UiT The Arctic University of Norway

Faculty of Health Sciences

Human Gut Colonisation by the *Klebsiella pneumoniae* Species Complex: Detection, Duration, Dynamics, and Microbiota Associations

Kenneth Lindstedt

A dissertation for the degree of Philosophiae Doctor (PhD)

A dissertation for the degree of Philosophiae Doctor

Human Gut Colonisation by the *Klebsiella pneumoniae* Species Complex: Detection, Duration, Dynamics, and Microbiota Associations

Kenneth Lindstedt

March 2024

Research group of Host-Microbe Interactions,
Department of Medical Biology,
Faculty of Health Sciences,
UiT – The Arctic University of Norway

Acknowledgements

First and foremost, I would like to extend my sincere gratitude to my supervisors Arnfinn Sundsfjord, Ørjan Samuelsen, Torunn Pedersen, and Erik Hjerde. Thank-you for giving me this opportunity to contribute to an exciting and important area of research, and for always making yourselves available to offer guidance and support along the way.

Most of all thanks to my main supervisor Arnfinn. You were always willing to give your time to offer advice, discuss ideas, impart wisdom (about science and life in general), and gave me space to develop as a researcher. You made both Lillan and I feel very at home here in Tromsø and I know Sonja considers you and Ann-Britt as an additional pair of besteforeldre.

Thanks to Ørjan, your valuable insights, willingness to share your expertise and experience, and for always calling a spade a spade, was greatly appreciated. Thank-you Torunn for your valuable feedback and comments and for helping me find my feet in the lab. Thanks, Erik, for helping get our data analysis off the ground. A big thanks also to Einar Holsbø for all your time, guidance, advice, and clear explanation of technical concepts that made paper III possible.

Thanks to all my friends and colleagues at HMI for creating such an open, friendly, and collaborative working environment. Thanks Wasifa Kabir, your excellent and tireless work has made much of this project possible and is greatly appreciated. Thanks also Dorota Buczek for so willingly sharing your bioinformatic skills and expertise. A special thanks also to Bhupender Singh, Kjersti Julin, Ellen Josefsen, and Bjørg Haldorsen for all the technical support.

To Lil, thank-you for taking the leap to move to the arctic. Thank-you for your continual support (especially in these last weeks!) and sharing in this amazing adventure. To our darling Sonja, thank-you for being such a source of joy and always brightening up the day with your smile.

To Mum, thanks for all your love and support throughout the years and for making the trip all the way from Australia to visit us (twice!) up here in the north.

And thanks Dad. You were always a solid rock of support throughout my life, education, and carrier. Your dedication and genuine care for your patients was my inspiration to pursue a career in medicine.

Table of Contents

List of Figures

Abbreviations

ARG: antimicrobial resistance gene

AMR: antimicrobial resistance

ANI: average nucleotide identity

AST: antimicrobial susceptibility testing

bp: base-pair

CCF: commensal colonisation factor

CG: clonal group

cgMLST: core-genome multi-locus sequence type

cKp: classical *Klebsiella pneumoniae*

CLSI: Clinical and Laboratory Standards Institute

CPKp: carbapenemase-producing *Klebsiella pneumoniae*

C_q: quantification cycle

DNA: deoxynucleic acid

DNB: DNA nanoball

ESBL: extended spectrum beta-lactamase

EUCAST: European Committee on Antimicrobial Susceptibility Testing

FWER: family-wise error rate

FDR: false discovery rate

FMT: faecal microbiota transplant

GIT: gastrointestinal tract

HAI: hospital-acquired infection

HRT: high-resolution melting

HvKp: hypervirulent *Klebsiella pneumoniae*

IEC: intestinal epithelial cells

Kp: *Klebsiella pneumoniae*

KpSC: the *Klebsiella pneumoniae* Species Complex

LAMP: loop-mediated amplification

LCA: lowest common ancestor

LIN: life identification number

LMIC: lower- and middle-income countries

LMM: linear mixed-effects model

LOD: limit of detection

LOQ: limit of quantification

MALDI-TOF: matrix-assisted laser desorption/ionization-time of flight

Mbp: megabase-pairs

MDR: multidrug-resistant

MIC: mean inhibitory concentration

MLSL: multi-level single linkage

MLST: multi-locus sequence type

NAAT: nucleic acid amplification technology

NASBA: nucleic acid sequence-based amplification

NeLS: Norwegian e-infrastructure for Life Sciences

NGS: next-generation sequencing

NRIS: Norwegian Research Infrastructure Service

ONT: Oxford Nanopore

PDR: pandrug-resistant

qPCR: quantitative polymerase chain reaction

RCA: rolling circle amplification

RNA: ribonucleic acid

RT-qPCR: reverse transcription quantitative polymerase chain reaction

SDD: selective decontamination of the digestive tract

SMRT: single molecule real-time technology

SNP: single nucleotide polymorphism

SL: sublineage

ST: sequence type

TMA: transcription mediated amplfication

WGS: whole genome sequencing

WHO: the World Health Organization

WMS: whole metagenomic sequencing

XDR: extensively drug-resistant

List of Papers

Paper I

Lindstedt K, Buczek D, Pedersen T, Hjerde E, Raffelsberger N, Suzuki Y, Brisse S, Holt K, Samuelsen Ø, Sundsfjord A. **Detection of *Klebsiella pneumoniae* human gut carriage: a comparison of culture, qPCR, and whole metagenomic sequencing methods**. *Gut Microbes.* 2022; 14(1): 14(1):2118500. doi: 10.1080/19490976.2022.2118500.

Paper II

Buszek D, **Lindstedt K**, Kabir W, Mäklin T, Thorpe HA, Suzuki Y, Corander J, Samuelsen Ø, Sundsfjord A. **Sequence type and strain level detection of *Klebsiella pneumoniae* in complex bacterial communities by metagenomics: comparative performance of mSWEEP and StrainGE bioinformatic tools**. (manuscript)

Paper III

Lindstedt K, Kabir W, Holsbø E, Buczek D, Pedersen T, Holt K, Brisse S, Samuelsen Ø, Sundsfjord A. **Longitudinal analysis of *Klebsiella pneumoniae* Species Complex human gut colonization identifies a subgroup of high-abundance persistent carriers with strong microbiota associations** (manuscript)

Abstract

The *Klebsiella pneumoniae* Species Complex (KpSC) is an important group of human pathogens, of which, *Klebsiella pneumoniae* is most clinically important and frequently associated with antimicrobial resistance (AMR). *K. pneumoniae* gastrointestinal tract (GIT) colonisation is a key risk factor for infection and hub for AMR dissemination.

Here, we investigated two molecular-based tools, the ZKIR-qPCR and whole metagenomic sequencing (WMS), to improve detection, quantification, and sequence type (ST) analysis of the KpSC from human faecal samples. These tools were then applied to longitudinally collected samples from 108 community-based adults recruited from the general population in Tromsø municipality to investigate key questions regarding KpSC ecology: the duration, dynamics, host- and microbiota-associations of KpSC GIT colonisation.

In **paper I**, the ZKIR-qPCR demonstrated the highest detection sensitivity, positive in 52/52 KpSC culture-positive samples, and 11/51 and 23/47 culture-negative samples, using a direct-faecal and culture-enrichment method, respectively. Using a 0.01% abundance cut-off, WMS detected the KpSC in 37/52 culture-positive samples but was inclined to false positives at low KpSC abundances. Both tools accurately quantified the KpSC across a range of abundances.

Paper II showed WMS performed accurate KpSC ST detection, agreeing with single colony whole genome sequencing in 44/49 and 46/49 culture-enriched faecal samples using two tools, StrainGE and mSWEEP, respectively. Both tools could detect within-sample ST diversity and StrainGE could recreate accurate phylogenetic relationships between closely related strains.

In **paper III**, we found 27.1%, 69.4%, and 3.5% of participants were persistent (positive in all samples), transient (positive one to five times), or non-carriers of the KpSC, respectively Persistent carriers had higher KpSC GIT abundance and tended to retain the same ST for multiple months. KpSC GIT abundance associated positively with *Bacteroides* and *Phocaeicola*, and negatively with *Bifidobacterium*, *Alistipes*, *Akkermansia*, and multiple Bacillota (Firmicutes). Older age, travel abroad, and diabetes mellitus were positively associated with KpSC abundance, while antimicrobial use was negatively associated.

This project shows that qPCR and WMS are valid tools for KpSC detection and analysis from human faecal samples, and demonstrated important findings regarding KpSC GIT colonisation duration, diversity, dynamics, and microbiota associations. This project lays the foundation for

future studies investigating mechanisms behind persistent KpSC GIT colonisation and the microbiota taxa influencing this, with potential to lead to important insights into decolonisation strategies.

Introduction

Klebsiella pneumoniae (Kp) is a non-motile, encapsulated, facultative anaerobic gram-negative bacteria belonging to the Enterobacterales order that is found ubiquitously in humans, animals, and the environment[1-3]. In addition to being a common and serious cause of infection in humans, Kp is presenting major challenges to public health due to its high propensity for the acquisition and dissemination of antimicrobial resistance (AMR)[2-4]. The spread of AMR by Kp has become a global problem, with much of the burden effecting lower income countries[3, 5, 6]. Consequently, the World Health Organisation (WHO) has classified Kp as a critical priority pathogen for further research and development of new treatments[5].

Recently, genomic insights have demonstrated Kp belongs to the broader *Klebsiella pneumoniae* Species Complex (KpSC), consisting of seven closely related phylogroups sharing 95-96% average nucleotide identity (ANI): Kp1 (*K. pneumoniae sensu stricto*), Kp2 and Kp4 (*K. quasipneumoniae* subsp. *quasipneumoniae* and subsp. *similipneumoniae*, respectively), Kp3 and Kp5 (*K. variicola* subsp. *variicola* and subsp. *tropica*, respectively), Kp6 (*K. quasivariicola*), and Kp7 (*K. africana*) (Figure 1)[3, 7-11]. In addition to Kp1 (hereafter Kp), other members of this complex, in particular Kp3 (hereafter *K. variicola*), are increasingly recognised as important human pathogens and have been associated with carriage of AMR, also warranting further research[12-14].

Key questions, however, remain regarding much of the ecology of this group of pathogens, particularly with respect to niche specialisation, inter-niche strain cross talk, and of particular relevance to this project, colonisation dynamics within the human gastrointestinal tract (GIT). Furthermore, current surveillance methods for the detection of Kp and associated AMR determinants, while highly robust and validated, have limitations that could be complimented by new and cutting-edge molecular-based culture-independent techniques.

This project evaluated two such molecular-based methods, quantitative polymerase chain reaction (qPCR) and the next-generation sequencing (NGS) technology, whole metagenomic sequencing (WMS). These methods were compared to conventional culture-based detection of the KpSC in complex microbiome samples from the human GIT. Analysis of WMS data was further evaluated for the ability to identify the KpSC in samples at the sequence type (ST) level. Comparative strengths and weaknesses of each method were identified. These methodologies were then applied to investigate a key knowledge gap in KpSC ecology with relevance to human

health and spread of AMR, namely, the duration, dynamics, microbiota- and host-associations of human GIT colonisation by the KpSC.

This general introduction will therefore discuss: 1) the genetics and phylogenetics of Kp and the KpSC, followed by the role of Kp and the KpSC in human disease, spread of AMR, and colonisation of the human GIT, and 2) current and emerging AMR and pathogen surveillance and diagnostic methodologies, with a focus on the molecular-based methods, qPCR and NGS, particularly WMS. As Kp is the most clinically relevant member of the KpSC, both in terms of total burden of disease and carriage of AMR, this species will be the focus of the introduction, however, other KpSC members are discussed where relevant.

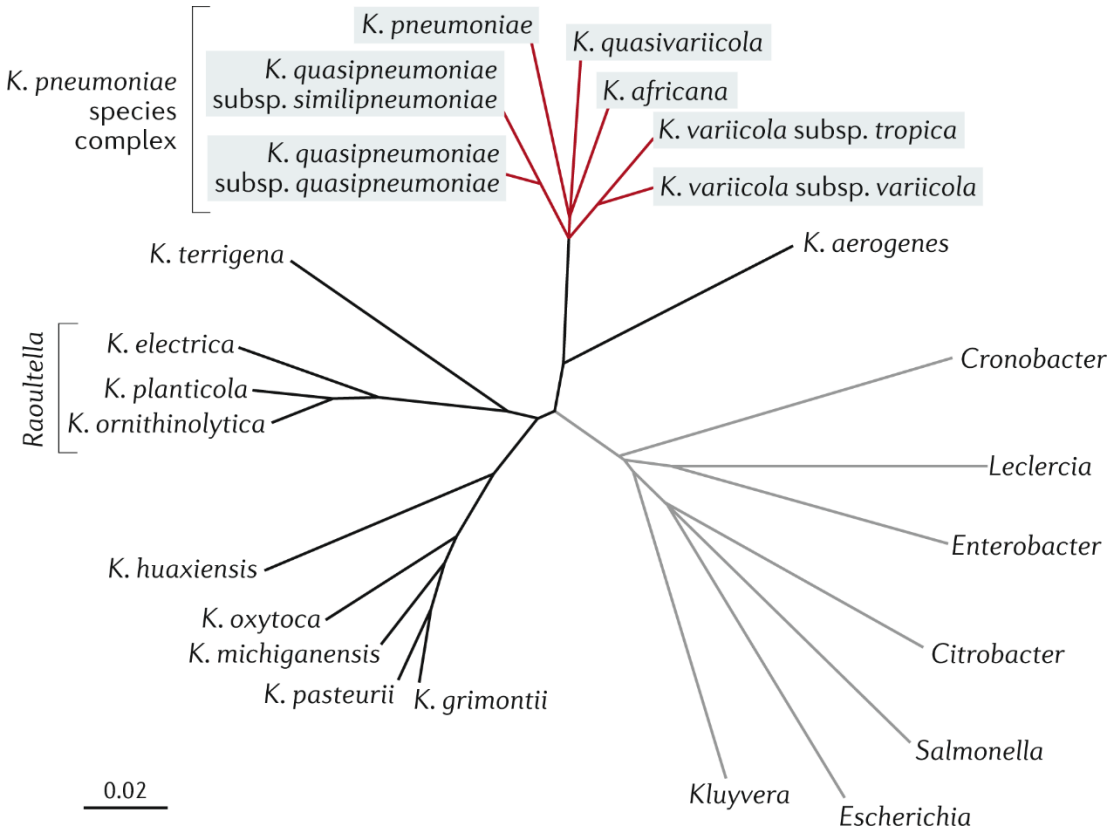


Figure 1. Taxonomic relationships of the *Klebsiella pneumoniae* Species Complex. Whole-genome-based phylogenetic tree displaying relationships between the KpSC (red branches), non-KpSC *Klebsiella* species (black branches), and other Enterobacterales (grey branches). Bar represents estimated average nucleotide divergence. Reprinted with permission from Nat. Rev. Microbiol[3].

1. *Klebsiella pneumoniae* and the *Klebsiella pneumoniae* Species Complex (KpSC)

1.1 The Genetics, Phylogenetics, and Lineage Classification of *K. pneumoniae*

The population structure of Kp is highly diverse, composed of hundreds of deep branching lineages that differ by ~0.5% ANI[15]. The size of each Kp genome is a typically around 5-6 Mbp and contains around 5,000-6,000 genes[3]. Comparative genomics has demonstrated the core genome of Kp, which are the genes found in all Kp lineages, comprises around 1,700 genes, while the remainder are composed of accessory genes shared variably across lineages[15]. Within the phylogenetic framework of Kp, the sum of all core and accessory genes, known as the pan-genome, is large and highly diverse, likely containing upwards of 100,000 genes with new gene continually being added[3, 15].

Lineage classification of Kp and the KpSC has classically been based on the multilocus sequence typing (MLST) scheme. This scheme uses the allele profiles of seven conserved house-keeping genes to cluster KpSC strains into sublineages known as sequence types (STs)[16]. While the MLST scheme is easy to use, well validated, and highly useful in epidemiology and public health surveillance, it can inaccurately classify lineages which have undergone recent large chromosomal recombination events[3, 17, 18]. Using comparative genomics, a core-genome multilocus sequence typing (cgMLST) scheme has therefore been developed, which currently consists of 629 core genome Kp genes and allows a more accurate definition of clonal groups (CGs) within the Kp phylogeny[19, 20].

Recently, Kp taxonomic classification has been further developed into a dual barcoding approach based on the cgMLST scheme[20]. This approach combines multilevel single linkage (MLSL) clustering and life identification numbers (LIN) to group KpSC genomes into clusters and subclusters, which includes subspecies, sublineages (SLs), and CGs, based on the number of cgMLST mismatches and applies a numerical LIN based on this grouping[20]. This approach, which is backwards compatible with the traditional MLST scheme, is designed to improve the accuracy of KpSC lineage classification and aid in strain identification for outbreak tracing, infection control, and epidemiological surveillance[20].

1.2.1 The KpSC in Human Disease

Kp is well known as an opportunistic pathogen and leading cause of hospital-acquired infections (HAIs) of the urinary tract, lungs, soft tissue, and blood stream, and is a frequent

cause of hospital outbreaks[1]. In this setting, Kp is a major cause of sepsis in vulnerable and high-risk patients, including neonates, the elderly, and the immunosuppressed, with a reported mortality rate of up to 20%[1, 21]. As of 2022, Kp accounted for approximately 7-8% of all bacteraemia cases in Norway, the third most prevalent behind *Escherichia coli* and *Staphylococcus aureus*[22].

In addition to the typical opportunistic nature of infections caused by most Kp strains, often referred to as classical *K. pneumoniae* (cKp), distinct lineages of hypervirulent Kp (HvKp) have also emerged[23, 24]. HvKp often possess a hypermucoviscous phenotype and are defined by their ability to cause invasive community-acquired infections (CAIs) in otherwise healthy, typically younger, immunocompetent hosts, with diabetes mellitus and Asian ethnic background being the major associated risk factors[23, 24]. In contrast to cKp, HvKp can cause a different spectrum of clinical syndromes, including pyogenic liver abscesses, meningitis, necrotizing fasciitis, and endopthalmitis[23-25]. Although HvKp lineages have spread globally, they are of particular concern in Asian countries strongly driven by expansion of CG23[24, 26, 27]. HvKp is now the major cause of pyogenic liver abscess in many parts of Asia and is often associated with severe disease and metastatic spread of infection to additional sites[25, 28-30].

1.2.2 Virulence factors in *K. pneumoniae*

Several acquired factors associated with HvKp have been shown to increase virulence of these lineages, including the iron-scavenging siderophores aerobactin (*iuc*), salmochelin (*iro*), and yersiniabactin (*ybt*), the regulators of mucoid phenotype *rmpA* and *rmpA2*, and the genotoxin colibactin (collectively known as virulence factors) (Figure 2)[15, 31-35]. In addition, HvKp are strongly associated with the polysaccharide capsule types K1 and K2, which are also associated with enhanced virulence, likely through resistance to phagocytosis[25, 36-38]. HvKp virulence factors are typically mobilised within the population in association with the integrative conjugative element ICE*Kp* or virulence plasmids, particularly KpVP-1 and KpVP-2[39, 40].

In contrast to HvKp, factors promoting infection in cKp lineages are not as well characterised and is an area requiring further study. Among those that have been identified are the *ybt* locus, which is associated with infections in both cKp and HvKp, and type 1 and 3 fimbriae, which are associated with urinary tract infections and biofilm formation, respectively (Figure 2)[40-42].

Evading and dampening host immune responses also appears to be an important strategy in Kp pathogenesis, particularly through attenuation of TLR-mediated inflammation by hijacking

various host anti-inflammatory pathways, such as by IL-10 induction[43, 44]. Moreover, in addition to the importance of K1 and K2 in HvKp lineages, the presence of the capsule in cKp strains has been shown to reduce phagocytosis and compliment mediated lysis/opsonization, and increase pathogenicity in murine models of pneumonia[45, 46]. Recently, a large case-control comparative genomics study of predominantly cKp lineages also identified several plasmid-borne stress resistance and regulatory genes that were associated with invasive infection[47]. Intriguingly, among these, several antimicrobial resistance genes (ARGs) were associated with an increased infection rate, despite adjustment for antimicrobial exposure[47]. While it was speculated ARGs may enhance fitness in Kp even in the absence of antimicrobial exposure, it was considered more likely these ARGs are linked to other genes enhancing infection risk carried on the same plasmid, thereby potentially complicating treatment once the infection was initiated[47].

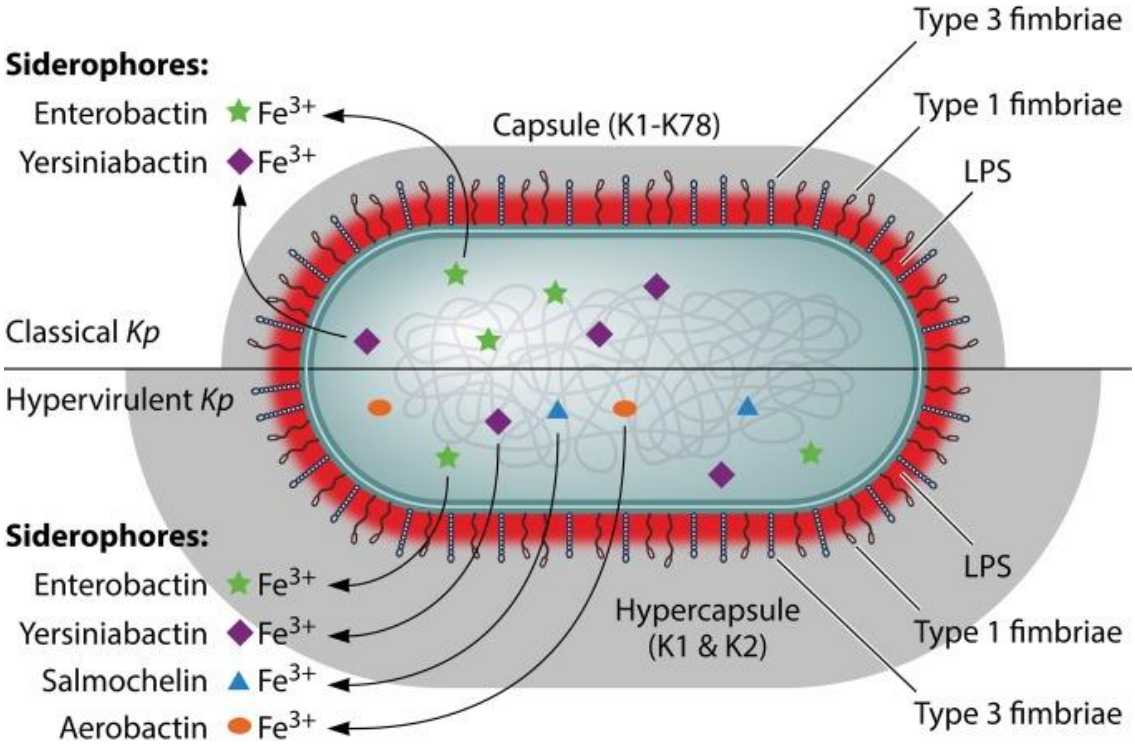


Figure 2. **Illustration of the well-characterised virulence factors for both cKp and HvKp.** All Kp lineages carry the chromosomally encoded siderophore Enterobactin, and acquisition of Yersiniabactin (cKp and HvKp), Salmochelin, and Aerobactin (HvKp) are associated with increased virulence. While cKp are associated with a variety of capsule types, HvKp are strongly associated with capsule types K1 and K2 and can produce a thicker 'hypercapsule'. Lipopolysaccharide (LPS), a component of the outer membrane and containing an O antigen, is also associated with increased virulence and inflammation. Reprinted with permission from Microbiol Mol Biol Rev[48].

