources. Kp carries the Na^+-dependent citrate carrier (citS) and oxaloacetate decarboxylase (oadGAB), which enables the use of citrate as the sole source of carbon [139, 140]. Most *E. coli* strains, however, lack the citrate transporter and can use citrate as the source of carbon only under anoxic conditions with the presence of oxidizable co-substrate [141]. KpSC colonies grown on SCAI have distinctive morphology of large, moist and yellowish dome shape [138]. The fermentation of inositol lowers the pH of the media due to acidic final product and the pH indicator bromothymol blue turns yellow. On the other hand, fermentation of citrate creates basic environment and the pH indicator turns blue. Most *E. coli* strains grown on SCAI have either small colonies due to limited growth using trace nutrients from faeces or pale-yellow flat colonies or blue colonies (when capable of fermenting citrate) [138]. However, *Klebsiella ozaenae* (now *Klebsiella terrigena*) and *Klebsiella rhinoscleromatis* (now *Klebsiella. planticola*) are also capable of growing on SCAI media as they can utilize citrate and inositol at a lower percentage [138]. Human bowel rarely has *K. ozaenae* and *K. rhinoscleromatis* limiting any influence on analysis of Kp phylogroup in faecal samples [138].

Faecal samples (100 µL) collected in ESwab stool collection kit, were plated on SCAI media (Sigma-Aldrich, cat # 85462– 500 G and I5125-500 G) and incubated for 48 hours at 37ºC. Single KpSC isolates were picked during the T7-study as described by Raffelsberger *et al.* [65]. For SCAI sweep samples, following incubation, samples were collected as described by Lindstedt *et al.* [137]. Briefly, everything grown on the plate was first 'swept' into the centre of the plate and collected with 10 µL inoculation loop (Figure 13). Sweep samples from the previous comparative Kp detection methods study were resuspended in Buffer S1 + 20mg/Ml lysozyme for in-house DNA extraction (see Section 3.3), while samples from the current Kp Longitudinal Carriage study were resuspended in DNA/RNA Shield (Zymo Research Cat. No R1100-250) stored at -80ºC for long term storage.

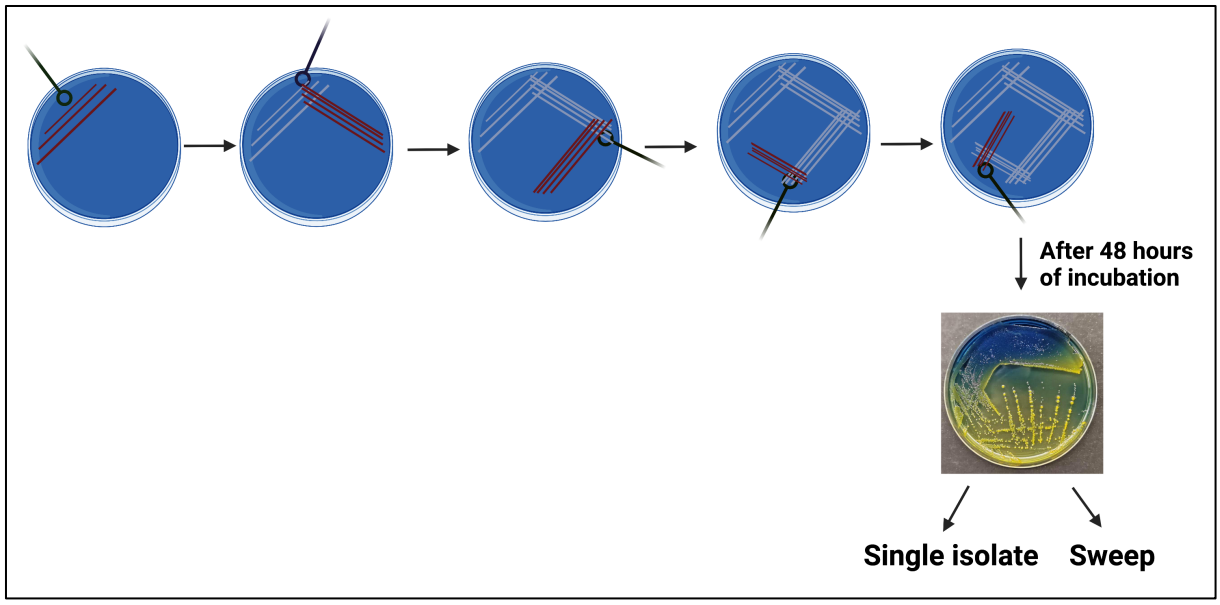


Figure 13. Streaking technique for bacterial culture on SCAI plates. *Created with BioRender.com.*

3.3 Isolation of DNA for qPCR

Extraction of total microbiome DNA by appropriate methodology is one of the crucial steps in the study of the human microbiota since biases introduced during the DNA extraction process can generate erroneous microbial representation [142]. Thus, for faecal samples from Norgen collection kit was used for metagenomic analysis in this study, the whole microbiome DNA was isolated with the Purelink™ Microbiome DNA Purification kit (Thermo Fisher Scientific, catalogue # A29790). This was performed according to the manufacturer's instructions, with some modifications as described by Lindstedt *et al.* [137]. Purelink™ Microbiome DNA Purification kit uses chemical, heat and mechanical (beads beating) methods to lyse the cells as efficient cell lysis ensures proper DNA extraction and has been shown previously to give a good representation of stool microbiota [142]. Faecal samples were thawed on ice and 600 µL of S1 lysozyme solution (Lysozyme, Thermo Fisher, catalogue # 89833 in S1 lysis buffer at final concentration of 20 mg/mL) was added to 200 µL - 300 µL of stool samples, vortexed briefly and incubated at 37 ºC for 10 minutes. This step was done to improve DNA extraction by breaking down bacterial cell wall and dissolving any precipitation developed in the S1 solution [143]. Samples were transferred to Matrix E bead tubes (MP Biomedicals, 6914050) followed by addition of 100 µL S2-Lysis Enhancer and vortexed briefly, incubated at 95 ºC for 10 minutes. At this stage, Matrix E bead tubes were used to homogenize cells rather than bead tubes provided with the kit due to recommended bead sizes required for bead beating by Precellys Evolution Homogeniser (Bertin Technologies, Montigny Le Bretonneux, France).

Bead beating was performed at 6500 rpm for 2x23 seconds followed by centrifugation at 14000g for 5 minutes. The supernatant (400 µL) was transferred to a clean Eppendorf tube avoiding any pellet and debris followed by re-centrifugation at 14000g for 5 minutes to remove residual debris. The supernatant was transferred to a clean eppendorf tube. RNAse A (2 µL) (10 mg/mL, Thermo Fisher, cat # EN0531) was added and incubated at room temperature for 5 minutes to remove any RNA contamination. Upon addition of S3-Cleanup Buffer (250µL) samples were vortexed immediately to remove inhibitors of downstream analysis, such as bile, bilirubin, digestive food and humic acid. Samples were centrifuged at 14000g for 2 minutes and 500 µL supernatant was transferred to clean eppendorf tubes, avoiding pellet and debris. The next step ensured binding of DNA to the column by adding 900 µL of S4-Binding Buffer to each sample and vortexing briefly followed by transfer of sample (700 µL) to spin column-tube assembly and centrifuge at 14000g for 1 minute. Binding buffer contains chaotropic salts, which facilitate binding of DNA to silica surfaces by disrupting the bonds between nucleic acids and protein [144]. Flow through was discarded and remainder of sample (approx. 700 µL) was added to spin column assembly and re-centrifuged at 14000g for 1 minute. The spin column was placed in clean collection tube and 500 µL S5-Wash Buffer (with added ethanol) was added to spin column. Samples were centrifuged at 14000g for 1 minute. Flow through was discarded and re-centrifuged at 14000g for 30 seconds to remove residual S5-Wash Buffer. Spin column was placed in LoBind microcentrifuge tube, followed by addition of S6-Elution Buffer (50 µL). Samples were incubated at room temperature for 1 minute and centrifuged at 14000g for 1 minute. Purified DNA was in collection tube and spin column was discarded. Extracted DNA was stored at -20ºC for further analysis. All centrifugation steps were done at room temperature.

3.3.1 Quantification of DNA by Qubit

The concentration of DNA was measured by Qubit™ 1X double-stranded DNA (dsDNA) high-sensitivity (HS) kit using the Qubit® 3.0 fluorometer (Thermo Fischer Scientific, catalogue # Q33231) according to the kit protocol. Target-selective DNA dye in the kit emits fluorescence only when bound to dsDNA. The Qubit™ dsDNA HS assay is designed to specifically detect double-stranded DNA over RNA or single-strand DNA (ssDNA) in samples with low concentrations of DNA (0.005–120 ng/µL). This allowed the use of only 1 µL of the sample for concentration measurement, minimizing the amount of sample needed for the assay [46]. Qubit was the choice for quantifying DNA due to its sensitivity (10 pg/mL) and specificity towards dsDNA[46].

3.3.2 Quality assessment of DNA by Nanodrop

The quality of the DNA was assessed by Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham USA) according to the manufacturer's instructions. The detection area for dsDNA by Nanodrop is 2 ng/µl to 15,000 ng/µl. Nanodrop reports the quality of the DNA as the ratio of A_{260}/A_{280} [145]. The maximum absorption occurs at around 260 nm for DNA and 280 nm for proteins. The ratio of absorbance at 260 nm and 280 nm for pure DNA is ~1.8 and ~2.0 for pure RNA [146]. Nanodrop also reports A_{260}/A_{230}, which indicates phenol, EDTA, or carbohydrate contamination. The standard range of A_{260}/A_{230} for DNA and RNA is 2.0 to 2.2 [146]. Interpretation of absorbance from Nanodrop is in Table 3 [147].

Table 3. Interpretation of absorbance from Nanodrop

Feature	Value	Indications
A_{260}/A_{280}	< 1.7 for DNA and <1.9 for RNA	Protein contamination
A_{260}/A_{230}	< 2.0 for DNA and RNA	phenol, EDTA, or carbohydrate contamination. Improper washing steps.

3.4 Detection of KpSC by ZKIR (zur-khe intergenic region) quantitative PCR (qPCR)

Quantitative PCR (qPCR) has been used as a powerful technique to rapidly detect and quantify microorganisms from complex compositions like environmental and faecal samples due to easy standardization, high sensitivity and high throughput [148]. qPCR gives information of the amplification of each cycle in real-time by using fluorescence reporters. Currently, there are several fluorescence reporters used in qPCR, such as TaqMan Probe [149], Hybridization probes [150], LightUp probes [151] and SYBR Green [152]. SYBR Green has become popular due to the simplicity of its application and cost-effectiveness. SYBR Green does not require probe incorporation into the primer design. SYBR Green binds only to the minor grove of double-strand DNA and is not target-specific (Figure 14a). The fluorescence signal increases with the increase of DNA amplicon in each cycle. The number of cycles in which the fluorescence signal reaches above the background signal is known as the threshold cycle (C_T) [153]. The absolute quantity of the target DNA can be obtained through C_T value and a standard curve (Figure 14b) of serially diluted standard samples with known concentrations or copy numbers [154, 155]. T_m, melting temperature (the temperature at which the double strand is separated and the dye dissociates resulting decrease in fluorescent signal) is an important

parameter in qPCR to check the specific amplification. Any deviation from the specific T_m range would indicate primer dimer formation or a non-specific amplification [156].

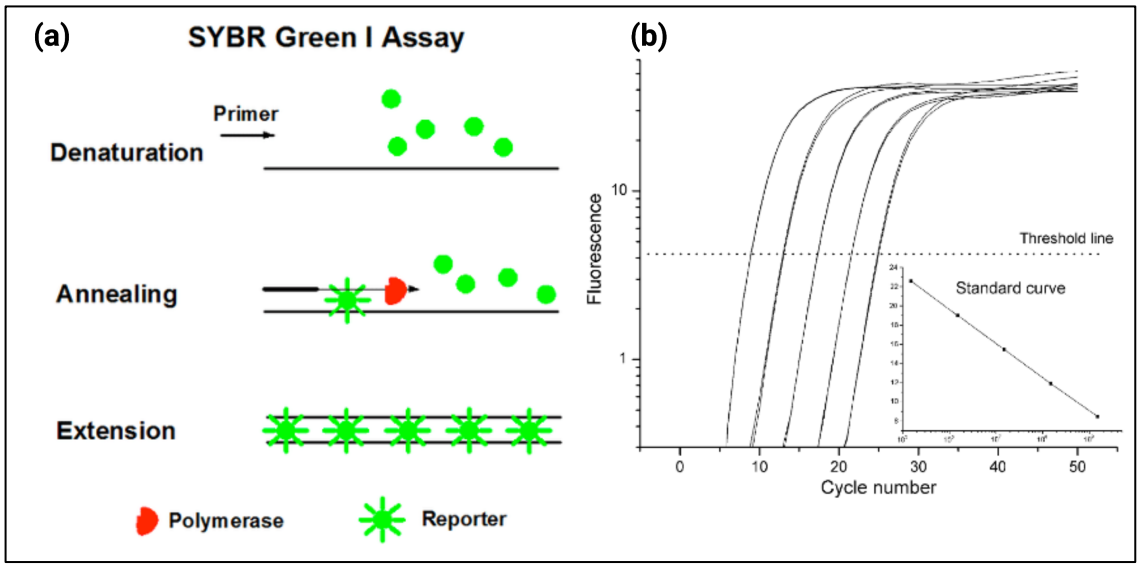


Figure 14. SYBR green assay and standard curve for qPCR. *(a) Mechanism of SYBR Green* [157]. *(b) Typical standard curve and amplification plot in qPCR.* [154]

ZKIR qPCR was selected for qPCR detection in our samples because ZKIR-qPCR is a novel qPCR approach to detect KpSC in environmental and faecal samples [137, 158, 159]. ZKIR-qPCR overcomes the existing challenge of specificity for the detection of KpSC by qPCR as it targets a unique region of KpSC known as zur-khe intergenic region (ZKIR). ZKIR is highly specific and conserved in all members of the KpSC and has not been identified in any other species, even closely related *Klebsiella* species [158]. ZKIR is 249 bp non-coding region residing in between *khe* (annotated as a putative hemolysin) and *zur* (zinc uptake regulator) genes. ZKIR-qPCR uses primer pairs that amplify a 78 bp sequence with melting temperature in the range between 79.9ºC to 80.7ºC (Figure 15) [137, 158].

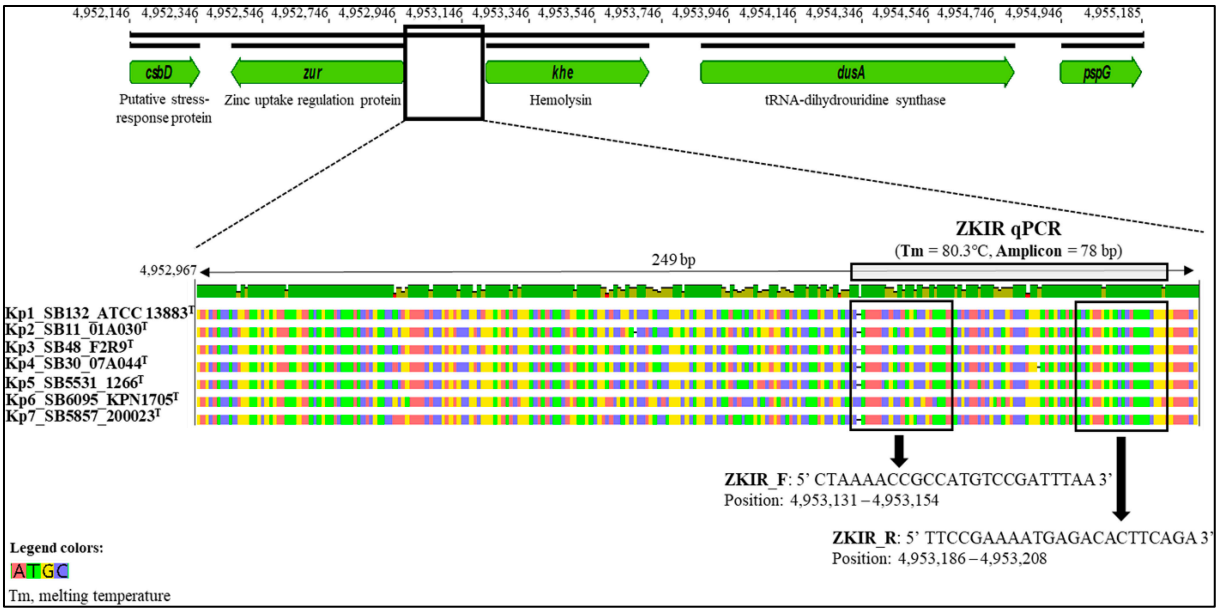


Figure 15. ZKIR on Kp genome (ATCC 13883) with primer and amplicon details. *Boxed area expands to show the detail of the location of ZKIR primers and amplicon region (78bp) specific for KpSC.* [158]

3.4.1 ZKIR-qPCR reaction conditions and primers

All qPCR reaction conditions and primers used were as described by Barbier *et al.* [158]. The final volume of each reaction mixture was 20 µL and contained 2 µL of each primer (final concentrations, 300 nM, ZKIR forward primer: 5′-CTA-AAA-CCG-CCA-TGT-CCG-ATT-TAA-3′, ZKIR reverse primer: 5′-TTC-CGA-AAA-TGA-GAC-ACT-TCAGA–3′), 10 µL of Takyon Low Rox SYBR MasterMix dTTP Blue (Eurogentec, Belgium), 2.5 µL (total 25 ng of human faecal microbiome DNA or 10ng/uL of human microbiome DNA per qPCR reaction) of template DNA, 0.5 µL of T4 gene 32 protein (Sigma-Aldrich) and 3.0 µL of PCR-grade water. Cycling conditions for qPCR were as described in Table 4. All the qPCR assays were performed on 7500 real-time thermocycler (Applied Biosystems, Life technologies, Waltham, USA). Samples were tested in technical triplicate and a minimum of two positive replicates with a *T*M between 78.3°C and 80.4°C and Cq < 40 were considered as positive to minimize false positive detection. Chronic carriers were defined as being KpSC positive by the ZKIR-qPCR for all of the six sampling points throughout the six months study period, transient carriers were defined as being KpSC positive for at least once, but less than six months and non-carriers were defined as being negative for KpSC for all six months.

Table 4. ZKIR-qPCR cycles and conditions

Steps	Temperature (ºC)	Time (minutes)	Purpose
Holding	95	03:00	Enzyme activation
Cycling x40	95	00:10	Denaturation
	60	01:00	Annealing (data collection)
Melt curve	95	00:15	Dissociation stage
	60	01:00	Program step and hold T increment
	95	00:30	+0.3C

3.4.2 Standard curve

Genomic DNA (gDNA) from K47-25 (GenBank assembly accession: GCA_001462555.1) was used to prepare the standard curve [160]. Seven five-fold dilutions of K47-25 gDNA with final genome copies per reaction: $2.5x10^5$, $5x10^4$, 10^4, $2x10^3$, 400, 80, 16 and 3 were used for the standard curve. Genome copies per reaction were calculated according to the following

$$genome\ copy\ number = \frac{(mass\ of\ input\ DNA\ in\ ng)\ x\ (6.022\ x\ 10^{23}\ molecules/mole)}{length\ of\ genome\ in\ bp\ x\ 660\ g/mol\ x\ 10^9\ \ ng/g}$$

[161]

$$where,\ length\ of\ Kp\ genome = 5.5\ x\ 10^6\ bp$$

All the samples and standards had three technical replicates. The threshold for fluorescence signal was set at 1.0 and analysis was done by 7500 real-time PCR Analysis Software v2.3 (Applied Biosystems, Life Technologies, Waltham, USA). The standard curve was considered acceptable if R^2 (coefficient of determination) was greater than 0.98 and E (amplification efficiency) was greater than 90%. R^2 value greater than 0.98 indicates that there was no pipetting error and E value greater than 90% indicates the proper doubling of targeted DNA during each cycle [162]. Melting temperature (T_m) was checked from the melting curve for the confirmation of specific target amplification.

3.5 WMS analysis from faecal and culture-enriched samples

WMS was used for the identification of KpSC-strains by StrainGE in the current KpSC longitudinal study from faecal samples (n=573) collected by the Norgen stool collection kit. Moreover, all the KpSC positive and selected KpSC negative samples, identified by ZKIR-qPCR, were grown on SCAI for culture-sweep enrichment (n=353) and sent for WMS analysis to extract as much KpSC-strain information as possible by StrainGE. Although, the sequencing

data of the sweep samples was not available during the master project and could not be analysed.

Isolation of DNA for WMS was performed using MagPure Stool DNA LQ Kit (catalogue # D6364-01) by MGI Tech Co (Riga, Latvia). WMS analysis of faecal and culture-enriched samples from the longitudinal study was sequenced on the G400 platform, MGI Tech Co (150bp short reads and sequencing depth of 20 million paired-end reads). MGI sequencing takes advantage of the DNBSEQ^{TM} technology, which uses DNA nanoballs prepared by rolling circle amplification (RCA). RCA uses high-fidelity polymerase and original DNA circle is amplified in each cycle ensuring lower amplification error rate. The use of RCA in DNBSEQ^{TM} technology provides advantages of low amplification error rates over the other existing sequencing platforms [163]. Prior to sequencing, samples went through library preparation by MGIEasy FS DNA Library Prep Set v2.1 (MGI Tech Co, catalogue# 1000005254) on the 7-MGISP-960 automated library preparation system (software version: V1.2.0.163, automation version: V1.0). MGIEasy FS DNA Library Prep Set uses advanced Adapter Ligation technology and High-fidelity PCR Enzymes to efficiently increase library yield and conversion rate [164].

3.6 Bioinformatic analysis

3.6.1 Pre-processing of the data

Raw data from the sequencing procedure needs to undergo thorough preprocessing steps before the downstream analysis can be performed. The raw reads contain adapters ligated during the library preparation step, host DNA and bad-quality reads, which might affect downstream analysis. Faecal metagenomic sequences used in this master project for strain analysis and sweep samples for phylogenetic analysis did not require any pre-processing. However, the single isolates from the cross-sectional study used for phylogenetic analysis in this project went through pre-processing and assembly as described by Raffelsberger *et al.* [65]. Additionally, samples from the current longitudinal study also went under pre-processing before taxonomic classification with Centrifuge (version 1.0.4) with the default database, p_compressed+h + v along with Kraken 2 (version 2.1.2) and Bracken (version 2.6.1) with the MiniKraken DB_8GB v202003, performed by one of our colleagues (Dorota Julia Buczek, Postdoc, UiT, Norway) as described by Lindstedt *et al.* [137, 165-167]. The taxonomic classification output file generated from Bracken was used in this master project for the calculation of alpha diversity index

(Shannon index). The pre-processing of the sequences from the current longitudinal study is described below.

3.6.1.1 Quality assessment and filtering by FastP

MGI sequencing platform provided sequences for forward and reverse reads (150bp) of each sample in FASTQ format. FASTQ is a text-based format storing information about the raw sequences and corresponding quality scores of each base [168]. Quality assessment and trimming of the bad-quality reads were performed by fastp (version 0.23.2) [169]. fastp is an automated quality control and preprocessing tool for FASTQ files [169]. It assesses the before and after filtering quality of the sequence, trims adapters, filters bad reads, corrects bases and performs ultra-fast deduplication. fastp is faster than other preprocessing tools like Trimmomatic and Cutadapt [169]. Preprocessing was performed by using the default parameters of fastp.

3.6.1.2 Removal of duplicates by Clumpify

During the sequencing process, duplicates may occur, including PCR duplicates, optical duplicates and tile-edge duplicates. These duplicate reads were removed by Clumpify, package of BBMap version 38.79 [170]. Clumpify groups overlapping reads into clumps and maximizes the compression of the file, making the downstream analysis faster [170].

3.6.1.3 Removal of host contamination by FastQ Screen

DNA was isolated from human faeces. Thus, the isolated DNA may contain residuals of human DNA. Human DNA was removed by FastQ Screen (version 0.14.0) to avoid confidentiality risk and any problem during downstream analysis by screening reads against the GRCh38 assembly (accession number GCF_000001405.39) [171].