1.2.3 Other KpSC Members in Human Disease

In addition to Kp, *K. variicola* and *K. quasipneumoniae* are also relatively common causes of infection, particularly in the hospital setting[10, 12, 13, 15]. Indeed, genomic analysis has revealed many infections previously attributed to Kp were in fact caused by these closely related species[10, 49, 50]. *K. variicola* and *K. quasipneumoniae* generally cause the same spectrum of disease as Kp, including HAIs of the respiratory tract, urinary tract, and bloodstream, and both have been known to cause community-acquired liver abscesses[10, 14, 51, 52]. *K. variicola* in particular has been reported to account for up to one quarter of infections caused by the KpSC and has been associated with a higher mortality rate than that of Kp[12, 13]. Moreover, Potter *et al*, recently demonstrated the ability of *K. variicola* to cause urinary tract infections varied considerably between different strains, with one isolate demonstrating higher uropathogenicity than that of Kp[53]. Like Kp, *K. variicola* has also been known to acquire siderophores and the regulators of mucoid phenotype *rmpA* and *rmpA2*[54]. Virulence determinants and pathogenic potential of different lineages within *K. variicola* and *K. quasipneumoniae*, however, remains an area in need of further study.

1.2.4 *K. pneumoniae* and AMR in the Global and Norwegian Context

Kp has emerged as one of the major perpetrators in the worldwide spread of AMR and is currently ranked third in terms of total burden of deaths attributed to or associated with resistant infections[3, 4, 6]. Kp is particularly strongly linked to the dissemination of extended spectrum beta-lactamases (ESBL) and carbapenemases, conferring resistance to third-generation cephalosporins and carbapenems, respectively (Figure 3)[4, 5]. ESBL- and carbapenemase-producing Kp (CPKp) clones have spread globally and are of high clinical importance due to their ability to cause infections with limited treatment options and associated mortality rates up to 33% and 42%, respectively[55, 56]. In 2015, it was estimated ESBL-Kp and CPKp together accounted for approximately 85,000 infections and 6,000 deaths in Europe[57]. Infections with CPKp are of particular concern due to very limited treatment options and increasing global prevalence, now accounting for up to 68-70% of all Kp infections in Taiwan, India, and Greece, and are the fastest growing cause of death due to AMR in Europe[57, 58].

In Norway, while the prevalence of both ESBL-Kp and CPKp remain comparatively low, an increase in ESBL-Kp isolated from blood from 0% to 6.8% has been observed between 2001 to 2022[22]. This increase was largely facilitated by the introduction and expansion of the clonal groups CG15 and CG307 carrying the $bla_{\text{CTX-M-15}}$ ESBL gene[12]. Both CG15 and CG307 are

recognised high-risk lineages which have spread throughout the world and are frequently associated with carriage of carbapenemases[59-61]. In addition, outbreaks of both ESBL-Kp and CPKp have been reported in Norway[62, 63].

Perhaps even more concerningly, 89 CPKp isolates were detected in Norway in 2022, representing a 216% increase from 2021[22]. These were predominantly imported isolates detected through screening in relation to recent international conflicts, although two cases of secondary transmission within Norway were confirmed, as well as two isolates linked to a hospital outbreak[22]. The introduction and potential establishment of high-risk Kp lineages such as these in Norway highlights the need for continuous and vigilant surveillance through programs such as the NORM and NORM-Vet, for monitoring AMR in humans and animals, respectively, as well as access to cutting-edge technologies for high-resolution and timely strain detection and tracking.

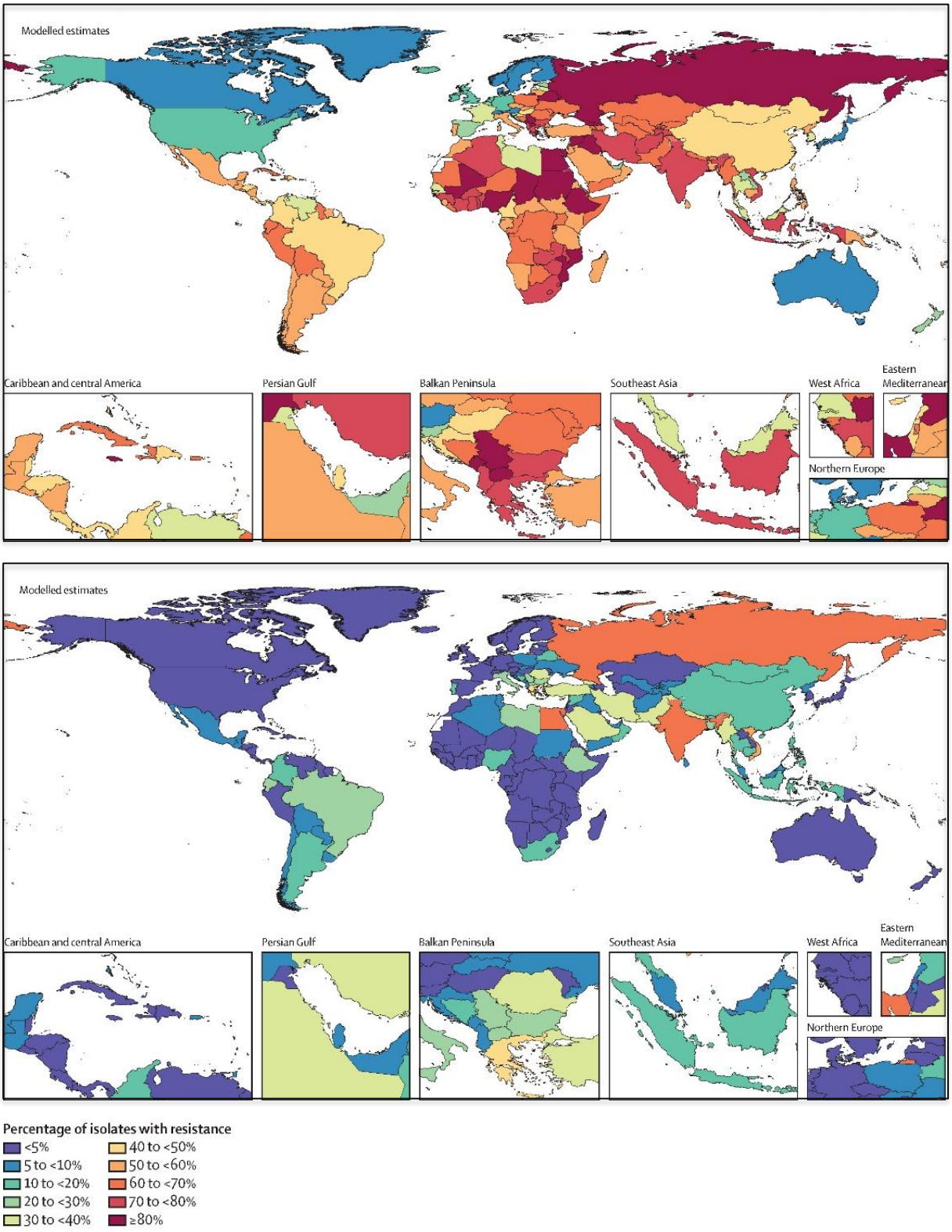


Figure 3. The global burden of carbapenem and cephalosporin resistant *Klebsiella pneumoniae*. Modelled estimates of the proportion of Kp isolates resistant to third generation cephalosporins (upper panel) and carbapenems (lower panel) by country and territory (2019). Reprinted with permission, modified from The Lancet[6].

1.2.5 Carriage and Spread of AMR by *K. pneumoniae*

In addition to the chromosomally encoded SHV beta-lactamase gene, which provides intrinsic resistance to aminopenicillins and carboxypenicillins, most acquisition and spread of AMR by Kp is mediated by plasmids and other mobile genetic elements (MGEs)[4, 15]. Large, conjugative resistance plasmids carried by Kp often contain multiple resistance elements, creating multi-drug resistant (MDR) strains, (defined as resistance toward three different antimicrobial classes)[3, 4, 64, 65]. In addition to third-generation cephalosporins and carbapenems, these MDR-plasmids commonly encode resistance to other clinically important antimicrobial classes, including aminoglycosides and fluoroquinolones[3, 4, 65]. Spread of AMR by Kp is strongly linked to the expansion and global dissemination of a relatively small number of clonal lineages that are adept at acquiring and maintaining these MDR-plasmids. In addition to CG15 and CG307 mentioned above, other high-risk lineages include CG258, CG20 (CG17), CG29, CG37, CG147, and CG101 (CG43)[4, 65]. These lineages predominantly spread within and between hospitals and are responsible for most of the global burden of carbapenem resistant Kp infections, with ST11 and ST258/512 (both belonging to CG258), ST15 (CG15), and ST101 (CG101) accounting for 69.9% of all CPKp infections in Europe[3, 66].

In parallel to the expansion of high-risk clones, horizontal gene transfer (HGT) of AMR determinants frequently occurs between Kp lineages, as well as from Kp to other species within the Enterobacterales order, and even to more distantly related taxa[2, 4, 10]. HGT thus also plays an important role in the dissemination of AMR by Kp, with a complex and overlapping network of transmissions occurring at the level of ARGs, plasmids, and strains. The complexity of the interplay between these different ARG transmission routes was recently demonstrated in a European-wide genomic survey by David *et al*[67]. Here, three distinct modes of transmission of carbapenemase genes in Kp were identified, firstly involving stable linkage of the $bla_{\text{OXA-48-like}}$ genes to a single epidemic pOXA-48-like plasmid that has spread to multiple different Kp lineages, secondly by transmission of the bla_{VIM} and bla_{NDM} in diverse plasmids and numerous lineages, and thirdly by stable association of bla_{KPC} to the single ST258/512 lineage, albeit on diverse plasmid backgrounds[67]. Similarly, León-Sampedro *et al*, recently demonstrated that within a single hospital, between-patient transfer of the high-risk clonal lineages ST11, ST307, and ST15, carrying the pOXA-48-like carbapenemase-encoding plasmid, as well as within-patient transfer of the pOXA-48-like plasmid between Kp and *E. coli*, were both occurring simultaneously[68]. These studies highlight the pivotal role of high-resolution genomic analyses

in teasing out these complex transmission networks, and potentially identifying targets to focus infection control efforts to slow the dissemination of AMR by Kp.

In addition to Kp, MDR strains of *K. variicola* and *K. quasipneumoniae* capable of acquiring both plasmid-borne ESBLs and carbapenemases have been reported, albeit at a lower frequency[10, 14, 69, 70]. Indeed Kp, *K. variicola*, and *K. quasipneumoniae* have large overlap in carriage of plasmids and MGEs encoding both AMR and virulence determinants, which are thought to be readily shared between these three species[10]. Like Kp, MDR-*K. variicola* strains have also been isolated worldwide, including Norway[12, 14]. Thus, in addition to the growing concern of MDR-Kp, MDR-*K. variicola* and MDR-*K. quasipneumoniae* are emerging as threats that will likely increasingly challenge patient care within healthcare systems.

1.2.6 Pandrug-resistance and Convergence of Hypervirulence and Multi-drug Resistance

Arguably one of the most concerning trends with respect to AMR in Kp and global health is the emergence of extensively drug-resistant (XDR) and pandrug-resistant (PDR) Kp isolates. These refer to strains that have become either non-susceptible to all but one or two antimicrobial classes (XDR) or are non-susceptible to all currently available antimicrobials (PDR)[64]. In Kp, this typically occurs when resistance develops toward the last-line antimicrobials polymyxins and/or tigecycline in combination with carbapenem resistance[71, 72]. PDR-Kp have been reported world-wide, including in Norway, with a reported mortality rate of up to 55% in patients presenting with sepsis[71-76]. The global dissemination of a plasmid-borne polymyxin resistance gene *mcr-1* raises further concerns of the potential for spread of XDR and PDR throughout the Kp population[77]. Newer conjugate antimicrobial strategies, such as the siderophore-beta-lactam conjugate cefiderocol, hold promise as alternate rescue therapies in the setting of PDR, however, resistance toward even this new antimicrobial has already been reported in Kp[78]. Thus, without the development of alternate treatment strategies to prevent infections and spread of AMR by Kp, PDR-Kp is likely to become an ever-increasing problem in hospitals world-wide.

Of similar concern to PDR-Kp is the convergence of hypervirulence traits and MDR within a single Kp strain. Historically, MDR-Kp and HvKp have remained two distinct evolutionary trajectories within different Kp clonal lineages[79]. Rates of acquisition of virulence plasmids by MDR-Kp and MDR-plasmids by HvKp, or alternatively, the convergence of virulence and AMR-determining factors on a single plasmid, however, are increasing[80-83]. Such merging of

resistance and virulence phenotypes in Kp often result in highly aggressive and difficult-to-treat infections which are associated with poor clinical outcomes[82]. Although most prevalent in Asia, especially China, hypervirulent MDR-Kp have been reported world-wide, including in Norway[81, 83]. Importantly, a recently released Rapid Risk Assessment by the ECDC highlighted the emergence of hypervirulent MDR-Kp in Europe, with increasing reports of HvKp ST23 isolates carrying carbapenemase genes from multiple EU/EEA countries, including an ongoing inter-hospital outbreak in Ireland[84]. Thus, the probability of widespread dissemination and establishment of hypervirulent MDR-Kp is now considered a very real threat requiring careful surveillance.

1.3.1 KpSC Colonisation of the Human Gastrointestinal Tract (GIT)

Members of the KpSC, in particular Kp, *K. variicola*, and *K. quasipneumoniae*, are common colonisers of the human GIT. The overall carriage prevalence of KpSC carriage within the GIT of adults varies considerably dependent on geographical setting, ranging from approximately 4-6% in the USA and Australia, 40% and 65% in Senegal and Madagascar, and up to 65-87% in Cambodia, Taiwan, and Malysia[85-88]. KpSC GIT carriage rates have also been shown to increase dramatically following hospitalisation[85]. Recently, Raffelsberger *et al*., detected a carriage rate of 16.3% in a cross-sectional study of 2975 community-based adults in Tromsø, Norway, as part of the previous Tromsø7 study[89]. In addition to adults, Kp has also been found to be a frequent early coloniser of the neonatal GIT, common following both vaginal and caesarean section deliveries[90]. Similar to differences seen in the prevalence of KpSC species in infections, Kp is the most common species isolated from human GIT samples, accounting for approximately 60-75%, followed by *K. variicola* (15-30%), and *K. quasipneumoniae* (10-30%)[87, 89]. Although the GIT is the major human reservoir for colonisation, Kp carriage has also been detected from the nares, nasopharynx, oropharynx, and skin[1, 86, 91, 92].

Several host-related factors have been identified that are associated with an increased prevalence of KpSC detection from GIT samples, which also vary depending on geographical location and population demographics. In the study by Raffelsberger *et al*, age over 60, travel to Greece or Asia, presence of inflammatory bowel disease, or use of antimicrobials, proton pump inhibitors, or non-steroidal anti-inflammatory drugs, were positive predictors of KpSC carriage[89]. Conversely, in a cross-sectional study of pregnant women in lower- and middle-income countries (LMIC) by Huynh, *et al*., antimicrobial therapy, dry fish consumption, and contact with chicken were associated with higher KpSC carriage in Madagascar, Cambodia,

and Senegal, respectively[87]. Furthermore, older age, alcohol, and smoking have been found to increase oropharyngeal Kp colonisation in adults in Vietnam[91]. Thus, efforts to reduce the burden of KpSC carriage through risk modification in hosts would clearly need to incorporate knowledge of predisposing factors at the local level.

1.3.2 GIT colonisation of the KpSC and Extra-intestinal Infection

Kp colonisation of the GIT is well-established as a major risk factor predisposing to HAIs, leading to a four-fold increase in risk of invasive extra-intestinal infection[85, 93, 94]. Genomic comparisons have demonstrated that Kp infections in hospitalized patients are caused by the patient's own colonising strain in at least 50% of occasions[85, 93]. Indeed, Martin *et al*. demonstrated perfect concordances between infecting and colonising strains in n = 7 and n = 4 patients with Kp-induced pneumonia and urinary tract infections, respectively[93]. Moreover, qPCR-based quantification studies have demonstrated a higher load of Kp in faecal samples predicts a higher infection risk[95, 96]. It has been postulated therefore, that KpSC overgrowth in the GIT microbiome, in combination with loss of immune control in the setting of old age or immunosuppressive therapy, may allow GIT escape followed by seeding and infection of extra-intestinal sites[3].

In addition to this observed association between Kp abundance and infection risk, Lerner *et al.* demonstrated high CPKp load in the GIT is associated with an increased frequency of strain transmission in hospital[97]. This study identified the presence of a sub-group of high-abundance CPKp carriers, termed 'super-spreaders', consisting of 18% of the colonised patient population that were responsible for 80% of all spread within the local hospital environment[97]. Thus, interventions reducing the carriage load of KpSC within the GIT has potential to reduce a large burden of disease caused by these opportunistic pathogens within the hospital setting.

1.3.3 Prolonged KpSC GIT Colonisation as a Reservoir for AMR

Understanding duration of carriage within the GIT, as well as factors influencing this, is a major area within KpSC ecology in need of further research. Recently, a longitudinal Kp GIT carriage study by Lepuschitz *et al*., which repeat sampled six community-based adults each week for one year, observed carriage to be intermittent and of high turn-over, with the no strains detected longer than two consecutive weeks[98]. Furthermore, two study participants, who were colleagues that often shared meals, were found to carry the same Kp strain on multiple occasions, suggesting food as a potential source of exposure to new Kp strains[98]. Other studies focusing on carriage of ESBL- or CPKp following large hospital outbreaks or in returned travellers,

however, have found while most patients undergo spontaneous decolonisation within months, carriage in these settings can last up to one to four years[99-103]. Moreover, following a large outbreak of an ESBL-Kp ST17 strain in the neonatal intensive care unit in Stavanger University Hospital, Norway, Lohr *et al*., observed multiple transmission events occurred from persistently colonised infants to close household contacts[99]. The genome of the ESBL-Kp ST17 clones and the associated ESBL-encoding resistance plasmid carried by these infants also remained remarkably stable during the two-year follow-up period of this study[99, 104]. Similarly, in a study following prolonged GIT colonisation of a CPKp ST258 strain in two adult patients post-discharge found high stability of the *bla*$_{KPC}$-encoding resistance plasmid, despite multiple rearrangements of additional plasmids carried by this strain, as well as multiple HGT events to an additional ST37 Kp strain, as well as an ST127 *E. coli* strain[100]. These important studies demonstrate the potential role of prolonged KpSC GIT carriage as a reservoir for the dissemination of high-risk MDR-strains and HGT of AMR genes in the community, highlighted the need for further understanding of KpSC colonisation in this setting.

1.3.4 KpSC Colonisation Factors

Several factors carried by the KpSC have been shown to be important in promoting colonisation fitness in the GIT. Inhibiting these processes may therefore have potential as strategies for inducing decolonisation of the KpSC. Recently, Merciecca *et al*. demonstrated the presence of the type VI secretion system (T6SS), a molecular syringe capable of injecting toxic effector molecules into target cells, was important in long-term Kp colonisation of mice following streptomycin treatment[105]. Genome-wide analysis revealed the T6SS is widely carried by Kp and *K. variicola* (98% and 95% of strains examined), however, was only present in 20% of *K. quasipneumoniae* subsp. *similipneumoniae* strains analysed[105]. The T6SS may therefore represent a key colonisation advantage in Kp and *K. variicola* that has contributed to them becoming the dominant human pathogens within the KpSC. Similarly, the thick polysaccharide capsule produced by Kp, known as the K-antigen, in addition to being a known virulence factor, was shown by Favre-Bonte *et al*. to be important in allowing persistence and even distribution in the mucous-layer within the GIT of streptomycin treated mice[106]. The K-antigen has been identified as a potential target for Kp-based vaccine strategies, targeting of which may therefore have potential to reduce both the virulence and colonisation capacity of Kp[107].

In addition to these factors, the plasmid-borne *ter* operon, which is strongly associated with Kp infection in humans, was shown to have its action through promoting colonisation fitness in the

GIT rather than increasing virulence itself[108]. Interestingly, the *ter* operon is thought to enhance Kp colonisation fitness by increasing stress tolerance in the presence of short-chain fatty acids (SCFAs, discussed below)[108]. Acquisition of this trait may thus represent an evolutionary response by Kp to oppose GIT colonisation inhibition imposed by SCFA-producing bacteria (see below). Strategies inhibiting *ter* could therefore hold potential to enhance the natural resistance to Kp colonisation by commensals within the GIT.

Finally, two studies generating large transposon mutagenesis libraries have also identified multiple genes important for colonisation by Kp[109, 110]. In both studies, several genes involved in metabolism of nutrients were shown to be crucial in competitive colonisation as well as achieving high density with the GIT. Moreover, different Kp strains have been found to rely on differing metabolic strategies for colonisation success in the GIT[111]. Understanding and inhibiting the metabolic pathways essential to colonisation by high-risk Kp clones may thus represent a potential strategy to preventing colonisation by these lineages. Despite the large therapeutic potential of identifying and targeting colonisation fitness factors in the KpSC, however, this remains an area in need of further study.

A summary of known predisposing factors and consequences of KpSC GIT colonisation are presented in Figure 4

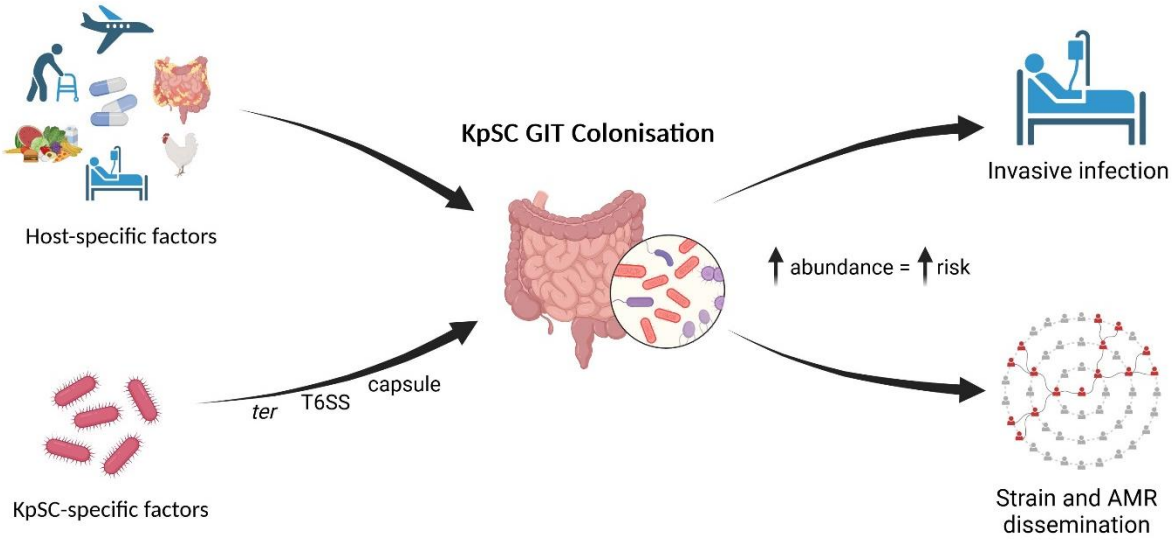


Figure 4. Summary of KpSC colonisation of the human GIT. Host-specific factors associated with increased KpSC colonisation includes older age, international travel, inflammatory bowel disease (IBD), hospitalisation, use of antimicrobials, proton pump inhibitors (PPIs), non-steroidal anti-inflammatory drugs (NSAIDs), contact with chickens, and dry fish consumption[85, 87, 89]. The capsule, type VI secretion system (T6SS), and *ter* operon in KpSC strains is associated with enhanced GIT colonisation[105, 106, 108]. GIT KpSC carriage is a risk factor for invasive extraintestinal infection, strain transmission, and HGT of AMR genes[85, 97, 99, 100]. A higher abundance of Kp in faecal samples has been associated with a higher risk of infection and strain transmission in hospitalised patients[95-97]. Created with BioRender.com

1.3.5 The KpSC and the GIT Microbiota

The human GIT microbiota, which is defined as the sum of all bacteria, archaea, and fungi in the GIT, is estimated to contain approximately 10^{13} bacteria[112, 113]. The total GIT microbiota plus all structural elements (proteins, lipids, polysaccharides, nucleic acids, viruses, bacteriophages), microbial metabolites, and physiochemical properties (pH, temperature, etc), then makes up the GIT microbiome[113]. Despite large interpersonal variation, a typical 'healthy' microbiota is dominated by the Bacteroidota (Bacteroidetes) and Bacillota (Firmicutes) phyla, which together make up approximately 80% of all bacterial species present[114, 115]. The remaining 20% is usually comprised of Actinomycetota (Actinobacteria), Verrucomicrobiota (Verrucomicrobia), and Pseudomonadota (Proteobacteria), which includes the Enterobacterales and the KpSC[114, 115]. Under normal circumstances, the KpSC is a very low abundance member of the GIT microbiota, typically making up less than 0.1%[86, 114]. Conditions which allow overgrowth of potential pathogens, with an associated loss of beneficial commensal species and an overall reduction in microbiota diversity is termed 'dysbiosis'[116].

Commensal bacteria within the GIT microbiota are thought to play a crucial role in limiting colonisation and expansion of potentially pathogenic species, such as the KpSC, through a process referred to a 'colonisation resistance' (Figure 5)[117]. Colonisation resistance can occur by direct mechanisms, such as through niche/nutrient competition, contact-inhibition, e.g. via the T6SS, and production of antimicrobial peptides and other inhibitory molecules, or through indirect mechanisms including stimulation of mucosal barrier production, oxygen limitation through production of SCFAs, and stimulation of the immune system[117]. Identifying commensal species able to inhibit KpSC colonisation, and the mechanisms they use, could potentially lead to microbiota-based decolonisation therapies, reducing the burden of disease and spread of AMR by this group of pathogens.

Interestingly, in contrast to other opportunistic pathogens such as *Clostrioides difficile*, previous studies in both mice and humans have found no relationship between microbiota alpha diversity, i.e., the overall diversity of species within the microbiota, and Kp colonisation[118-120]. This suggests the specific microbiota composition rather than overall diversity may be essential for effective colonisation resistance against the KpSC. This notion is supported by a recent study by Spragge *et al*, who demonstrated effective colonisation resistance against Kp *in vitro* and *in vivo* (gnotobiotic mice) only occurred in the presence of specific species within a diverse consortium of GIT commensals through a process of nutrient blocking[121].

To date, identification of most potential inhibitors of KpSC colonisation have been through *in vitro* studies and *in vivo* mouse models, many of which use antibiotic exposure to facilitate Kp colonisation, with a current lack of association studies performed in human subjects. Amongst these previous studies, members of the *Bifidobacterium* genus, part of the Actinomycetota (Actinobacteria) phylum, have been identified as potentially important inhibitors of Kp GIT colonisation[121, 122]. *Bifidobacterium* species were also able to reduce inflammation and improve mortality in murine models of Kp-induced pneumonia, suggesting this genus may have an important protective role against both colonisation and infection caused by the KpSC[123, 124]. Inverse relationships between Kp colonisation density and presence of various members of the Bacillota (Firmicutes) phylum, including the *Lactobacillus*, *Lachnoclostridium*, and *Roseburia* genera have also been observed, suggesting an inhibitory relationship with members of this phyla[119, 125]. Somewhat conflicting these results, however, Sequira *et al*. showed following antimicrobial-induced microbiota depletion, only a consortium consisting of Bacteroidetes (Bacteroidota), rather than Bacillota (Firmicutes), Actinomycetota (Actinobacteria), or Pseudomonadota (Proteobacteria), were able to induce clearance of Kp colonisation from the

GIT through a IL-36-induced macrophage dependent mechanism requiring commensal colonisation factor (CCF)[126]. These differing findings are possibly due to the highly complex nature of microbiota interactions and possible redundancy of taxa able to inhibit the KpSC which may vary under different experimental conditions.

Competition between members of the Enterobacterales order may also play an important role in limiting KpSC GIT colonisation. In particular, *E. coli* was identified as a keystone species within the Kp-inhibiting consortium identified by Spragge *et al*, likely resulting from the large metabolic overlap between these two species[121]. *E. coli* was also identified as a strong competitor of Kp colonisation in the developing neonatal microbiome, further suggesting an important inhibitory role for commensal strains of this species[90]. The non-KpSC species, *Klebsiella oxytoca*, has also been shown to induce decolonisation of Kp in mice with humanised microbiota, driven by competition for specific carbohydrates[127]. Interestingly, long-term clearance in this setting was only achieved in the presence of three additional species of the Bacillota (Firmicutes) phyla: *Blautia coccoides*, *Enterococcus faecalis*, and *Enterocloster clostridioformis*[127]. This finding gives further support to the notion that effective colonisation resistance against the KpSC likely results from the concerted action of multiple commensal species, with no one species sufficient in isolation. Given its importance and potential for therapeutic interventions, more studies in humans combining cutting-edge molecular and omics approaches, including WGS, WMS, metabolomics, and proteomics, are required to further understand the complex network of interactions between the KpSC and commensal microbiota and conditions required for effective colonisation resistance.

1.3.6 The Role of Short Chain Fatty Acids in Colonisation Resistance Against the KpSC

In addition to competition for nutrients and the induction of specific immune responses mentioned above, the potential role of SCFAs in colonisation resistance against the KpSC warrants specific mention. SCFAs are a group of carboxylic acids including acetate, butyrate, and propionate, produced by the fermentation of indigestible dietary polysaccharides by certain commensal bacteria, which have been shown to have multiple important roles in human health[128]. Major producers of SCFAs include members of the Bacillota (Firmicutes) phyla, in particular, *Faecalibacterium prausnitzii*, *Roseburia*, and *Eubacterium* species, as well as *Bifidobacterium* and members of the Bacteroidota (Bacteroidetes) phylum[129]. Among the proposed functions of SCFAs in the GIT is the ability to suppress the growth of multiple potential pathogens, including Kp, through several different mechanisms. Principally, SCFAs

can freely diffuse across cell membranes where they directly inhibit growth of Kp and other Enterobacterales by interfering with transmembrane potential and intracellular pH[130]. Butyrate is also aerobically metabolised as a major energy source by intestinal epithelial cells (IECs), leading to a reduction in oxygen concentration at the epithelial surface, creating an anaerobic environment that deprives facultative anaerobes like the Enterobacterales of this growth advantage[131]. Furthermore, SCFAs can enhance the integrity of tight junctions between IECs, preventing the translocation of pathogens into the extra-intestinal tissues, and are known to have anti-inflammatory effects by direct interaction with the immune system[132, 133]. Moreover, oral supplementation with SCFAs has been shown to improve survival rates of mice with Kp-induced pneumonia, suggesting the protective effects of SCFAs extend beyond the GIT[123]. Further studies using metabolomics-based approaches in combination with experimental models, are warranted to improve our understanding of the role of SCFAs in colonisation resistance against the KpSC in the human GIT and evaluate their use as possible alternate therapeutic strategies.

A summary of experimentally determined mechanisms of colonisation resistance against the KpSC is presented in Figure 5.

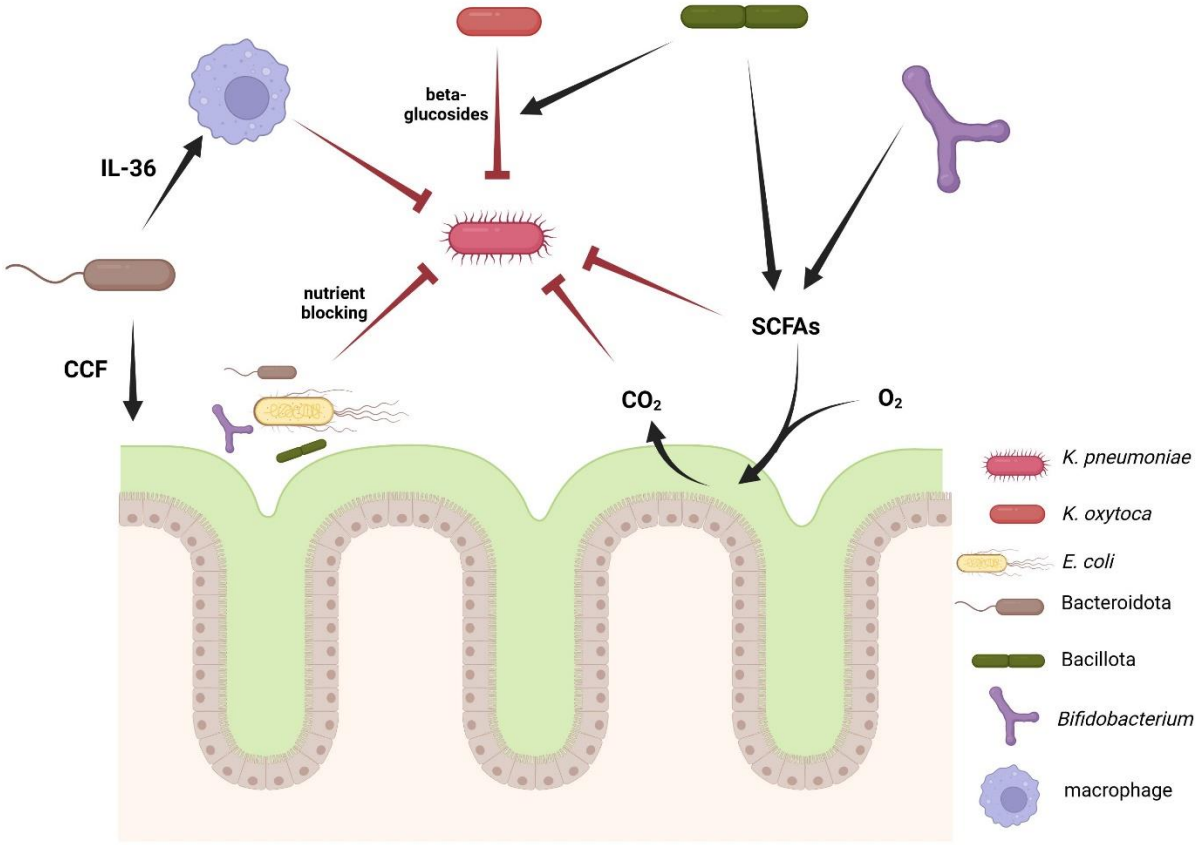


Figure 5. Summary of experimentally determined mechanisms of colonisation resistance against Kp within the GIT lumen. Briefly, from left to right: members of the Bacteroidota (Bacteroidetes) phyla inhibit Kp by a macrophage-dependent mechanism requiring IL-36 and CCF[126]. Specific microbiota consortia containing *E. coli* as a key species inhibit Kp colonisation by nutrient blocking[121]. *K. oxytoca*, in cooperation with several Bacillota (Firmicutes) inhibit Kp by outcompeting for beta-glucosides via CasA[127]. Short chain fatty acids (SCFAs), produced by several commensal microbiota species, inhibit Kp directly by interfering with transmembrane potential and intracellular pH, as well as indirectly by lowering oxygen concentration in the GIT through aerobic metabolism by colonic epithelial cells[130, 131]. Created with BioRender.com

1.3.7 GIT Decolonisation Strategies for MDR-Kp and other Enterobacterales

Due to the risks of infection and AMR dissemination associated with GIT colonisation of MDR-Kp and other MDR-carrying Enterobacterales, strategies to eradicate these pathogens from the GIT through inducing decolonisation have gained considerable interest[134-136]. Such strategies are aimed at eradicating potential pathogens through modification of the underlying microbiota and have potential to reduce reliance on antimicrobial use and prevent infections before they

occur. One such approach is selective decontamination of the digestive tract (SDD), which involves the oral administration of poorly absorbed antimicrobials, such as colistin, neomycin, streptomycin, and gentamicin[136]. SDD has been shown to be effective in preventing infections in critically ill ICU patients and may support a reduction in the burden of MDR-Enterobacterales colonisation, however, so far this effect appears to be only temporary[137-139].

Faecal microbiota transplant (FMT) is another promising decolonisation strategy that involves infusion of stool from a healthy donor into the GIT of a patient with dysbiosis[140]. FMT is already in routine use for the treatment of recurrent *Clostridioides difficile* infection and a recent systematic review by Macareño-Castro *et al*. found a decolonisation rate of 78.7% in patients colonised with carbapenem-resistant Enterobacterales after one year[140, 141]. It should be noted, however, of the ten studies reviewed by Macareño-Castro *et al*, three were retrospective and only one was a randomised clinical trial, indicating the need for more comprehensive clinical evaluation of FMT in this setting[141].

Use of probiotics, typically involving oral administration of live *Lactobacillus*, *Saccharomyces*, or *Bifidobacterium*, or a combination of these, has also shown promise at reducing the burden of MDR-Enterobacterales colonisation. Probiotic treatment has been shown to be effective in reducing MDR-Enterobacterales colonisation in infants and children as well as protecting against necrotizing enterocolitis in premature infants[142, 143]. In adults, although probiotics have been associated with a reduction in the abundance of MDR-Enterobacterales colonisation, clinical trials have not yet shown their efficacy in inducing complete eradication from the GIT[142, 144].

Similarly, bacteriophages, which are viruses that can specifically target and lyse bacteria at the species or even strain-level, are also gaining interest as potentially highly specific inducers of decolonisation[145]. Decolonisation by bacteriophages targeting specific MDR-Enterobacterales, including Kp, have shown promise in *in vivo* models as well as a limited number of case studies and clinical trials[145, 146]. While this form of treatment appears safe and well tolerated, bacterial resistance against bacteriophage therapy can develop quickly and is a potential limitation to treatment[145, 146].

While all these approaches have shown some efficacy in inducing decolonisation of MDR-Enterobacterales, including Kp, in case studies and a small number of clinical trials, results are variable, and there is a lack large high-quality randomised placebo controlled trials to support

their routine use in this setting[134, 136, 147]. These treatments are also not without risks. For example, selection of a colistin resistant CPKp has been reported following SDD[148]. Similarly, cases of *Lactobacillus* bacteraemia and endocarditis following probiotic administration in patients with underlying risk factors such as immunosuppression and inflammatory bowel disease has been reported[136]. Thus, further studies into the efficacy, safety, and better understanding of underlying mechanisms of these decolonisation strategies are required.

1.4 *K. pneumoniae* in a OneHealth context: A Proposed Role as a Major Trafficker of AMR Genes from Environment to Clinic

'OneHealth' is a holistic approach to improving health outcomes that recognises the interconnected nature of humans, animals, and the environment[149]. Examining KpSC within the OneHealth context, by investigating the movement of KpSC strains and associated AMR genes between different environmental, animal, and human niches, is critical to understanding major routes of transmission into the clinical and human niche, as well as the spread of AMR by this group of pathogens. It is well documented that Kp has a very wide ecological range, found in the terrestrial, freshwater, and marine environments, and has been isolated from mammals, birds, reptiles, insects, plants, soil, shellfish, and fish[2, 150-153]. This range is facilitated by the high genomic plasticity and large pan-genome of Kp, giving access to a wide range of genetic capabilities, including a large potential metabolic capacity, allowing survival and adaption in a diverse range of environmental niches[2, 15, 154]. In addition, Kp is known to carry a diverse plasmidome and is particularly adept at acquiring and maintaining high numbers of plasmids in comparison to other important AMR-associated pathogens, regularly carrying between two and five different plasmids, with up to ten reported[2]. Moreover, Kp is known to regularly engage in HGT and share ARGs with other members of the Enterobacterales order, as well as more distantly related taxa, including *Streptomyces*, *Acinetobacter*, *Bacteroides*, *Bacillus*, and *Pseudomonas*[4, 15].

It is this combination of wide ecological range, genetic flexibility, and ability to acquire, maintain, and transmit genetic material that has led to the proposed role of Kp as a major trafficker of important AMR genes from environmental bacteria to other clinically relevant pathogens[2] (Figure 6). Evidence for this comes from the observation that many currently clinically relevant mobile AMR genes, including ESBL variants of TEM, the carbapenemases KPC, NDM-1, and OXA-48, and the quinolone resistance genes *qnrA* and *qnrB*, were first detected in Kp before appearing in other Enterobacterales, or even *Acinetobacter* and *Pseudomona*[155-162]. Kp, along with *E. coli*, was also one of the earliest identified carriers of the

widespread ESBL gene CTX-M and the mobile colistin resistance gene *mcr-1*[77, 163]. Some of these AMR genes also have direct links to being mobilised from environmental bacteria, such as *qnrA* and OXA-48 from the marine species *Shewanella*, and CTX-M from *Kluyvera*[163-165]. The overlap in niches between Kp and these species further supports the proposed role of Kp as an important trafficker of AMR genes into humans and the clinical environment.

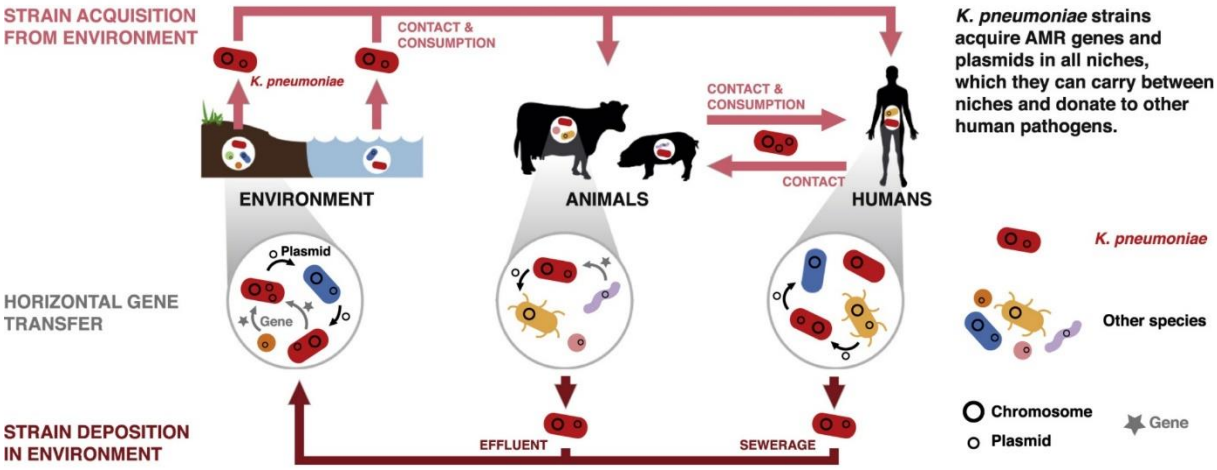


Figure 6. *Klebsiella pneumoniae* as a major trafficker of AMR. Proposed model of how Kp traffics AMR genes and plasmids from environmental sources, such as soil and waterways, and transmits these into the human and animal niches. Kp strains carrying AMR determinants then return to the environment via sewerage and effluent where it can form reservoirs for ongoing transmission. Reprinted with permission from Curr Opin Microbiol[2].

To identify the environmental niches which may be acting as potential sites for AMR acquisition and reservoirs for MDR-Kp, several OneHealth studies have been conducted sampling the KpSC from human, animal, and environmental niches in various geographical locations, including Italy, England, Brazil, French West Indies, Malawi, and Ghana[166-171]. A large OneHealth study is also currently underway in Norway, known as KLEB-GAP, investigating KpSC cross-talk in the human, animal, and marine environment, which also incorporates this project (https://www.nor-kleb.net/). Interestingly, these studies have shown ESBL-Kp is variably detected in the environment and animals, particularly livestock. As was found by Cocker *et al*., in a LMIC setting such as Malawi, where there can be increased direct contact with animals, higher use of antimicrobials in animal husbandry practises, as well as

poorer sanitation infrastructure, these can act as significant reservoirs for ESBL-Kp transmission to humans[171]. CPKp, however, appears largely restricted to the hospital environment, and overall limited movement of Kp strains from environmental or animal sources into humans or the clinical setting has been identified. The major reservoirs and sources of human acquisition of ESBL- and CPKp identified in these studies were in fact from other humans, i.e. person-to-person spread, or from the hospital environment itself, with hospital sinks identified as major potential reservoirs. Thus, although capture of AMR genes from environmental bacteria by Kp and spread of MDR-Kp from environmental and animal niches into both humans and hospitals are critical events with major clinical outcomes, these appear to occur only rarely and are highly context dependent. These observations support that surveillance and targeted interventions to prevent further spread of AMR by Kp may be best focused on the human niche and clinical setting.

Notably, to date there have been few large Kp OneHealth-based studies preformed in either India or China, both of which have a heavy burden of CPKp infections, increased direct human-animal contact, as well as pervasive use of antimicrobials in agriculture[58, 172]. These factors may create an environment favouring environmental acquisition of AMR genes by Kp and increased opportunities for movement of MDR-Kp strains from environmental reservoirs into the human niche. An important example of this which requires vigilant surveillance is spread of the plasmid-mediated colistin resistance *mcr-1* gene. This gene was most likely mobilised from bacteria in livestock in China resulting from the previous widespread use of colistin in agriculture[77]. Although a decline in the prevalence of *mcr-1* carrying *E. coli* in livestock, human carriage, and clinical infections has been reported since the ban of colistin in Chinese agriculture in 2018[173], large environmental reservoirs of this gene may still exist. Supporting this, a recent study from 2022 found highly similar plasmids carrying the *mcr-1* gene (>98.5% ANI) in Kp clinical isolates were also present in other Enterobacterales species in silver gulls, poultry, and wastewater in Australia, China, and Japan, respectively[174]. Thus, further OneHealth-based studies in these settings are warranted.