3.6.1.4 Synchronization of the reads by BBrepair

The paired reads can become unsynchronized during the previous processing steps. Unsynchronized paired-end files are often not accepted by different downstream analysis tools. Thus, unpaired reads were synchronized by the repair.sh tool of BBMap (version 38.79) [170].

3.6.2 Alpha diversity

Alpha diversity was analysed to observe the difference of microbiota diversity between KpSC non-carriers, transient carriers and chronic carriers in the longitudinal study. Alpha diversity termed by Whittaker (1972) is the diversity of taxa within a sample [172]. It is described by the richness (number of taxa) and/or evenness (the relative abundance of the taxa) [173]. In this

master project Shannon index was used for alpha diversity metrics as Shannon index measures both richness and evenness [174]. R package Vegan 2.6-4 was used to analyse Shannon index from the taxonomic classification output file at the species level generated from Bracken (version 2.6.1) with a cut off value of 0.1% (relative abundance) (Figure 16). Command lines used are in Appendix 2.

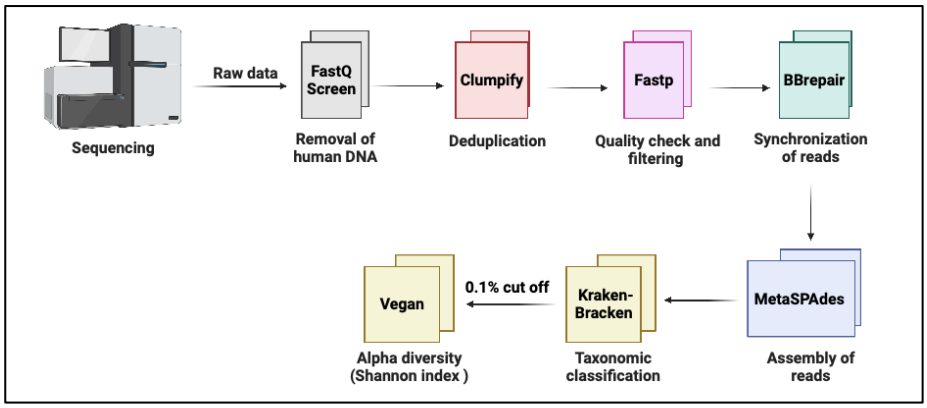


Figure 16. Workflow of alpha diversity analysis. *Created with BioRender.com.*

3.6.3 Evaluation of StrainGE for phylogenetic relationship analysis

Phylogenetic analysis was performed on single isolates from the T7-study and compared to phylogenetic analysis of SCAI-Sweep samples from the comparative study of KpSC detection methods by Lindstedt *et al.* (Figure 17, Table 5) [137]. This was performed to investigate the relationships between strains with the same sequence types and to assess the ability of StrainGE in tracking clonal transmission using WMS [65, 137]. Samples were selected based on the availability of both single isolate and sweep sample from the same participant and same sample in the T7-study. A phylogenetic tree gives information about the ancestor from which a descendent arises. A typical phylogenetic tree consists of tips, nodes and branches [175]. The clade, tips, nodes and branches represent the group of organisms with a common ancestor, the descendent taxa such as species, common ancestor of the descendants and relationship between ancestor and descendants, respectively [176].

Table 5. Sample details for phylogenetic relationship analysis

	ST20	ST25	ST26
SCAI-SWEEP	T7-012-Sweep[a]	T7-127-Sweep	S-26-Sweep
	T7-037-Sweep	T7-308-Sweep[b]	T7-19-Sweep
	T7-463-Sweep	T7-451-Sweep	T7-384-Sweep
			T7-408-Sweep[c]
Single Isolate (Reference)	T7-012-Isolate[a]	T7-308-Isolateb	T7-408-Isolate[c]

[a] Same sample from one individual, cultured on SCAI media and collected as single isolate and sweep.
[b, c] Similar condition as described above.

Phylogenetic tree for SCAI-Sweep samples (ST20, ST25 and ST26) from the comparative study of KpSC detection methods was constructed using the subcommand 'tree' of StrainGR. StrainGR constructs the phylogenetic tree based on distance matrix by neighbour joining (NJ) method [136, 137]. A core SNP alignment was also performed for single isolate (ST20) from T7-study using **Snippy** (version 4.6.0) (https://github.com/tseemann/snippy). SNP distance matrix was calculated by **snp-dists** (version 0.8.2) from the core alignment file (https://github.com/tseemann/snp-dists). Maximum likelihood (ML) phylogeny of the core genome alignment was constructed using **RAxML** (Randomized Axelerated Maximum Likelihood, version 8.2.12) with 1000 bootstrap [177]. All the phylogenetic trees were visualized by **FigTree** (version 1.4.4) (http://tree.bio.ed.ac.uk/software/figtree/). All the command lines used are in Appendix 2.

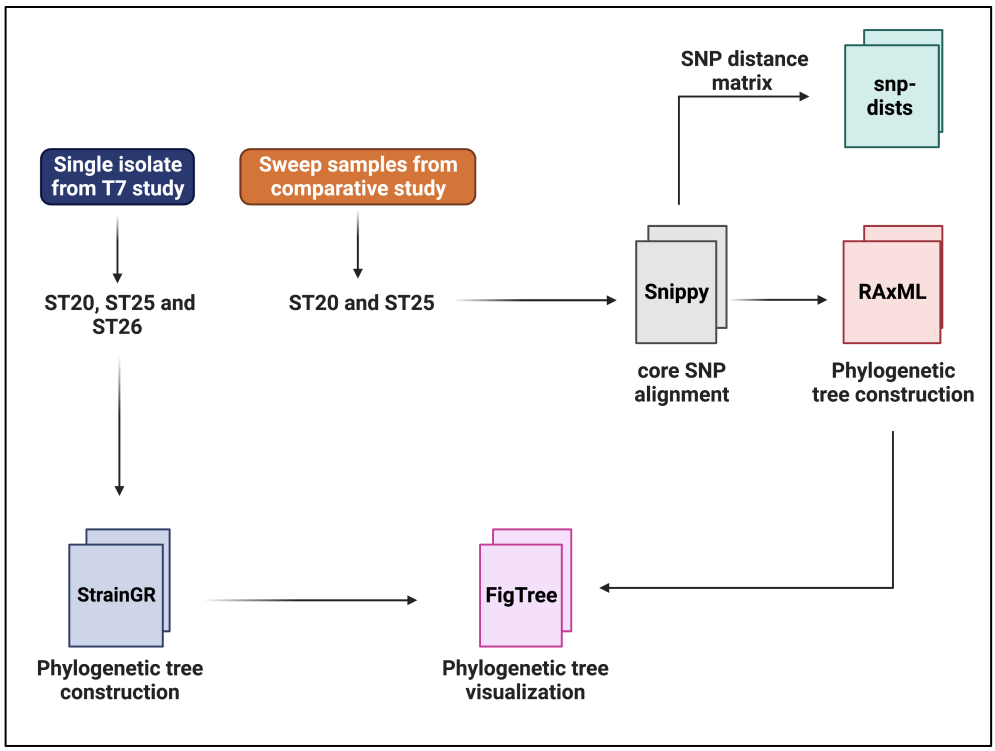


Figure 17. Workflow of phylogenetic relationship analysis. *Created with BioRender.com.*

3.6.4 Identification of strains by StrainGE and Kleborate

Strain identification for KpSC was performed by Strain Genome Search Tool (StrainGST), one of the key components of StrainGE, followed by Kleborate (version 0.4.0 beta) (Figure 18) [22, 136]. StrainGE is a toolkit for characterizing and tracking low abundant strains in complex samples like metagenome samples and is accurate to at least 0.1% relative abundance [136,

137]. StrainGST provides information about the nearest matched reference genome (s) from a customizable database.

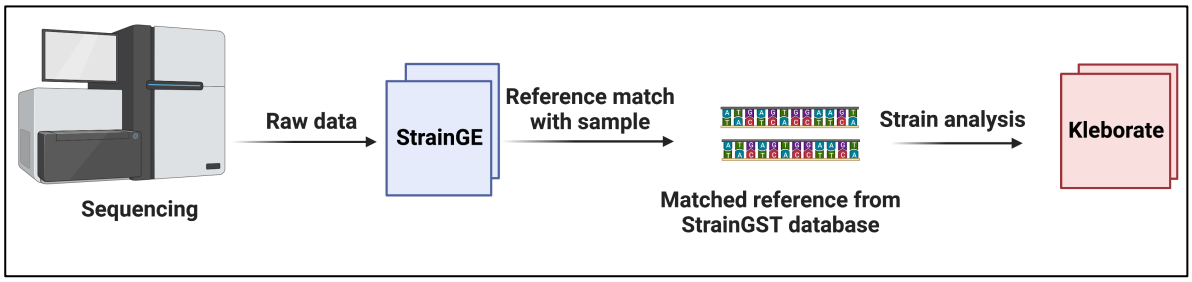


Figure 18. Workflow for identification of KpSC strains. *Created with BioRender.com.*

The database for this study included KpSC genomes (n = 3922). Kp refseq (NCBI) genomes (n = 1305) were downloaded on 18/10/2022 using ncbi-genome-download (version 0.3.1) tool (Appendix 2). Additionally, KpSC genomes from the T7 study (n=484) and the SPARK study (n = 2109) were added to build the customized database for StrainGST (Table 6) [65, 72, 178]. Reference sequences were kmerized (k-mer size = 23) and kmer sets were compared for similarity followed by clustering of references. The threshold value of clustering was set to 0.95 according to previous study by Lindstedt *et al.* to avoid co-clustering of related ST-types such as ST11 and ST258 [137]. StrainGST is designed to select only one representative reference genome from the cluster. At the end of StrainGST pipeline, pangenome of k-mer database was created from the selected reference genomes. Detailed command lines used are in Appendix 2. WMS raw reads were kmerized (k-mer size = 23) and StrainGST reported the closest reference genome (s) based on the comparison of kmers. StrainGR, the second component of StrainGST, can characterize the strains within metagenome samples. StrainGR reports the ST and nucleotide level comparison of the strain (s) present in the sample with the reference including single nucleotide variant (SNV) [136].

Table 6. Details of additional genomes to create the in-house database of StrainGST

KpSC	Number of genomes	
	T7	SPARK
Kp	303	1705
Kp2	31	76
Kp3	134	279
Kp4	16	49

However, the main purpose of this master project was strain analysis at the ST level in the longitudinal samples and strain analysis by Kleborate was faster than StrainGR. Kleborate is a genotyping tool which uses seven locus MLST scheme to strain identification [22]. Strain analysis by Kleborate can also be performed from metagenome samples, however, it requires metagenome-assembled genomes (MAGs) [22]. Hence, matched references from our custom StrainGST database were run through Kleborate (version 0.4.0 beta) for identifying ST to reduce analysis time in this master project. All the command lines used are in Appendix 2.

Samples were selected based on the cut-off value of 0.05% relative abundance of Kp from the taxonomic classification by Centrifuge (version 1.0.4) and participants who delivered samples for at least five months were only selected for further ST analysis.

3.7 Data visualization and Statistical analyses

All the figures and statistical analyses were done by R studio (version 4.2.1). The normality of the dataset was analysed by density plot and Shapiro-Wilk test (Figure S2). A linear mixed effects analysis of the relationship between the Shannon index and KpSC relative abundances (from ZKIR-qPCR) was done by using the lme4 package in R (Appendix 2) [179]. The random effect in the model was participants and fixed effects were Shannon index and KpSC relative abundances. A similar model was used while comparing different KpSC carriage groups where the random effect in the model was participants and fixed effects were Shannon index and carriage groups (Appendix 2). Visual inspection of residual vs fitted values did not show any noticeable deviations from homoscedasticity (Figures S3 and S4). p-values were calculated for the mixed model linear regression models by Likelihood Ratio Tests of the full model with the effect in comparison to the reduced model without the effect using the anova function. The nonparametric Kruskal-Wallis test was done for the association of age between different KpSC carriage types. p-values were adjusted by Dunn's posthoc test with the Bonferroni method for the association of age between different KpSC carriage types. p-value <0.05 was considered significant. All the command lines used are in Appendix 2.

3.8 Computational analysis

All the bioinformatic analyses were performed through Saga supercomputer on the Norwegian academic high-performance computing and storage services maintained by the Sigma2 Norwegian Research Infrastructure Service (NRIS) [180].

3.9 Ethical considerations

The project was approved by the Regional Committee for Medical and Health Research Ethics, North Norway (REK North reference 137064/2020).

4 Results

4.1 Gastrointestinal carriage of KpSC

Gastrointestinal carriage of KpSC was assessed by ZKIR-qPCR targeting KpSC in faecal samples from 108 participants (62% female, median age of 69.0 years with interquartile range (IQR) of 74.0-65.0 years). A total of 595 samples were received for examination from September 2021 to March 2022) 47 years and older. KpSC prevalence varied from month to month, with the lowest prevalence in September (49.5%) and the highest prevalence in December (69.1%) during the six-month study period (Figure 19a, Table S1). Qualitative analysis of the prevalence data showed no strong visual correlation between the prevalence of KpSC in September 2021 and the T7-study (unpublished data, Lindstedt *et al.*, Figure S5). A total of 85 participants delivered all six samples throughout the six months period. The majority were transient carriers (76.5%, n=65) (Figure 19b). However, there were also 9.4% (n=8) non-carriers (defined by ZKIR-qPCR negative for all six months) and 14.1% (n=12) chronic carriers (defined by ZKIR-qPCR positive for all six months) among the 85 participants (Figure 19b and 19c). KpSC abundance significantly increased with carriage duration (unpublished data, Lindstedt *et al.*, Figure S6) and chronic carriers had high and relatively stable KpSC abundances throughout the study period (Figure 19c).

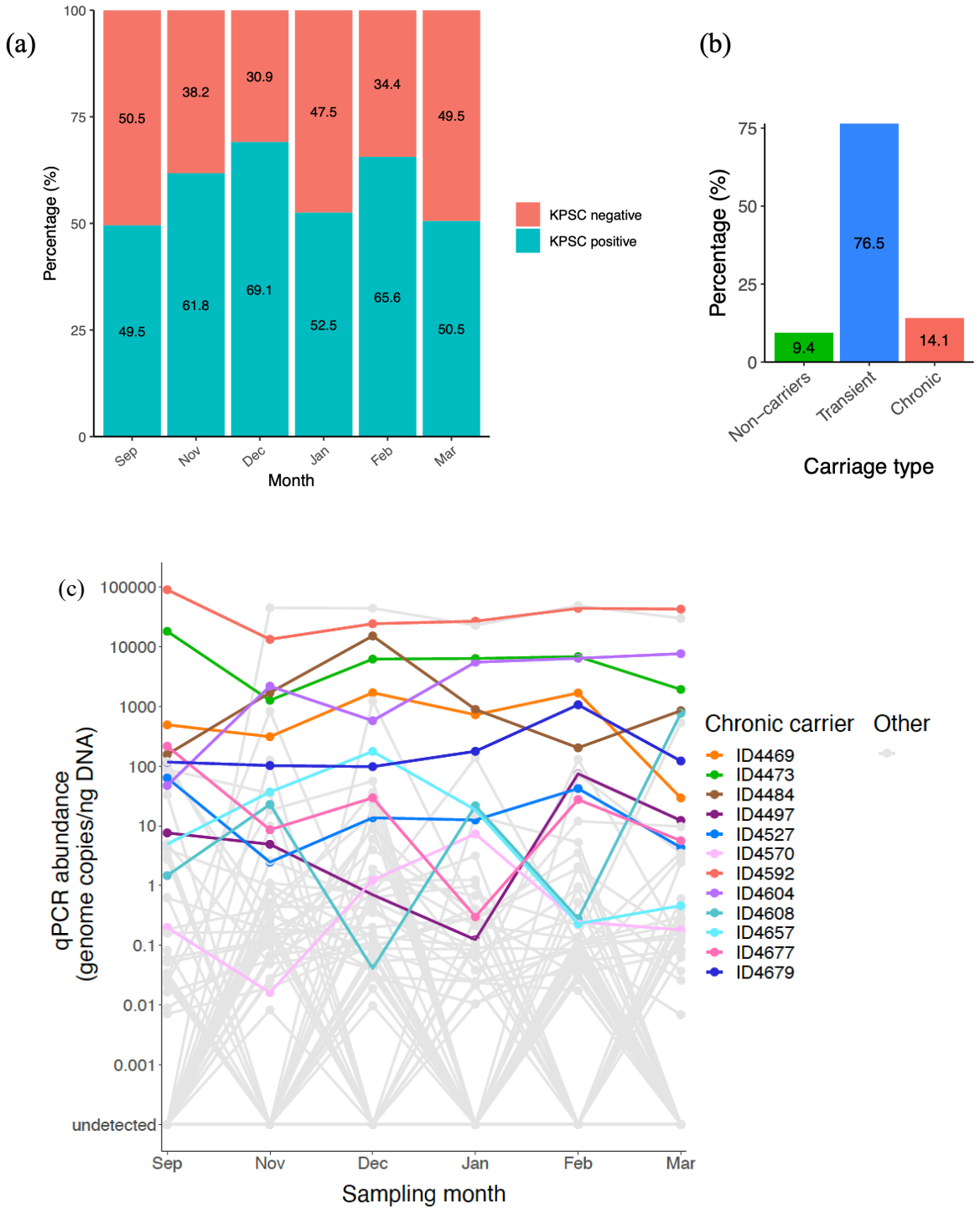
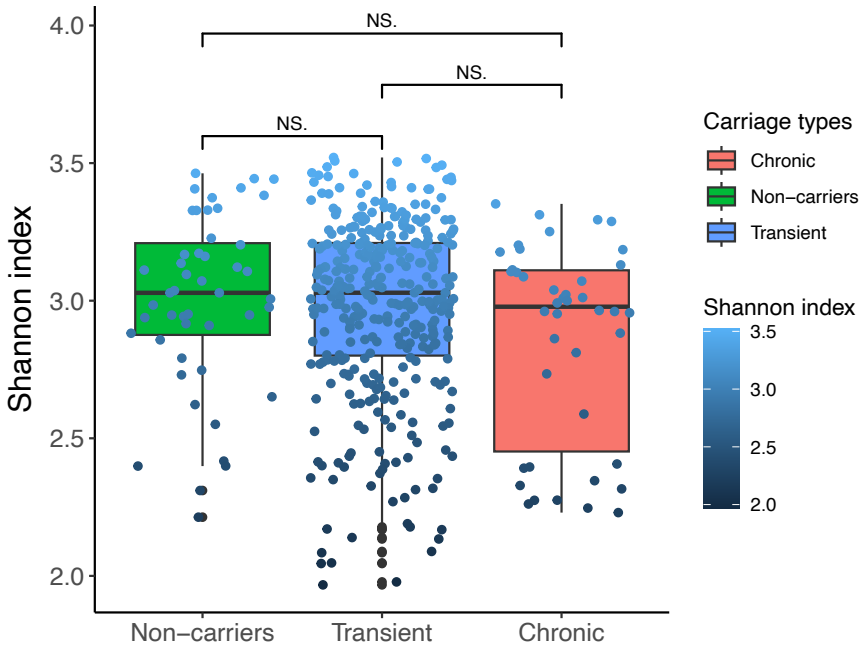


Figure 19. Gastrointestinal prevalence of KpSC. *(a)KpSC prevalence by month (From September 2021 to March 2022) among 108 participants[a] (b) Percentage of non-carriers, transient carriers and chronic carriers during the six-month study period in a total of 85 participants who delivered samples all six months[b] (c) qPCR abundance of chronic, transient and non-carriers. Chronic carriers (n=12) are highlighted in different colour lines and non- and transient carriers are shown in grey colour lines.*
[a]Number of samples analysed was 591 from a total of 595 received, as 4 of the samples failed DNA extraction.
[b]Number of samples analysed was 510 as only 85/108 participants delivered samples for all six months

4.2 Correlation of human gut microbiota diversity with KpSC carriage duration

Members of the human gut microbiota have a complex relationship with each other. Some compete and some help each other for their successful colonisation and survival. Correlation between human gut microbiota diversity and KpSC gut carriage duration was assessed in this study by comparing the Shannon index of each participant within non-carriers, transient carriers and chronic carriers to investigate whether there is any significant difference in gut microbial diversity between the groups. Median Shannon index of non-carriers, transient and chronic carriers was 3.03 (IQR: 2.87-3.22), 3.03 (IQR: 2.80-3.21) and 2.98 (IQR: 2.41-3.11), respectively (Figure 20, Table S2).



Median	3.03	3.03	2.98
IQR	2.87-3.22	2.80-3.21	2.41-3.11

Figure 20. Alpha diversity (Shannon index) in three different KpSC carriage groups*. Shannon diversity index for non-carriers, transient carriers and chronic carriers shows no significant difference in the gut microbial diversity between the groups[a]. The lower border of the box indicates the 25th percentile, the upper border of the box indicates the 75th percentile and the median is indicated by the black line within the box. The top and bottom whiskers indicate the 90th and 10th percentile, respectively. Points outside the whiskers are outliers beyond the 10th and 90th percentile. NS=non-significant (Likelihood Ratio Test by anova function). The median and IQR are shown below the plot.*
[a]Number of samples analysed was 439 (73 participants)/ 510 (85 participants) due to at least one out of the six samples failing the sequencing QC for 12/85 participants. NS = non-significant

Chronic carriers had a slightly lower median Shannon index than non-carriers and transient carriers. However, the Likelihood Ratio Test did not show any significant difference in the Shannon index between the groups (*p*-value: 0.3685 for chronic carriers and non-carriers, 0.3076 for chronic carriers and transient carriers and 0.8252 for non-carriers and transient carriers) (Figure 20).

4.3 Correlation of human gut microbiota diversity with KpSC abundance

The correlation between KpSC abundance (from ZKIR-qPCR) and alpha diversity (Shannon index) was analysed to further observe the effect of gut microbiota diversity on KpSC abundance. Correlation analysis showed a negative correlation between KpSC abundance and Shannon index with a slope of -0.007 (Figure 21a). However, the correlation was non-significant (*p*-value = 0.148) (Figure 21a).

As age over 60 is a known risk factor for KpSC carriage and age also influences microbiome diversity, we looked at the correlation between the Shannon index and KpSC abundance in two age subgroups (age group of <60 y and ≥60 y) [65, 181]. There was a stronger negative association between the Shannon index and KpSC abundance in the age group of <60 y than in the age group of ≥60 y, with slopes of -0.032 and -0.004, respectively (Figure 21b). Although there was no significant association of the Shannon index and KpSC abundance in the age group ≥60 y (*p*-value = 0.423), the age group <60 y interestingly demonstrated a significant negative relationship (*p*-value = 0.005) (Figure 21b).