2 Methods for Detection of Pathogens and Associated AMR

Bacterial pathogen and AMR detection in clinical microbiology encompasses surveillance programs as well as clinical diagnostics. Surveillance is aimed at the detection and monitoring of high-risk pathogens, such as Kp, and associated AMR and virulence determinants, and are essential for outbreak tracing, infection control, and understanding the local and global dissemination[175]. Clinical diagnostics involves identification of causative species and associated antimicrobial susceptibilities +/- virulence determinants from clinical samples[176]. Both surveillance and clinical diagnostics require up-to-date, robust, and validated detection methods that are highly sensitive and specific yet timely and cost-effective.

2.1 Culture-based Methods

Currently, culture-based and phenotypic methods are predominantly used in pathogen and AMR detection both in surveillance and clinical diagnostics[177, 178]. Culture-based methods have several advantages, including being relatively simple, inexpensive, and allowing recovery of viable isolates of further testing and analysis, including antimicrobial susceptibility testing (AST) and strain typing by whole genome sequencing (WGS)[177, 179-181]. Use of selective media, through addition of elements that are essential for growth of desired species and inhibition of undesired species, can facilitate accurate identification of suspect pathogens[182]. An advance on this is the development of chromogenic media, which targets organisms through specific metabolism of substrate which releases a chromogen, allowing faster identification of pathogens and AMR-phenotypes, and are particularly useful for surveillance and screening for MDR-pathogens[178, 183]. The recent widespread use of matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry has further revolutionized species identification in this setting[176]. As has been exemplified by the KpSC, however, MALDI-TOF has limitations in discriminating between closely related bacteria at the species level[10, 49].

Coupled to culture-detection is phenotypic-based AST, which principally employs disc diffusion assays, gradient tests, and broth macro- and microdilution assays to give a phenotypic measure of the susceptibility of bacteria to different antimicrobials[178]. Broth dilution assays also allow quantification of the minimum inhibitory concentration (MIC) of a given antimicrobial, defined as the lowest concentration required to inhibit bacterial growth[178]. Further testing can then give insights into the underlying resistance mechanisms, for example, combination testing, colorimetric tests, detection of hydrolysis with MALDI-TOF, and lateral

flow assays can be used to confirm the presence of carbapenemases[184-187]. Thanks to thorough validation by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) and the Clinical and Laboratory Standards Institute (CLSI), clinical breakpoints have been established that predict effectiveness of specific antimicrobials in treating infections caused by particular pathogens[188, 189]. The standardisation of these methods means AST results can be directly compared between different microbiology laboratories.

While the information provided through these culture-based methods is critical for pathogen and AMR detection in both a surveillance and clinical setting, they have several shortcomings. Principally among these are time to diagnosis in culture-based detection, typically taking at least two days for species identification and AST results[190]. For fastidious organisms these times can extend to days or weeks, while many bacteria are also unable to be cultured by current methods[191, 192]. Delays in pathogen identification and AST can be critical in the clinical setting, leading to inappropriate and prolonged use of broad-spectrum antimicrobials or even ineffective therapy[193]. Since every hour before commencement of appropriate antimicrobial therapy is associated with an increased mortality in the setting of sepsis, such delays can have serious impacts on patient outcomes[194]. Automated systems that use optical changes for pathogen growth and AST, such as Vitek2 and Phoenix, have been shown to reduce turn-around times, capable of producing clinically usable results in as little as 4-18 hours[178]. Furthermore, the detection sensitivity and ability to capture the full microbial diversity, particularly in complex bacterial samples, is also an issue with culture-based detection[195-197]. Indeed, it has been shown culture-based detection of CPKp GIT carriage lacks sensitivity compared to molecular-based methods[198].

2.2 Molecular-based Methods

Molecular-based, culture-independent methods have large potential to enhance pathogen and AMR detection, both in surveillance and the clinical setting, by complimenting many of the shortcomings of conventional culture-based detection methods[181, 197, 199, 200]. Typically, molecular-based methods are fast, highly sensitive and specific, and can give accurate gene-level information[199, 200]. Major drawbacks to these methods include higher costs, as well as generally targeting only known pathogens and AMR genes means novel pathogens and resistance mechanisms can be difficult to detect[178]. AMR in pathogens may also be overestimated since the presence of an AMR gene does not always translate to phenotypic resistance[178]. Three major categories of molecular detection methods are currently in use today:

nucleic acid amplification technologies (NAAT), hybridisation-based microarray, and NGS. The following sections will give and overview of each of these categories, including a more detailed discussion of qPCR and NGS, particularly WMS, due to the relevance to this project. A more detailed description of the methods used in this study will then follow in the Choice of Methods section.

2.2.1 Nucleic Acid Amplification Technology

The most common NAAT in use currently are polymerase chain reaction (PCR)-based methods, which amplify target DNA sequences via specific DNA primers and DNA polymerase through thermocycling for successive rounds of amplification and denaturation[201] (Figure 7). PCR-based methods are widely used for pathogen and AMR detection, capable of giving results with high sensitivity in as little as one hour[178, 201, 202]. Major PCR methods include qPCR, digital PCR (dPCR), and high-resolution melting (HRM), each of which incorporate fluorescent dyes into the PCR reaction to allow high-sensitivity detection +/- quantification in real time[199, 200]. Isothermal amplification is an additional emerging NAAT technology, which includes loop-mediated isothermal amplification (LAMP), nucleic acid sequence-based amplification (NASBA), and transcription-mediated amplification (TMA). These methods allow nucleic acid amplification at a constant temperature, negating the need for expensive thermocyclers, and have become standard for diagnosis of gonorrhoea and chlamydial infection[199, 200]. One major disadvantage of NAAT are the limited number of pathogen or AMR targets that can be amplified in a single assay. Development of PCR-based panel arrays such as the BioFire FilmArray Panel improves this range by simultaneously assaying for up to 33 pathogens and 10 associated AMR genes in selected panels of clinically important pathogens causing infections of the respiratory tract, bloodstream, and meningitis-encephalitis[203-205].

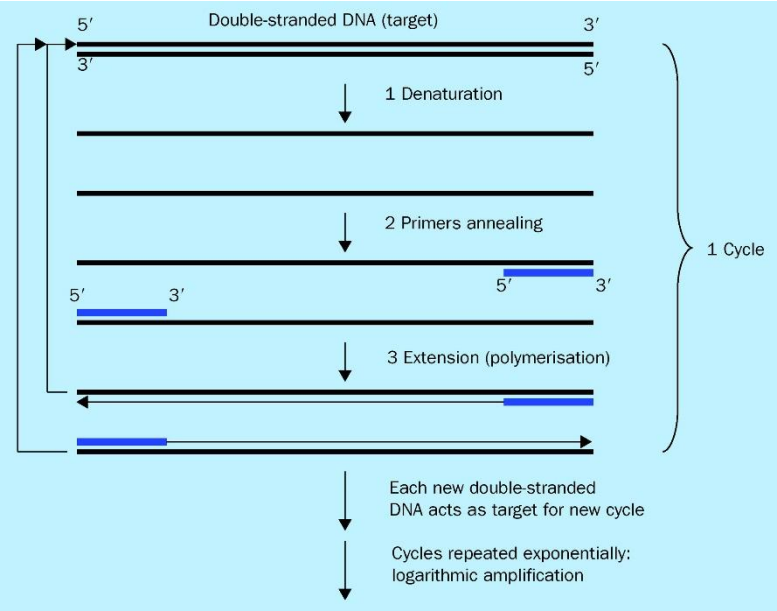


Figure 7. Schematic representation of PCR. Double-stranded DNA is first denatured by heating to 95ºC (1). This is followed by primer annealing to specific complementary target sequences on the forward and reverse single stranded DNA molecules (2). DNA polymerase incorporates nucleotides to create a new copy of the targeted DNA (3). Newly synthesised double-stranded DNA then forms the template for the next cycle, resulting in exponential amplification of target DNA. Reprinted with permission from The Lancet[201].

The most widespread and validated NAAT in use in clinical microbiology is qPCR (also referred to as real-time PCR or RT-PCR), which has been shown to be highly rapid, accurate, and sensitive in detecting a variety of pathogens and ARGs[202, 206]. qPCR assays detect specific DNA sequences in samples, while a modification of qPCR, which first involves a reverse transcription step (known as RT-qPCR) is used to detect RNA[201]. Studies have shown qPCR has equal or greater sensitivity in detecting a number of important pathogens, such as Group A Strep, *Haemophilus influenzae*, *Legionella pneumoniae*, and *Neisseria meningiditis*, and ARGs including ESBLs, than conventional methods with faster turnaround times[206]. Two major qPCR methods currently exist: (i) fluorescent dyes that intercalate non-specifically to the newly synthesised double stranded DNA molecules and (ii) sequence-specific fluorescently labelled probes which emit fluorescence once the probe is incorporated into the synthesised double-stranded DNA[201]. For both methods, a positive sample is reported once the fluorescence crosses above a threshold distinguishable from background[207]. The cycle at which this occurs, known as the quantification cycle (C_q) is directly proportional to the number of copies of target

sequence in the sample, and can be quantified by comparison to standard curve of known copy number[207]. Intercalating dye-based qPCR has the additional quality control step of a melting cycle, which measures the temperature at which the amplified double-stranded DNA denatures (melts)[208]. Although qPCR typically amplifies a single target sequence, multiplexing can be used to simultaneously amplify multiple targets[201].

Limitations of qPCR in clinical microbiology include false positives due to contamination, primer dimer formation, and incorrect baseline settings, as well as false negatives resulting from presence of inhibitors, incorrect primer sequences or annealing temperatures[199]. The potential for false positives through DNA contamination is a major potential drawback of qPCR and is directly related to its high sensitivity. Such false positives can easily occur since qPCR is capable of amplifying as little as a single copy of DNA target sequence[207, 209]. Indeed, DNA contamination was a large problem encountered in the early stages of this project. In a microbiology diagnostic setting, this could lead to important clinical implications such as incorrect diagnoses and implementation of inappropriate therapies. Major strategies to prevent contamination are use of UV-radiation and sodium hypochlorite to destroy any contaminating DNA in the environment, as well as treatment of reagents with DNA destroying enzymes such as DNase I which can then be heat inactivated prior to qPCR assay, although none of these are 100% effective[209]. This risk of false positives in qPCR highlights the need for robust negative controls in all assay set-ups.

2.2.2 Hybridisation-based Microarray Technologies

Hybridisation-based microarrays detect gene sequences in samples through hybridisation to probes fixed on the surface of the microarray[178, 199, 200]. Like NAAT, hybridisation-based microarrays can detect specific pathogens and AMR determinants rapidly and with high sensitivity[178, 199, 200]. Due the proximity of fixed probes, hybridisation-based microarrays have the added advantage of simultaneous detection of a much larger spectrum of pathogens and AMR determinants than NAAT methods, including Kp strains and associated AMR determinants[210]. These methods, which include both solid-phase and liquid-phase arrays, are useful in rapid surveillance and screening for a large number of targets in complex samples, such respiratory and GIT samples[199, 211, 212]. The high cost and complexity of these methods, however, can be a limited factor in the widespread utilisation[200].

2.2.3 Next-Generation Sequencing

NGS refers to high throughput, massively parallel sequencing of millions of DNA fragments[213]. These technologies are rapidly advancing, becoming more readily available with reducing costs, making them increasingly attractive for use in pathogen and AMR surveillance and clinical microbiology laboratories[214]. NGS is generally divided into short-read and long-read technologies, also referred to as second and third generation sequencing, respectively, both with differing advantages and disadvantages[213].

Short-read NGS techniques sequence DNA fragments typically several hundred base pairs in length[213]. The major advantage of short-read NGS is higher base-call accuracy than that of long-read sequencing, allowing for more accurate detection of single nucleotide polymorphisms (SNP) and small insertions and deletions (indels), which can be critical in identifying sources and transmission chains in clonal outbreaks[215-217]. The major short-read sequencing platform is Illumina, which uses adapter-ligated DNA fragments (known as DNA libraries) bound to the surface of a flow cell, followed by solid-phase bridge amplification creating clusters of clonal populations of forward and reverse DNA fragments (referred to as paired-end reads)[213]. Fluorescently labelled nucleotides are then added in successive sequencing cycles, known as 'Sequencing by Synthesis', and base calls are made by signal intensity measurements during each cycle[213].

Another emerging short-read platform is MGIseq which has been shown to be highly comparable to Illumina in sequencing accuracy with considerably lower sequencing costs[218-220]. MGIseq uses a similar approach of Sequencing by Synthesis, however, instead of sequencing DNA fragments bound to a solid flow cell, DNA fragments are circularised and amplified into DNA nanoballs (DNB)[221]. Each DNB consists of between 300-500 copies of the original DNA fragment amplified by rolling circle amplification (RCA)[222]. By using the original DNA circle as template for each successive amplification cycle, this approach reportedly reduces the exponential accumulation of errors seen in other amplification methods[222]. RCA also does not require precise titration of template concentrations preventing stochastic inefficiencies seen other sequencing platforms[221, 222].

In contrast to short-read sequencing, long-read sequencing can sequence DNA fragments up to thousands or tens of thousands of base-pairs[223]. The major advantage of this is the ability to sequence long repetitive DNA sequences that are not possible with short-read technologies, allowing sequencing and *de novo* assembly of entire bacterial genomes and plasmids[224]. The

two major long-read sequencing platforms are PacBio and Oxford Nanopore Technology (ONT), both of which use single molecule sequencing, albeit utilising different approaches. PacBio uses Single Molecule Real-Time (SMRT) technology to form large, closed hairpins from each DNA fragment that are individually immobilised in cells and sequenced[213]. Conversely, ONT feeds individual DNA fragments through thousands of nanopores embedded in a membrane and measures changes in ion current as different nucleic acids pass through the pore[213]. Two major advantages of ONT are the small size and portability of the sequencing devices, and the real-time generation of sequence data, opening the possibility for rapid point-of-care clinical diagnostics and outbreak surveillance the can be performed either at the bedside or in the field[225, 226]. Although the base-call accuracy of ONT is lower than other sequencing platforms, this is continually improving with advancements in pore technology, base-calling algorithms, and sequence corrections, making ONT a very promising tool in the future of pathogen surveillance and clinical diagnostics[213, 227, 228].

2.2.4 Application of NGS in Clinical Microbiology I: Whole Genome Sequencing

A major application of NGS in clinical microbiology is in WGS of bacterial pathogens for strain typing, phylogenetic analysis, and detection of associated AMR and virulence determinants[229]. WGS involves sequencing the genome of single cultured bacterial isolates, followed by bioinformatic processing to filter-out low-quality reads and remove adapters[213, 214]. Analysis can then be performed on the unassembled sequenced reads or joining reads together to form larger continuous DNA fragments known as ‘contigs’, and generally involves a combination of mapping reads to a reference genome or database of reference genes, and/or creating *de novo* genome assemblies from contigs[214, 229]. Combining both short-read and long-read technologies on a single isolate enables entire high-quality genomes and associated plasmids to be assembled, referred to as hybrid assemblies[214]. WGS thereby provides highly detailed strain-level identification and SNP profiling, as well as analysis of the entire genetic content of bacteria including AMR and virulence determinants.

The recent SARS-CoV-2 pandemic demonstrated the power of WGS in enabling rapid surveillance of viral spread and monitoring for emergence of new variants on a global scale[230, 231]. In the setting of bacterial pathogen and AMR detection, WGS also has a major role in global epidemiological surveillance, allowing a precise picture and phylogenetic analysis of the spread and prevalence of high-risk MDR-bacterial lineages and associated AMR determinants at a national and international level[232]. Indeed, two large multi-country WGS-based epidemiologic studies investigating the prevalence and dissemination of high-risk Kp lineages discussed in the

previous section has recently demonstrated the power of WGS in this setting. The first of these was a recent study investigating CPKp dissemination in Europe by David *et al*., which identified key clonal lineages responsible for the majority of CPKp infections, predominantly driven by inter-hospital spread within, rather than between, countries[66]. Similarly, Wyres *et al*. demonstrated an alarmingly high prevalence hypervirulence and MDR convergence in diverse Kp lineages causing bloodstream infections in South Asia, including several examples of co-carriage of hypervirulent and MDR determinants on the same plasmid[80]. Both studies have provided key insights into the global prevalence and trends in dissemination of high-risk Kp lineages that can directly inform public health measures.

At the local level, WGS has also shown great promise in hospital outbreak tracing and infection control through high-resolution SNP tracking in strains. In this setting it has been employed to delineate sources, reservoirs, transmission chains, within and between patient plasmid and clonal transfer events, and identify high-risk wards to directly inform infection control measures[217, 233]. Indeed, WGS has been used to gain important insights and help control several CPKp hospital outbreaks in a variety of geographical settings[234-236]. Thus, with the continually improving availability and affordability of WGS, this technology is likely going to be increasingly integrated into routine clinical microbiology and public health laboratories.

2.2.5 Application of NGS in Clinical Microbiology II: Whole Metagenomic Sequencing

WMS is a rapidly developing NGS technology which involves the sequencing of the entire genomic content of complex microbial communities that is also showing promise at improving detection of bacterial pathogens and associated AMR determinants in both the surveillance and clinical diagnostic setting[237, 238]. The typical workflow of a WMS analysis parallels that of WGS and is outlined in Figure 8. Typically an WMS analysis pipeline involves processing of reads, mapping of unassembled reads or assembled contigs for detection of genes of interest such as AMR and virulence determinants, as well as creating *de novo* genome assemblies, known as metagenome-assembled genomes (MAGs), through a process of clustering reads belonging to a single genome (known as 'binning')[239]. Taxonomy profilers are also used to detect the presence and abundance of all species present in samples, with newer tools allowing accurate strain-level detection to relative abundances as low as 0.1%[239, 240].

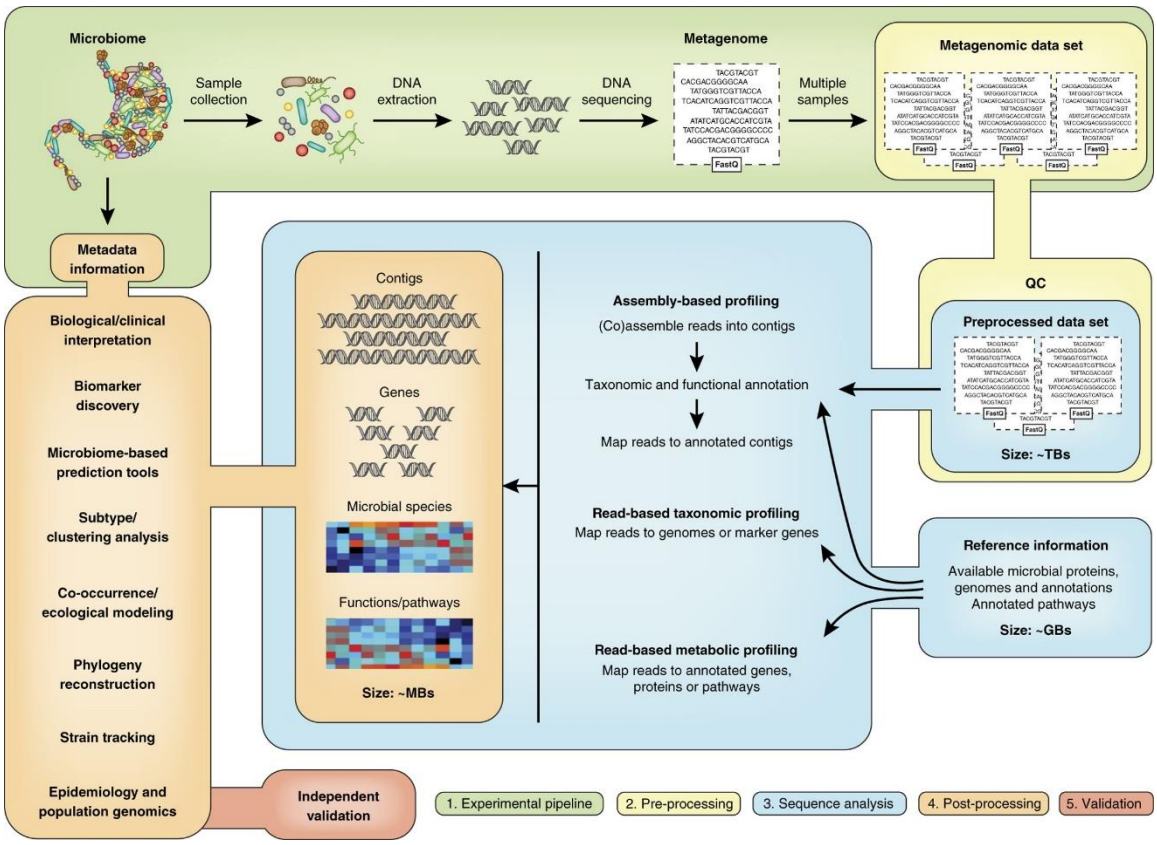


Figure 8. Workflow summary of WMS. (1) sample collection, DNA extraction and sequencing. (2) quality control, adapter removal, and read trimming. (3) read-based and assembly-based sequence analysis, including mapping reads to reference genomes or reference genes, and taxonomy profiling. (4) Data analysis/interpretation. Reprinted with permission, Nat. Biotechnol[239].

A key advantage of WMS over traditional culture-based and other targeted molecular-based methods is its 'shotgun'-based culture-independent approach. This allows unbiased detection of bacterial pathogens, both culturable and unculturable, as well as associated AMR and virulence determinants, holding promise as a universal pathogen detection test[241, 242]. This has potential for diagnosis of causative infectious agents that have otherwise eluded detection by conventional culture or other molecular methods. Several real-world examples of the use of WMS in clinical diagnosis have already been demonstrated, including in the setting of meningoencephalitis, neuroleptospirosis, sepsis, and respiratory infections[243-246]. Furthermore, WMS has potential to greatly enhance diagnosis of polymicrobial infections[247]. The combination of WMS and the portability and real-time availability of sequencing results produced by ONT discussed above further supports the promise of rapid direct-from-sample diagnosis with clinically actionable results performed at the patient bedside. This possibility

was evidenced by a recent proof-of-concept study in neonates with necrotising enterocolitis by Leggett *et al.*, which demonstrated diagnosis of causative pathogens, including Kp, and associated resistance profiles, could be achieved in less than five hours from sample collection by WMS, compared to 36 – 48 hours by conventional culture-based detection[248].

Similar to applications in clinical diagnostics, the untargeted, culture independent nature of WMS has potential to facilitate pathogen and AMR surveillance and inform public health and infection control measures. This has been demonstrated by previous WMS studies which have identified reservoirs for potential pathogens and AMR determinants in wastewater, coastal areas, livestock, and vegetables, with potential for movement into the human niche[249-252]. WMS surveillance of wastewater in particular has potential as an efficient method to monitor the total burden of AMR (known as the resistome) in different geographical regions and identify contributing factors. Illustrating this, a recent global sewage surveillance project identified AMR abundance, which included detection of the high-risk ARGs NDM and *mcr*, correlated with socioeconomic factors including a lower ranking in the human development index[253]. Another WMS-based study of wastewater treatment also found conventional treatments may not efficiently remove AMR-carrying bacteria and associated ARGs, and may even enrich for these, suggesting wastewater may be an important source for the dissemination of AMR[254]. Monitoring the microbiome and resistome within hospitals by metagenomics can also be used to identify and monitor reservoirs for ARGs and has demonstrated significant potential transfer of pathogens and AMR elements, including Kp, between the hospital environments, particularly sinks and other surfaces, patients and healthcare workers[255-257]. Additionally, WMS has potential to improve surveillance of high-risk non-MDR pathogens, such as HvKp. Such pathogens can be difficult to detect by conventional culture-based surveillance methods since they lack the AMR-markers that allows convenient selective enrichment of their MDR counterparts. Thus, WMS has large potential to support local and international surveillance efforts and enhance infection control.

In addition to detection and surveillance, WMS can give insights into the ecology of pathogens and the microbiota as well as their influence on human health in ways are not possible by other methods. For example, differences in GIT microbiota composition have already been linked to multiple human diseases, including diabetes, cardiovascular disease, depression/anxiety, and bowel cancer, suggesting new avenues for treatment of many chronic diseases could exist through modulation of the microbiota[258]. Detecting potentially pathogenic bacteria in the context of the wider microbiota by WMS also has potential to allow better understanding of

pathogen ecology and identify key inhibitors and promotors of colonisation through investigating associations with other taxa[259]. Furthermore, a key advantage of WMS over single-isolate WGS is the ability to detect multiple strains of the same species within a single sample[90]. This ability to analyse within-sample strain diversity has potential to allow better understanding of within-species competition, cooperation, and interactions including HGT of AMR determinants.

Despite the potential advantages of WMS, there are several limitations that exist preventing widespread application of this technology, including increased costs relative to other methods, lack of standardisation, and limitations in detection sensitivity, which makes positive pathogen identification challenging[238]. Detection sensitivity in WMS is directly related to sequencing depth, pathogen abundance, and presence of related species leading to misassignment of sequence reads[260-262]. As many pathogens are found at low abundance in microbiomes, particularly if strain-level detection is required, reliable detection by WMS in these settings can often be challenging. The development of targeted metagenomics, however, through methods such as probe-hybridisation capture, has greatly improved detection sensitivity of both low abundance species and ARGs in complex metagenomes[263, 264]. Although the obvious trade-off to this approach is the loss of the unbiased and untargeted nature of WMS. Furthermore, due to the cell lysis and DNA fragmentation that occurs during metagenomic sample processing, placing ARGs in their genetic context and linking plasmids and other MGEs to their bacterial hosts is a major challenge in WMS[265]. Since knowledge of the bacterial hosts and chromosomal- or plasmid/MGE-association of ARGs are essential in assessing their clinical relevance and dissemination potential, this represents a major limitation of WMS[266]. Recent developments in chromosome conformation capture methods, known as Hi-C, which involves forming a covalent linkage between DNA fragments in close proximity (e.g. chromosome and plasmid) prior to bacterial cell lysis, shows promise at linking ARGs to plasmids and to bacterial hosts within metagenomes[267]. This, however, is an area of ongoing research.

Objectives of the Study

The first objective of this study was to develop and validate qPCR and WMS-based methods to improve the detection and analysis of the KpSC, including to the level of ST, in complex microbial samples from the human GIT (**Papers I and II**). The second objective was to use these validated methods to investigate the duration and dynamics of KpSC colonisation of the human GIT, as well as associations with the GIT microbiota and host metadata (**Paper III**).

Paper I

Hypothesis: qPCR and WMS have high detection sensitivity of the KpSC from human GIT samples and have potential to supplement culture-based detection.

Specific Objectives:

- Determine the sensitivity and efficiency of KpSC detection by the ZKIR-qPCR in human faecal samples, including determination of the limit of detection (LOD) and limit of quantification (LOQ).
- Determine the sensitivity of KpSC detection in human faecal samples by WMS using standard taxonomy profilers as well as detection of MLST genes and the ZKIR sequence.
- Determine the effect of different faecal sample collection methods on taxonomy profile by WMS.
- Assess KpSC detection by the ZKIR-qPCR and WMS, both direct from human faecal samples as well as following culture-enrichment, and compare to culture-based detection.
- Compare quantification of KpSC abundance in human faecal samples by the ZKIR-qPCR v. WMS
- Explore KpSC ST-level detection by WMS direct from faecal samples.

Paper II

Hypothesis: KpSC ST typing by WMS from culture-enriched faecal metagenomes is equivalent to colony isolate WGS and allows investigation of within-sample strain diversity.

Specific Objectives:

- Compare the accuracy of KpSC ST typing in culture-enriched faecal samples by two recently developed WMS tools, mSWEEP and StrainGE, to single colony WGS.
- Evaluate the detection of within-sample KpSC ST diversity by mSWEEP and StrainGE.
- Investigate the accuracy of StrainGE to reconstruct phylogenetic relationships between the same KpSC STs detected in faecal samples from different hosts compared to single colony WGS.

Paper III

Hypothesis: KpSC GIT carriage is typically of short duration and associates with KpSC abundance, carriage ST/s, other taxa within the microbiota, and host-related factors.

Specific Objectives:

- Determine the GIT carriage duration of the KpSC within a cohort of community-based adults over a six-month period.
- Investigate the association between KpSC carriage duration and abundance within the GIT.
- Investigate the duration and turn-over of KpSC GIT carriage at the ST-level and its association to KpSC carriage duration and GIT abundance.
- Explore the GIT microbiota for taxa that positively and negatively associate with KpSC carriage.
- Investigate the association of selected host-related metadata to KpSC GIT carriage abundance and duration.

Choice of Methods

Appropriate experimental methodologies are a cornerstone of high quality, reliable, and reproducible scientific research. Selecting the correct methods to investigate a specific research question requires careful consideration of strengths and weaknesses of each. Experimental design should be logical, as straight forward as possible, with attention paid to good fundamental design, including robust positive and negative controls. The following is a description of the main methods used in this study, including a rationale for inclusion. More detailed descriptions of each method are subsequently included in **papers I**, **II**, and **III**. A summary of the project workflow is presented in Figure 9.

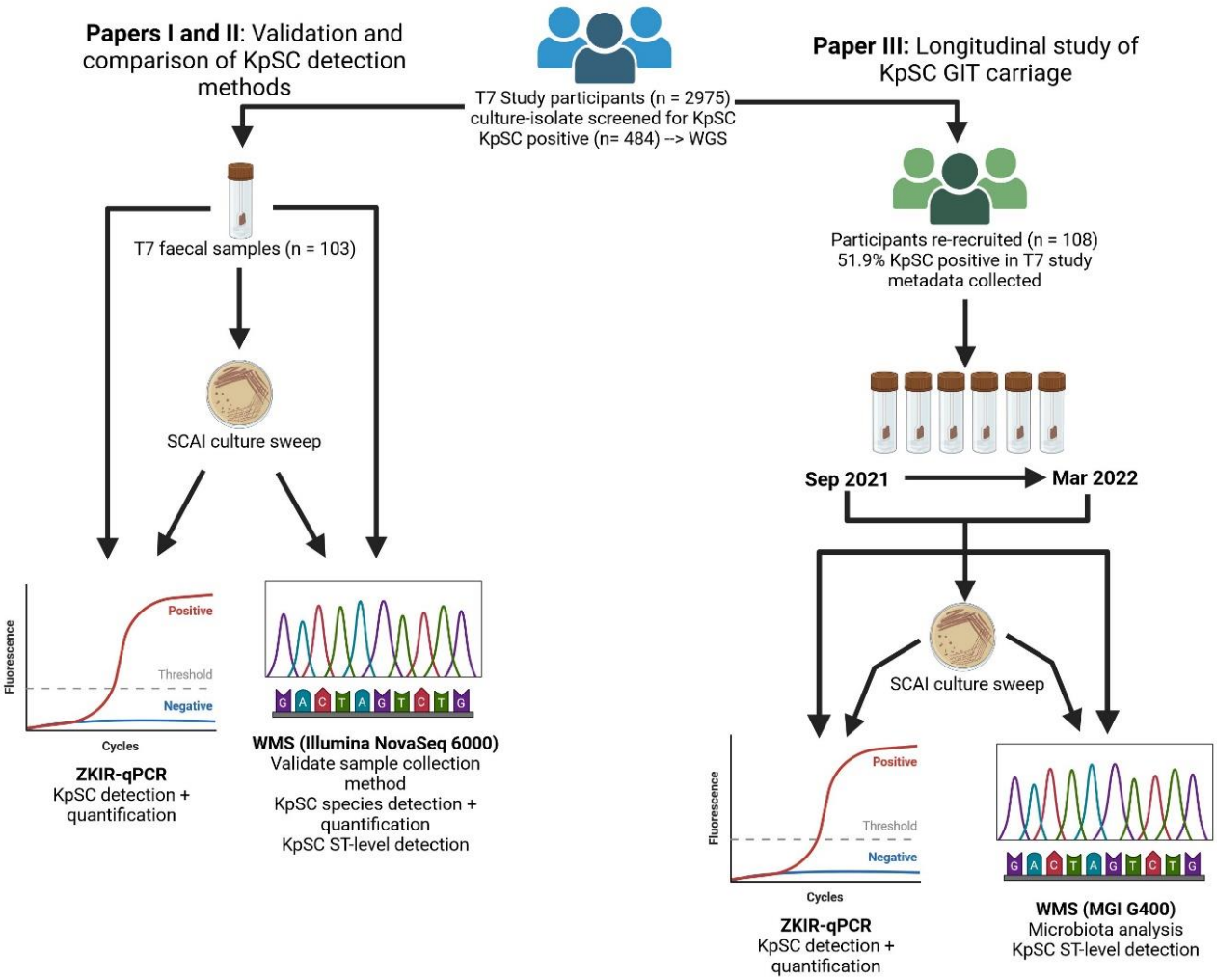


Figure 9. Overview of the project workflow. Created with BioRender.com

1.1 Study Participants and Samples

All participants and samples in this study came from a previous cross-sectional culture-based study of KpSC GIT carriage, conducted from March 2015 to October 2016[89]. This formed part of the seventh survey of the Tromsø Study (Tromsø7, https://uit.no/research/tromsostudy). The Tromsø Study is a repeat cross-sectional, epidemiological population-based study in the municipality of Tromsø, Norway, which is considered representative of a Northern European urban population[268]. As part of the previous cross-sectional KpSC GIT carriage study, faecal samples from 2975 participants were screened for KpSC carriage by culture-based detection, and all positive carriers underwent single colony WGS[89]. This therefore represented a highly suitable sample collection and study population for both the comparison of detection methods and investigation of the ecology of the KpSC within the GIT of community-based adults. For **papers I** and **II**, faecal samples from this previous cross-sectional study were re-analysed by qPCR and WMS (n = 103 for both papers), while for **paper III**, which was a longitudinal prospective study, participants were re-recruited (n = 108) and repeat sampled each month for six months (September 2021 to March 2022). Since the KpSC are primarily known as hospital-associated pathogens, it could be argued that a hospital-based cohort would have been more a more clinically relevant study population. As exposure to hospitals is associated with changes to the GIT microbiome, however, including loss of diversity and accumulation of potential pathogens, such as Kp[269], this would have been addressing a fundamentally different research question. Moreover, KpSC strains carried within the community likely represent a reservoir for KpSC infections once in hospital[85].

All KpSC detection and analysis were performed on faecal samples. Use of faecal samples in WMS-based studies represent a highly convenient, inexpensive, and non-invasive method of studying the GIT microbiota that is highly comparable to rectal swabbing/biopsy[270, 271]. It should be noted, however, the GIT microbiota is not uniform along its length, and significant difference have been found between faecal samples and samples taken from different regions of the colon, particularly the colonic mucosa[270-272]. Thus, faecal sampling should be regarded as a proxy for the GIT microbiota, particularly the colon/rectum. For **papers I** and **II**, faecal samples had been collected in nylon flocked ESwab 490CE.A tubes (Copan, Brescia, Italy). These are convenient sampling devices that can be self-sampled and transported at room temperature and are validated for viable collection of both aerobic and anaerobic bacteria for use in culture-based studies[273]. The results of **paper I**, however, demonstrated this collection method was not suitable for WMS-based studies due to bacterial overgrowth, particularly of *E.*

coli, biasing taxonomy profiles. Therefore, for **paper III**, both ESwabs and Norgen Nucleic Acid Preservation system (Norgen Biotek, Ontario, Canada) were used to allow both culture- and WMS-based studies to be performed. The Norgen Nucleic Acid Preservation system contains a preservative solution preventing further bacterial growth and stabilising genomic material and have been validated for use in WMS-based studies, including as part of the results of **paper I**[274].

Selection of samples and participants for **papers I**, **II**, and **III** was also not random but chosen to have an over-representation of KpSC culture-positive samples. During the previous cross-sectional KpSC carriage study, 16.3% of the study population were KpSC positive by culture-based detection[89]. For **papers I** and **II**, however, 50.4% (n = 52) of selected samples were KpSC positive, while for **paper III**, of the 108 participants recruited to the study, 51.9% had screened positive for the KpSC during the previous cross-sectional study. This increased proportion of KpSC positive samples in each of the studies was done for comparative purposes to improve study power. One drawback of this, however, was that the proportion of persistent carriers among the cohort we found in **paper III** could not be directly extrapolated back onto the underlying population.

1.2 Sample Processing and DNA extraction

All faecal samples collected during this study were self-sampled by participants. Upon arrival at the laboratory 200 µL 85ºC glycerol was added to ESwab tubes prior to storage at -80ºC, a method that has been validated previously[89]. Although Norgen samples can reportedly be stored at room temperature for up to two years (https://norgenbiotek.com/product/stool-nucleic-acid-collection-and-preservation-system), these samples were also directly stored at -80ºC until further processing to ensure stability of the genomic material.

After thawing, most samples underwent DNA extraction directly from faecal material (referred to as Direct samples or direct fecal samples). In addition, ESwab samples were cultured on the *Klebsiella*-selective Simmon's citrate agar with inositol (SCAI) for 48 hours at 37ºC, after which a culture-sweep of all growth was taken (referred to as Sweep samples). SCAI media contains citrate and inositol as its sole carbon sources, both of which can be metabolised by the KpSC resulting in large yellow dome-shaped colonies[275]. This allows differentiation of the KpSC from many other potential competitors, particularly *E. coli*, which typically appear as small, greyish-blue colonies due to inability to utilise one or both carbon sources[275]. Thus, due to the generally low abundance of the KpSC in faecal samples, culture-sweep enrichment with

SCAI media was investigated as a potential method to increase detection sensitivity by both qPCR and WMS.

All samples underwent whole microbiome DNA extraction and purification using the PureLink Microbiome DNA Purification Kit (Thermo Fisher Scientific, Massachusetts, USA). DNA extraction has shown to be a major potential biasing step in WMS-based studies due to differential bacterial cell lysis between different taxa[276, 277]. The PureLink Microbiome DNA Purification Kit uses a combination of heat, chemical, and mechanical (bead-beating) induced cell lysis and has been shown to perform well at giving an unbiased representation of the microbiota, comparable to the International Human Microbiota Consortium (IHMC) Protocol Q[278]. To ensure maximum cell lysis of both gram-negative and gram-positive bacteria was achieved, we made additional modifications to the protocol by: (i) adding a lysozyme-digestion step, (ii) increasing the temperature of the heat-induced lysis step to 95^{o}C, and (iii) bead-beating samples using the Precellys Evolution tissue homogenizer (Bertin Technologies, Montigny Le Bretonneux, France).

1.3 qPCR analysis of Samples

The ZKIR-qPCR was chosen as a rapid, high sensitivity method to screen samples for the presence of the KpSC. The ZKIR-qPCR is a recently developed qPCR assay that amplifies a 78 base-pair intergenic region (the *zur-khe* intergenic region) that was demonstrated to be highly conserved in the KpSC but absent in 88 non-KpSC species analysed, including the closely related non-KpSC *Klebsiella* species *K. oxytoca* and *K. aerogenes*[150]. This qPCR assay has been demonstrated to have very high sensitivity in detecting the KpSC in soil, plant, chicken meat, and salad samples[150, 279]. We defined limit of detection (LOD) and limit of quantification (LOQ) of the ZKIR-qPCR within human faecal samples as the lowest number of KpSC genome copies that could be detected in ten out of ten technical replicates (LOD) and quantified with a coefficient of variation (CV) $\leq$ 35% (LOQ) in accordance with previously published guidelines[207, 280, 281]. LOD and LOQ were determined for the four major human GIT colonising phylogroups within the KpSC (i.e. Kp1, Kp2, Kp3, and Kp4) to ensure the assay would not bias detection toward different KpSC members. Due to the very high sensitivity of qPCR, robust negative controls are essential to control for contamination with exogenous DNA during all qPCR assays. We therefore used non-template controls in all assays, and additionally assayed an *E. coli* strain that had been through all processing steps in parallel with Direct samples, and a *K. oxytoca* strain that was processed in parallel with Sweep samples, to control for DNA

contamination in all sample processing steps. All samples were assayed in technical triplicate to ensure reproducibility of the method and reporting of results was performed in line with the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) Guidelines[280].

1.4 WMS of Samples

Short read sequencing was chosen as our sequencing method for this study due to its well documented use in WMS studies and validated bioinformatic analysis tools. For **papers I** and **II** the Illumina NovaSeq 6000 (Illumina, San Diego, USA) platform were used, however, for **paper III**, as well as our comparison of microbiome profiles from ESwabs v. Norgen samples in **paper I**, the MGI G400 (MGI Tech Co, Shenzen, China) was used due to its equivalent quality of sequencing data and lower per sample sequencing costs[218, 220]. Sequencing depth in WMS studies is an important consideration, since deeper sequencing of samples improves detection sensitivity[261, 262]. This has to be balanced, however, against per sample sequencing costs, as well as data storage and analysis requirements, which also increase with sequencing depth. We chose 20 million 150bp paired-end reads as a trade-off between detection sensitivity and cost. Based on previously published comparative studies this would be expected to give sufficient species-level identification and reasonable ARG coverage[261, 262, 282]. Sample library preparation, sequencing, and demultiplexing were all performed as per the operating protocols of the respective sequencing facilities.

1.5 Bioinformatic Sample Processing

The raw short-read sequencing output, which comes in a FASTQ file format, must undergo several processing and quality control steps before analysis can be performed. Although a range of bioinformatic tools exist for these tasks, processing generally follows the same or similar steps. These will be briefly outlined.

An initial important step in WMS sample processing is removal of human host DNA. In human faecal samples this mostly arises from desquamation of GIT epithelial cells and typically accounts for <1% total DNA content[283]. Human host DNA must be removed not only to improve down-stream analysis but also for participant safety and confidentiality, due to the potential risk that participant identification and important health-related information could be gleaned from this data[284]. Contaminating human DNA can be removed either pre-sequencing, using selective human cell lysis and DNase treatment or by DNA methylation-based target-capture, as well as post-sequencing by selectively mapping and removing human reads from

samples[285, 286]. Several tools are available for the latter purpose, such as FastQ Screen, DeconSeq, Bowtie 2, and BWA[287-290]. We used FastQ Screen (version 0.14.0) against the GRCh39 human reference assembly to remove human DNA from our samples following sequencing[288]. FastQ Screen uses either Bowtie, Bowtie 2, or BWA to map reads against pre-specified genomes, presenting mapping results in graphical format, and has been shown to be efficient at identifying and removing contaminating sequences of interest from samples[288].

Due to the high density of clusters of amplified DNA fragments that are generated during Illumina sequencing, these can sometimes be separated by the sequencer and misidentified as two or more individual clusters[291]. These artifacts are referred to as 'optical duplicates' and must be removed prior to down-stream analysis to prevent bias. For this task we used the Clumpify tool of the BBMap version 38.79 package which has the added advantage of grouping overlapping reads into 'clumps' to increase file compression and accelerate down-stream analysis[292].

Poor-quality sequences that occur during the sequencing process, as well as adapters used for DNA fragment amplification, must be also removed prior to down-stream analysis of sequenced reads[213, 214]. The FASTQ file format of raw sequenced reads contains ASCII-coded information on the quality and confidence of each base call[293]. The probability that a base call is correct (referred to as a Quality-score or Q-score) is expressed in Phred numerical format. Illumina reads typically have Q-scores of 30+ corresponding to a base call accuracy of 99.9% (i.e. probability of an incorrect base call is 1 in 1000)[213]. Several bioinformatic tools, such as FastQC, Trimmomatic and TrimGalore can decode this information and identify and remove ('trim') low quality sequences from reads[294-296]. Other tools such as CutAdapt and AdapterRemoval can be used for adapter removal[297, 298]. For our sample processing we used the package fastp (version 0.20.1) which can perform a quality check, adapter removal, and trim poor-quality reads as a single integrated pipeline, subsequently producing an easy-to-read quality control report for each sequence[299].

During the above processing steps, paired reads can become disordered. Down-stream analysis tools often will not accept these unsynchronized reads. Resynchronisation of paired reads was therefore performed using the Repair tool of the BBMap version 38.79 package[292].

1.6 Bioinformatic Analysis

Like sequence read processing, a large number of tools are available for analysis of processed metagenomic reads. Additionally, analysis can be performed on unassembled reads or by first assembling reads into larger contigs. In general, while both approaches can produce high quality, accurate information, metagenomic assemblies have the advantage of encompassing entire genes, with the trade-off that information can be lost during the assembly process, while unassembled reads retain all sequence information but may be prone to increased false positives[239]. Moreover, the assembly process itself is quite computationally intensive, making this step difficult in larger sample collections[239]. For our study, we predominantly analysed samples as unassembled reads due to the size of our sample collection, the potential for higher detection sensitivity, and the acceptable reported level of precision of this approach[239]. The only exception was in **paper I**, in which we investigated KpSC identification in metagenomic samples through detecting of the ZKIR sequence and the seven genes from the KpSC MLST scheme, which first required an assembly to be performed.

A key step in most WMS studies is taxonomy profiling to obtain an accurate picture of the presence and relative abundance of different taxa within samples. One of the advantages of WMS over 16S rRNA metagenomics, which is an additional metagenomic-based approach involving targeted sequencing of 16S rRNA hypervariable regions, is significantly higher accuracy of species-level identification even at lower sequencing depths[282]. Unsurprisingly, there are a considerable number of taxonomic profiling tools available for short-read WMS taxonomic analysis, with some of the most widely used and validated being Kraken2/Bracken, Centrifuge, MetaPhlAn4, CLARK, and PathSeq[300-305]. For **papers I** and **II**, Kraken2/Bracken and Centrifuge were used, while Kraken2/Bracken only was used in **paper III**. Centrifuge uses Ferrangina-Manzini (FM) and Burrows-Wheeler transform (BWT) to create a highly compressed reference database that can rapidly and accurately perform species identification[303]. Kraken2 is a k-mer based taxonomy profiler that uses a lowest common ancestor (LCA) assignment approach to assign reads to the lowest taxonomic level within a certain level of confidence[305]. The Bayesian Re-estimation of Abundance with Kraken (Bracken) is an extension of Kraken which can then be used to improve abundance estimates[304]. A key issue with all WMS taxonomy profilers, however, is the generation of false positive assignments that occur at low species abundances[306]. This phenomenon was also demonstrated in the results of **paper I**. In this paper we saw the number of Kp false positives was low above 0.1% relative

abundance. This was therefore used as a cut-off for species identification when performing microbiota association studies in **paper III** to balance detection sensitivity against specificity.