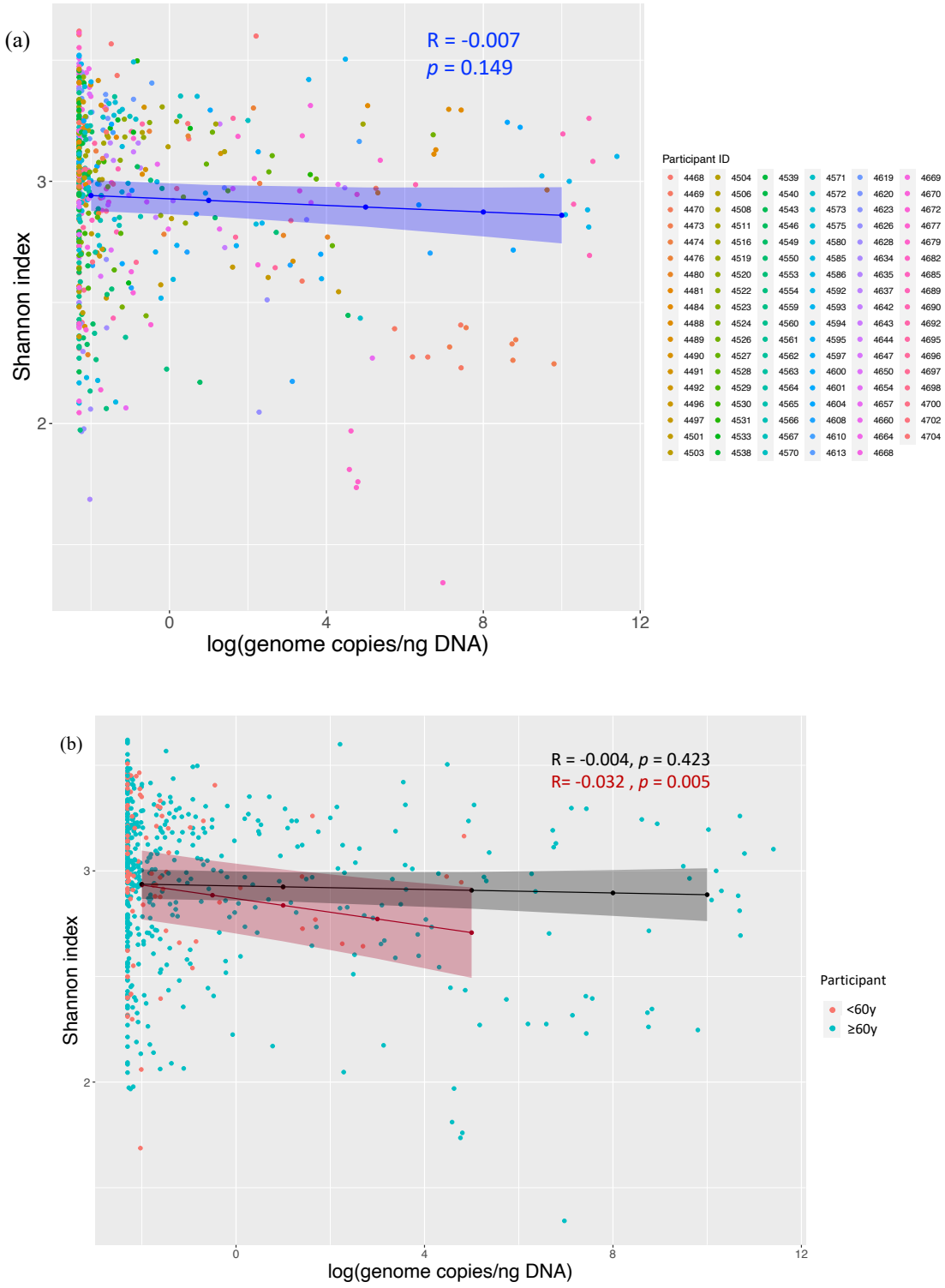
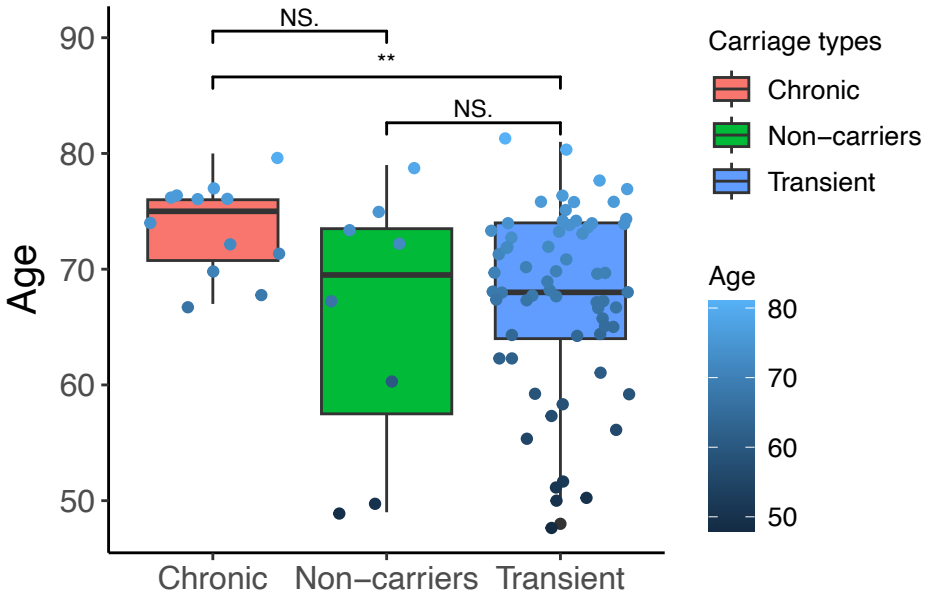


Figure 21. Correlation between alpha diversity (Shannon index) and KpSC abundance in the human gut. *(a) Shannon index vs KpSC relative abundance (genome copies/ng DNA)[a]. The blue line represents the regression line and the blue shaded area indicates the upper and lower 95% confidence interval. (b) Shannon index vs KpSC relative abundance (genome copies/ng DNA) among the age group of <60 y (n=94) and ≥60 y (n=474). The black and red lines represent the regression line and the shaded areas indicate the upper and lower 95% confidence interval. R is the slope. (Likelihood Ratio Test)*
[a]Number of samples analysed was 568/595 received due to 21 samples failing sequencing QC and six samples had insufficient amounts of DNA.for qPCR

4.4 Effect of age on KpSC carriage duration

Age has already been established as an associated risk factor for KpSC gut carriage [65]. Thus, the relationship between age and KpSC gut carriage duration was assessed in this study by comparing the age of each participant between three groups, including non-carriers, transient carriers and chronic carriers, to investigate whether there is any significant difference in age between the groups. The median age of non-carriers, transient carriers and chronic carriers were 69.5 (IQR: 55-74), 68 (IQR: 64-74) and 75 (IQR: 70.5-76) (Figure 22, Table S3). Chronic carriers had higher age than non-carriers and transient carriers. However, a significant difference in age was only observed between chronic and transient carriers (p-value = 0.011). in contrast to chronic carriers versus non-carriers (p-value = 0.068) and non-carriers versus transient carriers (p-value = 1.0).



Median	75	69.5	68
IQR	70.5-76	55-74	64-74

Figure 22. Age in three different carriage groups. *The lower border of the box indicates the 25th percentile, the upper border of the box indicates the 75th percentile and the median is indicated by the black line within the box. The top and bottom whiskers indicate the 90th and 10th percentile, respectively. Points outside the whiskers are outliers beyond 10th and 90th percentile. Median and IQR are shown below the plot[a]. NS=non-significant and ** indicates p value < 0.01 (Kruskal-Wallis with Dunn's post hoc adjusted with Bonferroni method).*

[a] *Number of samples analysed were 510 as only 85/108 participants delivered samples all six months*

4.5 Evaluation of the application of StrainGE in analysing phylogenetic relationships from metagenome sequence

StrainGE based phylogenetic relationships from culture-enriched metagenome samples (Sweep) were analysed to evaluate the suitability and accuracy of StrainGE in strain identification before applying this tool for strain identification. Ten sweep samples representing three different STs (ST20: T7-012-Sweep, T7-037-Sweep and T7-463-Sweep; ST25: T7-308-Sweep, T7-451-Sweep and T7-127-Sweep; and ST26: T7-408-Sweep, T7-19-Sweep, T7-384-Sweep and S26-Sweep) were selected from the comparative study of KpSC detection methods and used to construct phylogenetic relationship by StrainGE. The corresponding single isolates taken from the same sample as on the sweep samples from each ST (ST20: T7-012-Isolate, ST25: T7-308-Isolate and ST26: T7-408-Isolate) from the T7-study were used as references. This was performed to observe whether the single isolate and the same sweep sample from the same individual could be correctly identified from WMS as belonging to the same clade.

Phylogeny of ST20 samples confirmed the presence of T7-012-Sweep and T7-012-Isolate samples in the same clade with no difference in the horizontal distance from the recent common ancestor (Figure 23). The other ST20-containing sweep samples, T7-037-Sweep and T7-463-Sweep, belonged to a different clade than T7-012-Sweep and T7-012-Isolate samples (Figure 23). Phylogeny of ST25 samples revealed the presence of T7-308-Sweep and T7-308-Isolate samples in different clades, although the samples are from the same individual (Figure 24). T7-308-Isolate was placed in the same clade as T7-127-Sweep despite not corresponding to the single isolate clone (Figure 24). T7-451-Sweep belonged to a different clade than the other three samples (Figure 24). Phylogeny of ST26 samples confirmed the presence of T7-408 sweep and isolate samples in the same clade (Figure 25). S26-Sweep was located further away in the tree from the rest of the samples. T7-408-Sweep and T7-408-Isolate were more closely related to T7-19-Sweep and T7-384-Sweep than S26-Sweep (Figure 25).

Phylogeny of core genome alignment from single isolates of ST20 samples (T7-012, T7-037 and T7-463) was analysed to compare the accuracy of the phylogeny of corresponding ST20 sweep samples (T7-012-Sweep, T7-037-Sweep and T7-463-Sweep) created by StrainGE. ST25 (T7-308-Isolate) was also included in the analysis as an outlier to verify the phylogenetic relationship. The core genome phylogeny of ST20 isolates revealed similar phylogeny structures compared to the phylogeny of ST20 metagenome samples created by StrainGE (Figure 26). Single isolates from T7-037 and T7-463 were more closely related to each other than single isolate from T7-012 depicted from the location of the clades (Figure 26). ST25 (T7-

308-Isolate) was included in the analysis as an outlier to verify the phylogenetic relationship. The SNP distance matrices also confirmed fewer SNPs (n= 605) between T7-037-Isolate and T7-463-Isolate than T7-012-Isolate and T7-037-Isolate (n=1678) or T7-012-Isolate and T7-463-Isolate (n=1432) (Table S4). The core genome alignment of single isolates of ST25 and ST26 could not be analysed due to the unavailability of corresponding single isolates of the sweep samples. Phylogenetic relationships of the ST20 single isolates using core genome alignment corresponded to the phylogenetic relationships of the ST20 sweep samples by StrainGE.

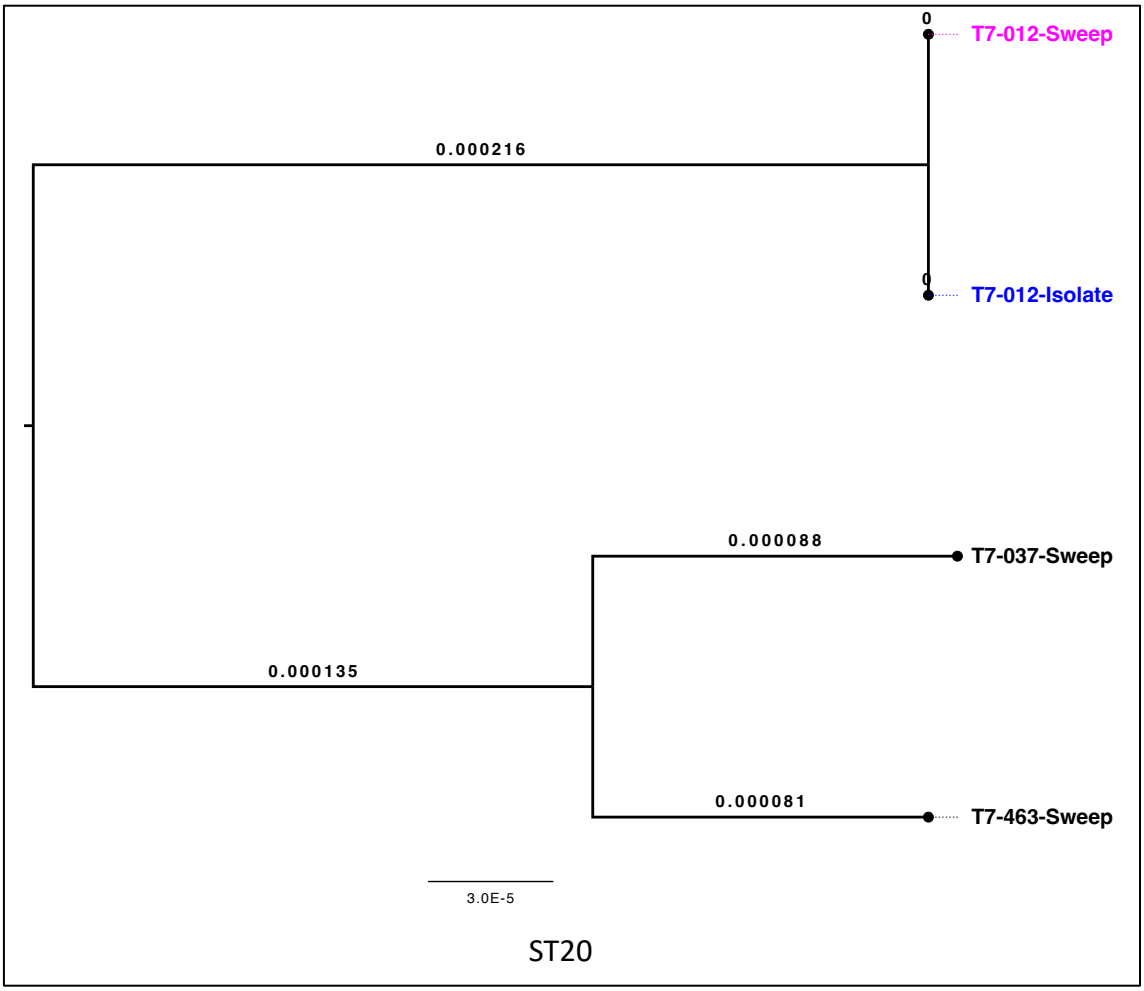


Figure 23. Phylogenetic tree reconstruction of Kp-ST20 from the sweep metagenome and the corresponding WGS single ST20 isolate. *NJ tree constructed by StrainGE. Trees were rooted arbitrarily. Scale bar at the bottom represents the horizontal distance scale.* ■ *Reference* ■ *Sweep sample from same participant as the corresponding single isolate.*

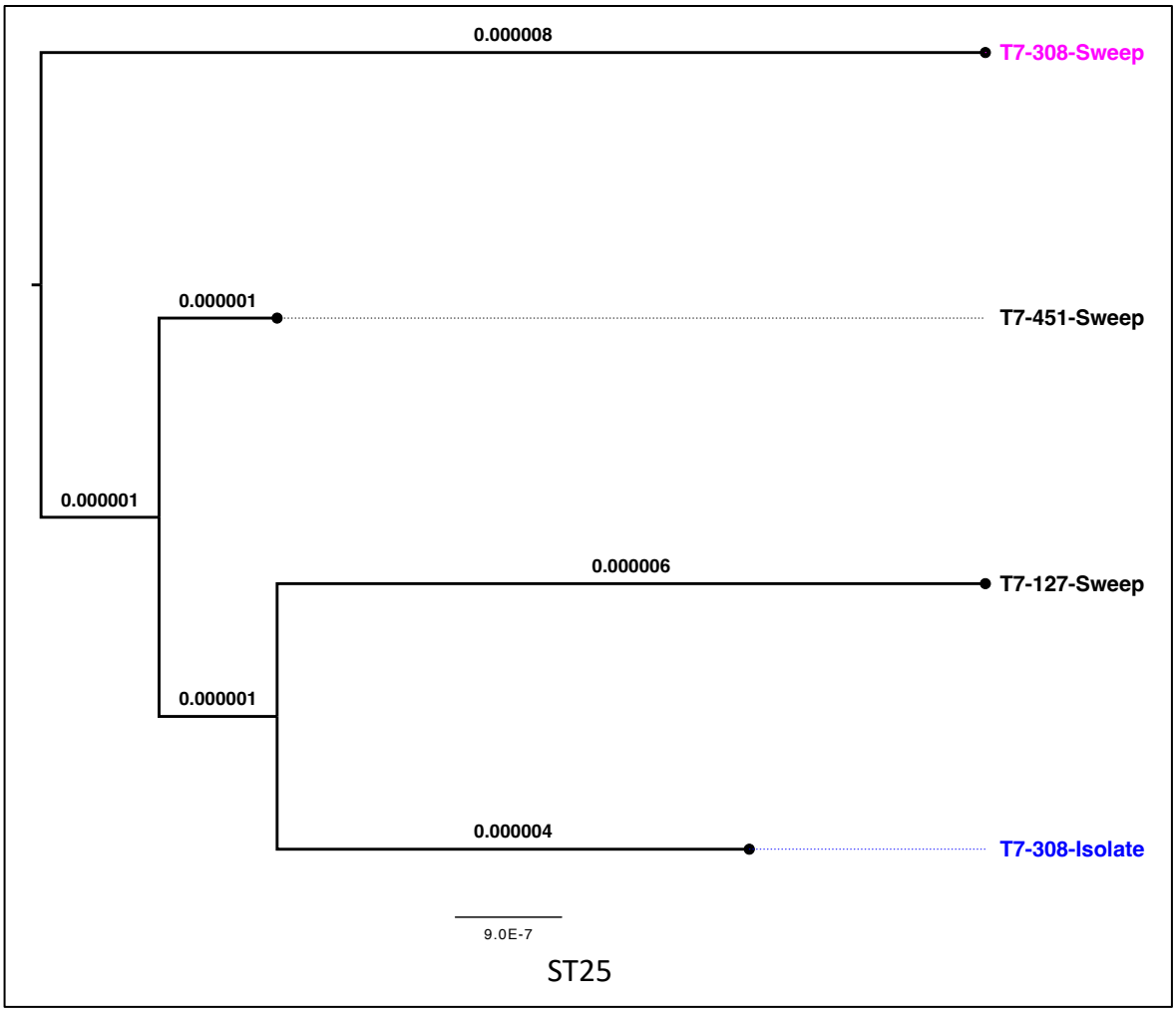


Figure 24. Phylogenetic tree reconstruction of Kp-ST25 from the sweep metagenome and the corresponding WGS single ST25 isolate. *NJ tree constructed by StrainGE. Trees were rooted arbitrarily. Scale bar at the bottom represents the horizontal distance scale.* ■ *Reference* ■ *Sweep sample from same participant as the corresponding single isolate.*

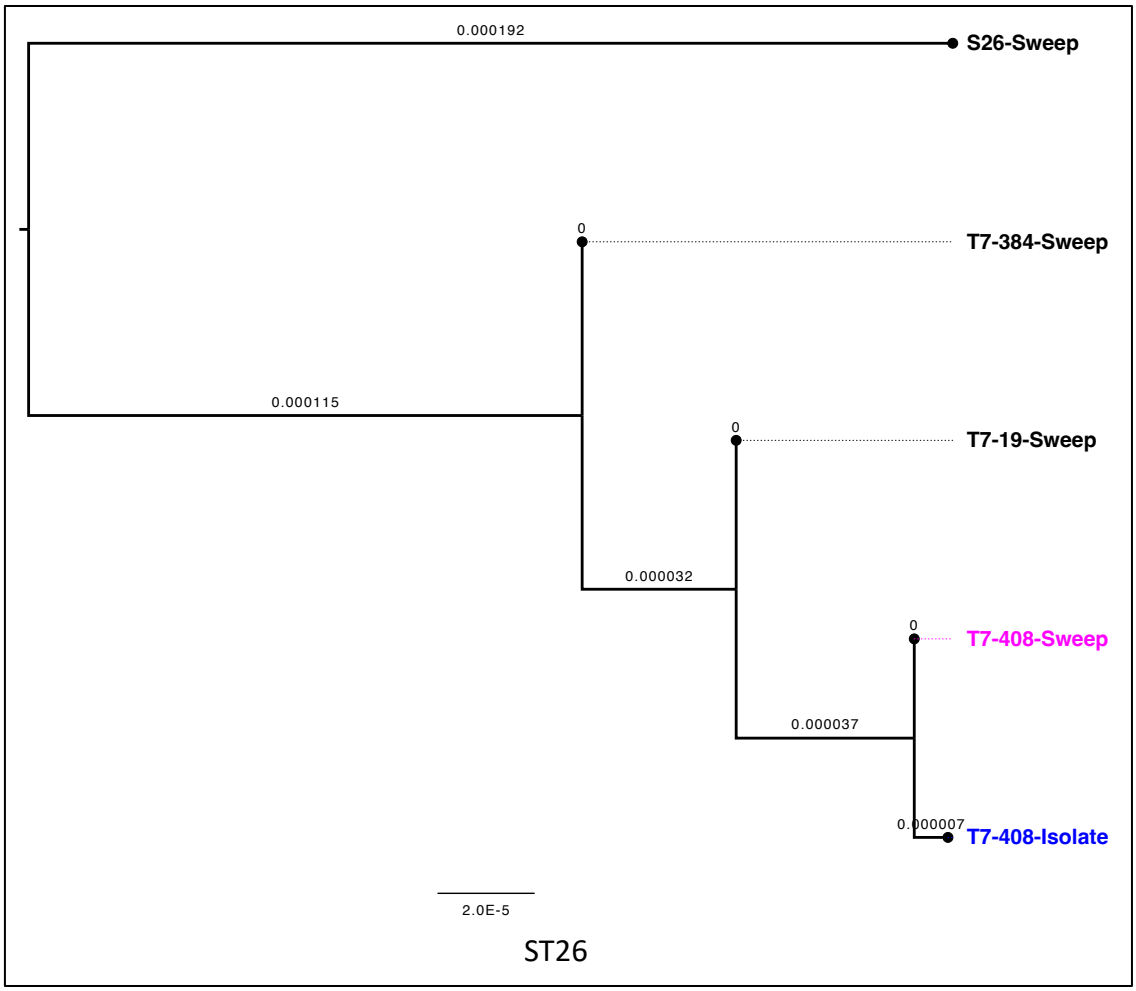


Figure 25. Phylogenetic tree reconstruction of Kp-ST26 from the sweep metagenome and the corresponding WGS single ST26 isolate. *NJ tree constructed by StrainGE. Trees were rooted arbitrarily. Scale bar at the bottom represents the horizontal distance scale.* ■ *Reference* ■ *Sweep sample from same participant as the corresponding single isolate.*

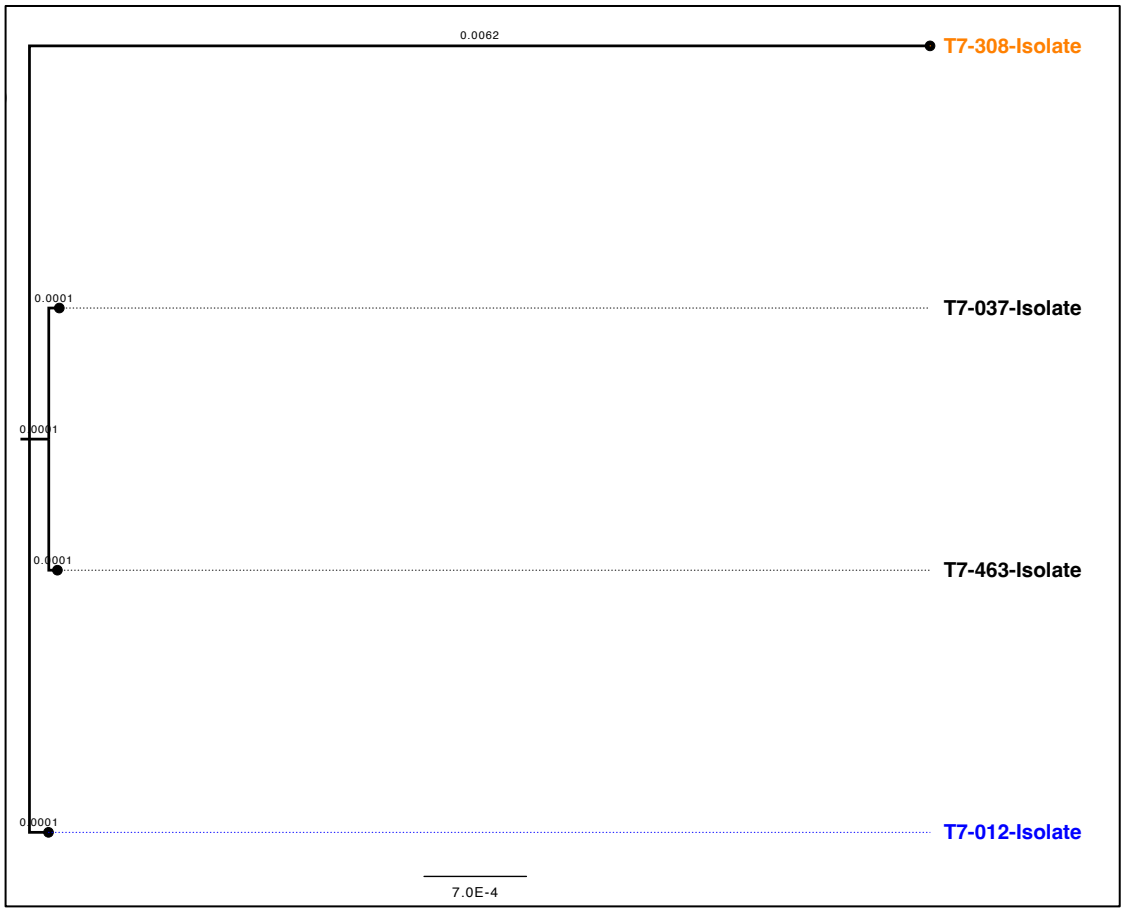


Figure 26. Phylogenetic tree reconstruction of Kp-ST20 from the WGS single ST20 isolates. *ML tree from the core genome alignment of the single ST20 isolates. Trees were rooted arbitrarily. Scale bar at the bottom represents the horizontal distance scale.* *Reference* *Outlier*

4.6 KpSC ST diversity

ST analysis of metagenomes from the seven most KpSC abundant participants (ID: 4469, 4473, 4484, 4592, 4604, 4679 and 4689) identified 13 different STs in 41 samples distributed in two phylogroups (Kp1 and Kp3) (Figures 27 and 28, Table S5). All the participants were positive for KpSC detected by ZKIR-qPCR throughout the six months, except participant ID4689 was negative for KpSC during September 2021. Samples were selected based on ≥0.05% relative abundance of Kp. ID 4484 had Kp1(ST12) and Kp3 (ST360) in September and December 2021. ID 4484 had Kp3 (ST360 and ST2386) also in January and February 2022. Most of the STs were unique for the individual participant except ST12, which was found in two participants (ID 4484 and ID 4604) (Figure 28, Table 7). Three participants (ID: 4469, 4592 and 4473) had single ST over the six-month study period and three (ID: 4689, 4604 and 4484) had multiple STs over the six-month study period. Multiple STs within a single sample were also detected in two participants (ID 4484 and ID 4689). StrainGE could not identify any STs in ID 4679 for

two samples (November 2021 and January 2022) out of six samples and ID 4604 for one sample (November 2021), despite these samples were positive by the ZKIR-qPCR, as the abundances were below the detection limit of StrainGE (0.1% relative abundance). ID 4689 was initially qPCR negative during September, but three different STs (ST643 with one locus variant (LV), ST502 and ST3544) were identified in the following months, with ST643-1LV identified in all of the five following samples. ID 4604 had two different STs (ST12 and ST20) and ID4484 had five different STs (ST12, ST36, ST253, ST360 and ST2386) (Figure 28).