A key focus and aim of this study was to perform WMS-based ST-level detection of the KpSC both in Direct and Sweep metagenomic samples. Due to the high level of sequence similarity that occurs between two strains of the same species, differentiation of these by WMS methods is challenging[307]. Despite this, several sophisticated tools capable of performing accurate strain-level detection in WMS have recently been developed. In this project we investigated and compared two such tools: StrainGST (part of the StrainGE toolkit) and mSWEEP[240, 308]. While a more detailed explanation and comparison of these tools is presented in **paper II**, briefly, StrainGST is a k-mer based tool that rapidly detects closest reference genomes in a customised database using relatively low computing resources, while mSWEEP identifies strains through pseudoalignments to genomes in a clustered database, infers probabilistic relative abundances, and adjusts for false positive detections[240, 308]. Although mSWEEP is comparatively slower and more resource intensive to run, its major advantage is a binning step that allows recovery of strain-specific reads for downstream analysis[308]. Similar to using a species detection limit for taxonomy profiling, in **paper II** we took an additional step of implementing a 5% relative abundance cut-off for strain identification for both tools to reduce false positives.

1.7 Statistical Analysis

Appropriate and accurate statistical analyses are essential in research to draw valid conclusions and generate tenable hypotheses from collected data. For **papers I** and **II** we used various standard non-parametric statistical tests such as the Mann Whitney U test and Spearman correlations depending on the specific research question and characteristics of the underlying data. The data collected for **paper III**, however, had a more complex structure due to the repeat sampling of participants. This required a more advanced statistical analytic approach due to violation of the assumption of independence between samples required for most conventional statistical tests. For this paper we employed linear mixed effects models (LMMs) to investigate associations between KpSC GIT abundance and predictor variables in our dataset. LMMs are powerful statistical modelling tools that allow investigation between response and predictor variables of interest (known as fixed effects) while accounting for non-independence between samples induced by a grouping or clustering variable, which in our case was repeat sampling of the same participants (known as random effects)[309]. In other words, LMMs can estimate variance in the response variable both within and between groups within a dataset[309]. Use of

random effects in LMMs control for non-independence typically by constraining groups of samples to have either the same intercept (known as a random intercept model), or the same intercept and the same slope (known as a random intercept and random slope model), with both approaches having their own pros and cons[309]. In general, random intercept models are simpler and easier to interpret but may have increased type I (false positives) and type II errors (false negatives), whereas combined random intercept and random slope models are more complex, which, while allowing for more accurate estimates, also risks overfitting data[309, 310]. Overfitting is a phenomenon where a model starts to fit the noise inherent to the dataset and therefore will not generalise well to new data[310]. As **paper III** is primarily aimed at exploring associations with KpSC GIT carriage to generate hypotheses for future testing, we chose random intercept LMMs to model our data due to simplicity and reduced risk of data overfitting.

Another important consideration with relevance to **paper III** is multiple testing correction. This is a method of adjusting *p*-values to correct for the increased probability of observing type I errors when multiple statistical tests are performed[311, 312]. Two common methods for performing multiple testing correction are the Bonferroni Correction, which controls the probability of at least one type I error occurring across all statistical tests performed, known as the family-wise error rate (FWER), and the Benjamini-Hochberg Procedure, which controls the expected proportion of type I errors among all rejected null hypotheses, known as the false discovery rate (FDR)[311, 312]. In general, the Bonferroni Correction is more conservative and most appropriate when type I errors can't be tolerated, e.g. in studies where type I errors can have serious consequences, but may produce more type II errors, whereas the Benjamini-Hochberg Procedure is preferred in studies where maximising discovery of true positives is important and some false discoveries can be tolerated[311, 312]. Like the choice of LMM above, since **paper III** was an exploratory study of associations with KpSC GIT carriage, we wished to maximise true positive associations and therefore the Benjamini-Hochberg Procedure was selected for multiple testing correction.

1.8 Ethical Considerations

Proper consideration of ethical issues is important to ensure participant safety as well as maintain public trust in scientific research. Two important considerations in studies involving participant sampling are maintaining privacy and confidentiality, as well as risks involved in the sampling process. To mitigate potential risks in this study, enrolled participants were de-identified and given unique four-digit IDs, in addition, samples and WMS sequences were given

different unique codes. All data except WMS sequence data was stored electronically on the UiT OneDrive, protected with two-step authentication. WMS sequences were stored separately on the Norwegian academic high-performance computing and storage services maintained by the Sigma2 Norwegian Research Infrastructure Service (NRIS), as well as the Norwegian e-infrastructure for Life Sciences (NeLS) maintained by ELIXIR Norway. Access to data storage was restricted to direct project members only. As discussed above, human DNA was removed upon receipt of all WMS sequence data. Collecting faecal samples is also non-invasive and posed no risk to participants. Finally, all participants were provided with detailed information regarding the purpose and importance of the study, the faecal sampling procedure, and the planned use of faecal material provided. Signed consent was obtained from all enrolled participants. Enrolment in the study was voluntary and participants could withdraw at any time. This study was approved by the Regional Committee for Medical and Health Research Ethics, North Norway (REC North reference: 2016/1799, 2014/940, and 137064).

Summary of Results

Paper I

Detection of *Klebsiella pneumoniae* human gut carriage: a comparison of culture, qPCR, and whole metagenomic sequencing methods

- We found the ZKIR-qPCR had an amplification efficiency > 90% and R^2 > 0.99 for representative strains of Kp1, Kp2, Kp3, and Kp4 in a linear dynamic range of 250,000 to 3 genome copies per reaction in the presence of 25 ng human faecal microbiome DNA.
- All four KpSC phylogroups could be detected to three genome copies per reaction (LOD) and quantified to 16 genome copies per reaction with CV ≤ 35% (LOQ) in the presence of 25 ng human faecal microbiome DNA.
- The ZKIR-qPCR was positive in 52/52 Kp culture-positive samples and 11/51 (22%) and 23/47 (49%) KpSC culture-negative Direct and Sweep samples, respectively. Culture positive samples had a significantly higher abundance of Kp than culture negative samples when quantified by the ZKIR-qPCR ($p < .001$).
- Faecal samples collected in ESwabs had a significantly higher abundance of Enterobacterales and *E. coli* compared to a sample from the same faecal material from the same participant collected in Norgen collection devices (median 38.2% vs 0.62% $p = .002$, and median 26.2% vs 0.16% $p = .002$, respectively).
- When measured by WMS, the relative abundance of Kp was significantly higher in ZKIR-qPCR positive samples than negative samples for both Direct and Sweep samples (median 0.027% vs 0.00035% $p < .001$, and 6.05% vs 0.063% $p < .001$, respectively) but had considerable overlap.
- Using a 0.01% relative abundance cutoff, WMS was positive in 37/52 (71.2%) of culture positive Direct samples but identified multiple qPCR- and culture negative samples as positive.
- Kp false positives detected by WMS correlated strongly with the abundance of other Enterobacterales present in *in silico* binary species mixes.
- The ZKIR sequence and 4/7 Kp MLST genes were detected directly from WMS assemblies in 20/103 and 17/103 Direct samples and 54/99 and 52/99 Sweep samples, respectively. All Direct ZKIR and MLST WMS positive samples were also positive by the ZKIR-qPCR and culture. Similarly, all Sweep ZKIR and MLST positive WMS

samples were also positive by the ZKIR-qPCR and 50/54 and 49/52 were positive by culture, respectively.

- Kp relative abundance measured by WMS correlated strongly with abundance measured by the ZKIR-qPCR for both Direct and Sweep samples (Spearman's $R = 0.91$ $p < .001$ and $R = 0.96$ $p < 0.001$, respectively).
- WMS performed accurate ST-level detection, agreeing with culture-isolate WGS detection in 16/19 Direct samples, and detected within sample ST diversity.

Paper II

Sequence type and strain level detection of *Klebsiella pneumoniae* in complex bacterial metagenomes: comparative performance of mSWEEP and StrainGE bioinformatic tools

- In samples with a single Kp ST detected by single colony WGS we found the ST detected by both StrainGE and mSWEEP was concordant in 44/49 (89.8%) and 46/49 (93.9%) Sweep samples, respectively.
- In samples with two STs detected by WGS, StrainGE and mSWEEP detected the same two STs as single colony WGS in 2/3 (66.7%) samples. In the sample that was discrepant, only one of the STs detected by both StrainGE and mSWEEP agreed with single colony WGS (ST151).
- Within sample ST diversity was detected in 15 and 19 samples by StrainGE and mSWEEP, respectively (max 3 STs/sample). Additional STs detected were concordant between StrainGE and mSWEEP in ten out of twelve samples.
- In phylogenetic analysis, the ST detected by single colony WGS and the corresponding ST detected by StrainGE were assigned to the same clade for ST20 and ST26. For ST25 the single colony WGS detected ST and StrainGE detected ST were assigned to different clades despite coming from the same sample.

Paper III

Longitudinal analysis of *Klebsiella pneumoniae* Species Complex human gut colonization identifies a subgroup of high-abundance persistent carriers with strong microbiota associations

- Of the 85 participants that delivered all six samples, we found 69.4% (59/85) carried the KpSC between 1/6 - 5/6 months (collectively termed transient carriers), 27.1%

(23/85) were positive in all 6/6 months (persistent carriers), and 3.5% (3/85) were negative in all 6/6 months (non-carriers).

- We found the abundance of the KpSC in samples from persistent carriers was significantly higher than all carriage durations of transient carriers (*p*-values .008, <.001, .008, .004, and .03, for 5/6, 4/6, 3/6, 2/6, and 1/6 months, respectively, Walt t-distribution approximation). Persistent carriers also made up 14 of the top 15 participants with the highest median KpSC carriage abundance.
- The same KpSC ST was detected between 3/6 – 6/6 months in 15/21 (65.2%) persistent carriers. Four persistent carriers had the same ST was detected in all six months (ST405, ST26, ST13, and ST14) and five had the same ST detected in 5/6 months (ST27, ST643-1LV, ST20, ST360, and ST876). Eight STs carried 3/6 – 6/6 months were among the 3.5% most prevalent in the previous Tromsø7 KpSC cross-sectional carriage study.
- Of the 25 transient carriers with an ST detected in at least two months, the same ST was detected two and three times in n = 7 and n = 3 participants, respectively. The remaining 15 did not have the same ST detected in repeated samples.
- Of the 21 STs detected in a participant between 3/6 – 6/6 months, 16 were Kp, three were *K. variicola*, and two were *K. quasipneumoniae* subsp. *similipneumoniae*.
- We did not find a significant association between KpSC carriage abundance and alpha diversity (measured by the Shannon Index) by LMM analysis among our samples (β -0.008, 95% CI: -0.02 – 0.002, *p* = .107, Walt t-distribution approximation).
- Beta-diversity analysis using Bray-Curtis dissimilarity at the genus level and visualisation by Principal coordinate analysis (PCoA) demonstrated samples with high KpSC abundance were more predominant in the positive half of PCoA axis 1.
- Biplot analysis of Bray-Curtis dissimilarities and hierarchical clustering demonstrated samples grouped into three distinct clusters based on different taxa abundances. Microbiome Cluster I was significantly higher in KpSC abundance and was higher in *Bacteroides* and *Phocaeicola*, while Microbiome Cluster II and III were higher in *Bifidobacterium*, *Alistipes*, *Akkermansia*, as well as multiple genera belonging to Bacillota (Firmicutes).
- LMM analysis identified 103 microbiota species with significant positive or negative associations with the KpSC. There was strong agreement between positively and negatively associated species and the genera identified by biplot analysis and hierarchical clustering. Notably, five of the fifteen strongest negatively associated species all belonged to the *Bifidobacterium* genus.

- Significant associations were detected between the KpSC and ten Enterobacterales species. Of these five had positive associations (including the non-KpSC *Klebsiella* species *K. michiganesis*, *K. grimontii*, and *K. oxytoca,* as well as *Citrobacter freundii* and *Enterobacter roggenkampii*), and five had negative associations to the KpSC (*Enterobacter asburiae*, *Salmonella enterica*, *Enterobacter kobei*, *Escherichia alberti*, *and Enterobacter cloacae*).
- Further investigation of the twenty microbiota species with the strongest negative associations to the KpSC demonstrated negative associations could also be observed dynamically within an average of 70.9% of participants (range 60.9% - 77.8%) that had at least one sample with KpSC abundance over 3 genome copies/ng DNA.
- LMM analysis demonstrated a positive association between KpSC abundance and age, travel abroad, and diabetes. Use of antimicrobials within four weeks of sample taking was negatively associated with KpSC abundance. Microbiota species significantly associated with both diabetes and antimicrobial use strongly overlapped with microbiota species significantly associated with KpSC abundance.
- A spike in KpSC abundance of at least three-fold was observed in all ten participants that reported at least one incident of antimicrobial use and delivered at least one sample post exposure. One participant underwent a one-hundred-fold increase in GIT abundance of the Kp strain ST643-1LV (from 0.35% to 32.3%) following penicillin exposure and sustained this high abundance for the following five months.

General Discussion

Ever since the first reports of acquired carbapenem resistance during the 1990s, Kp has steadily emerged as a major threat to global health through the widespread acquisition and dissemination of resistance toward clinically important and last-line antimicrobials[4]. The resulting difficult-to-treat nosocomial infections and frequent hospital outbreaks have resulted in serious consequences for patients, particularly those most vulnerable, and created a large burden on healthcare systems[5, 6, 57]. Coupled to this has been the establishment of HvKp lineages, able to cause aggressive community-acquired infections in younger hosts lacking the typical risk factors associated with opportunistic nosocomial Kp infections[23, 24]. The concerning rise of XDR- and PDR-Kp, the convergence of hypervirulence and AMR in single Kp lineages, as well as the emergence of other KpSC members, in particular *K. variicola*, as important human pathogens capable of acquiring AMR elements, further underlines the importance of this threat[14, 71, 80]. To meet this challenge, fast and accurate methods are required to improve our detection, surveillance, and understanding of this group of pathogens, as well as identify new targets and strategies for intervention.

Kp colonisation of the GIT represents a major target for surveillance and intervention due to its importance as a risk factor for invasive infection and potential as a silent reservoir for dissemination of high-risk clones and AMR determinants[85, 93, 99, 100]. Despite this, much remains unknown regarding the ecology of KpSC within the GIT, and there is a need for new tools allowing more accurate detection and analysis of the KpSC in this setting. The aim of this study was therefore to investigate the molecular-based methods, the ZKIR-qPCR and WMS, to detect and analyse the KpSC from human faecal samples compared to culture-based detection (**paper I**), assess ST-level detection of the KpSC by WMS in culture-enriched faecal samples compared to single isolate WGS (**paper II**), and use the methods developed in **papers I and II** to investigate the duration, ST dynamics, microbiota and host associations of KpSC GIT colonisation in a longitudinal setting (**paper III**).

qPCR has previously been demonstrated to be a highly sensitive method for the detection of pathogens from a range of clinical samples[206]. Accordingly, of the methods we investigated in **paper I**, our results demonstrated the ZKIR-qPCR had the highest KpSC detection sensitivity, with a remarkably low LOD of just three genome copies per reaction. We found the ZKIR-qPCR was able to detect the KpSC directly from faecal material in all culture-positive samples, as well as 11/51 culture-negative samples. This sensitivity was further enhanced by performing

SCAI-culture sweep enrichment, which increased the detection to 23/47 culture-negative samples. In-line with our findings, Barbier *et al*. and Rodrigues *et al*. have also previously demonstrated a higher KpSC detection sensitivity by the ZKIR-qPCR compared to culture in soil, plant, chicken meat, and salad samples[150, 279]. Together, these findings indicate the ZKIR-qPCR is a highly sensitive tool suitable for the rapid and accurate detection of the KpSC in a range of complex microbial samples.

To date most studies investigating the KpSC GIT colonisation have used conventional single-colony, culture-based approaches. While these studies have yielded invaluable insights into the ecology of the KpSC, several studies, including our work in **paper I**, have suggested such culture-based detection may have reduced detection sensitivity in complex microbial communities such as faecal samples[195-198]. Indeed, the results of **paper I** strongly suggest single-colony culture-based detection substantially underestimates the true prevalence of KpSC GIT carriage in faecal samples. Moreover, we demonstrated detection by culture was directly related to the abundance of the KpSC, suggesting an abundance threshold exists, below which detection by culture become increasingly difficult. Beyond studying the ecology KpSC GIT colonisation, the potential for reduced sensitivity of culture-based detection can have important public health implications. This was illustrated by Singh *et al*., in which culture-based screening of rectal swabs in a long-term acute care hospital failed to identify GIT carriage of CPKp in 9 out of 66 patients[198]. This highlights the potential utility for high sensitivity detection methods such as the ZKIR-qPCR for infection control and screening purposes.

The unbiased and untargeted nature of WMS has led to its proposed use as a universal test for detection of pathogens, as well as associated AMR and virulence determinants, in both clinical diagnostics and surveillance[237, 241, 242]. Indeed, studies have suggested WMS is potentially faster and has a higher overall detection rate than conventional methods, with key advantages being enhanced detection of fastidious/unculturable organisms, polymicrobial infections, and in the setting of prior antibiotic exposure[247, 313, 314]. Major hurdles preventing more widespread use of WMS, however, include high costs, uncertainties regarding sensitivity and specificity, and lack of standardisation, particularly regarding criteria for pathogen detection[238, 314]. A key aim of **paper I** was to therefore to assess the sensitivity and specificity of Kp detection by WMS in human faecal samples. Using a standard taxonomy profiler, we found WMS had strong agreement with both culture- and qPCR-detection at higher relative abundances, however, below approximately 0.1% abundance Kp detection by WMS became increasing difficult to distinguish from background noise. Moreover, using in silico binary species mixes we showed

the false positive rate of Kp detection by WMS was directly proportional to the abundance and relatedness of other Enterobacterales present. Performing a SCAI-sweep enrichment step also did not markedly improve detection specificity, presumably because other non-KpSC *Klebsiella* species were also enriched, leading to a proportional increase in false positives. Targeting KpSC-specific sequences, the ZKIR-regions and MLST genes, in assembled metagenomes did improve this specificity, with sensitivity similar to that of culture. Given that Kp abundance within the GIT has been shown to be predominantly below 0.1%, including in **paper III**, this is an important consideration that may limit the use of WMS with standard taxonomy profilers as a surveillance tool for monitoring GIT carriage of the KpSC[86, 114].

Previous studies have shown an increased abundance of the Kp within the GIT microbiota infers a higher infection risk and is associated with an increased rate of strain transmission in hospitalised patients[95-97]. Infection in this setting is likely driven by translocation of bacteria from the GIT lumen across the intestinal epithelial layer into extraintestinal tissues[315]. As this is thought to be a mostly stochastic event in susceptible hosts, the probability of a translocation event occurring increases with increasing pathogen abundance in the GIT[315]. In addition to this, in **paper III** we found GIT abundance likely has an important role in determining success and longevity of KpSC colonisation (discussed below). These findings suggest an important role of high abundance GIT colonisation in both the pathogenesis and ecology of the KpSC. Our results of **paper I** indicate that while both WMS and the ZKIR-qPCR can be used to perform accurate KpSC quantification from faecal material, the ZKIR-qPCR could perform this faster and was accurate to lower abundances, with a CV $\leq$ 35% at just 16 genome copies per reaction. Thus, the rapid and accurate quantification of KpSC by the ZKIR-qPCR could have potential roles in both the research setting, investigating KpSC colonisation ecology and responses to interventions and exposures, as well as the clinical setting, as an efficient means of identifying and stratifying patients at risk of KpSC infection and strain transmission.

The global spread of AMR by Kp is predominantly driven by the expansion and dissemination of a relatively small number of high-risk CGs[3]. Moreover, within hospitals, Kp outbreaks are typically clonal in nature and can rapidly spread between hosts and form reservoirs in the hospital environment that hamper infection control efforts and facilitate ongoing transmission[234-236]. Identification and analysis of Kp at the sub-species and strain-level is therefore essential to understand colonisation dynamics and spread, identify reservoirs, and target infection control interventions to prevent further dissemination. Our results indicate the accuracy of ST detection by WMS, using the tools StrainGE and mSWEEP, is comparable to

that of single colony WGS. As shown in **paper I**, however, ST detection by WMS direct from faecal material was limited by the generally low abundance of the KpSC (<0.1%). **Paper II** demonstrated ST detection was greatly enhanced by performing an SCAI-sweep culture-enrichment step. Using this method, ST detection agreed with single colony WGS detection in 44/49 and 46/49 samples analysed by StrainGST (part of StrainGE) and mSWEEP, respectively. Moreover, in samples that had two KpSC STs detected by WGS, WMS correctly identified both in 2/3 and 2/2 samples in **paper II** and **paper III**, respectively. Notably in the sample that was not concordant, one ST detected by both StrainGST and mSWEEP agreed with culture (ST151). Furthermore, in this sample, StrainGST detected an additional ST that was the same phylogroup as that detected by single colony WGS (*K. variicola*) but different ST. Moreover, in **paper II**, STs detected in addition to the WGS detected ST were concordant between StrainGST and mSWEEP in ten out of twelve samples. The ability to detect multiple STs within the same sample is a key potential advantage of WMS-based strain detection over single colony detection by WGS. Indeed, in **paper III**, using StrainGST we were able to show multiple pairs of KpSC STs, including ST20/ST1832, ST36/ST360, ST507/ST5084, and ST10/ST641, were capable of stable co-colonisation within the same participant for prolonged periods, demonstrating the advantage of WMS in understanding ST colonisation dynamics in this setting.

Paper II also demonstrated the relative strengths and weaknesses of StrainGE and mSWEEP in ST detection by WMS. StrainGST was able to perform rapid ST detection (within hours) using low computing resources. Furthermore, StrainGE could reconstruct accurate phylogenetic relationships between the same STs found in different samples for 2/3 STs analysed (ST20 and ST25 but not ST26). Conversely, while mSWEEP was slower and more resource intensive, the inclusion of a binning step means mSWEEP has the potential for recovery of complete or near complete strain-level genomes for in-depth down-stream analysis. Despite the many barriers already outlined preventing the current use of WMS in clinical microbiology, this work nonetheless demonstrates the potential of this technology as a useful tool in the future of pathogen and AMR surveillance and infection control. In particular, the ability to perform rapid ST detection with tools like StrainGST, combined with the real-time availability of sequencing data from ONT platforms, hints at the exciting potential for future point-of-care outbreak screening and infection control, producing actionable results in a clinically relevant timeframe.

The duration of KpSC GIT colonisation represents a large knowledge gap in our current understanding of KpSC ecology. Previous small sample-sized studies have suggested GIT colonisation by the KpSC is of short-duration with high strain turn-over[98, 153]. Studies of MDR-Kp acquired in hospital or during international travel, however, have demonstrated prolonged GIT colonisation up to four years is possible[99-103]. Furthermore, during prolonged colonisation, strain transmission of MDR-Kp to household contacts as well as HGT of AMR-contained plasmids to other Kp species, as well as *E. coli*, has been observed[99, 100]. The aim of **paper III** was to therefore utilise the methods we had developed in **papers I and II** to answer key questions regarding the duration of KpSC GIT colonisation. By analysing faecal samples from 85 community-based participants collected monthly for six months using the ZKIR-qPCR, we demonstrated KpSC GIT colonisation is indeed predominantly transient, lasting between one and five months in 69.4% (n = 59) of participants, with only 3.5% (n = 3) of participants persistently negative for the KpSC in all samples. The remaining 27.1% (n = 23) of participants, however, were persistently positive for the KpSC in all six months, indicating prolonged GIT colonisation of the KpSC is possible even in the absence of predisposing risk factors such as recent hospitalisation or international travel. Further to this, quantifying KpSC by the ZKIR-qPCR demonstrated persistent GIT colonisation was significantly associated with a higher KpSC abundance compared to transient carriage.

Performing ST-detection using StrainGST, we found the same ST was present in 3 – 6 out of 6 months in 65.2% (15/21) of persistent carriers. This contrasted to transient carriers, in which only 40% (10/25) of participants analysed had the same ST detected either two or three times (n = 7 and 3, respectively), while the remaining 15 did not have the same ST detected in repeated samples. In addition, we found a significantly higher abundance of KpSC among participants that maintained the same ST between 4 – 6 months compared to those that had highest ST turnover (at least three different STs in three different months). These findings indicate a strong trend toward ST retention among high-abundance persistent KpSC carriers and suggests a role for high abundance carriers as reservoirs for long-term ST maintenance within the community.

Interestingly, of the four STs detected in all six participant samples, three were global high-risk lineages associated with the carriage of ESBLs and carbapenemases (ST405, ST13, and ST14) [316-318]. This suggests these STs may be particularly well adapted to prolonged colonisation of the human GIT and may explain at least part of the global success of these lineages by providing more frequent opportunities for dissemination and acquisition of AMR +/- virulence determinants. Further studies elucidating the strategies used to achieve prolonged colonisation

may therefore offer potential targets for intervention to slow the dissemination and spread of AMR by high-risk lineages such as these.

Despite the therapeutic potential of colonisation resistance to induce decolonisation of Kp and other Enterobacterales from the GIT through microbiota-modulating therapies, current approaches to utilise this, such as through FMT and probiotics, have not yet proved efficacious[134, 136, 147]. One possible reason for this is the lack of fundamental studies identifying the combinations of commensal species required to induce efficient decolonisation of Kp and other Enterobacterales. For example, most probiotics used in clinical trials are *Saccharomyces*, *Lactobacillus*, and *Bifidobacterium*, or a combination of these, however, Spragge *et al*. demonstrated colonisation resistance of Kp requires the collective action of a diverse community of specific bacteria, with single species in isolation unable to induce efficient decolonisation[121, 136]. To further investigate this, we found the microbiota of participants in **paper III** could be grouped into three distinct clusters based on taxa abundance. Of these, Microbiota Clusters II and III, which were abundant in *Bifidobacterium*, *Akkermansia*, *Alistipes*, and multiple Bacillota (Firmicutes) including *Faecalibacterium*, *Christensenella*, *Oscillibacter*, *Intestinimonas*, and *Eubacterium*, were associated with a lower abundance of the KpSC compared to Microbiome Cluster I. This suggests the taxa abundant in these clusters may be able to collectively suppress KpSC expansion in the GIT. These observations were supported by univariate linear regression investigating individual KpSC-microbiota associations at the species level, both across the cohort and within individual participants. Notably, many of the taxa we identified are major producers of SCFAs in the adult GIT, supporting the proposed role for this group of molecules in inhibiting KpSC expansion[129]. Future studies investigating colonisation resistance against the KpSC involving combinations of these taxa, and the mechanisms involved, such SCFA production, may therefore have potential to lead to more tailored and effective microbiota-based decolonisation strategies.

Competition for nutrients is also a major mechanism driving colonisation resistance[117]. Moreover, most nutrient competition occurs between closely related species due to a larger potential metabolic overlap[117]. Nutrient competition between members of the Enterobacterales is therefore thought to be important in inhibiting colonisation by several pathogens within this order, including Kp[127, 319, 320]. Conversely, it is conceivable that Enterobacterales species that can stably co-inhabit the same nutrient niche for extended periods may be more likely to engage in HGT through prolonged shared proximity. Thus, understanding the colonisation dynamics between the KpSC and other Enterobacterales members may have potential to improve our

understanding of mechanisms that inhibit KpSC colonisation, as well as conditions in which the KpSC may be more likely to participate in HGT of AMR determinants. In **paper III**, we found positive associations between the KpSC and the non-KpSC *Klebsiella* species *K. michiganesis*, *K. oxytoca*, and *K. grimontii* within microbiome samples. This was somewhat surprising, given the close evolutionary relationship between these species, it may have been predicted that negative associations would have predominated due to the potentially large overlap in shared nutrient requirements. Indeed, Osbelt *et al*. previously demonstrated *K. oxytoca* can induce decolonisation of Kp in mice by outcompeting for specific carbohydrates[127]. One possible explanation for these conflicting results is that while antagonistic interactions such as these may be occurring at the micro-level, on the macro scale, a microbiome with sufficient nutrients could plausibly support multiple different *Klebsiella* species. Additionally, although Osbelt *et al*. observed this inhibitory interaction for two different Kp STs, the high metabolic diversity of the KpSC may mean other strains are able to overcome this form of nutrient inhibition.

Another potential inhibitor of the KpSC is *E. coli*, which was identified as a key species in blocking Kp colonisation by Spragge *et al*., while Mäklin *et al*. found *E. coli* to be an important competitor of Kp in the developing infant microbiome[90, 121]. Although we did not identify *E. coli* as significantly associated with the KpSC in **paper III**, five Enterobacterales species were identified as being negatively associated with the KpSC suggesting other species within the order may be able to fill this antagonistic role in the adult GIT microbiota. It should be noted, however, as we demonstrated in **paper I**, WMS-based detection of closely related low abundance species such as the Enterobacterales may have limitations in accuracy due to read misassignment between species. Future studies investigating the dynamics between the KpSC and other Enterobacterales using more sensitive methods such as species-specific qPCR in addition to the ZKIR-qPCR are therefore warranted.

Although several cross-sectional studies have identified host factors that are associated with an increased prevalence of KpSC detection in GIT samples, the underlying mechanisms by which these factors may promote KpSC colonisation are not well understood[87, 89]. Investigation of previously identified KpSC carriage risk factors in **paper III** demonstrated positive associations between KpSC GIT abundance and age, international travel, and diabetes. This may suggest achieving high abundance within the GIT microbiota is an important mechanism in increased KpSC colonisation in the presence of these factors. Additionally, investigating associations within the microbiota demonstrated species negatively associated with diabetes

strongly overlapped with those negatively associated with the KpSC, including *Bifidobacterium* and multiple Bacillota (Firmicutes). Similar microbiota changes have previously been described in the presence of diabetes and are thought to contribute to disease progression through promoting a chronic inflammatory state within the GIT[321]. It is possible that this chronic inflammation may also promote the increased KpSC colonisation seen in diabetics. Moreover, probiotic supplementation with *Bifidobacterium, Clostridium,* and *Akkermansia* has been shown to improve glycaemic control in mice[322]. This suggests that appropriate probiotic therapy may have potential to provide additional health benefits beyond promoting decolonisation of the KpSC.

Antimicrobial therapy has been associated with an increase in the prevalence of KpSC GIT carriage and is a risk factor for Kp HAIs[87, 89, 323, 324]. In **paper III**, we observed a significantly lower KpSC abundance within four weeks of antimicrobial exposure, with an associated reduction in taxa negatively associated with the KpSC, including *Akkermansia mucinophila*, *Bifidobacterium, Alistipes,* and multiple Bacillota (Firmicutes). This was followed by a relative 'spike' in KpSC in the subsequent months. Although this observation was based on only ten participants in our cohort that delivered at least one sample post antimicrobial exposure, it was nonetheless suggestive of a post-antimicrobial bloom, which have been reported previously for several Enterobacterales, including Kp[325, 326]. Such blooms are thought to be the result of increased nutrient availability, and reduction in SCFAs, leading to increase oxygen availability[131, 326, 327]. Importantly, while the abundance of KpSC returned to pre-spike levels in most participants, one participant underwent a large increase from <1% to 32% relative abundance of the Kp strain ST643-1LV, which was sustained for the following five months. Although further studies are required to confirm and further investigate this phenomenon, this suggests antimicrobial exposure may represent an important source of high abundance persistent carriage of the KpSC.

This project had several key strengths, including the use of the highly sensitive and specific ZKIR-qPCR, allowing accurate detection and quantification of the KpSC within samples, and was a robust tool for analysing carriage duration, abundance, and microbiota associations in **paper III**. The recent cross-sectional KpSC carriage study, part of the Tromsø7 Study, also allowed access to a large number of samples already analysed by culture and WGS. This was an excellent basis for our development and comparison of KpSC detection methods in **papers I** and **II**. It also facilitated recruitment of participants for **paper III** for further analysis of KpSC GIT carriage.

The study had some important limitations. In **paper I**, Centrifuge was used as our standard taxonomy profiler to investigate the KpSC detection sensitivity of WMS. Although Centrifuge is well validated and has performed well in comparative studies in terms of detection accuracy[303, 306, 328], it is possible using another profiler with a different database may have proved more sensitive and specific at detecting the KpSC. Additionally, in **paper II** our investigation of accuracy of StrainGE to recreate phylogenetic strain relationships was limited by the number of samples we had containing the same ST. Moreover, the SNP distances of 605 to 1678 between ST20 isolates suggested these strains were not especially closely related. **Paper III** was also limited by non-random participant selection, potentially being underpowered to find associations between KpSC carriage and participant factors, use of questionnaire data, and being a single population-based study of older age of participants (median age 69 y), which are all discussed in more detail in this paper.

Conclusions and Future Prospectives

This study has demonstrated the ZKIR-qPCR and WMS are powerful tools to detect and analyse the KpSC in human faecal samples that can complement culture-based detection in studying the ecology of KpSC GIT colonisation and may have potential to support more clinical applications. The ZKIR-qPCR and WMS were complimentary in their respective possibilities and limitations. This was evidenced by the speed and accuracy of the ZKIR-qPCR in detecting and quantifying the KpSC, allowing more in-depth microbiota analysis and ST-detection by WMS. Applying these methods to the study of KpSC GIT carriage in a longitudinal setting demonstrated a pervasive, predominantly transient pattern of carriage, with a subset of the population susceptible to high-abundance prolonged colonisation and ST retention. Our identification of taxa consistently associated with KpSC abundance, both positively, including *Bacteroides* and *Phocaeacola*, and negatively, including *Bifidobacterium*, *Alistipes*, *Akkermansia*, and multiple Bacillota (Firmicutes), suggests these may be important facilitators and inhibitors of KpSC colonisation, respectively. Understanding the mechanisms by which these taxa may influence KpSC colonisation has potential to inform novel decolonisation strategies that could reduce the burden of infection and spread of AMR by this important group of pathogens.

This study has generated a number of hypotheses which has created the foundation for future studies. In **paper II**, while we showed both StrainGE and mSWEEP had good accuracy at KpSC ST detection from culture-enriched faecal samples, it would be very interesting to apply

these tools to samples from a real-life Kp hospital outbreak as a 'proof of principle' study. This would allow direct comparison of the accuracy and speed of outbreak strain detection, as well as the ability to contribute to accurate elucidation of transmission chains and identify outbreak sources compared to conventional methods, allowing further evaluation of the future potential and utility of WMS in the clinical setting. Moreover, investigation into the use of probe-target enrichment WMS is currently underway in our lab, which is showing promise for further improvement of KpSC and associated AMR detection by WMS in the surveillance and clinical setting.

Our findings and dataset we have created in **paper III** have also laid the basis for further investigations, many of which are already underway. Currently, we have begun the process of isolating KpSC strains we identified as both persistent and transient colonisers. Using comparative genomics and potentially *in vitro* or even *in vivo* studies, we hope to identify and test acquired traits within these strains that may predispose to prolonged GIT colonisation. Moreover, analysis of within-host evolution of persistently colonising strains across the six-month study period may provide insights into microbiome adaption allowing their prolonged colonisation. Additionally, *in vitro* and *in vivo* competition assays using the taxa identified as either positively or negatively associated with KpSC colonisation may allow identification of the minimum consortia required for effective colonisation resistance and the mechanisms by which this is achieved.

In combination with the ongoing and planned studies outlined above, metabolomic-based studies, both at the level of individual KpSC strains and the microbiota/microbiome are also underway to identify the metabolic conditions that may promote or inhibit KpSC GIT colonisation. Here, *in silico* metabolic profiling of KpSC strains and comparison between persistent and transient colonisers, using tools such as the recently developed Bactabolize[329], may identify key KpSC metabolic traits that promote GIT colonisation persistence, which could be tested *in vitro* and *in vivo*. *In silico* metabolomic profiling of the GIT microbiota in high KpSC abundance compared to low KpSC abundance samples is also ongoing to test whether the microbiota of low abundance KpSC samples are more able to consume nutrients required by the KpSC, and thus blocking nutrient access and restricting GIT expansion of the KpSC as hypothesised by Spragge *et al*[121]. Lastly, metabolomic profiling of the faecal microbiome itself is being performed to identify metabolites that associate positively and negatively with KpSC presence and abundance, with a particular focus on SCFAs, to investigate the role of different microbiome metabolites as potential promoters and inhibitors of KpSC colonisation.

References

1. Podschun R, Ullmann U. Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clin Microbiol Rev 1998; 11:589-603.
2. Wyres KL, Holt KE. Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria. Curr Opin Microbiol 2018; 45:131-9.
3. Wyres KL, Lam MMC, Holt KE. Population genomics of Klebsiella pneumoniae. Nat Rev Microbiol 2020; 18:344-59.
4. Navon-Venezia S, Kondratyeva K, Carattoli A. Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance. Fems Microbiol Rev 2017; 41:252-75.
5. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect Dis 2018; 18:318-27.
6. Antimicrobial Resistance C. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 2022; 399:629-55.
7. Rosenblueth M, Martinez L, Silva J, Martinez-Romero E. Klebsiella variicola, a novel species with clinical and plant-associated isolates. Syst Appl Microbiol 2004; 27:27-35.
8. Brisse S, Passet V, Grimont PAD. Description of Klebsiella quasipneumoniae sp. nov., isolated from human infections, with two subspecies, Klebsiella quasipneumoniae subsp. quasipneumoniae subsp. nov. and Klebsiella quasipneumoniae subsp. similipneumoniae subsp. nov., and demonstration that Klebsiella singaporensis is a junior heterotypic synonym of Klebsiella variicola. Int J Syst Evol Microbiol 2014; 64:3146-52.
9. Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, et al. Whole-Genome Sequencing of a Human Clinical Isolate of the Novel Species Klebsiella quasivariicola sp. nov. Genome Announc 2017; 5.
10. Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, et al. Whole-Genome Sequencing of Human Clinical Klebsiella pneumoniae Isolates Reveals Misidentification and Misunderstandings of Klebsiella pneumoniae, Klebsiella variicola, and Klebsiella quasipneumoniae. mSphere 2017; 2.
11. Rodrigues C, Passet V, Rakotondrasoa A, Diallo TA, Criscuolo A, Brisse S. Description of Klebsiella africanensis sp. nov., Klebsiella variicola subsp. tropicalensis subsp. nov. and Klebsiella variicola subsp. variicola subsp. nov. Res Microbiol 2019; 170:165-70.
12. Fostervold A, Hetland MAK, Bakksjo R, Bernhoff E, Holt KE, Samuelsen O, et al. A nationwide genomic study of clinical Klebsiella pneumoniae in Norway 2001-15: introduction

and spread of ESBLs facilitated by clonal groups CG15 and CG307. J Antimicrob Chemother 2022; 77:665-74.

13. Maatallah M, Vading M, Kabir MH, Bakhrouf A, Kalin M, Naucler P, et al. 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:e113539.

14. Rodriguez-Medina N, Barrios-Camacho H, Duran-Bedolla J, Garza-Ramos U. Klebsiella variicola: an emerging pathogen in humans. Emerg Microbes Infect 2019; 8:973-88.

15. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, 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:E3574-81.

16. Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. J Clin Microbiol 2005; 43:4178-82.

17. Breurec S, Guessennd N, Timinouni M, Le TA, Cao V, Ngandjio A, et al. Klebsiella pneumoniae resistant to third-generation cephalosporins in five African and two Vietnamese major towns: multiclonal population structure with two major international clonal groups, CG15 and CG258. Clin Microbiol Infect 2013; 19:349-55.

18. Brisse S, Fevre C, Passet V, Issenhuth-Jeanjean S, Tournebize R, Diancourt L, et al. Virulent clones of Klebsiella pneumoniae: identification and evolutionary scenario based on genomic and phenotypic characterization. PLoS One 2009; 4:e4982.

19. Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Delannoy-Vieillard AS, et al. Genomic definition of hypervirulent and multidrug-resistant Klebsiella pneumoniae clonal groups. Emerg Infect Dis 2014; 20:1812-20.

20. Hennart M, Guglielmini J, Bridel S, Maiden MCJ, Jolley KA, Criscuolo A, et al. A Dual Barcoding Approach to Bacterial Strain Nomenclature: Genomic Taxonomy of Klebsiella pneumoniae Strains. Mol Biol Evol 2022; 39.

21. Meatherall BL, Gregson D, Ross T, Pitout JD, Laupland KB. Incidence, risk factors, and outcomes of Klebsiella pneumoniae bacteremia. Am J Med 2009; 122:866-73.

22. NORM/NORM-VET 2022. Usage of Antimicrobial Agents and Occurance of Antimicrobial Resistance in Norway. Tromsø / Oslo 2023. ISSN: 1502-2307 (print) / 1890-9965 (electronic).

23. Russo TA, Marr CM. Hypervirulent Klebsiella pneumoniae. Clin Microbiol Rev 2019; 32.

24. Shon AS, Bajwa RP, Russo TA. Hypervirulent (hypermucoviscous) Klebsiella pneumoniae: a new and dangerous breed. Virulence 2013; 4:107-18.

25. Siu LK, Yeh KM, Lin JC, Fung CP, Chang FY. Klebsiella pneumoniae liver abscess: a new invasive syndrome. Lancet Infect Dis 2012; 12:881-7.
26. Lam MMC, Wyres KL, Duchene S, Wick RR, Judd LM, Gan YH, et al. Population genomics of hypervirulent Klebsiella pneumoniae clonal-group 23 reveals early emergence and rapid global dissemination. Nat Commun 2018; 9:2703.
27. Siu LK, Fung CP, Chang FY, Lee N, Yeh KM, Koh TH, et al. Molecular typing and virulence analysis of serotype K1 Klebsiella pneumoniae strains isolated from liver abscess patients and stool samples from noninfectious subjects in Hong Kong, Singapore, and Taiwan. J Clin Microbiol 2011; 49:3761-5.
28. Chung DR, Lee SS, Lee HR, Kim HB, Choi HJ, Eom JS, et al. Emerging invasive liver abscess caused by K1 serotype Klebsiella pneumoniae in Korea. J Infect 2007; 54:578-83.
29. Wang JH, Liu YC, Lee SS, Yen MY, Chen YS, Wang JH, et al. Primary liver abscess due to Klebsiella pneumoniae in Taiwan. Clin Infect Dis 1998; 26:1434-8.
30. Qian Y, Wong CC, Lai S, Chen H, He X, Sun L, et al. A retrospective study of pyogenic liver abscess focusing on Klebsiella pneumoniae as a primary pathogen in China from 1994 to 2015. Sci Rep 2016; 6:38587.
31. Bachman MA, Oyler JE, Burns SH, Caza M, Lepine F, Dozois CM, et al. Klebsiella pneumoniae yersiniabactin promotes respiratory tract infection through evasion of lipocalin 2. Infect Immun 2011; 79:3309-16.
32. Russo TA, Olson R, Macdonald U, Metzger D, Maltese LM, Drake EJ, et al. Aerobactin mediates virulence and accounts for increased siderophore production under iron-limiting conditions by hypervirulent (hypermucoviscous) Klebsiella pneumoniae. Infect Immun 2014; 82:2356-67.
33. Fischbach MA, Lin H, Zhou L, Yu Y, Abergel RJ, Liu DR, et al. The pathogen-associated iroA gene cluster mediates bacterial evasion of lipocalin 2. Proc Natl Acad Sci U S A 2006; 103:16502-7.
34. Lu MC, Chen YT, Chiang MK, Wang YC, Hsiao PY, Huang YJ, et al. Colibactin Contributes to the Hypervirulence of pks(+) K1 CC23 Klebsiella pneumoniae in Mouse Meningitis Infections. Front Cell Infect Microbiol 2017; 7:103.
35. Hu D, Chen W, Wu J, Luo X, Yu L, Qu Y, et al. Coexistence of c-rmpA with p-rmpA and p-rmpA2 rather than excessive siderophores confers higher virulence in K1 Klebsiella pneumoniae. Pathology 2023; 55:1004-12.

36. Fang CT, Lai SY, Yi WC, Hsueh PR, Liu KL, Chang SC. Klebsiella pneumoniae genotype K1: an emerging pathogen that causes septic ocular or central nervous system complications from pyogenic liver abscess. Clin Infect Dis 2007; 45:284-93.
37. Yu WL, Chan KS, Ko WC, Lee CC, Chuang YC. Lower prevalence of diabetes mellitus in patients with Klebsiella pneumoniae primary liver abscess caused by isolates of K1/K2 than with non-K1/K2 capsular serotypes. Clin Infect Dis 2007; 45:1529-30; author reply 32-3.
38. Kabha K, Nissimov L, Athamna A, Keisari Y, Parolis H, Parolis LA, et al. Relationships among capsular structure, phagocytosis, and mouse virulence in Klebsiella pneumoniae. Infect Immun 1995; 63:847-52.
39. Lam MMC, Wyres KL, Judd LM, Wick RR, Jenney A, Brisse S, et al. Tracking key virulence loci encoding aerobactin and salmochelin siderophore synthesis in Klebsiella pneumoniae. Genome Med 2018; 10:77.
40. Lam MMC, Wick RR, Wyres KL, Gorrie CL, Judd LM, Jenney AWJ, et al. Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICEKp in Klebsiella pneumoniae populations. Microb Genom 2018; 4.
41. Struve C, Bojer M, Krogfelt KA. Identification of a conserved chromosomal region encoding Klebsiella pneumoniae type 1 and type 3 fimbriae and assessment of the role of fimbriae in pathogenicity. Infect Immun 2009; 77:5016-24.
42. 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. Infect Immun 2008; 76:4055-65.
43. Feriotti C, Sa-Pessoa J, Calderon-Gonzalez R, Gu L, Morris B, Sugisawa R, et al. Klebsiella pneumoniae hijacks the Toll-IL-1R protein SARM1 in a type I IFN-dependent manner to antagonize host immunity. Cell Rep 2022; 40:111167.
44. Bengoechea JA, Sa Pessoa J. Klebsiella pneumoniae infection biology: living to counteract host defences. Fems Microbiol Rev 2019; 43:123-44.
45. Domenico P, Salo RJ, Cross AS, Cunha BA. Polysaccharide capsule-mediated resistance to opsonophagocytosis in Klebsiella pneumoniae. Infect Immun 1994; 62:4495-9.
46. Cortes G, Borrell N, de Astorza B, Gomez C, Sauleda J, Alberti S. Molecular analysis of the contribution of the capsular polysaccharide and the lipopolysaccharide O side chain to the virulence of Klebsiella pneumoniae in a murine model of pneumonia. Infect Immun 2002; 70:2583-90.