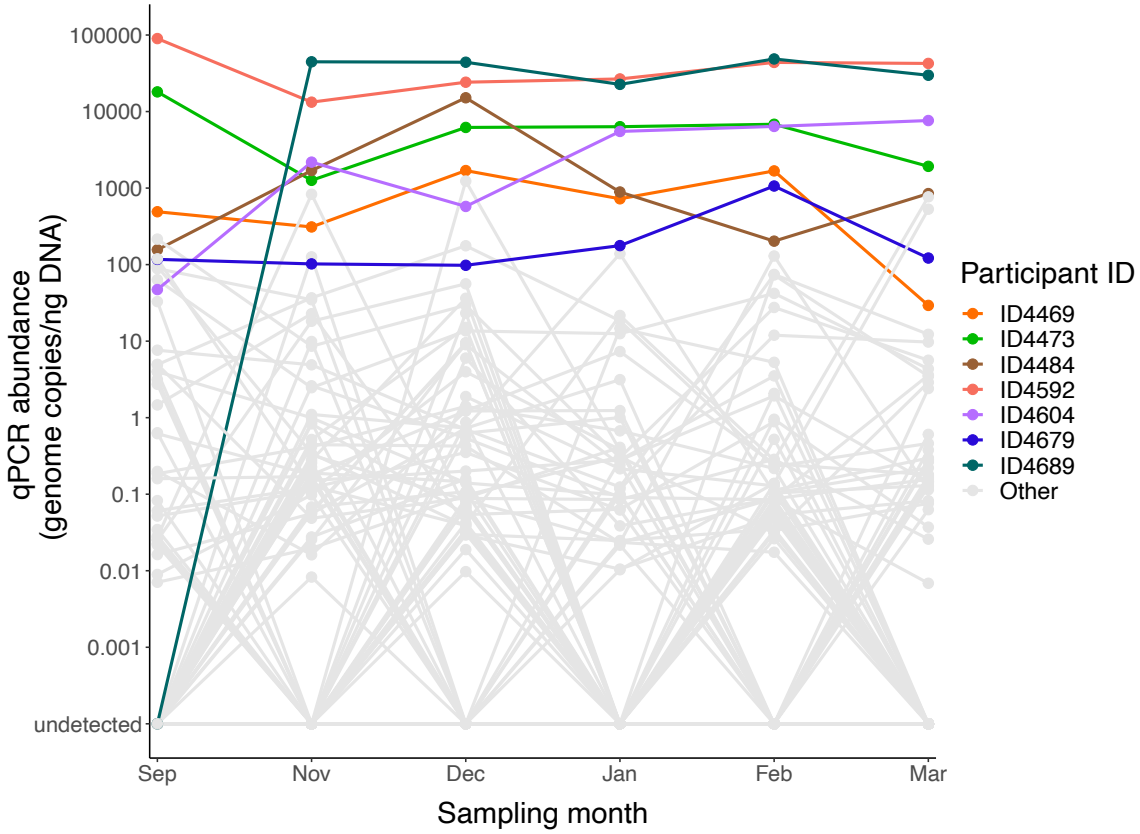


Figure 27. KpSC qPCR abundance in seven highest abundant participants selected for ST analysis by StrainGE and Kleborate. *Seven highest abundant participants are highlighted in different colour lines and others are shown in grey colour lines.*

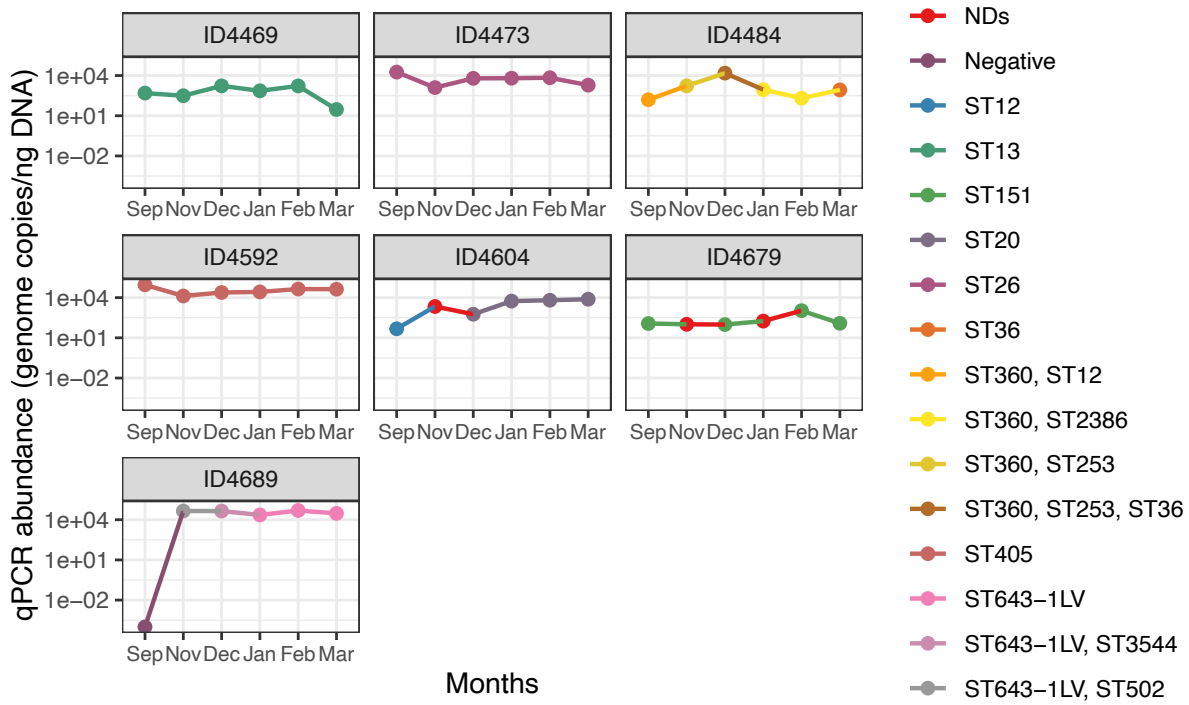


Figure 28. ST diversity in seven participants with highest KpSC abundance. *ST was detected by StrainGST and Kleborate. NDs = ST not detected by StrainGE, Negative = ZKIR-qPCR negative for KpSC.*

Table 7. **Relative abundance of each STs by StrainGE and metadata in the seven highest KpSC abundant participants**

	September	November	December	January	February	March
ID4469	ST13 (0.238%%)	ST13 (0.119%)	ST13 (0.404%)	ST13 (0.1155%)	ST13 (0.598%)	ST13 (0.311%)
ID4473	ST26 (4.72%)	ST26 (0.39%)	ST26 (0.617%)	ST26 (1.084%)	ST26 (1.831%)	ST26 (0.311%)
ID4484	ST360 (0.099%) ST12 (0.196%)	ST36 (0.358%) ST253 (0.43%)	ST360 (0.228%) ST253 (1.822%) ST36 (4.683%)	ST360 (0.149%) ST2386 (0.192%)	ST360 (0.174%) ST2386 (0.31%)	ST36 (0.366%)
ID4592	ST405 (5.749%)	ST405 (9.51%)	ST405 (9.21%)	ST405 (5.231%)	ST405 (9.493%)	ST405 (11.589%)
ID4604	ST12 (0.089%)	Too low to be detected by StrainGE	ST20 (0.598%)	ST20 (1.175%)	ST20 (1.045%)	ST20 (2.099%)
ID4679	ST151 (0.294%)	Too low to be detected by StrainGE	ST151 (0.106%)	Too low to be detected by StrainGE	ST151 (0.316%)	ST151 (0.385%)
ID4689	ZKIR-qPCR negative	ST643-1LV (11.025%) ST502 (6.374%)	ST643-1LV (3.118%) ST3544 (3.312%)	ST643-1LV (6.353%)	ST643-1LV (7.823%)	ST643-1LV (11.226%)

PPI[a] NSAIDs[a] Antibiotic[a] Diabetes IBD Travel to Spain[a]

Hospitalization: Prior to study Last 4 weeks Work at healthcare [a] Within the last 4 weeks prior to sample collection

Two of the participants (ID 4469 and ID 4604) took PPI during September and November 2021, ID 4473 and ID 4592 took NSAIDs during March 2022, ID 4592 and ID 4689 took antibiotics during March 2022 and September 2021, respectively (Table 7). Only one of the participants (ID4473) had diabetes and was hospitalized prior to the study and during the study (November 2021) as well (Table 7). Only one (ID 4679) had IBD and one (ID 4689) travelled to Spain (November and December 2021) during the study (Table 7). Qualitative assessment of the metadata showed that there was no effect of hospitalization (during the last 4 weeks) on ST diversity in ID 4473 (Table 7). There were no STs changes in ID 4469 after taking PPI, ID4473 after taking NSAIDs and ID4593 after taking both antibiotic and NSAIDs (Table 7). There was a change of ST from ST12 to ST20 in ID 4604 after taking PPI and new STs were also detected in ID 4689 after travelling to Spain (ST643-1LV, ST502 and ST3544) (Table 7).

5 Discussion

In this master's project, we have studied the duration and dynamics of the KpSC gut carriage in an adult community-based cohort of 108 individuals over a period of six months from September 2021 to March 2022.

KpSC identification was performed by ZKIR-qPCR in 108 individuals (September 2021 to March 2022) recruited from the T7-study [158]. The percentage of KpSC varied over the six months of the period from 49.5% (September) to 69.1% (December). The culture-based KpSC carriage prevalence in the cross-sectional study of healthy Chinese adults in Asian countries (Taiwan, Japan, Hong Kong, China, Thailand, Malaysia, Singapore and Vietnam) was (19–88%) and the community-based general adult population from Norway was 16.3% [65, 82]. Furthermore, variation in carriage prevalence in pregnant women in low-income countries (40–66%), hospital admission study in Australia (6–19%) and hospitalized patients in the USA (23%) indicates that the prevalence depends on the context, including socio-economic situation, geographical location and study populations [77, 79, 89]. The rate of KpSC carriage from the current study could not be compared to the T7-study due to the selection criteria for the participants and method of KpSC detection. Participants were recruited randomly in the T7-study, whereas in the current study, the participants were invited among KpSC culture positive (50%) and KpSC culture negative (50%) from the T7-study. The ZKIR-qPCR detection method for KpSC used in this study is also more sensitive than the culture-based detection method, which may at least partly explain the higher prevalence of KpSC in this study compared to the T7-study [137]. Additionally, a previous study based on hospital data from four different

continents confirmed increased Kp BSI during the summer season, but one longitudinal study on six healthy adults in Austria found no impact of seasonal variation on Kp carriage prevalence [69, 182]. However, due to time constraints, the current study lacks the correlation of seasonal variation on KpSC carriage prevalence and will require further statistical analysis.

Among the 108 study subjects, 85 delivered samples for all six months, hence were included in the investigation of KpSC carriage duration. KpSC carriage duration was categorized into three different groups based on the detection of KpSC detection by ZKIR-qPCR: non-carriers (positive 0/6 months), transient carriers (positive 1 or 5 times out of 6 months) and chronic carriers (positive 6/6 months). Most of the population fell into the transient carrier category (76.5%), which is in concordance with the previous culture-based longitudinal study on a limited number of community-based healthy adults (n=6), showing that faecal colonisation is intermittent [69]. However, subsets of the study participants were either negative (9.4%) for KpSC or consecutively positive for six months (14.1%). A possible explanation for this colonisation failure or success might be due to colonisation resistance due to a particular gut microbiota composition that could prevent or assist KpSC colonisation, respectively. In accordance with the explanation, a study in mice has shown that Kp is outcompeted in gut colonisation by *K. oxytoca* with the help of other commensals [121]. Difference between gut microbiota composition was also reported between Kp colonised and non-colonised ICU patients based on 16s rRNA gene-based analysis from rectal swab culture in one of the hospitals in the USA, which was due to the presence or absence of Kp [183]. However, at the time of this writing, there was no available study related to microbial composition in people with long-term KpSC colonisation (stated as chronic carriers in the study). Investigation of the influence of gut microbiota composition is a major aim of the larger ongoing longitudinal KpSC gut carriage study.

Chronic carriers had a relatively higher relative abundance of KpSC than non-carriers and transient carriers (unpublished data, Lindstedt *et al.*). The relative abundance of KpSC in chronic carriers also remained high throughout the six months study period.

The high relative abundance of KpSC in chronic carriers suggested a possible decrease in the overall gut microbial diversity consistent with dysbiosis providing KpSC with an advantage to overcome colonisation resistance in the gut. Hence, further investigation of alpha diversity using the Shannon index was done to analyse the difference in microbial diversity among the non-, transient and chronic carriers. A high Shannon index value indicates high microbial

diversity and people with dysbiosis are associated with a low Shannon index and decreased microbial diversity [184, 185]. However, a previous cross-sectional study in the USA has shown no significant difference in the overall microbial diversity between Kp colonised and non-colonised ICU patients [183]. This also corresponds to the microbiota analysis in this study showing no significant difference in the overall gut microbial diversity among different carriage durations (non-, transient and chronic carriers). However, the number of non-carriers and chronic carriers was low compared to transient carriers, which might have limitations in reaching the statistical power of the analysis. One of the possible explanations for the lack of significant difference in microbial diversity among different carriage groups is the general adult elderly population in the current study cohort, who did not have a dramatic change in their gut microbial diversity. Microbial composition instead microbial diversity in elderly people might affect the KpSC colonisation as the elderly people lose the beneficial commensals with age, which can result in decreased colonization resistance towards KpSC[186].

Chronic carriers had high KpSC abundance (unpublished data, Lindstedt *et al*.), thus we next investigated the association between alpha diversity and the relative abundance of KpSC. In accordance with the previous experiment, there was also no significant correlation found between the relative abundance of KpSC and the Shannon index. One of the possible explanations behind the lack of statistical significance of the negative correlation between the relative abundance of KpSC and the Shannon index was likely the wide age range of the study population (47-81y), as previous studies have shown that the microbiome diversity is strongly influenced by age [181]. Therefore, the correlation of gut microbial diversity is better explained if the study population is analysed in different age groups [181]. Additionally, age ≥60y has been established as an independent risk factor for KpSC carriage [65]. Thus, the next analysis was performed to observe the correlation of gut microbial diversity with the relative abundance of KpSC within two different age groups (ages <60y and ≥60y). Interestingly, the age group <60y showed a significant negative correlation of the Shannon index with the relative abundance of KpSC, however, no significant association was found in the age group ≥60y. Previous studies have revealed that alpha diversity increases with age, stabilizes around 40 years and then starts to increase again after the seventh decade of life. [187-189]. The relatively higher stability of microbiome diversity within the age group <60y, therefore, could be a barrier to KpSC colonisation and requires a reduction in diversity for the successful colonisation of KpSC. The lack of a significant negative correlation between the Shannon index and the relative abundance of KpSC in the age group ≥60y suggests that KpSC does not require low alpha

diversity lacking in the number and varieties of species in the gut microbiome for colonisation within this group. A possible mechanism for colonisation resistance inhibiting successful colonisation of potential pathogens like KpSC is interactions with other specific commensal bacteria in the gut [121, 123]. Therefore, the loss of such beneficial commensals with increased age might facilitate the colonisation of KpSC within the elderly group in our study, which needs further investigation [186]. Thus, analysis of the beta diversity index which reflects the similarity or differences of microbial communities between groups, is currently ongoing.

Age (≥60y) is already a well-known risk factor for KpSC carriage and advanced age (≥80y) was associated with Kp bacteremia [65, 190]. Elderly people have a higher prevalence of underlying diseases like diabetes, hypertension and immunosuppression along with exposure to antibiotics, other therapeutics and surgical operations [183]. Therefore, we next investigated whether the chronic carriers of KpSC belonged to the elderly group and whether there was any significant difference in the age between different groups of carriage duration. The current study showed a significant difference in age between chronic and transient carriers, further strengthening the association of increased age with KpSC carriage. Although, there was no significant difference in age neither between chronic carriers and non-carriers nor non-carriers and transient carriers. However, the difference in age between non-carriers and chronic carriers was close to significance (p-value 0.068). The lack of statistical significance might be due to the small size of the sample. These observations support the notion that increased age is a risk factor for KpSC carriage. However, age in itself might not be the most influencing factor for the colonisation failure of KpSC. This is also in line with a study showing that other factors, such as distinct microbial composition in non-colonised ICU patients compared to colonised patients, might be associated with colonisation failure of KpSC [183].

An integral part of epidemiological typing and pathogen monitoring during outbreaks is ST typing and phylogenetic analysis of related strains [191]. Identification of the sources of infection, characterization of transmission dynamics, allocations of adequate prevention resources and control measures are essential during outbreaks [192]. With the increasing availability and affordability of WGS, it has become a choice of method for pathogen sequencing and is widely used for ST typing [193]. WMS has the potential to complement WGS and will probably be used in future outbreak investigations [193]. WMS can detect unculturable bacteria and provides a gain of time due to culture independence [191]. An ongoing comparative analysis of tools has confirmed the accuracy of strain identification by StrainGE from metagenome samples (unpublished data, Buczek *et al.*). In this master project, we also

evaluated the suitability of StrainGE for the analysis of phylogenetic relationships from metagenome samples. Phylogenetic relationship analysis by StrainGE of single WGS-isolates and corresponding SCAI-sweep samples of the same ST from the same individual revealed that StrainGE analysis of metagenomes could identify two (ST20 and ST26) out of the three STs correctly. Furthermore, the phylogenetic relationship from the core genome alignment of single WGS-isolates belonging to ST20 (taken from the same faecal samples as the SCAI-sweep samples) by Snippy also agreed with the phylogenetic relationship built by StrainGE. However, StrainGE failed to predict the correct phylogenetic relationship of the single WGS-isolate and SCAI-sweep samples belonging to ST25 from the same individual. Notably, the pairwise distances generated by StrainGE at lower coverage might not produce an accurate phylogenetic relationship which might account for the incorrect phylogenetic relationship prediction in the samples containing ST25 [136]. Although StrainGE showed promise in correctly identifying 2/3 phylogenetic relationships, it will be impossible to conclude based on only three sets of STs containing 3-4 samples each, whether StrainGE can index case and transmission chains in a clonal outbreak or not. Larger studies are required to determine the accuracy of StrainGE for the analysis of phylogenetic relationships from metagenome samples more conclusively. However, StrainGE might be a potential tool for rapid analysis of clonal transmission or to index clones from metagenome samples.

The molecular epidemiology of KpSC is important in the assessment of KpSC infections from a public health perspective. For example, high-risk clones of Kp are either prone to develop MDR causing severe HAI or cause CAI mostly in younger and healthy individuals (hvKp strains) [28, 61]. Worldwide, certain STs of Kp belong to high-risk clones and can also colonise individuals outside the hospital [89, 194]. ST analysis from the metagenomic sample of the seven most KpSC-abundant participants (six chronic carriers and one transient carrier) in this study revealed highly diverse STs (n=13) distributed in two phylogroups (Kp1 and Kp3). Although Kp1 is the most prevalent clinical species among the KpSC, Kp3 is considered an emerging human pathogen [195]. Kp3 has been isolated more frequently in UTI than Kp1 and Kp2 by culture-based MALDI-TOF detection in a retrospective study in the USA[195]. Moreover, proportion of Kp3 is highly variable in clinical settings, with the highest prevalence of 24.4% reported in KpSC-BSI in the University Hospital in Solna, Sweden [66]. There were no overlapping STs found between participants except one, ST12 was present in two participants. A previous longitudinal study on healthy adults in Austria showed that colonisation with the same specific ST was found for no longer than two consecutive weeks

[69]. In contrast, the data from our study revealed that four of the six chronic carriers maintained single STs for six months and three of the six chronic carriers maintained single STs for four months, which strongly suggests the ST maintenance by chronic carriers throughout the study period. A previous study has suggested that bacteriocin produced by commensal strain Kp MV91-1 inhibited the growth of wide-spectrum *Klebsiella* species [196]. This in line with our observation, suggests the possibility of colonisation dominance of single KpSC ST type in chronic carriers inhibiting colonization by other KpSC STs, which needs further investigation. It was not possible to conclude about the ST diversity of transient carriers due to small sample size (n=1). However, a larger dataset is required to precisely conclude whether chronic carriers tend to maintain a single ST over time and transient carriers undergo higher STs turnover. The observed ST diversity might also be an underestimation due to the analysis of direct metagenomic samples, which has limited sensitivity. Instead, culture-enriched metagenomic samples (SCAI-sweep) facilitate the detection of less abundant STs providing greater ST detection sensitivity. The analysis tool StrainGE can only detect and analyse STs from the samples with 0.1% relative abundance of the taxa, which limited the selection of samples for the analysis in this study [136]. Upon future analysis of the STs from sequencing data of SCAI sweep samples, it will be possible to fill in the blank spots where StrainGE could not detect the ST directly from faecal metagenomes and more precise information on ST diversity will be available.

Questionnaires answered by participants were analyzed qualitatively to observe the effect on KpSC ST diversity. Diabetes has been associated as a risk factor for Kp liver abscess in a nationwide Korean study [197]. Although a chronic carrier in our study had diabetes and was hospitalized before the study and during the study, there was no change in the STs from this participant. One of the chronic carriers with Crohn´s disease also had a single ST in four months, although it was undetectable by StrainGE in two months. Interestingly ID4689 was qPCR negative in the first month but became heavily colonised the second month after taking antibiotics and travelling to Spain. These single case observations in our study are consistent with identified risk factors in the T7-study showing Crohn´s disease, use of antibiotic and travelling to certain places (Greece/Asia) as risk factors of KpSC carriage [65]. It is also noteworthy that ID4484 worked in healthcare and had the highest number of different STs (n=5) belonging to two different phylogroups (Kp1 and Kp3), which emphasizes the possible occupational exposure of a variety of STs. It has already been reported in a previous study in Baltimore, USA, that 14% of the healthcare workers had contamination of gloves or gowns

with Kp after patient care [198]. The presence of Kp1 and Kp3 at the same time in ID4484 also indicates that there might be a synergistic relationship between Kp1 and Kp3, which needs further investigation.

The major strengths of this study include a large cohort in a community setting, the longitudinal methodology with six-time points for sample collection and metadata provided by the participants. The elderly age group of the study population also added valuable knowledge which will aid in understanding KpSC ecology and colonisation in the most vulnerable group prone to KpSC infections. The choice of method for sample collection, including collection kits suitable for microbiome analysis (Norgen) and viable bacteria for culture enrichment (ESwab), also provided additional strength for future analysis of ST diversity. Use of ZKIR-qPCR ensured the accurate detection and quantification of KpSC. Furthermore, WMS was one of the strengths in providing the maximum amount of information for microbiome analysis.

However, the age of the study population was one of the limiting factors, as the majority of people were elderly. It would be interesting to include different age groups, such as young adults and early middle-aged people with less exposure to antibiotics, disease and hospitals to correlate the carriage duration with age. Although metadata including different questions such as age, sex, cohabitation, hospitalization, travelling, medicines etc. were included, it lacked some other features like diet, oral hygiene, alcohol consumption, other co-morbidities and body mass index (BMI), which are established factors influencing the gut microbial composition. The outcome of the analysis based on the longitudinal data might also be different depending on the criteria of carriage durations for chronic carriers (qPCR positive for 6/6 months) and transient carriers (qPCR positive for (1 to 5)/6 months) used in this study. According to U.S. National Center for Health Statistics and CDC the minimum duration of a disease to be called chronic is three months and one year, respectively [199]. Although colonisation with KpSC is not considered a disease condition, further investigation would be needed to observe if the inclusion criteria for chronic carriers is set to qPCR positive for ≥3 out of six months and has any effect on the outcome of the overall analysis. One of the limitations also included a lack of adjustment of the data for all the confounding variables during the correlation analysis in the study. It is noteworthy that the sample size of non-carriers and chronic carriers was only eight and twelve, respectively, which might also affect the statistical power. An increased sample size will provide more precise results. Further statistical analysis is required to investigate the association of metadata with carriage durations and KpSC abundance. Moreover, the key limitation of the study was the time limit for the master's degree, which was reflected by the

unavailability of the ST diversity from SCAI-Sweep samples as the sequencing data was not available in time.

6 Conclusions and future perspectives

To our knowledge, this is the first study investigating the duration and dynamics of KpSC in general community-based adults using WMS. This is also the first study investigating the ability of StrainGE to analyse KpSC phylogenetic relationships from metagenome samples. Findings from this study revealed that:

- In a community-based adult elderly cohort, we could identify chronic, transient (dominant) and non-carriers of KpSC in faecal samples.
- There were no significant differences in overall gut microbiome diversity among different KpSC carriage groups.
- There was a significant decrease in the overall microbiome diversity with increasing relative abundance of KpSC in the age group of <60y, indicating the mechanism of colonisation by KpSC involves interaction with other members of the human gut.
- In chronic carriers, age was significantly higher than in transient carriers, supporting previous observations that the elderly people are more prone to be colonised with KpSC for a longer duration.
- StrainGE reported two out of three phylogenetic relationships correctly from metagenome samples, suggesting the potential use of WMS in future clonal transmission investigations.
- STs detected (n=13) were highly diverse and distributed in two phylogroups (Kp and Kp3). Chronic carriers (4/6) maintained only a single ST throughout the study period, indicating that long-term gut carriage of KpSC might be associated and dominated by single STs.

Future studies:

- Beta diversity needs to be studied to explore the effect of specific bacteria on KpSC gut colonisation.
- All metadata needs to be analysed in relation to KpSC carriage duration to identify potential significant variables.

- Statistical methods adjusting all the variables influencing the dataset will be required for the association of factors from the metadata.
- More metagenome samples need to be analysed with StrainGE to confirm the possibility of its use in future analysis of clonal transmission.
- Analysis of the SCAI-Sweep samples is required for extracting more information on ST diversity and dynamics from samples with a low abundance of KpSC.

7 References

1. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Aguilar GR, Gray A, Han C, Bisignano C, Rao P, Wool E: **Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis**. *The Lancet* 2022, **399**(10325):629-655.
2. Browne K, Chakraborty S, Chen R, Willcox MD, Black DS, Walsh WR, Kumar N: **A New Era of Antibiotics: The Clinical Potential of Antimicrobial Peptides**. *Int J Mol Sci* 2020, **21**(19).
3. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, Colomb-Cotinat M, Kretzschmar ME, Devleesschauwer B, Cecchini M: **Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis**. *The Lancet infectious diseases* 2019, **19**(1):56-66.
4. Organization WH: **Global antimicrobial resistance and use surveillance system (GLASS) report: 2022**. 2022.
5. Wyres KL, Lam M, Holt KE: **Population genomics of Klebsiella pneumoniae**. *Nature Reviews Microbiology* 2020, **18**(6):344-359.
6. Navon-Venezia S, Kondratyeva K, Carattoli A: **Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance**. *FEMS microbiology reviews* 2017, **41**(3):252-275.
7. Podschun R, Ullmann U: **Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors**. *Clinical microbiology reviews* 1998, **11**(4):589-603.
8. Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B, Bartlett J: **Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America**. *Clinical infectious diseases* 2009, **48**(1):1-12.
9. Who W: **Priority Pathogens List for R&D of New Antibiotics**. In.: WHO, Geneva, Switzerland; 2017.
10. Guo Y, Cen Z, Zou Y, Fang X, Li T, Wang J, Chang D, Su L, Liu Y, Chen Y: **Whole-genome sequence of Klebsiella pneumonia strain LCT-KP214**. In.: Am Soc Microbiol; 2012.
11. Ashurst JV, Dawson A: **Klebsiella pneumonia**. 2018.
12. Wyres KL, Holt KE: **Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria**. *Current opinion in microbiology* 2018, **45**:131-139.
13. Brisse S, Grimont F, Grimont PA: **The genus klebsiella**. *Prokaryotes* 2006, **6**(chapter 3.3. 8):159-196.
14. Vanhooren P, De Baets S, Bruggeman G, Vandamme E: **Klebsiella**. 1999.
15. Guilhen C, Miquel S, Charbonnel N, Joseph L, Carrier G, Forestier C, Balestrino D: **Colonization and immune modulation properties of Klebsiella pneumoniae biofilm-dispersed cells**. *npj Biofilms and Microbiomes* 2019, **5**(1):1-11.
16. Dong N, Yang X, Chan EW-C, Zhang R, Chen S: **Klebsiella species: Taxonomy, hypervirulence and multidrug resistance**. *EBioMedicine* 2022, **79**:103998.
17. Brisse S, Verhoef J: **Phylogenetic diversity of Klebsiella pneumoniae and Klebsiella oxytoca clinical isolates revealed by randomly amplified polymorphic DNA, gyrA and parC genes sequencing and automated ribotyping**. *Int J Syst Evol Microbiol* 2001, **51**(Pt 3):915-924.

18. Rodrigues C, Passet V, Rakotondrasoa A, Brisse S: **Identification of Klebsiella pneumoniae, Klebsiella quasipneumoniae, Klebsiella variicola and Related Phylogroups by MALDI-TOF Mass Spectrometry**. *Front Microbiol* 2018, **9**:3000.
19. Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, Olsen RJ: **Whole-genome sequencing of human clinical Klebsiella pneumoniae isolates reveals misidentification and misunderstandings of Klebsiella pneumoniae, Klebsiella variicola, and Klebsiella quasipneumoniae**. *Msphere* 2017, **2**(4):e00290-00217.
20. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J *et al*: **Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health**. *Proc Natl Acad Sci U S A* 2015, **112**(27):E3574-3581.
21. Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Delannoy-Vieillard AS, Garin B, Le Hello S, Arlet G, Nicolas-Chanoine MH *et al*: **Genomic definition of hypervirulent and multidrug-resistant Klebsiella pneumoniae clonal groups**. *Emerg Infect Dis* 2014, **20**(11):1812-1820.
22. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE: **A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex**. *Nat Commun* 2021, **12**(1):4188.
23. Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S: **Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates**. *J Clin Microbiol* 2005, **43**(8):4178-4182.
24. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL, Lam MM, Duchêne S, Jenney A, Holt KE: **Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of Klebsiella pneumoniae**. *PLoS genetics* 2019, **15**(4):e1008114.
25. Zhu J, Wang T, Chen L, Du H: **Virulence Factors in Hypervirulent Klebsiella pneumoniae**. *Front Microbiol* 2021, **12**:642484.
26. Gonzalez-Ferrer S, Penaloza HF, Budnick JA, Bain WG, Nordstrom HR, Lee JS, Van Tyne D: **Finding Order in the Chaos: Outstanding Questions in Klebsiella pneumoniae Pathogenesis**. *Infect Immun* 2021, **89**(4).
27. Russo TA, Olson R, Fang CT, Stoesser N, Miller M, MacDonald U, Hutson A, Barker JH, La Hoz RM, Johnson JR: **Identification of Biomarkers for Differentiation of Hypervirulent Klebsiella pneumoniae from Classical K. pneumoniae**. *J Clin Microbiol* 2018, **56**(9).
28. Paczosa MK, Mecsas J: **Klebsiella pneumoniae: Going on the Offense with a Strong Defense**. *Microbiol Mol Biol Rev* 2016, **80**(3):629-661.
29. Henson SP, Boinett CJ, Ellington MJ, Kagia N, Mwarumba S, Nyongesa S, Mturi N, Kariuki S, Scott JAG, Thomson NR: **Molecular epidemiology of Klebsiella pneumoniae invasive infections over a decade at Kilifi County Hospital in Kenya**. *International Journal of Medical Microbiology* 2017, **307**(7):422-429.
30. Gorrie CL, Mirceta M, Wick RR, Judd LM, Wyres KL, Thomson NR, Strugnell RA, Pratt NF, Garlick JS, Watson KM: **Antimicrobial-resistant Klebsiella pneumoniae carriage and infection in specialized geriatric care wards linked to acquisition in the referring hospital**. *Clinical infectious diseases* 2018, **67**(2):161-170.
31. Russo TA, Marr CM: **Hypervirulent Klebsiella pneumoniae**. *Clin Microbiol Rev* 2019, **32**(3).
32. Tang M, Kong X, Hao J, Liu J: **Epidemiological Characteristics and Formation Mechanisms of Multidrug-Resistant Hypervirulent Klebsiella pneumoniae**. *Front Microbiol* 2020, **11**:581543.

33. Turton JF, Payne Z, Coward A, Hopkins KL, Turton JA, Doumith M, Woodford N: **Virulence genes in isolates of Klebsiella pneumoniae from the UK during 2016, including among carbapenemase gene-positive hypervirulent K1-ST23 and 'non-hypervirulent'types ST147, ST15 and ST383**. *Journal of medical microbiology* 2018, **67**(1):118-128.
34. Villa L, Feudi C, Fortini D, Brisse S, Passet V, Bonura C, Endimiani A, Mammina C, Ocampo AM, Jimenez JN: **Diversity, virulence, and antimicrobial resistance of the KPC-producing Klebsiella pneumoniae ST307 clone**. *Microbial genomics* 2017, **3**(4).
35. Peirano G, Chen L, Kreiswirth BN, Pitout JDD: **Emerging Antimicrobial-Resistant High-Risk Klebsiella pneumoniae Clones ST307 and ST147**. *Antimicrob Agents Chemother* 2020, **64**(10).
36. Peirano G, Pitout JDD: **Extended-Spectrum beta-Lactamase-Producing Enterobacteriaceae: Update on Molecular Epidemiology and Treatment Options**. *Drugs* 2019, **79**(14):1529-1541.
37. Giske CG, Sundsfjord AS, Kahlmeter G, Woodford N, Nordmann P, Paterson DL, Canton R, Walsh TR: **Redefining extended-spectrum beta-lactamases: balancing science and clinical need**. *J Antimicrob Chemother* 2009, **63**(1):1-4.
38. Brolund A: **Overview of ESBL-producing Enterobacteriaceae from a Nordic perspective**. *Infection ecology & epidemiology* 2014, **4**(1):24555.
39. Sawa T, Kooguchi K, Moriyama K: **Molecular diversity of extended-spectrum beta-lactamases and carbapenemases, and antimicrobial resistance**. *J Intensive Care* 2020, **8**:13.
40. Chen Q, Wang M, Han M, Xu L, Zhang H: **Molecular basis of Klebsiella pneumoniae colonization in host**. *Microb Pathog* 2023, **177**:106026.
41. Follador R, Heinz E, Wyres KL, Ellington MJ, Kowarik M, Holt KE, Thomson NR: **The diversity of Klebsiella pneumoniae surface polysaccharides**. *Microb Genom* 2016, **2**(8):e000073.
42. Lee IR, Molton JS, Wyres KL, Gorrie C, Wong J, Hoh CH, Teo J, Kalimuddin S, Lye DC, Archuleta S *et al*: **Differential host susceptibility and bacterial virulence factors driving Klebsiella liver abscess in an ethnically diverse population**. *Sci Rep* 2016, **6**:29316.
43. Brisse S, Fevre C, Passet V, Issenhuth-Jeanjean S, Tournebize R, Diancourt L, Grimont P: **Virulent clones of Klebsiella pneumoniae: identification and evolutionary scenario based on genomic and phenotypic characterization**. *PloS one* 2009, **4**(3):e4982.
44. Wyres KL, Hawkey J, Hetland MAK, Fostervold A, Wick RR, Judd LM, Hamidian M, Howden BP, Lohr IH, Holt KE: **Emergence and rapid global dissemination of CTX-M-15-associated Klebsiella pneumoniae strain ST307**. *J Antimicrob Chemother* 2019, **74**(3):577-581.
45. Siegel SJ, Weiser JN: **Mechanisms of bacterial colonization of the respiratory tract**. *Annual review of microbiology* 2015, **69**:425-444.
46. **Qubit™ 1X dsDNA High Sensitivity (HS) and Broad Range (BR) Assay Kits** [https://www.thermofisher.com/order/catalog/product/Q33231]
47. Struve C, Bojer M, Krogfelt KA: **Characterization of Klebsiella pneumoniae type 1 fimbriae by detection of phase variation during colonization and infection and impact on virulence**. *Infection and immunity* 2008, **76**(9):4055-4065.
48. Murphy CN, Mortensen MS, Krogfelt KA, Clegg S: **Role of Klebsiella pneumoniae type 1 and type 3 fimbriae in colonizing silicone tubes implanted into the bladders**

of mice as a model of catheter-associated urinary tract infections. *Infection and immunity* 2013, **81**(8):3009-3017.
49. Young TM, Bray AS, Nagpal RK, Caudell DL, Yadav H, Zafar MA: **Animal model to study Klebsiella pneumoniae gastrointestinal colonization and host-to-host transmission**. *Infection and immunity* 2020, **88**(11):e00071-00020.
50. Paczosa MK, Mecsas J: **Klebsiella pneumoniae: going on the offense with a strong defense**. *Microbiology and molecular biology reviews* 2016, **80**(3):629-661.
51. Weinberg ED: **Iron availability and infection**. *Biochim Biophys Acta* 2009, **1790**(7):600-605.
52. Holden VI, Breen P, Houle S, Dozois CM, Bachman MA: **Klebsiella pneumoniae siderophores induce inflammation, bacterial dissemination, and HIF-1α stabilization during pneumonia**. *MBio* 2016, **7**(5):e01397-01316.
53. Abergel RJ, Moore EG, Strong RK, Raymond KN: **Microbial evasion of the immune system: structural modifications of enterobactin impair siderocalin recognition**. *J Am Chem Soc* 2006, **128**(34):10998-10999.
54. Walker KA, Miner TA, Palacios M, Trzilova D, Frederick DR, Broberg CA, Sepúlveda VE, Quinn JD, Miller VL: **A Klebsiella pneumoniae regulatory mutant has reduced capsule expression but retains hypermucoviscosity**. *MBio* 2019, **10**(2):e00089-00019.
55. Ramirez MS, Iriarte A, Reyes-Lamothe R, Sherratt DJ, Tolmasky ME: **Small Klebsiella pneumoniae plasmids: neglected contributors to antibiotic resistance**. *Frontiers in microbiology* 2019, **10**:2182.
56. Lam MMC, Wyres KL, Duchene S, Wick RR, Judd LM, Gan YH, Hoh CH, Archuleta S, Molton JS, Kalimuddin S *et al*: **Population genomics of hypervirulent Klebsiella pneumoniae clonal-group 23 reveals early emergence and rapid global dissemination**. *Nat Commun* 2018, **9**(1):2703.
57. Herridge WP, Shibu P, O'Shea J, Brook TC, Hoyles L: **Bacteriophages of Klebsiella spp., their diversity and potential therapeutic uses**. *Journal of medical microbiology* 2020, **69**(2):176.
58. Shen J, Zhou J, Xu Y, Xiu Z: **Prophages contribute to genome plasticity of Klebsiella pneumoniae and may involve the chromosomal integration of ARGs in CG258**. *Genomics* 2020, **112**(1):998-1010.
59. Marques AT, Tanoeiro L, Duarte A, Goncalves L, Vitor JMB, Vale FF: **Genomic Analysis of Prophages from Klebsiella pneumoniae Clinical Isolates**. *Microorganisms* 2021, **9**(11).
60. Olszak T, Latka A, Roszniowski B, Valvano MA, Drulis-Kawa Z: **Phage Life Cycles Behind Bacterial Biodiversity**. *Curr Med Chem* 2017, **24**(36):3987-4001.
61. Marques C, Belas A, Aboim C, Cavaco-Silva P, Trigueiro G, Gama LT, Pomba C: **Evidence of sharing of Klebsiella pneumoniae strains between healthy companion animals and cohabiting humans**. *Journal of Clinical Microbiology* 2019, **57**(6):e01537-01518.
62. Zadoks R, Griffiths H, Munoz M, Ahlstrom C, Bennett G, Thomas E, Schukken Y: **Sources of Klebsiella and Raoultella species on dairy farms: be careful where you walk**. *Journal of dairy science* 2011, **94**(2):1045-1051.
63. Laxminarayan R, Duse A, Wattal C, Zaidi AK, Wertheim HF, Sumpradit N, Vlieghe E, Hara GL, Gould IM, Goossens H *et al*: **Antibiotic resistance-the need for global solutions**. *Lancet Infect Dis* 2013, **13**(12):1057-1098.
64. Rosenblueth M, Martinez L, Silva J, Martinez-Romero E: **Klebsiella variicola, a novel species with clinical and plant-associated isolates**. *Syst Appl Microbiol* 2004, **27**(1):27-35.

65. Raffelsberger N, Hetland MAK, Svendsen K, Småbrekke L, Löhr IH, Andreassen LLE, Brisse S, Holt KE, Sundsfjord A, Samuelsen Ø: **Gastrointestinal carriage of Klebsiella pneumoniae in a general adult population: a cross-sectional study of risk factors and bacterial genomic diversity**. *Gut microbes* 2021, **13**(1):1939599.
66. Maatallah M, Vading M, Kabir MH, Bakhrouf A, Kalin M, Naucler P, Brisse S, Giske CG: **Klebsiella variicola is a frequent cause of bloodstream infection in the Stockholm area, and associated with higher mortality compared to K. pneumoniae**. *PloS one* 2014, **9**(11):e113539.
67. Fostervold A, Hetland MA, Bakksjø R, Bernhoff E, Holt KE, Samuelsen Ø, Simonsen GS, Sundsfjord A, Wyres KL, Löhr IH: **A nationwide genomic study of clinical Klebsiella pneumoniae in Norway 2001–15: introduction and spread of ESBLs facilitated by clonal groups CG15 and CG307**. *Journal of Antimicrobial Chemotherapy* 2022, **77**(3):665-674.
68. Håkonsholm F, Hetland MA, Svanevik CS, Lunestad BT, Löhr IH, Marathe NP: **Insights into the genetic diversity, antibiotic resistance and pathogenic potential of Klebsiella pneumoniae from the Norwegian marine environment using whole-genome analysis**. *International Journal of Hygiene and Environmental Health* 2022, **242**:113967.
69. Lepuschitz S, Hauser K, Schriebl A, Schlagenhaufen C, Stöger A, Chakeri A, Vötsch K, Pekard-Amenitsch S, Springer B, Allerberger F: **Fecal Klebsiella pneumoniae carriage is intermittent and of high clonal diversity**. *Frontiers in Microbiology* 2020, **11**:581081.
70. Leangapichart T, Lunha K, Jiwakanon J, Angkititrakul S, Jarhult JD, Magnusson U, Sunde M: **Characterization of Klebsiella pneumoniae complex isolates from pigs and humans in farms in Thailand: population genomic structure, antibiotic resistance and virulence genes**. *J Antimicrob Chemother* 2021, **76**(8):2012-2016.
71. Franklin-Alming FV, Kaspersen H, Hetland MA, Bakksjø R-J, Nesse LL, Leangapichart T, Löhr IH, Telke AA, Sunde M: **Exploring Klebsiella pneumoniae in healthy poultry reveals high genetic diversity, good biofilm-forming abilities and higher prevalence in turkeys than broilers**. *Frontiers in Microbiology* 2021, **12**:725414.
72. Thorpe HA, Booton R, Kallonen T, Gibbon MJ, Couto N, Passet V, López-Fernández S, Rodrigues C, Matthews L, Mitchell S: **A large-scale genomic snapshot of Klebsiella spp. isolates in Northern Italy reveals limited transmission between clinical and non-clinical settings**. *Nature Microbiology* 2022, **7**(12):2054-2067.
73. Martin RM, Bachman MA: **Colonization, infection, and the accessory genome of Klebsiella pneumoniae**. *Frontiers in cellular and infection microbiology* 2018, **8**:4.
74. Dani A: **Colonization and infection**. *Central European journal of urology* 2014, **67**(1):86.
75. Russo TA, Marr CM: **Hypervirulent klebsiella pneumoniae**. *Clinical microbiology reviews* 2019, **32**(3):e00001-00019.
76. Calderon-Gonzalez R, Lee A, Lopez-Campos G, Hancock SJ, Sa-Pessoa J, Dumigan A, McMullan R, Campbell EL, Bengoechea JA: **Modelling the Gastrointestinal Carriage of Klebsiella pneumoniae Infections**. *mBio* 2023:e0312122.
77. Gorrie CL, Mirčeta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, Pratt NF, Garlick JS, Watson KM, Pilcher DV: **Gastrointestinal carriage is a major reservoir of Klebsiella pneumoniae infection in intensive care patients**. *Clinical infectious diseases* 2017, **65**(2):208-215.
78. Conlan S, Kong HH, Segre JA: **Species-level analysis of DNA sequence data from the NIH Human Microbiome Project**. *PLoS One* 2012, **7**(10):e47075.

79. Martin RM, Cao J, Brisse S, Passet V, Wu W, Zhao L, Malani PN, Rao K, Bachman MA: **Molecular Epidemiology of Colonizing and Infecting Isolates of Klebsiella pneumoniae**. *mSphere* 2016, **1**(5).
80. Ludden C, Moradigaravand D, Jamrozy D, Gouliouris T, Blane B, Naydenova P, Hernandez-Garcia J, Wood P, Hadjirin N, Radakovic M: **A One Health study of the genetic relatedness of Klebsiella pneumoniae and their mobile elements in the East of England**. *Clinical Infectious Diseases* 2020, **70**(2):219-226.
81. Chung DR, Lee H, Park MH, Jung SI, Chang HH, Kim YS, Son JS, Moon C, Kwon KT, Ryu SY *et al*: **Fecal carriage of serotype K1 Klebsiella pneumoniae ST23 strains closely related to liver abscess isolates in Koreans living in Korea**. *Eur J Clin Microbiol Infect Dis* 2012, **31**(4):481-486.
82. Lin YT, Siu LK, Lin JC, Chen TL, Tseng CP, Yeh KM, Chang FY, Fung CP: **Seroepidemiology of Klebsiella pneumoniae colonizing the intestinal tract of healthy Chinese and overseas Chinese adults in Asian countries**. *BMC Microbiol* 2012, **12**:13.
83. Dorman MJ, Short FL: **Genome watch: Klebsiella pneumoniae: when a colonizer turns bad**. *Nature Reviews Microbiology* 2017, **15**(7):384-385.
84. Donskey CJ: **The role of the intestinal tract as a reservoir and source for transmission of nosocomial pathogens**. *Clinical infectious diseases* 2004, **39**(2):219-226.
85. Shimasaki T, Seekatz A, Bassis C, Rhee Y, Yelin RD, Fogg L, Dangana T, Cisneros EC, Weinstein RA, Okamoto K: **Increased relative abundance of Klebsiella pneumoniae carbapenemase-producing Klebsiella pneumoniae within the gut microbiota is associated with risk of bloodstream infection in long-term acute care hospital patients**. *Clinical Infectious Diseases* 2019, **68**(12):2053-2059.
86. Willing BP, Russell SL, Finlay BB: **Shifting the balance: antibiotic effects on host-microbiota mutualism**. *Nat Rev Microbiol* 2011, **9**(4):233-243.
87. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H *et al*: **Extensive impact of non-antibiotic drugs on human gut bacteria**. *Nature* 2018, **555**(7698):623-628.
88. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M: **The treatment-naive microbiome in new-onset Crohn's disease**. *Cell host & microbe* 2014, **15**(3):382-392.
89. Huynh BT, Passet V, Rakotondrasoa A, Diallo T, Kerleguer A, Hennart M, Lauzanne A, Herindrainy P, Seck A, Bercion R *et al*: **Klebsiella pneumoniae carriage in low-income countries: antimicrobial resistance, genomic diversity and risk factors**. *Gut Microbes* 2020, **11**(5):1287-1299.
90. Lohr IH, Rettedal S, Natas OB, Naseer U, Oymar K, Sundsfjord A: **Long-term faecal carriage in infants and intra-household transmission of CTX-M-15-producing Klebsiella pneumoniae following a nosocomial outbreak**. *J Antimicrob Chemother* 2013, **68**(5):1043-1048.
91. Mo Y, Hernandez-Koutoucheva A, Musicha P, Bertrand D, Lye D, Tek NO, Fenlon SN, Chen SL, Lin LM, Tang WY: **Carriage duration of carbapenemase-producing Enterobacteriaceae in a hospital cohort-implications for infection control measures**. *medRxiv* 2019:19001479.
92. Lipworth S, Vihta KD, Chau KK, Kavanagh J, Davies T, George S, Barker L, Vaughan A, Andersson M, Jeffery K *et al*: **Ten Years of Population-Level Genomic Escherichia coli and Klebsiella pneumoniae Serotype Surveillance Informs Vaccine Development for Invasive Infections**. *Clin Infect Dis* 2021, **73**(12):2276-2282.

93. King CH, Desai H, Sylvetsky AC, LoTempio J, Ayanyan S, Carrie J, Crandall KA, Fochtman BC, Gasparyan L, Gulzar N *et al*: **Baseline human gut microbiota profile in healthy people and standard reporting template**. *PLoS One* 2019, **14**(9):e0206484.
94. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI: **The human microbiome project**. *Nature* 2007, **449**(7164):804-810.
95. Zhu B, Wang X, Li L: **Human gut microbiome: the second genome of human body**. *Protein & cell* 2010, **1**(8):718-725.
96. Lederberg J, McCray AT: **Ome SweetOmics--A genealogical treasury of words**. *The scientist* 2001, **15**(7):8-8.
97. Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, Chen X, Cocolin L, Eversole K, Corral GH *et al*: **Microbiome definition re-visited: old concepts and new challenges**. *Microbiome* 2020, **8**(1):103.
98. Radjabzadeh D, Boer CG, Beth SA, van der Wal P, Kiefte-De Jong JC, Jansen MA, Konstantinov SR, Peppelenbosch MP, Hays JP, Jaddoe VW: **Diversity, compositional and functional differences between gut microbiota of children and adults**. *Scientific reports* 2020, **10**(1):1040.
99. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C: **The NIH human microbiome project**. *Genome research* 2009, **19**(12):2317-2323.
100. Liang D, Leung RK, Guan W, Au WW: **Involvement of gut microbiome in human health and disease: brief overview, knowledge gaps and research opportunities**. *Gut Pathog* 2018, **10**:3.
101. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T *et al*: **A human gut microbial gene catalogue established by metagenomic sequencing**. *Nature* 2010, **464**(7285):59-65.
102. Vaga S, Lee S, Ji B, Andreasson A, Talley NJ, Agreus L, Bidkhori G, Kovatcheva-Datchary P, Park J, Lee D *et al*: **Compositional and functional differences of the mucosal microbiota along the intestine of healthy individuals**. *Sci Rep* 2020, **10**(1):14977.
103. Maklin T, Thorpe HA, Pontinen AK, Gladstone RA, Shao Y, Pesonen M, McNally A, Johnsen PJ, Samuelsen O, Lawley TD *et al*: **Strong pathogen competition in neonatal gut colonisation**. *Nat Commun* 2022, **13**(1):7417.
104. Worby CJ, Schreiber HLt, Straub TJ, van Dijk LR, Bronson RA, Olson BS, Pinkner JS, Obernuefemann CLP, Munoz VL, Paharik AE *et al*: **Longitudinal multi-omics analyses link gut microbiome dysbiosis with recurrent urinary tract infections in women**. *Nat Microbiol* 2022, **7**(5):630-639.
105. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, Goedert JJ, Shi J, Bork P, Sinha R: **Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing**. *PLoS One* 2016, **11**(5):e0155362.
106. Vich Vila A, Imhann F, Collij V, Jankipersadsing SA, Gurry T, Mujagic Z, Kurilshikov A, Bonder MJ, Jiang X, Tigchelaar EF *et al*: **Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome**. *Sci Transl Med* 2018, **10**(472).
107. Kuang Y, Lu J, Li S, Li J, Yuan M, He J, Chen N, Xiao W, Shen S, Qiu L: **Connections between the human gut microbiome and gestational diabetes mellitus. GigaScience 6, 1–12**. In.; 2017.
108. Shanahan F, Ghosh TS, O'Toole PW: **The healthy microbiome—what is the definition of a healthy gut microbiome?** *Gastroenterology* 2021, **160**(2):483-494.

109. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C: **Metagenomic microbial community profiling using unique clade-specific marker genes**. *Nat Methods* 2012, **9**(8):811-814.
110. Donaldson GP, Lee SM, Mazmanian SK: **Gut biogeography of the bacterial microbiota**. *Nat Rev Microbiol* 2016, **14**(1):20-32.
111. Risely A: **Applying the core microbiome to understand host-microbe systems**. *J Anim Ecol* 2020, **89**(7):1549-1558.
112. Cheng J, Ringel-Kulka T, Heikamp-de Jong I, Ringel Y, Carroll I, de Vos WM, Salojarvi J, Satokari R: **Discordant temporal development of bacterial phyla and the emergence of core in the fecal microbiota of young children**. *ISME J* 2016, **10**(4):1002-1014.
113. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM *et al*: **Enterotypes of the human gut microbiome**. *Nature* 2011, **473**(7346):174-180.
114. Buffie CG, Pamer EG: **Microbiota-mediated colonization resistance against intestinal pathogens**. *Nat Rev Immunol* 2013, **13**(11):790-801.
115. Kristensen NB, Bryrup T, Allin KH, Nielsen T, Hansen TH, Pedersen O: **Alterations in fecal microbiota composition by probiotic supplementation in healthy adults: a systematic review of randomized controlled trials**. *Genome Med* 2016, **8**(1):52.
116. Caballero-Flores G, Pickard JM, Nunez G: **Microbiota-mediated colonization resistance: mechanisms and regulation**. *Nat Rev Microbiol* 2022.
117. Kamada N, Kim YG, Sham HP, Vallance BA, Puente JL, Martens EC, Nunez G: **Regulated virulence controls the ability of a pathogen to compete with the gut microbiota**. *Science* 2012, **336**(6086):1325-1329.
118. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, Naidu N, Choudhury B, Weimer BC, Monack DM *et al*: **Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens**. *Nature* 2013, **502**(7469):96-99.
119. Fang K, Jin X, Hong SH: **Probiotic Escherichia coli inhibits biofilm formation of pathogenic E. coli via extracellular activity of DegP**. *Sci Rep* 2018, **8**(1):4939.
120. Lam LH, Monack DM: **Intraspecies competition for niches in the distal gut dictate transmission during persistent Salmonella infection**. *PLoS Pathog* 2014, **10**(12):e1004527.
121. Osbelt L, Wende M, Almasi E, Derksen E, Muthukumarasamy U, Lesker TR, Galvez EJC, Pils MC, Schalk E, Chhatwal P *et al*: **Klebsiella oxytoca causes colonization resistance against multidrug-resistant K. pneumoniae in the gut via cooperative carbohydrate competition**. *Cell Host Microbe* 2021, **29**(11):1663-1679 e1667.
122. Vornhagen J, Bassis CM, Ramakrishnan S, Hein R, Mason S, Bergman Y, Sunshine N, Fan Y, Holmes CL, Timp W *et al*: **A plasmid locus associated with Klebsiella clinical infections encodes a microbiome-dependent gut fitness factor**. *PLoS Pathog* 2021, **17**(4):e1009537.
123. Vieira AT, Rocha VM, Tavares L, Garcia CC, Teixeira MM, Oliveira SC, Cassali GD, Gamba C, Martins FS, Nicoli JR: **Control of Klebsiella pneumoniae pulmonary infection and immunomodulation by oral treatment with the commensal probiotic Bifidobacterium longum 5(1A)**. *Microbes Infect* 2016, **18**(3):180-189.
124. Sorbara MT, Dubin K, Littmann ER, Moody TU, Fontana E, Seok R, Leiner IM, Taur Y, Peled JU, van den Brink MRM *et al*: **Inhibiting antibiotic-resistant Enterobacteriaceae by microbiota-mediated intracellular acidification**. *J Exp Med* 2019, **216**(1):84-98.

125. Karaliute I, Ramonaite R, Bernatoniene J, Petrikaite V, Misiunas A, Denkovskiene E, Razanskiene A, Gleba Y, Kupcinskas J, Skieceviciene J: **Reduction of gastrointestinal tract colonization by Klebsiella quasipneumoniae using antimicrobial protein KvarIa**. *Gut Pathog* 2022, **14**(1):17.
126. Sana TG, Flaugnatti N, Lugo KA, Lam LH, Jacobson A, Baylot V, Durand E, Journet L, Cascales E, Monack DM: **Salmonella Typhimurium utilizes a T6SS-mediated antibacterial weapon to establish in the host gut**. *Proceedings of the National Academy of Sciences* 2016, **113**(34):E5044-E5051.
127. Sequeira RP, McDonald JA, Marchesi JR, Clarke TB: **Commensal Bacteroidetes protect against Klebsiella pneumoniae colonization and transmission through IL-36 signalling**. *Nature microbiology* 2020, **5**(2):304-313.
128. Semenec L, Cain AK, Dawson CJ, Liu Q, Dinh H, Lott H, Penesyan A, Maharjan R, Short FL, Hassan KA: **Cross-protection and cross-feeding between Klebsiella pneumoniae and Acinetobacter baumannii promotes their co-existence**. *Nature Communications* 2023, **14**(1):702.
129. Hudson AW, Barnes AJ, Bray AS, Ornelles DA, Zafar MA: **Klebsiella pneumoniae l-Fucose Metabolism Promotes Gastrointestinal Colonization and Modulates Its Virulence Determinants**. *Infect Immun* 2022, **90**(10):e0020622.
130. Hsieh PF, Lu YR, Lin TL, Lai LY, Wang JT: **Klebsiella pneumoniae Type VI Secretion System Contributes to Bacterial Competition, Cell Invasion, Type-1 Fimbriae Expression, and In Vivo Colonization**. *J Infect Dis* 2019, **219**(4):637-647.
131. Merciecca T, Bornes S, Nakusi L, Theil S, Rendueles O, Forestier C, Miquel S: **Role of Klebsiella pneumoniae Type VI secretion system (T6SS) in long-term gastrointestinal colonization**. *Sci Rep* 2022, **12**(1):16968.
132. Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC, Huttenhower C: **Sequencing and beyond: integrating molecular 'omics' for microbial community profiling**. *Nat Rev Microbiol* 2015, **13**(6):360-372.
133. Oniciuc EA, Likotrafiti E, Alvarez-Molina A, Prieto M, Santos JA, Alvarez-Ordonez A: **The Present and Future of Whole Genome Sequencing (WGS) and Whole Metagenome Sequencing (WMS) for Surveillance of Antimicrobial Resistant Microorganisms and Antimicrobial Resistance Genes across the Food Chain**. *Genes (Basel)* 2018, **9**(5).
134. Chiu CY, Miller SA: **Clinical metagenomics**. *Nat Rev Genet* 2019, **20**(6):341-355.
135. d'Humieres C, Salmona M, Delliere S, Leo S, Rodriguez C, Angebault C, Alanio A, Fourati S, Lazarevic V, Woerther PL *et al*: **The Potential Role of Clinical Metagenomics in Infectious Diseases: Therapeutic Perspectives**. *Drugs* 2021, **81**(13):1453-1466.
136. van Dijk LR, Walker BJ, Straub TJ, Worby CJ, Grote A, Schreiber HL, Anyansi C, Pickering AJ, Hultgren SJ, Manson AL: **StrainGE: a toolkit to track and characterize low-abundance strains in complex microbial communities**. *Genome biology* 2022, **23**(1):1-27.
137. Lindstedt K, Buczek D, Pedersen T, Hjerde E, Raffelsberger N, Suzuki Y, Brisse S, Holt K, Samuelsen O, Sundsfjord A: **Detection of Klebsiella pneumoniae human gut carriage: a comparison of culture, qPCR, and whole metagenomic sequencing methods**. *Gut Microbes* 2022, **14**(1):2118500.
138. Van Kregten E, Westerdaal NA, Willers JM: **New, simple medium for selective recovery of Klebsiella pneumoniae and Klebsiella oxytoca from human feces**. *J Clin Microbiol* 1984, **20**(5):936-941.
139. Pos KM, Dimroth P, Bott M: **The Escherichia coli citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate**

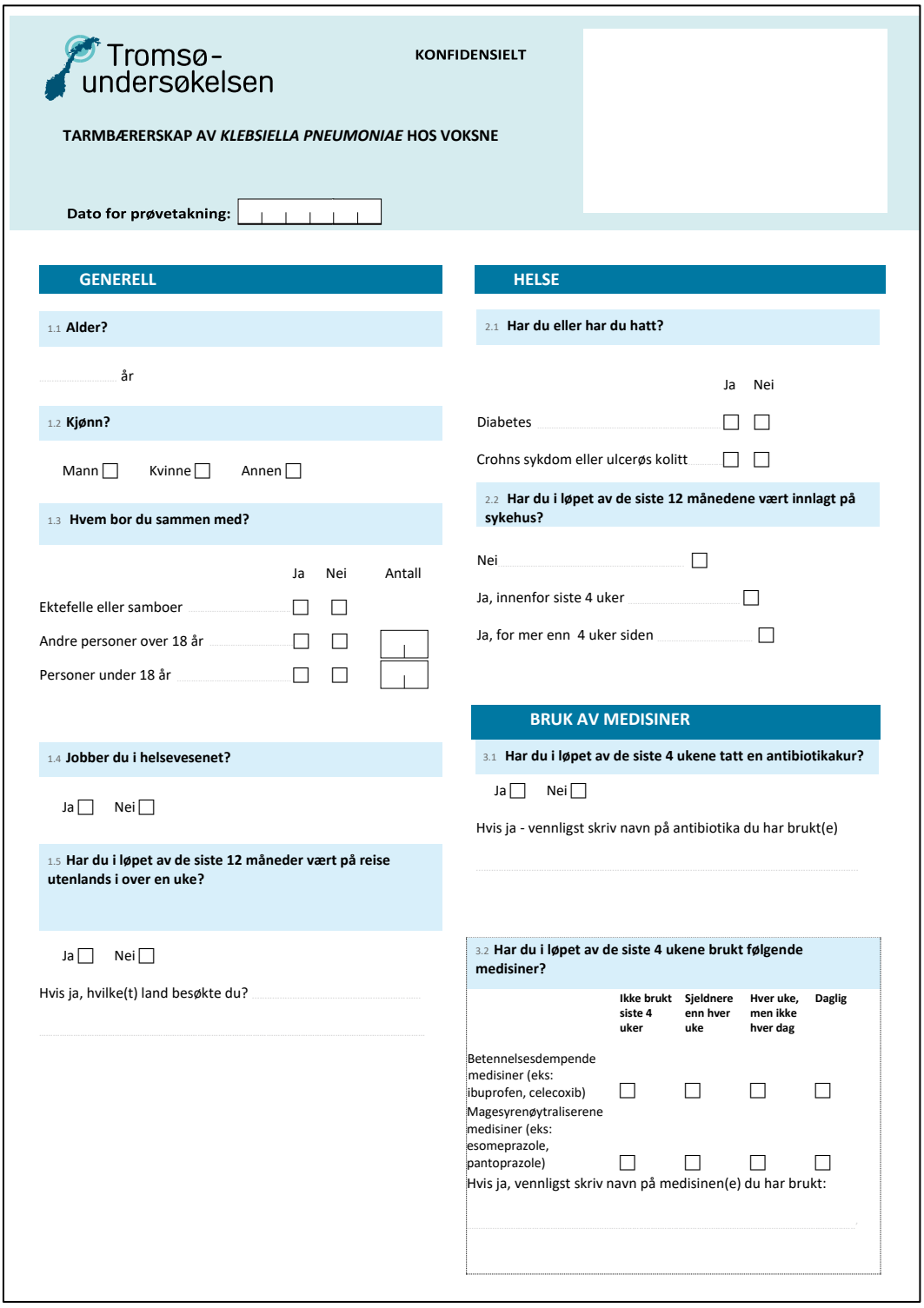
translocator from spinach chloroplasts. *Journal of Bacteriology* 1998, **180**(16):4160-4165.
140. van der Rest ME, Siewe R, Abee T, Schwarz E, Oesterhelt D, Konings W: **Nucleotide sequence and functional properties of a sodium-dependent citrate transport system from Klebsiella pneumoniae**. *Journal of Biological Chemistry* 1992, **267**(13):8971-8976.
141. Lütgens M, Gottschalk G: **Why a co-substrate is required for anaerobic growth of Escherichia coli on citrate**. *Microbiology* 1980, **119**(1):63-70.
142. Fiedorova K, Radvansky M, Nemcova E, Grombirikova H, Bosak J, Cernochova M, Lexa M, Smajs D, Freiberger T: **The Impact of DNA Extraction Methods on Stool Bacterial and Fungal Microbiota Community Recovery**. *Front Microbiol* 2019, **10**:821.
143. Gautam A: **DNA Isolation by Lysozyme and Proteinase K**. In: *DNA and RNA Isolation Techniques for Non-Experts.* Springer; 2022: 85-88.
144. Vandeventer PE, Lin JS, Zwang TJ, Nadim A, Johal MS, Niemz A: **Multiphasic DNA adsorption to silica surfaces under varying buffer, pH, and ionic strength conditions**. *J Phys Chem B* 2012, **116**(19):5661-5670.
145. **NanoDrop 2000 user manual** [https://assets.thermofisher.com/TFS-Assets/CAD/manuals/NanoDrop-2000-User-Manual-EN.pdf]
146. Desjardins P, Conklin D: **NanoDrop microvolume quantitation of nucleic acids**. *J Vis Exp* 2010(45).
147. **Purity Ratios Explained** [https://www.denovix.com/tn-130-purity-ratios-explained/]
148. Bilodeau GJ: **Quantitative polymerase chain reaction for the detection of organisms in soil**. *CABI Reviews* 2011(2011):1-14.
149. Holland PM, Abramson RD, Watson R, Gelfand DH: **Detection of specific polymerase chain reaction product by utilizing the 5'----3'exonuclease activity of Thermus aquaticus DNA polymerase**. *Proceedings of the National Academy of Sciences* 1991, **88**(16):7276-7280.
150. Caplin B, Rasmussen R, Bernard P, Wittwer C: **Lightcycler™ hybridization probes—the most direct way to monitor PCR amplification and mutation detection**. *Biochemica* 1999, **1**:5-8.
151. Svanvik N, Westman G, Wang D, Kubista M: **Light-up probes: thiazole orange-conjugated peptide nucleic acid for detection of target nucleic acid in homogeneous solution**. *Analytical biochemistry* 2000, **281**(1):26-35.
152. Zipper H, Brunner H, Bernhagen J, Vitzthum F: **Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications**. *Nucleic acids research* 2004, **32**(12):e103-e103.
153. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL: **The MIQE Guidelines: M inimum I nformation for Publication of Q uantitative Real-Time PCR E xperiments**. In.: Oxford University Press; 2009.
154. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonak J, Lind K, Sindelka R, Sjoback R, Sjogreen B, Strombom L *et al*: **The real-time polymerase chain reaction**. *Mol Aspects Med* 2006, **27**(2-3):95-125.
155. Yang S, Rothman RE: **PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings**. *The Lancet infectious diseases* 2004, **4**(6):337-348.
156. Ririe KM, Rasmussen RP, Wittwer CT: **Product differentiation by analysis of DNA melting curves during the polymerase chain reaction**. *Analytical biochemistry* 1997, **245**(2):154-160.

157. Cao Y, Yu M, Dong G, Chen B, Zhang B: **Digital PCR as an emerging tool for monitoring of microbial biodegradation**. *Molecules* 2020, **25**(3):706.
158. Barbier E, Rodrigues C, Depret G, Passet V, Gal L, Piveteau P, Brisse S: **The ZKIR assay, a real-time PCR method for the detection of Klebsiella pneumoniae and closely related species in environmental samples**. *Applied and environmental microbiology* 2020, **86**(7):e02711-02719.
159. Rodrigues C, Hauser K, Cahill N, Ligowska-Marzeta M, Centorotola G, Cornacchia A, Garcia Fierro R, Haenni M, Nielsen EM, Piveteau P *et al*: **High Prevalence of Klebsiella pneumoniae in European Food Products: a Multicentric Study Comparing Culture and Molecular Detection Methods**. *Microbiol Spectr* 2022, **10**(1):e0237621.
160. Samuelsen O, Naseer U, Tofteland S, Skutlaberg DH, Onken A, Hjetland R, Sundsfjord A, Giske CG: **Emergence of clonally related Klebsiella pneumoniae isolates of sequence type 258 producing plasmid-mediated KPC carbapenemase in Norway and Sweden**. *J Antimicrob Chemother* 2009, **63**(4):654-658.
161. Clifford RJ, Milillo M, Prestwood J, Quintero R, Zurawski DV, Kwak YI, Waterman PE, Lesho EP, Mc Gann P: **Detection of bacterial 16S rRNA and identification of four clinically important bacteria by real-time PCR**. *PloS one* 2012, **7**(11):e48558.
162. Johnson G, Nolan T, Bustin SA: **Real-time quantitative PCR, pathogen detection and MIQE**. In: *PCR detection of microbial pathogens.* Springer; 2013: 1-16.
163. MGI: **MGISP-100: Compatible with Multiple Sequencing Libraries, Enabling Automated Library Conversion**. 2022.
164. **MGIEasy FS DNA Library Prep Set.** [https://en.mgi-tech.com/products/reagents_info/id/7]
165. Lu J, Breitwieser FP, Thielen P, Salzberg SL: **Bracken: estimating species abundance in metagenomics data**. *PeerJ Computer Science* 2017, **3**:e104.
166. Wood DE, Lu J, Langmead B: **Improved metagenomic analysis with Kraken 2**. *Genome biology* 2019, **20**:1-13.
167. Kim D, Song L, Breitwieser FP, Salzberg SL: **Centrifuge: rapid and sensitive classification of metagenomic sequences**. *Genome Res* 2016, **26**(12):1721-1729.
168. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM: **The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants**. *Nucleic Acids Res* 2010, **38**(6):1767-1771.
169. Chen S, Zhou Y, Chen Y, Gu J: **fastp: an ultra-fast all-in-one FASTQ preprocessor**. *Bioinformatics* 2018, **34**(17):i884-i890.
170. Bushnell B: **BBMap: a fast, accurate, splice-aware aligner**. In.: Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States); 2014.
171. Wingett SW, Andrews S: **FastQ Screen: A tool for multi-genome mapping and quality control**. *F1000Res* 2018, **7**:1338.
172. Whittaker RH: **Evolution and measurement of species diversity**. *Taxon* 1972, **21**(2-3):213-251.
173. Willis AD: **Rarefaction, alpha diversity, and statistics**. *Frontiers in microbiology* 2019, **10**:2407.
174. Shannon CE: **A mathematical theory of communication**. *The Bell system technical journal* 1948, **27**(3):379-423.
175. McLennan DA: **How to read a phylogenetic tree**. *Evolution: Education and Outreach* 2010, **3**(4):506-519.
176. de Queiroz K: **Nodes, branches, and phylogenetic definitions**. *Systematic biology* 2013, **62**(4):625-632.

177. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics* 2014, **30**(9):1312-1313.
178. Thorpe H, Booton R, Kallonen T, Gibbon MJ, Couto N, Passet V, Fernandez JSL, Rodrigues C, Matthews L, Mitchell S: **One Health or Three? Transmission modelling of Klebsiella isolates reveals ecological barriers to transmission between humans, animals and the environment**. *BioRxiv* 2021.
179. Bates D, Mächler M, Bolker B, Walker S: **Fitting linear mixed-effects models using lme4**. *arXiv preprint arXiv:14065823* 2014.
180. HPC: **The Norwegian academic high-performance computing and storage services**. 2019.
181. de la Cuesta-Zuluaga J, Kelley ST, Chen Y, Escobar JS, Mueller NT, Ley RE, McDonald D, Huang S, Swafford AD, Knight R *et al*: **Age- and Sex-Dependent Patterns of Gut Microbial Diversity in Human Adults**. *mSystems* 2019, **4**(4).
182. Anderson DJ, Hervé R, Chen LF, Spelman DW, Hung Y-J, Huang AT, Sexton DJ, Raoult D: **Seasonal variation in Klebsiella pneumoniae bloodstream infection on 4 continents**. *The Journal of infectious diseases* 2008, **197**(5):752-756.
183. Collingwood A, Blostein F, Seekatz AM, Wobus CE, Woods RJ, Foxman B, Bachman MA: **Epidemiological and microbiome associations between Klebsiella pneumoniae and vancomycin-resistant Enterococcus colonization in intensive care unit patients**. In: *Open forum infectious diseases: 2020*. Oxford University Press US: ofaa012.
184. Yin L, Wan Y-D, Pan X-T, Zhou C-Y, Lin N, Ma C-T, Yao J, Su Z, Wan C, Yu Y-W: **Association between gut bacterial diversity and mortality in septic shock patients: a cohort study**. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research* 2019, **25**:7376.
185. Mirsepasi-Lauridsen HC, Vrankx K, Engberg J, Friis-Moller A, Brynskov J, Nordgaard-Lassen I, Petersen AM, Krogfelt KA: **Disease-Specific Enteric Microbiome Dysbiosis in Inflammatory Bowel Disease**. *Front Med (Lausanne)* 2018, **5**:304.
186. Ghosh TS, Shanahan F, O'Toole PW: **The gut microbiome as a modulator of healthy ageing**. *Nature Reviews Gastroenterology & Hepatology* 2022, **19**(9):565-584.
187. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE: **Succession of microbial consortia in the developing infant gut microbiome**. *Proceedings of the National Academy of Sciences* 2011, **108**(supplement_1):4578-4585.
188. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP: **Human gut microbiome viewed across age and geography**. *nature* 2012, **486**(7402):222-227.
189. Odamaki T, Kato K, Sugahara H, Hashikura N, Takahashi S, Xiao J-z, Abe F, Osawa R: **Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study**. *BMC microbiology* 2016, **16**(1):1-12.
190. Meatherall BL, Gregson D, Ross T, Pitout JD, Laupland KB: **Incidence, risk factors, and outcomes of Klebsiella pneumoniae bacteremia**. *Am J Med* 2009, **122**(9):866-873.
191. Bertelli C, Greub G: **Rapid bacterial genome sequencing: methods and applications in clinical microbiology**. *Clin Microbiol Infect* 2013, **19**(9):803-813.
192. Nutman A, Marchaim D: **How to: molecular investigation of a hospital outbreak**. *Clinical Microbiology and Infection* 2019, **25**(6):688-695.

193. Werner G, Couto N, Feil EJ, Novais A, Hegstad K, Howden BP, Friedrich AW, Reuter S: **Taking hospital pathogen surveillance to the next level**. *Microb Genom* 2023, **9**(4).
194. Roe CC, Vazquez AJ, Esposito EP, Zarrilli R, Sahl JW: **Diversity, Virulence, and Antimicrobial Resistance in Isolates From the Newly Emerging Klebsiella pneumoniae ST101 Lineage**. *Front Microbiol* 2019, **10**:542.
195. Rodriguez-Medina N, Barrios-Camacho H, Duran-Bedolla J, Garza-Ramos U: **Klebsiella variicola: an emerging pathogen in humans**. *Emerg Microbes Infect* 2019, **8**(1):973-988.
196. Gomez M, Valverde A, Del Campo R, Rodriguez JM, Maldonado-Barragan A: **Phenotypic and Molecular Characterization of Commensal, Community-Acquired and Nosocomial Klebsiella spp**. *Microorganisms* 2021, **9**(11).
197. Kim JK, Chung DR, Wie SH, Yoo JH, Park SW, Korean Study Group for Liver A: **Risk factor analysis of invasive liver abscess caused by the K1 serotype Klebsiella pneumoniae**. *Eur J Clin Microbiol Infect Dis* 2009, **28**(1):109-111.
198. Rock C, Thom KA, Masnick M, Johnson JK, Harris AD, Morgan DJ: **Frequency of Klebsiella pneumoniae carbapenemase (KPC)-producing and non-KPC-producing Klebsiella species contamination of healthcare workers and the environment**. *Infect Control Hosp Epidemiol* 2014, **35**(4):426-429.
199. Bernell S, Howard SW: **Use your words carefully: what is a chronic disease?** *Frontiers in public health* 2016, **4**:159.

Appendix 1: Supplementary figures



Tromsø-undersøkelsen

KONFIDENSIELT

TARMBÆRERSKAP AV *KLEBSIELLA PNEUMONIAE* HOS VOKSNE

Dato for prøvetakning: | | | | | | |

GENERELL

1.1 **Alder?**

..... år

1.2 **Kjønn?**

Mann ☐ Kvinne ☐ Annen ☐

1.3 **Hvem bor du sammen med?**

	Ja	Nei	Antall
Ektefelle eller samboer	☐	☐	
Andre personer over 18 år	☐	☐	☐☐
Personer under 18 år	☐	☐	☐☐

1.4 **Jobber du i helsevesenet?**

Ja ☐ Nei ☐

1.5 **Har du i løpet av de siste 12 måneder vært på reise utenlands i over en uke?**

Ja ☐ Nei ☐

Hvis ja, hvilke(t) land besøkte du? .....

.....

HELSE

2.1 **Har du eller har du hatt?**

	Ja	Nei
Diabetes	☐	☐
Crohns sykdom eller ulcerøs kolitt	☐	☐

2.2 **Har du i løpet av de siste 12 månedene vært innlagt på sykehus?**

Nei ..... ☐

Ja, innenfor siste 4 uker ..... ☐

Ja, for mer enn 4 uker siden ..... ☐

BRUK AV MEDISINER

3.1 **Har du i løpet av de siste 4 ukene tatt en antibiotikakur?**

Ja ☐ Nei ☐

Hvis ja - vennligst skriv navn på antibiotika du har brukt(e)

.....

3.2 **Har du i løpet av de siste 4 ukene brukt følgende medisiner?**

	Ikke brukt siste 4 uker	Sjeldnere enn hver uke	Hver uke, men ikke hver dag	Daglig
Betennelsesdempende medisiner (eks: ibuprofen, celecoxib)	☐	☐	☐	☐
Magesyrenøytraliserene medisiner (eks: esomeprazole, pantoprazole)	☐	☐	☐	☐

Hvis ja, vennligst skriv navn på medisinen(e) du har brukt:

.....

Figure S1. Sample questionaries

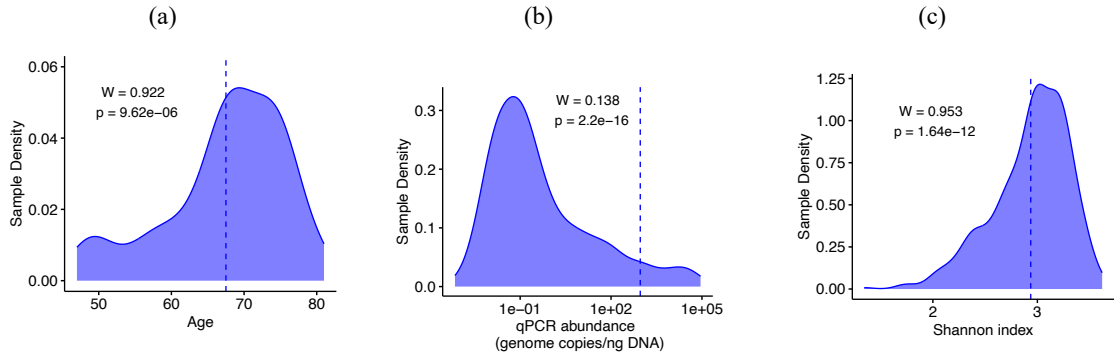


Figure S2. Density plot for normality of the dataset. *(a), (b) and (c) represents dataset of age, qPCR abundance and Shannon index. p-value >0.05 is considered a normal distribution of the data. W value <1.0 indicates difference from a normal distribution.*

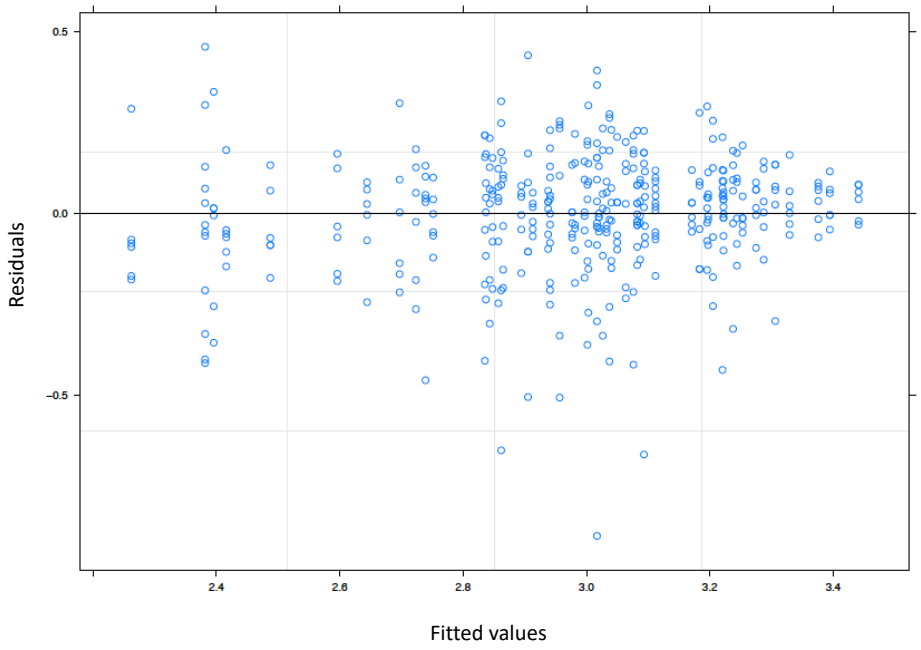


Figure S3. Homoscadascity of the model for linear mixed effects analysis of the relationship between the Shannon index and KpSC carriage duration.

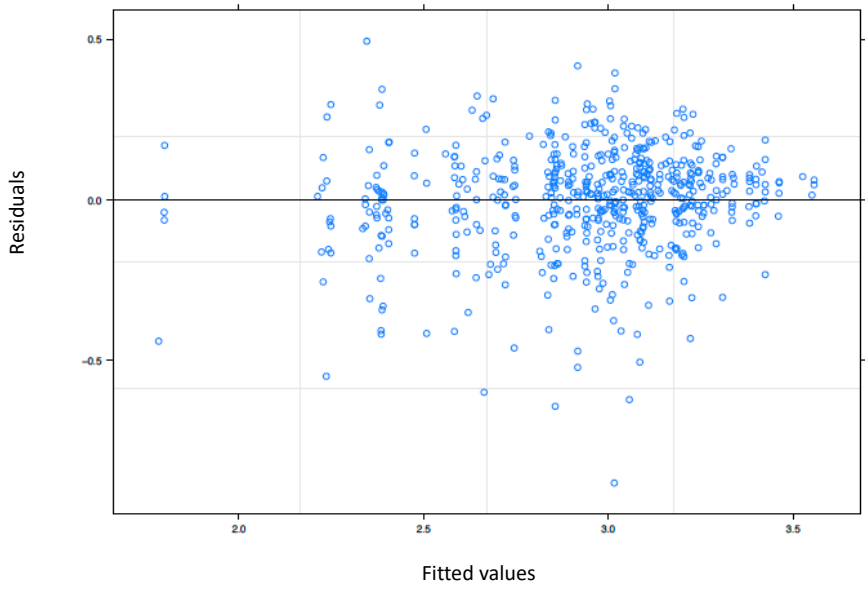


Figure S4. Homoscadascity of the model for linear mixed effects analysis of the relationship between the Shannon index and KpSC relative abundance.

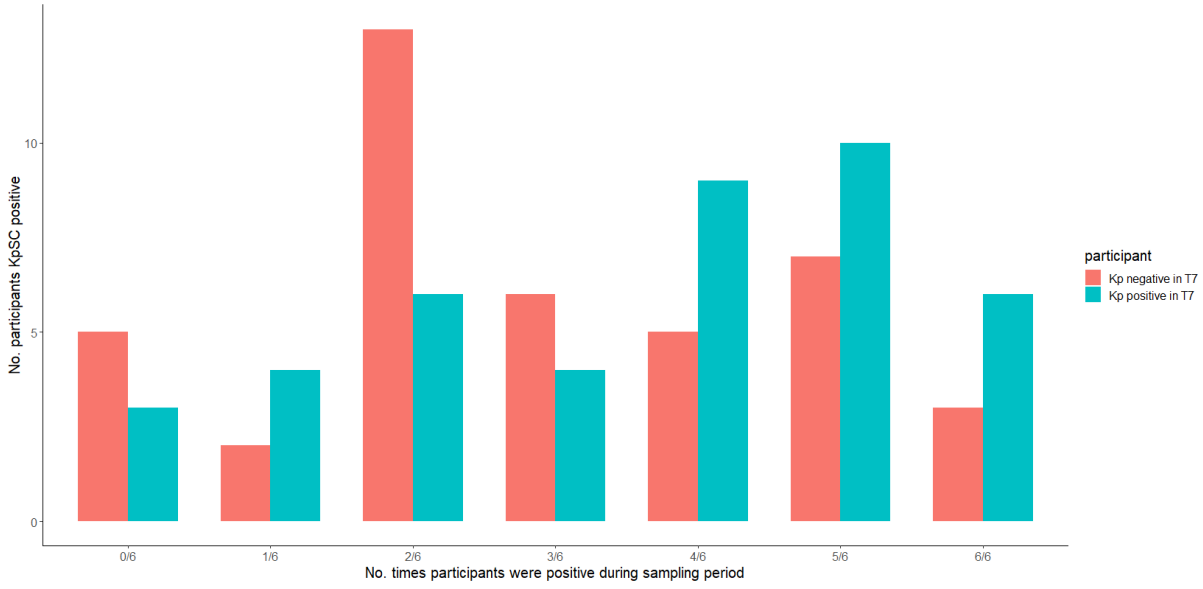


Figure S5. KpSC status in the T7 study and first month of the current longitudinal study *(unpublished data, Lindstedt et al.)*

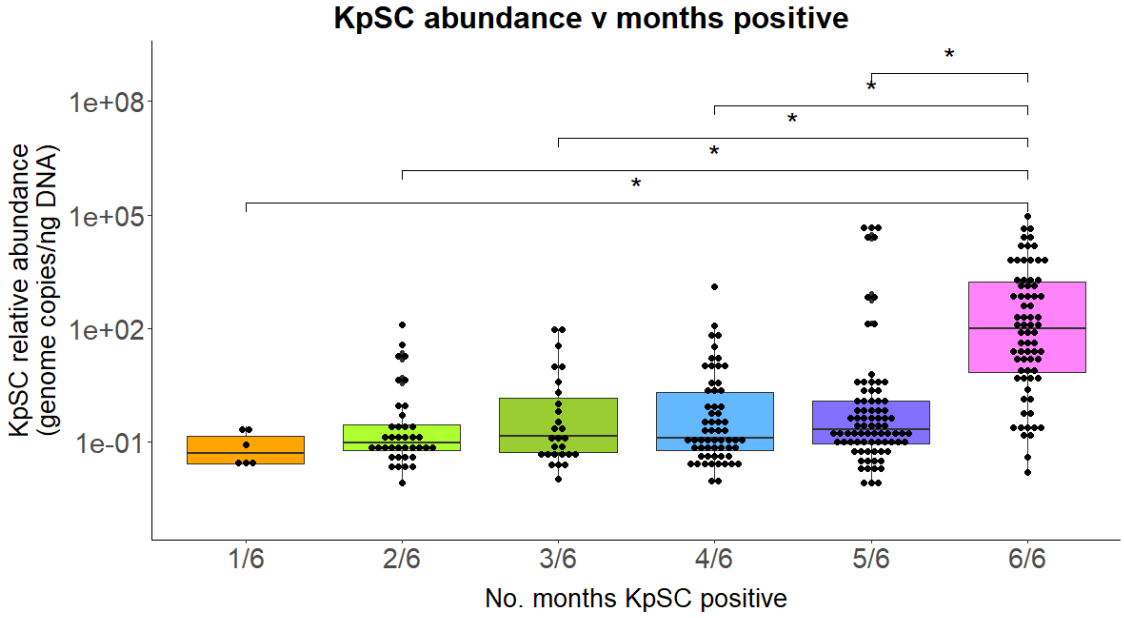


Figure S6. KpSC relative abundance in different carriage groups. ** indicates p-value < 0.05 (Likelihood Ratio Test by anova function) (unpublished data, Lindstedt et al.)*

Appendix 2: Bioinformatic syntax

Analysis of alpha diversity index the Shannon index

```
library(vegan)

##remove the input files from the list

rm(list = ls())

##define dataframe
input_data = "Bracken_all_min0-1.tsv"

df_input_data = read.table(file = input_data, header = TRUE, sep = "\t",
                           row.names = 1,
                           stringsAsFactors = FALSE)


##calculate shanon index
data_shannon <- diversity(df_input_data, index = "shannon")
```

####StrainGST database creation

As described in https://strainge.readthedocs.io/en/latest/create-database.html

Downloading high quality reference genome

ncbi-genome-download bacteria -l complete -T 573 -H -F all -o ref_genomes

Organization of references

python3 prepare_strainge_db.py ref_genomes/human_readable -s -o strainge_db > strainge_db/references_meta.tsv

k-merization of each genome

straingst kmersim --all-vs-all -t 4 -S jaccard -S subset /cluster/work/users/wka004/StrainGE/Kmer/*.hdf5 > similarities.tsv

Cluster of references

straingst cluster -i /cluster/work/users/wka004/StrainGE/Kmer/similarities.tsv -d -C 0.99 -c 0.95 --clusters-out clusters.tsv /cluster/work/users/wka004/StrainGE/Kmer/*.hdf5 > references_to_keep.txt

Creation of pan-genome k-mer database

straingst createdb -f /cluster/work/users/wka004/StrainGE/Cluster/references_to_keep.txt -o pan-genome-db.hdf5

k-merize samples by StrainGST

```
#!/bin/bash
#SBATCH --account=nn9794k --job-name=strainge_run_wka004
#SBATCH --partition=bigmem
#SBATCH --time=0-24:0:0
#SBATCH --ntasks=4
#SBATCH --mem-per-cpu=100G

#SBATCH --output=/cluster/work/users/wka004/Dorota/out_straingst
#SBATCH --mail-type=ALL
#SBATCH --mail-user=wka004@uit.no

## Set up job environment:
set -o errexit  # Exit the script on any error
set -o nounset  # Treat any unset variables as an error

module --quiet purge
module load Anaconda3/2019.03
export PS1=\$
source ${EBROOTANACONDA3}/etc/profile.d/conda.sh
conda init bash
conda activate /cluster/projects/nn9794k/Dorota/envs/StrainGE/

## Do some work:
DATADIR=""/cluster/projects/nn9794k/Dorota/PROJECTS/Repair_5a/Extracted/SWEEP_all"
for f1 in "${DATADIR}"/*_R1_repair.fastq.gz; do
    f2="${f1/_R1_repair.fastq.gz/_R2_repair.fastq.gz}"
    OUT="$(basename ${f1/_R1_repair.fastq.gz/})"
    straingst kmerize -k 23 -o ${OUT} ${f1} ${f2}
done
exit 0
```

Sample run by StrainGST

```
#!/bin/bash
#SBATCH --account=nn9794k --job-name=strainge_run_wka004
#SBATCH --partition=bigmem
#SBATCH --time=0-23:0:0
#SBATCH --ntasks=4
#SBATCH --mem-per-cpu=32G

#SBATCH --output=/cluster/work/users/wka004/Dorota/out_strainge_run_5
#SBATCH --mail-type=ALL
#SBATCH --mail-user=wka004@uit.no

## Set up job environment:
set -o errexit  # Exit the script on any error
set -o nounset  # Treat any unset variables as an error

module --quiet purge  # Reset the modules to the system default
module load Anaconda3/2019.03
export PS1=\$
source ${EBROOTANACONDA3}/etc/profile.d/conda.sh
conda init bash
conda activate /cluster/projects/nn9794k/Dorota/envs/StrainGE/

## Do some work:
DATADIR="/cluster/work/users/wka004/Dorota/"
for f1 in "${DATADIR}"/*.hdf5; do
    OUT="$(basename ${f1/.hdf5/.tsv})"
    straingst run -o ${OUT} /cluster/projects/nn9794k/DB/StrainGE_DB_Kp_NCBI_T7_SPARK/pan-genome-db-T7-SPARK.hdf5 ${f1}
done
exit 0
```

Concatenated reference FASTA by StrainGR

```
straingr prepare-ref -s /cluster/work/users/wka004/StrainGST/StrainGST_result/*.tsv -p
"/cluster/work/users/wka004/StrainGE/strainge_db/{ref}" -S /cluster/work/users/wka004/StrainGE/Kmer/similarities.tsv
-o refs_concat.fasta
```

Align reads by StrainGR

```
#!/bin/bash
#SBATCH --account=nn9794k --job-name=straingr_run_wka004
#SBATCH --partition=bigmem
#SBATCH --time=0-23:0:0
#SBATCH --ntasks=4
#SBATCH --mem-per-cpu=100G

#SBATCH --output=/cluster/projects/nn9794k/Wasifa/SCAI_sweep/StrainGR/Direct_S_35_45_46/out_straingr_allign
#SBATCH --mail-type=ALL
#SBATCH --mail-user=wka004@uit.no

## Set up job environment:
set -o errexit  # Exit the script on any error
set -o nounset  # Treat any unset variables as an error

module --quiet purge  # Reset the modules to the system default
module load Anaconda3/2019.03
export PS1=\$
source ${EBROOTANACONDA3}/etc/profile.d/conda.sh
conda init bash
conda activate /cluster/projects/nn9794k/Dorota/envs/StrainGE/

## Do some work:

DATADIR="/cluster/projects/nn9794k/Wasifa/SCAI_sweep/Direct_Fastq"
REFDIR="/cluster/projects/nn9794k/Wasifa/SCAI_sweep/StrainGR/Direct_S_35_45_46/"
for f1 in "${DATADIR}"/*_R1_001.fastq.gz; do
    f2="${f1/_R1_001.fastq.gz/_R2_001.fastq.gz}"
    OUT="$(basename ${f1/_R1_001.fastq.gz/.sam})"
bwa mem -I 300 -t 2 "${REFDIR}"/refs_concat.fasta ${f1} ${f2} > ${OUT}
done
exit 0
```

Call variants by StrainGR

```
DATADIR="/cluster/projects/nn9794k/Wasifa/SCAI_sweep/StrainGR/Direct_S_35_45_46"
straingr call "${DATADIR}"/refs_concat.fasta D-45_S217_L004.bam --hdf5-out D-45_S217_L004.hdf5 --
summary D-45_S217_L004.tsv --tracks all
```

Distance matrix by StrainGR

straingr dist -r /cluster/projects/nn9794k/DB/StrainGE/strainge_db/T7_012.fa.gz 51-S-36_S162_L004.hdf5 86-S-72_S202_L004.hdf5 -d jc -o matrix.txt

NJ phylogenetic tree by StrainGR

straingr tree matrix.txt -o tree.txt

Call SNPs by Snippy

snippy --ref /cluster/projects/nn9794k/Wasifa/SCAI_sweep/Snippy/ref/T7_308.fa --ctgs K66_45_ST11.fasta --outdir out_dir_K66_45_ST11 --cleanup –report

Core SNPs by Snippy core

snippy-core --ref /cluster/projects/nn9794k/Wasifa/SCAI_sweep/Snippy/ref/T7_308.fa out_dir_T7_451 out_dir_T7_127 out_dir_K66_45_ST11

Calculation of SNP distance

snp-dists /cluster/projects/nn9794k/Wasifa/SCAI_sweep/Snippy/snp_core_T7/core.full.aln > matrix_snp_core_T7.tab

Phylogenetic analysis RaxML

raxmlHPC -m GTRGAMMA -p 12345 -s /cluster/projects/nn9794k/Wasifa/SCAI_sweep/Snippy/T7_451_127_308/snp_core_T7/core.aln -n Raxml_wasifa -f a -x 12345 -N 1000 -T8

ST analysis by Kleborate

Kleborate –all -o high_abundance_carriers_Kleborate_results.txt -a *.fa.gz

Linear mixed effects analysis of the relationship between the Shannon index and KpSC carriage duration

```r
library(lme4)

data<- read.csv("Box_plot.csv")
df <- as.data.frame(data)

##Non-carriers vs transient

model1 = lmer((df[,2]) ~ carriage_type +
                (1|participant), subset=(df$carriage_type == "Non-carriers" | df$carriage_type == "Transient"),
              data=df, REML = FALSE)

model1_null = lmer((df[,2]) ~
                     (1|participant), subset=(df$carriage_type == "Non-carriers" | df$carriage_type == "Transient"),
                   data=df, REML = FALSE)

model1
anova(model1_null,model1)


##Chronic vs transient

model2 = lmer(shannon_index ~ carriage_type +
                (1|participant), subset=(df$carriage_type == "Chronic" | df$carriage_type == "Transient"),
              data=df, REML = FALSE)

model2_null = lmer((df[,2]) ~
                     (1|participant), subset=(df$carriage_type == "Chronic" | df$carriage_type == "Transient"),
                   data=df, REML = FALSE)

model2
anova(model2_null,model2)


##Chronic vs non-carriers

model3 = lmer(shannon_index ~ carriage_type +
                (1|participant), subset=(df$carriage_type == "Chronic" | df$carriage_type == "Non-carriers"),
              data=df, REML = FALSE)

model3_null = lmer((df[,2]) ~
                     (1|participant), subset=(df$carriage_type == "Chronic" | df$carriage_type == "Non-carriers"),
                   data=df, REML = FALSE)

model3
anova(model3_null,model3)
```

Linear mixed effects analysis of the relationship between the Shannon index and KpSC relative abundances

```
###linear mixed model regression analysis of shannon index vs KpSC abundance
library(lme4)

df= read.csv("alpha_diversity1.csv")
df_59= read.csv("59.csv")
df_60= read.csv("60.csv")

abundance <-  log(df[,6] + .1)
abundance_59 <-  log(df_59[,6] + .1)
abundance_60 <-  log(df_60[,6] + .1)

model1 <-  lmer(shannon_index ~ abundance +
                  (1|participant), data=df, REML = FALSE)

model1_null <-  lmer(shannon_index ~
                  (1|participant), data=df, REML = FALSE)

model1
anova(model1_null,model1)

model59 <-  lmer(shannon_index ~ abundance_59 +
                   (1|participant), data=df_59, REML=FALSE)

model59_null= lmer(shannon_index ~
                  (1|participant), df_59, REML=FALSE)

model59
anova(model59_null,model59)

model60 <-  lmer(shannon_index ~ abundance_60 +
                   (1|participant), data=df_60, REML=FALSE)

model60_null= lmer(shannon_index ~
                  (1|participant), df_60, REML=FALSE)

model60
anova(model60_null,model60)

```

Appendix 3: Supplementary tables

Table S1. Relative abundance of KpSC from ZKIR-qPCR

Prticpant ID	Month	Average genome copies/ng	Prticpant ID	Month	Average genome copies/ng
4468	Sep	0.009	4570	Jan	7.300
4468	Nov	0.098	4570	Mar	0.180
4468	Dec	0.011	4571	Sep	0.000
4468	Jan	0.100	4571	Nov	0.000
4468	Feb	0.000	4571	Dec	0.000
4468	Mar	0.164	4571	Jan	0.000
4469	Sep	490.667	4571	Feb	0.026
4469	Nov	310.533	4571	Mar	0.007
4469	Dec	1693.333	4572	Sep	0.000
4469	Jan	724.100	4572	Nov	0.324
4469	Feb	1676.462	4572	Dec	0.084
4469	Mar	29.308	4572	Jan	0.023
4470	Nov	74.364	4572	Feb	129.562
4470	Sep	192.933	4572	Mar	0.114
4473	Sep	18066.667	4573	Sep	0.000
4473	Nov	1262.667	4573	Nov	0.008
4473	Dec	6200.000	4573	Dec	0.000
4473	Jan	6340.200	4573	Jan	0.000
4473	Feb	6820.869	4573	Feb	0.068
4473	Mar	1922.589	4573	Mar	0.027
4474	Sep	0.000	4575	Sep	0.015
4474	Nov	9.890	4575	Nov	0.151
4474	Dec	0.000	4575	Dec	0.201
4474	Jan	0.000	4575	Jan	0.300
4474	Feb	0.054	4575	Feb	1.926
4474	Mar	0.082	4575	Mar	0.143
4476	Sep	1.510	4580	Sep	0.007
4476	Nov	0.770	4580	Nov	0.060
4476	Dec	8.405	4580	Dec	8.994
4476	Jan	0.500	4580	Jan	0.200
4480	Sep	0.025	4580	Feb	0.035
4480	Nov	18.178	4580	Mar	2.405
4480	Dec	56.133	4585	Sep	0.000
4480	Jan	0.000	4585	Nov	0.007
4480	Feb	0.064	4585	Dec	0.628
4480	Mar	0.000	4585	Feb	0.024
4481	Sep	0.014	4585	Jan	1.000

Prticpant ID	Month	Average genome copies/ng	Prticpant ID	Month	Average genome copies/ng
4481	Nov	0.028	4585	Mar	0.146
4481	Dec	0.158	4586	Sep	0.000
4481	Jan	0.000	4586	Nov	0.154
4481	Feb	0.015	4586	Dec	0.021
4481	Mar	0.018	4586	Jan	0.000
4484	Sep	156.267	4586	Feb	0.061
4484	Nov	1695.477	4586	Mar	0.000
4484	Dec	15133.333	4592	Sep	89866.667
4484	Jan	888.500	4592	Nov	13266.667
4484	Feb	202.578	4592	Dec	24176.541
4484	Mar	847.184	4592	Jan	26675.800
4488	Sep	0.035	4592	Feb	43846.785
4488	Nov	0.000	4592	Mar	42528.920
4488	Dec	1234.614	4593	Sep	10.495
4488	Jan	0.400	4593	Nov	2.940
4488	Feb	0.000	4593	Dec	0.706
4488	Mar	0.174	4593	Jan	1.800
4489	Sep	0.075	4594	Sep	88.533
4489	Nov	0.063	4594	Nov	34.741
4489	Jan	0.000	4594	Dec	0.006
4489	Feb	0.026	4594	Jan	0.000
4490	Sep	0.016	4594	Feb	0.000
4490	Nov	0.000	4594	Mar	0.036
4490	Dec	23.203	4595	Sep	0.602
4490	Jan	0.000	4595	Nov	1.109
4490	Feb	0.058	4595	Jan	5.400
4490	Mar	0.000	4595	Mar	48.151
4491	Sep	0.008	4597	Sep	0.000
4491	Nov	0.444	4597	Nov	0.057
4491	Dec	0.425	4597	Dec	9.513
4491	Jan	0.000	4597	Jan	0.000
4491	Feb	0.088	4597	Feb	0.000
4491	Mar	3.589	4597	Mar	0.284
4492	Sep	0.000	4600	Sep	0.000
4492	Nov	0.000	4601	Sep	2.729
4492	Dec	0.011	4601	Nov	0.000
4492	Jan	0.000	4601	Dec	1.890
4492	Feb	0.000	4601	Jan	0.200
4492	Mar	0.000	4601	Feb	0.128
4496	Sep	0.000	4601	Mar	0.013
4496	Nov	0.000	4604	Nov	2182.768
4496	Dec	0.069	4604	Sep	47.200

Prticpant ID	Month	Average genome copies/ng	Prticpant ID	Month	Average genome copies/ng
4496	Jan	0.000	4604	Dec	575.667
4496	Feb	0.080	4604	Jan	5499.000
4496	Mar	0.000	4604	Feb	6392.746
4497	Sep	7.616	4604	Mar	7629.783
4497	Nov	4.903	4608	Sep	1.468
4497	Dec	0.701	4608	Nov	22.811
4497	Jan	0.100	4608	Dec	0.041
4497	Feb	74.422	4608	Jan	21.600
4497	Mar	12.277	4608	Feb	0.278
4501	Sep	0.015	4608	Mar	768.784
4503	Sep	0.019	4610	Sep	0.000
4503	Nov	0.006	4610	Nov	126.267
4503	Dec	0.010	4610	Dec	0.000
4503	Jan	0.000	4610	Jan	0.000
4503	Feb	0.954	4610	Feb	0.137
4503	Mar	0.037	4610	Mar	0.017
4504	Sep	0.014	4613	Sep	0.000
4504	Nov	0.866	4613	Nov	2.607
4504	Dec	0.000	4613	Dec	0.538
4504	Jan	0.000	4613	Jan	0.100
4504	Feb	0.080	4613	Feb	0.084
4504	Mar	0.027	4613	Mar	0.007
4506	Sep	0.052	4619	Sep	0.007
4506	Nov	0.110	4619	Nov	0.020
4506	Dec	3.972	4619	Dec	0.097
4506	Jan	0.400	4619	Jan	0.300
4506	Feb	0.000	4619	Feb	0.031
4506	Mar	0.024	4619	Mar	0.039
4508	Sep	0.017	4620	Sep	3.956
4508	Nov	0.011	4620	Nov	0.003
4508	Dec	0.000	4620	Dec	0.067
4508	Jan	0.000	4620	Jan	0.000
4508	Feb	0.000	4620	Feb	0.108
4508	Mar	0.000	4620	Mar	0.000
4511	Sep	3.157	4623	Sep	0.000
4511	Nov	0.000	4623	Nov	0.000
4511	Dec	6.007	4623	Dec	0.003
4511	Jan	0.000	4623	Jan	0.000
4511	Feb	0.087	4623	Feb	0.006
4511	Mar	0.156	4623	Mar	0.009
4516	Sep	0.018	4626	Sep	32.884
4516	Nov	1.102	4626	Nov	0.007

Prticpant ID	Month	Average genome copies/ng	Prticpant ID	Month	Average genome copies/ng
4516	Dec	0.731	4626	Dec	0.013
4516	Jan	138.900	4626	Jan	0.100
4516	Feb	0.104	4626	Feb	11.954
4516	Mar	0.221	4626	Mar	9.731
4519	Sep	0.012	4628	Sep	0.000
4519	Nov	0.000	4628	Nov	0.034
4519	Dec	0.000	4628	Dec	0.102
4519	Jan	0.000	4628	Jan	0.000
4519	Feb	0.017	4628	Mar	0.005
4519	Mar	0.000	4634	Sep	0.031
4520	Sep	0.185	4634	Nov	0.000
4520	Nov	0.417	4634	Dec	0.010
4520	Dec	1.404	4635	Sep	87.200
4520	Jan	0.700	4637	Sep	0.061
4520	Feb	0.217	4637	Nov	0.121
4520	Mar	0.000	4637	Dec	0.351
4522	Sep	0.005	4637	Jan	0.100
4522	Nov	0.000	4637	Feb	0.000
4522	Dec	0.000	4637	Mar	0.377
4522	Jan	0.000	4642	Sep	3.999
4522	Feb	0.503	4642	Nov	0.996
4523	Sep	0.000	4642	Dec	0.140
4523	Nov	0.068	4642	Jan	0.100
4523	Dec	0.108	4642	Feb	0.000
4523	Jan	0.000	4642	Mar	0.605
4523	Feb	0.013	4643	Sep	0.000
4523	Mar	0.000	4643	Nov	0.000
4524	Sep	0.005	4643	Dec	0.003
4524	Nov	1.327	4643	Jan	0.300
4524	Dec	0.451	4643	Feb	3.515
4524	Jan	0.200	4643	Mar	0.003
4524	Feb	0.000	4644	Sep	0.000
4526	Sep	0.027	4644	Nov	0.256
4526	Nov	830.667	4644	Dec	0.012
4526	Dec	0.065	4644	Jan	0.000
4526	Jan	0.000	4644	Feb	0.000
4526	Feb	0.031	4644	Mar	0.001
4526	Mar	2.891	4647	Sep	0.000
4527	Sep	63.467	4647	Nov	0.000
4527	Nov	2.465	4647	Dec	0.031
4527	Dec	13.621	4647	Jan	0.000
4527	Jan	12.500	4647	Feb	0.000

Prticpant ID	Month	Average genome copies/ng	Prticpant ID	Month	Average genome copies/ng
4527	Feb	42.117	4647	Mar	0.007
4527	Mar	4.340	4650	Sep	0.000
4528	Sep	0.000	4650	Nov	0.000
4528	Nov	0.189	4650	Dec	0.073
4528	Dec	0.145	4650	Jan	0.000
4528	Jan	0.000	4650	Feb	0.109
4528	Feb	0.060	4650	Mar	0.028
4528	Mar	0.027	4654	Sep	0.083
4529	Sep	0.027	4654	Nov	0.000
4530	Sep	0.016	4654	Dec	0.000
4530	Nov	0.000	4654	Jan	0.000
4530	Dec	0.088	4654	Feb	0.000
4530	Jan	0.100	4654	Mar	0.126
4530	Feb	0.286	4657	Sep	4.933
4530	Mar	0.026	4657	Nov	36.805
4531	Sep	0.339	4657	Dec	176.267
4531	Nov	36.123	4657	Jan	18.500
4533	Sep	0.121	4657	Feb	0.230
4533	Nov	0.253	4657	Mar	0.461
4533	Dec	1.598	4660	Sep	0.635
4533	Jan	0.300	4660	Nov	0.000
4533	Feb	0.000	4660	Dec	0.872
4538	Sep	0.008	4660	Jan	0.000
4538	Nov	0.169	4660	Feb	73.326
4538	Dec	0.013	4660	Mar	3.486
4538	Jan	0.000	4664	Sep	0.028
4538	Feb	0.000	4664	Nov	0.000
4538	Mar	0.000	4664	Dec	0.030
4539	Sep	0.025	4664	Jan	0.000
4539	Nov	0.000	4664	Feb	0.035
4539	Dec	0.000	4664	Mar	0.020
4539	Jan	0.000	4668	Sep	0.024
4539	Feb	0.000	4668	Nov	0.000
4539	Mar	0.000	4668	Jan	0.000
4540	Sep	0.006	4669	Sep	0.008
4540	Nov	0.000	4669	Nov	0.278
4540	Dec	0.000	4669	Dec	36.301
4540	Jan	0.000	4669	Jan	0.000
4540	Feb	0.000	4669	Feb	0.000
4540	Mar	0.000	4669	Mar	0.000
4543	Sep	0.027	4670	Sep	0.007
4543	Nov	0.015	4670	Nov	0.026

Prticpant ID	Month	Average genome copies/ng	Prticpant ID	Month	Average genome copies/ng
4543	Dec	0.025	4670	Dec	9.429
4543	Jan	0.000	4670	Jan	0.000
4543	Feb	0.000	4670	Feb	0.113
4543	Mar	0.000	4670	Mar	0.297
4546	Sep	94.667	4672	Sep	0.000
4546	Nov	0.341	4672	Nov	0.000
4546	Dec	0.038	4672	Dec	0.074
4546	Jan	0.000	4672	Jan	0.000
4546	Feb	0.014	4672	Feb	0.522
4546	Mar	0.000	4672	Mar	0.019
4549	Sep	0.023	4677	Sep	215.467
4549	Nov	0.007	4677	Nov	8.592
4549	Dec	0.042	4677	Dec	29.434
4549	Jan	0.000	4677	Jan	0.300
4549	Feb	2.075	4677	Feb	27.542
4549	Mar	0.012	4677	Mar	5.631
4550	Sep	0.016	4679	Sep	117.067
4550	Nov	0.056	4679	Nov	102.157
4550	Dec	0.727	4679	Dec	97.972
4550	Jan	3.100	4679	Jan	177.100
4550	Feb	0.000	4679	Feb	1063.900
4550	Mar	0.063	4679	Mar	121.817
4553	Sep	0.005	4682	Sep	0.000
4553	Nov	0.519	4682	Nov	0.000
4553	Dec	15.108	4682	Dec	0.000
4553	Jan	0.000	4682	Jan	0.000
4553	Feb	0.104	4682	Mar	0.000
4553	Mar	0.130	4684	Jan	0.200
4554	Sep	0.003	4684	Feb	3.146
4554	Nov	0.831	4684	Mar	0.550
4554	Dec	0.054	4685	Sep	119.067
4554	Jan	0.100	4685	Nov	0.128
4554	Feb	0.886	4685	Dec	0.039
4554	Mar	0.000	4685	Jan	0.100
4559	Sep	0.000	4685	Feb	0.079
4559	Nov	0.000	4685	Mar	0.001
4559	Dec	0.000	4689	Sep	0.009
4559	Jan	0.000	4689	Nov	44666.667
4559	Feb	0.086	4689	Dec	44111.920
4559	Mar	0.000	4689	Jan	22647.400
4560	Sep	0.000	4689	Feb	48654.176
4560	Nov	0.000	4689	Mar	29807.114

Prticpant ID	Month	Average genome copies/ng	Prticpant ID	Month	Average genome copies/ng
4560	Dec	0.019	4690	Sep	0.159
4560	Jan	0.000	4690	Nov	0.170
4560	Feb	0.032	4690	Dec	0.000
4560	Mar	0.002	4690	Jan	0.000
4561	Sep	0.000	4690	Feb	0.139
4561	Nov	0.227	4690	Mar	530.101
4561	Dec	0.003	4692	Sep	0.000
4561	Jan	0.000	4692	Nov	0.000
4561	Feb	0.048	4692	Dec	0.000
4561	Mar	0.000	4692	Jan	0.000
4562	Sep	0.618	4692	Feb	0.005
4562	Nov	0.179	4692	Mar	0.014
4562	Dec	1.224	4695	Sep	4.966
4562	Jan	1.200	4695	Nov	0.049
4562	Feb	0.000	4695	Dec	0.090
4562	Mar	0.147	4695	Jan	0.400
4563	Sep	0.003	4695	Feb	0.010
4563	Nov	0.010	4695	Mar	0.279
4563	Dec	0.029	4696	Sep	0.000
4563	Jan	0.000	4696	Nov	0.012
4563	Feb	0.051	4696	Dec	0.011
4563	Mar	0.019	4696	Jan	14.700
4564	Sep	0.000	4696	Feb	5.291
4564	Nov	0.000	4696	Mar	0.000
4564	Dec	0.000	4697	Sep	0.328
4564	Jan	0.000	4697	Jan	0.100
4564	Feb	0.000	4697	Feb	0.073
4564	Mar	0.000	4697	Mar	0.003
4565	Sep	0.022	4697	Dec	0.646
4565	Nov	0.016	4698	Sep	0.008
4565	Dec	0.010	4698	Nov	0.000
4565	Jan	0.000	4698	Dec	0.000
4565	Feb	0.033	4698	Jan	0.000
4565	Mar	0.076	4698	Feb	0.000
4566	Sep	0.378	4698	Mar	0.013
4566	Nov	0.010	4700	Sep	0.017
4566	Dec	0.062	4700	Nov	0.028
4566	Feb	0.072	4700	Dec	0.014
4566	Mar	0.178	4700	Jan	22.000
4567	Sep	0.000	4700	Mar	0.041
4567	Nov	0.000	4702	Sep	1.106
4567	Dec	0.034	4702	Nov	1.497

Prticpant ID	Month	Average genome copies/ng	Prticpant ID	Month	Average genome copies/ng
4567	Jan	0.000	4702	Dec	0.139
4567	Feb	0.000	4702	Mar	1.538
4567	Mar	0.265	4704	Sep	0.000
4570	Sep	0.198	4704	Nov	8.997
4570	Nov	0.016	4704	Jan	0.000
4570	Dec	1.227	4704	Feb	0.127
4570	Feb	0.248			

Table S2. Statistical parameters for the comparison between Shannon index and KpSC carriage types

	Non-carriers	**Transient**	**Chronic**
Minimum	2.4	2.18	2.23
First Quartile	2.87	2.80	2.41
Median	3.03	3.03	2.98
Third Quartile	3.22	3.21	3.11
Maximum	3.46	3.52	3.35

Table S3. Statistical parameters for the comparison between age and KpSC carriage types

	Chronic	**Non-carriers**	**Transient**
Min	67	49	50
First Quartile	70.5	55	64
Median	75	69.5	68
Third Quartile	76	74	74
Maximum	80	79	81

Table S4. SNP distances among isolates belonging to ST20 by SNP-dist

	T7-012-Isolate	**T7-037-Isolate**	**T7-463-Isolate**	**T7-308-Isolate**
T7-012-Isolate	0	1678	1432	29866
T7-037-Isolate	1678	0	605	30136
T7-463-Isolate	1432	605	0	30065
T7-308-Isolate	29866	30136	30065	0

Table S5. Relative abundance of the KpSC in the seven highest abundant participants

Participant ID	Month	Kp (%) Centrifuge	KpSC (%) StrainGE	Participant ID	Month	Kp (%) Centrifuge	KpSC (%) StrainGE
4469	Sep	0.446	0.238	**4592**	Dec	16.292	9.21
4469	Nov	0.146	0.119	**4592**	Jan	8.256	5.231
4469	Dec	0.625	0.404	**4592**	Feb	11.376	9.493
4469	Jan	0.200	0.155	**4592**	Mar	18.456	11.589
4469	Feb	1.165	0.598	**4604**	Sep	0.053	0.089
4469	Mar	0.498	0.311	**4604**	Dec	0.324	0.598
4473	Sep	6.583	4.72	**4604**	Jan	1.048	1.175
4473	Nov	0.610	0.39	**4604**	Feb	1.044	1.045
4473	Dec	0.941	0.617	**4604**	Mar	2.179	2.099
4473	Jan	1.634	1.084	**4679**	Sep	0.076	0.294
4473	Feb	3.127	1.831	**4679**	Nov	0.062	undetectable
4473	Mar	0.416	0.311	**4679**	Dec	0.159	0.106
4484	Sep	0.283	0.295	**4679**	Feb	0.793	0.316
4484	Nov	1.845	0.788	**4679**	Mar	0.310	0.385
4484	Dec	13.879	6.734	**4689**	Sep	0.001	undetectable
4484	Jan	0.353	0.34	**4689**	Nov	29.655	17.399
4484	Feb	0.461	0.484	**4689**	Dec	20.116	6.43
4484	Mar	1.699	0.366	**4689**	Jan	17.104	6.353
4592	Sep	15.055	5.749	**4689**	Feb	16.874	7.823
4592	Nov	21.696	9.51	**4689**	Mar	20.847	11.226