47. Vornhagen J, Roberts EK, Unverdorben L, Mason S, Patel A, Crawford R, et al. Combined comparative genomics and clinical modeling reveals plasmid-encoded genes are independently associated with Klebsiella infection. Nat Commun 2022; 13:4459.
48. Paczosa M, J M, Email Mecsas J, Joan, Mecsas@tufts, edu. Klebsiella pneumoniae: Going on the offense with a strong defense. 2016.
49. 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.
50. Seki M, Gotoh K, Nakamura S, Akeda Y, Yoshii T, Miyaguchi S, et al. Fatal sepsis caused by an unusual Klebsiella species that was misidentified by an automated identification system. J Med Microbiol 2013; 62:801-3.
51. Breurec S, Melot B, Hoen B, Passet V, Schepers K, Bastian S, et al. Liver Abscess Caused by Infection with Community-Acquired Klebsiella quasipneumoniae subsp. quasipneumoniae. Emerg Infect Dis 2016; 22:529-31.
52. Zhang PJ, Lu ZH, Cao LJ, Chen H, Sun Y. Successful treatment of invasive liver abscess syndrome caused by Klebsiella variicola with intracranial infection and septic shock: A case report. World J Gastrointest Surg 2023; 15:2938-44.
53. Potter RF, Lainhart W, Twentyman J, Wallace MA, Wang B, Burnham CA, et al. Population Structure, Antibiotic Resistance, and Uropathogenicity of Klebsiella variicola. mBio 2018; 9.
54. Lu Y, Feng Y, McNally A, Zong Z. Occurrence of colistin-resistant hypervirulent Klebsiella variicola. J Antimicrob Chemother 2018; 73:3001-4.
55. Scheuerman O, Schechner V, Carmeli Y, Gutierrez-Gutierrez B, Calbo E, Almirante B, et al. Comparison of Predictors and Mortality Between Bloodstream Infections Caused by ESBL-Producing Escherichia coli and ESBL-Producing Klebsiella pneumoniae. Infect Control Hosp Epidemiol 2018; 39:660-7.
56. Xu L, Sun X, Ma X. Systematic review and meta-analysis of mortality of patients infected with carbapenem-resistant Klebsiella pneumoniae. Ann Clin Microbiol Antimicrob 2017; 16:18.
57. Cassini A, Hogberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. Lancet Infect Dis 2019; 19:56-66.

58. Lin XC, Li CL, Zhang SY, Yang XF, Jiang M. The Global and Regional Prevalence of Hospital-Acquired Carbapenem-Resistant Klebsiella pneumoniae Infection: A Systematic Review and Meta-analysis. Open Forum Infect Dis 2024; 11:ofad649.
59. Villa L, Feudi C, Fortini D, Brisse S, Passet V, Bonura C, et al. Diversity, virulence, and antimicrobial resistance of the KPC-producing Klebsiella pneumoniae ST307 clone. Microb Genom 2017; 3:e000110.
60. Wyres KL, Hawkey J, Hetland MAK, Fostervold A, Wick RR, Judd LM, et al. Emergence and rapid global dissemination of CTX-M-15-associated Klebsiella pneumoniae strain ST307. J Antimicrob Chemother 2019; 74:577-81.
61. Rodrigues C, Lanza VF, Peixe L, Coque TM, Novais A. Phylogenomics of Globally Spread Clonal Groups 14 and 15 of Klebsiella pneumoniae. Microbiol Spectr 2023; 11:e0339522.
62. Rettedal S, Lohr IH, Natas O, Giske CG, Sundsfjord A, Oymar K. First outbreak of extended-spectrum beta-lactamase-producing Klebsiella pneumoniae in a Norwegian neonatal intensive care unit; associated with contaminated breast milk and resolved by strict cohorting. APMIS 2012; 120:612-21.
63. Tofteland S, Naseer U, Lislevand JH, Sundsfjord A, Samuelsen O. A long-term low-frequency hospital outbreak of KPC-producing Klebsiella pneumoniae involving Intergenus plasmid diffusion and a persisting environmental reservoir. PLoS One 2013; 8:e59015.
64. Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect 2012; 18:268-81.
65. Wyres KL, Holt KE. Klebsiella pneumoniae Population Genomics and Antimicrobial-Resistant Clones. Trends Microbiol 2016; 24:944-56.
66. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, et al. Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. Nat Microbiol 2019; 4:1919-29.
67. David S, Cohen V, Reuter S, Sheppard AE, Giani T, Parkhill J, et al. Integrated chromosomal and plasmid sequence analyses reveal diverse modes of carbapenemase gene spread among Klebsiella pneumoniae. Proc Natl Acad Sci U S A 2020; 117:25043-54.
68. Leon-Sampedro R, DelaFuente J, Diaz-Agero C, Crellen T, Musicha P, Rodriguez-Beltran J, et al. Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. Nat Microbiol 2021; 6:606-16.

69. Barrios-Camacho H, Aguilar-Vera A, Beltran-Rojel M, Aguilar-Vera E, Duran-Bedolla J, Rodriguez-Medina N, et al. Molecular epidemiology of Klebsiella variicola obtained from different sources. Sci Rep 2019; 9:10610.

70. Mathers AJ, Crook D, Vaughan A, Barry KE, Vegesana K, Stoesser N, et al. Klebsiella quasipneumoniae Provides a Window into Carbapenemase Gene Transfer, Plasmid Rearrangements, and Patient Interactions with the Hospital Environment. Antimicrob Agents Chemother 2019; 63.

71. Papadimitriou-Olivgeris M, Bartzavali C, Georgakopoulou A, Kolonitsiou F, Papamichail C, Spiliopoulou I, et al. Mortality of Pandrug-Resistant Klebsiella pneumoniae Bloodstream Infections in Critically Ill Patients: A Retrospective Cohort of 115 Episodes. Antibiotics (Basel) 2021; 10.

72. Zhang R, Dong N, Huang Y, Zhou H, Xie M, Chan EW, et al. Evolution of tigecycline- and colistin-resistant CRKP (carbapenem-resistant Klebsiella pneumoniae) in vivo and its persistence in the GI tract. Emerg Microbes Infect 2018; 7:127.

73. de Man TJB, Lutgring JD, Lonsway DR, Anderson KF, Kiehlbauch JA, Chen L, et al. Genomic Analysis of a Pan-Resistant Isolate of Klebsiella pneumoniae, United States 2016. mBio 2018; 9.

74. Bartolleti F, Seco BM, Capuzzo Dos Santos C, Felipe CB, Lemo ME, Alves Tda S, et al. Polymyxin B Resistance in Carbapenem-Resistant Klebsiella pneumoniae, Sao Paulo, Brazil. Emerg Infect Dis 2016; 22:1849-51.

75. Gomes MZR, de Lima EM, Martins Aires CA, Pereira PS, Yim J, Silva FH, et al. Outbreak report of polymyxin-carbapenem-resistant Klebsiella pneumoniae causing untreatable infections evidenced by synergy tests and bacterial genomes. Sci Rep 2023; 13:6238.

76. Radisic V, Grevskott DH, Lunestad BT, Ovreas L, Marathe NP. Sewage-based surveillance shows presence of Klebsiella pneumoniae resistant against last resort antibiotics in the population in Bergen, Norway. Int J Hyg Environ Health 2023; 248:114075.

77. Schwarz S, Johnson AP. Transferable resistance to colistin: a new but old threat. J Antimicrob Chemother 2016; 71:2066-70.

78. Lan P, Lu Y, Chen Z, Wu X, Hua X, Jiang Y, et al. Emergence of High-Level Cefiderocol Resistance in Carbapenem-Resistant Klebsiella pneumoniae from Bloodstream Infections in Patients with Hematologic Malignancies in China. Microbiol Spectr 2022; 10:e0008422.

79. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL, et al. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of Klebsiella pneumoniae. PLoS Genet 2019; 15:e1008114.
80. Wyres KL, Nguyen TNT, Lam MMC, Judd LM, van Vinh Chau N, Dance DAB, et al. Genomic surveillance for hypervirulence and multi-drug resistance in invasive Klebsiella pneumoniae from South and Southeast Asia. Genome Med 2020; 12:11.
81. Lam MMC, Wyres KL, Wick RR, Judd LM, Fostervold A, Holt KE, et al. Convergence of virulence and MDR in a single plasmid vector in MDR Klebsiella pneumoniae ST15. J Antimicrob Chemother 2019; 74:1218-22.
82. Gu D, Dong N, Zheng Z, Lin D, Huang M, Wang L, et al. A fatal outbreak of ST11 carbapenem-resistant hypervirulent Klebsiella pneumoniae in a Chinese hospital: a molecular epidemiological study. Lancet Infect Dis 2018; 18:37-46.
83. Han YL, Wen XH, Zhao W, Cao XS, Wen JX, Wang JR, et al. Epidemiological characteristics and molecular evolution mechanisms of carbapenem-resistant hypervirulent Klebsiella pneumoniae. Front Microbiol 2022; 13:1003783.
84. European Centre for Disease Prevention and Control. Emergence of hypervirulent Klebsiella pneumoniae ST23 carrying carbapenemase genes in EU/EEA countries, first update. 14 February 2024. ECDC: Stockholm; 2024.
85. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, et al. Gastrointestinal Carriage Is a Major Reservoir of Klebsiella pneumoniae Infection in Intensive Care Patients. Clin Infect Dis 2017; 65:208-15.
86. Conlan S, Kong HH, Segre JA. Species-level analysis of DNA sequence data from the NIH Human Microbiome Project. PLoS One 2012; 7:e47075.
87. Huynh BT, Passet V, Rakotondrasoa A, Diallo T, Kerleguer A, Hennart M, et al. Klebsiella pneumoniae carriage in low-income countries: antimicrobial resistance, genomic diversity and risk factors. Gut Microbes 2020; 11:1287-99.
88. Lin YT, Siu LK, Lin JC, Chen TL, Tseng CP, Yeh KM, et al. Seroepidemiology of Klebsiella pneumoniae colonizing the intestinal tract of healthy Chinese and overseas Chinese adults in Asian countries. BMC Microbiol 2012; 12:13.
89. Raffelsberger N, Hetland MAK, Svendsen K, Smabrekke L, Lohr IH, Andreassen LLE, et al. Gastrointestinal carriage of Klebsiella pneumoniae in a general adult population: a cross-sectional study of risk factors and bacterial genomic diversity. Gut Microbes 2021; 13:1939599.
90. Maklin T, Thorpe HA, Pontinen AK, Gladstone RA, Shao Y, Pesonen M, et al. Strong pathogen competition in neonatal gut colonisation. Nat Commun 2022; 13:7417.

91. Dao TT, Liebenthal D, Tran TK, Ngoc Thi Vu B, Ngoc Thi Nguyen D, Thi Tran HK, et al. Klebsiella pneumoniae oropharyngeal carriage in rural and urban Vietnam and the effect of alcohol consumption. PLoS One 2014; 9:e91999.
92. Kloos WE, Musselwhite MS. Distribution and persistence of Staphylococcus and Micrococcus species and other aerobic bacteria on human skin. Appl Microbiol 1975; 30:381-5.
93. Martin RM, Cao J, Brisse S, Passet V, Wu W, Zhao L, et al. Molecular Epidemiology of Colonizing and Infecting Isolates of Klebsiella pneumoniae. mSphere 2016; 1.
94. Selden R, Lee S, Wang WL, Bennett JV, Eickhoff TC. Nosocomial klebsiella infections: intestinal colonization as a reservoir. Ann Intern Med 1971; 74:657-64.
95. Shimasaki T, Seekatz A, Bassis C, Rhee Y, Yelin RD, Fogg L, et al. 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. Clin Infect Dis 2019; 68:2053-9.
96. Sun Y, Patel A, SantaLucia J, Roberts E, Zhao L, Kaye K, et al. Measurement of Klebsiella Intestinal Colonization Density To Assess Infection Risk. mSphere 2021; 6:e0050021.
97. Lerner A, Adler A, Abu-Hanna J, Cohen Percia S, Kazma Matalon M, Carmeli Y. Spread of KPC-producing carbapenem-resistant Enterobacteriaceae: the importance of super-spreaders and rectal KPC concentration. Clin Microbiol Infect 2015; 21:470 e1-7.
98. Lepuschitz S, Hauser K, Schriebl A, Schlagenhaufen C, Stoger A, Chakeri A, et al. Fecal Klebsiella pneumoniae Carriage Is Intermittent and of High Clonal Diversity. Front Microbiol 2020; 11:581081.
99. 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; 68:1043-8.
100. Conlan S, Park M, Deming C, Thomas PJ, Young AC, Coleman H, et al. Plasmid Dynamics in KPC-Positive Klebsiella pneumoniae during Long-Term Patient Colonization. mBio 2016; 7.
101. Feldman N, Adler A, Molshatzki N, Navon-Venezia S, Khabra E, Cohen D, et al. Gastrointestinal colonization by KPC-producing Klebsiella pneumoniae following hospital discharge: duration of carriage and risk factors for persistent carriage. Clin Microbiol Infect 2013; 19:E190-6.

102. Lubbert C, Lippmann N, Busch T, Kaisers UX, Ducomble T, Eckmanns T, et al. Long-term carriage of Klebsiella pneumoniae carbapenemase-2-producing K pneumoniae after a large single-center outbreak in Germany. Am J Infect Control 2014; 42:376-80.
103. OstholmBalkhed A, Tarnberg M, Nilsson M, Nilsson LE, Hanberger H, Hallgren A, et al. Duration of travel-associated faecal colonisation with ESBL-producing Enterobacteriaceae - A one year follow-up study. PLoS One 2018; 13:e0205504.
104. Hetland MAK, Hawkey J, Bernhoff E, Bakksjo RJ, Kaspersen H, Rettedal SI, et al. Within-patient and global evolutionary dynamics of Klebsiella pneumoniae ST17. Microb Genom 2023; 9.
105. Merciecca T, Bornes S, Nakusi L, Theil S, Rendueles O, Forestier C, et al. Role of Klebsiella pneumoniae Type VI secretion system (T6SS) in long-term gastrointestinal colonization. Sci Rep 2022; 12:16968.
106. Favre-Bonte S, Licht TR, Forestier C, Krogfelt KA. Klebsiella pneumoniae capsule expression is necessary for colonization of large intestines of streptomycin-treated mice. Infect Immun 1999; 67:6152-6.
107. Opoku-Temeng C, Kobayashi SD, DeLeo FR. Klebsiella pneumoniae capsule polysaccharide as a target for therapeutics and vaccines. Comput Struct Biotechnol J 2019; 17:1360-6.
108. Vornhagen J, Bassis CM, Ramakrishnan S, Hein R, Mason S, Bergman Y, et al. A plasmid locus associated with Klebsiella clinical infections encodes a microbiome-dependent gut fitness factor. PLoS Pathog 2021; 17:e1009537.
109. Jung HJ, Littmann ER, Seok R, Leiner IM, Taur Y, Peled J, et al. Genome-Wide Screening for Enteric Colonization Factors in Carbapenem-Resistant ST258 Klebsiella pneumoniae. mBio 2019; 10.
110. Maroncle N, Balestrino D, Rich C, Forestier C. Identification of Klebsiella pneumoniae genes involved in intestinal colonization and adhesion using signature-tagged mutagenesis. Infect Immun 2002; 70:4729-34.
111. Cheung BH, Alisoltani A, Kochan TJ, Lebrun-Corbin M, Nozick SH, Axline CMR, et al. Genome-wide screens reveal shared and strain-specific genes that facilitate enteric colonization by Klebsiella pneumoniae. mBio 2023:e0212823.
112. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol 2016; 14:e1002533.
113. Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. Microbiome 2020; 8:103.

114. King CH, Desai H, Sylvetsky AC, LoTempio J, Ayanyan S, Carrie J, et al. Baseline human gut microbiota profile in healthy people and standard reporting template. PLoS One 2019; 14:e0206484.
115. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, et al. Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun 2020; 11:5206.
116. Hrncir T. Gut Microbiota Dysbiosis: Triggers, Consequences, Diagnostic and Therapeutic Options. Microorganisms 2022; 10.
117. Caballero-Flores G, Pickard JM, Nunez G. Microbiota-mediated colonization resistance: mechanisms and regulation. Nat Rev Microbiol 2023; 21:347-60.
118. Collingwood A, Blostein F, Seekatz AM, Wobus CE, Woods RJ, Foxman B, et al. Epidemiological and Microbiome Associations Between Klebsiella pneumoniae and Vancomycin-Resistant Enterococcus Colonization in Intensive Care Unit Patients. Open Forum Infect Dis 2020; 7:ofaa012.
119. Juhasz J, Ligeti B, Gajdacs M, Makra N, Ostorhazi E, Farkas FB, et al. Colonization Dynamics of Multidrug-Resistant Klebsiella pneumoniae Are Dictated by Microbiota-Cluster Group Behavior over Individual Antibiotic Susceptibility: A Metataxonomic Analysis. Antibiotics (Basel) 2021; 10.
120. Vazquez-Cuesta S, Villar L, Garcia NL, Fernandez AI, Olmedo M, Alcala L, et al. Characterization of the gut microbiome of patients with Clostridioides difficile infection, patients with non-C. difficile diarrhea, and C. difficile-colonized patients. Front Cell Infect Microbiol 2023; 13:1130701.
121. Spragge F, Bakkeren E, Jahn MT, E BNA, Pearson CF, Wang X, et al. Microbiome diversity protects against pathogens by nutrient blocking. Science 2023; 382:eadj3502.
122. Calderon-Gonzalez R, Lee A, Lopez-Campos G, Hancock SJ, Sa-Pessoa J, Dumigan A, et al. Modelling the Gastrointestinal Carriage of Klebsiella pneumoniae Infections. mBio 2023; 14:e0312122.
123. Wu T, Xu F, Su C, Li H, Lv N, Liu Y, et al. Alterations in the Gut Microbiome and Cecal Metabolome During Klebsiella pneumoniae-Induced Pneumosepsis. Front Immunol 2020; 11:1331.
124. Vieira AT, Rocha VM, Tavares L, Garcia CC, Teixeira MM, Oliveira SC, et al. Control of Klebsiella pneumoniae pulmonary infection and immunomodulation by oral treatment with the commensal probiotic Bifidobacterium longum 5(1A). Microbes Infect 2016; 18:180-9.

125. Zheng Y, Ding Y, Xu M, Chen H, Zhang H, Liu Y, et al. Gut Microbiota Contributes to Host Defense Against Klebsiella pneumoniae-Induced Liver Abscess. J Inflamm Res 2021; 14:5215-25.
126. Sequeira RP, McDonald JAK, Marchesi JR, Clarke TB. Commensal Bacteroidetes protect against Klebsiella pneumoniae colonization and transmission through IL-36 signalling. Nat Microbiol 2020; 5:304-13.
127. Osbelt L, Wende M, Almasi E, Derksen E, Muthukumarasamy U, Lesker TR, et al. Klebsiella oxytoca causes colonization resistance against multidrug-resistant K. pneumoniae in the gut via cooperative carbohydrate competition. Cell Host Microbe 2021; 29:1663-79 e7.
128. Xiong RG, Zhou DD, Wu SX, Huang SY, Saimaiti A, Yang ZJ, et al. Health Benefits and Side Effects of Short-Chain Fatty Acids. Foods 2022; 11.
129. Mirzaei R, Dehkhodaie E, Bouzari B, Rahimi M, Gholestani A, Hosseini-Fard SR, et al. Dual role of microbiota-derived short-chain fatty acids on host and pathogen. Biomed Pharmacother 2022; 145:112352.
130. Chang KC, Nagarajan N, Gan YH. Short-chain fatty acids of various lengths differentially inhibit Klebsiella pneumoniae and Enterobacteriaceae species. mSphere 2024:e0078123.
131. Rivera-Chavez F, Lopez CA, Baumler AJ. Oxygen as a driver of gut dysbiosis. Free Radic Biol Med 2017; 105:93-101.
132. Ney LM, Wipplinger M, Grossmann M, Engert N, Wegner VD, Mosig AS. Short chain fatty acids: key regulators of the local and systemic immune response in inflammatory diseases and infections. Open Biol 2023; 13:230014.
133. Perez-Reytor D, Puebla C, Karahanian E, Garcia K. Use of Short-Chain Fatty Acids for the Recovery of the Intestinal Epithelial Barrier Affected by Bacterial Toxins. Front Physiol 2021; 12:650313.
134. Roson-Calero N, Balleste-Delpierre C, Fernandez J, Vila J. Insights on Current Strategies to Decolonize the Gut from Multidrug-Resistant Bacteria: Pros and Cons. Antibiotics (Basel) 2023; 12.
135. Mascolo A, Carannante N, Mauro GD, Sarno M, Costanzo M, Licciardi F, et al. Decolonization of drug-resistant Enterobacteriaceae carriers: A scoping review of the literature. J Infect Public Health 2023; 16:376-83.
136. Campos-Madueno EI, Moradi M, Eddoubaji Y, Shahi F, Moradi S, Bernasconi OJ, et al. Intestinal colonization with multidrug-resistant Enterobacterales: screening, epidemiology,

clinical impact, and strategies to decolonize carriers. Eur J Clin Microbiol Infect Dis 2023; 42:229-54.

137. Wittekamp BHJ, Oostdijk EAN, Cuthbertson BH, Brun-Buisson C, Bonten MJM. Selective decontamination of the digestive tract (SDD) in critically ill patients: a narrative review. Intensive Care Med 2020; 46:343-9.

138. Huttner B, Haustein T, Uckay I, Renzi G, Stewardson A, Schaerrer D, et al. Decolonization of intestinal carriage of extended-spectrum beta-lactamase-producing Enterobacteriaceae with oral colistin and neomycin: a randomized, double-blind, placebo-controlled trial. J Antimicrob Chemother 2013; 68:2375-82.

139. Bar-Yoseph H, Hussein K, Braun E, Paul M. Natural history and decolonization strategies for ESBL/carbapenem-resistant Enterobacteriaceae carriage: systematic review and meta-analysis. J Antimicrob Chemother 2016; 71:2729-39.

140. Kim K, M G, NameId Gluck MORCIDho, org, Email Gluck M, Michael, et al. Fecal microbiota transplantation: An update on clinical practice. 2019.

141. Macareno-Castro J, Solano-Salazar A, Dong LT, Mohiuddin M, Espinoza JL. Fecal microbiota transplantation for Carbapenem-Resistant Enterobacteriaceae: A systematic review. J Infect 2022; 84:749-59.

142. Hung YP, Lee CC, Lee JC, Tsai PJ, Hsueh PR, Ko WC. The Potential of Probiotics to Eradicate Gut Carriage of Pathogenic or Antimicrobial-Resistant Enterobacterales. Antibiotics (Basel) 2021; 10.

143. Meyer MP, Chow SSW, Alsweiler J, Bourchier D, Broadbent R, Knight D, et al. Probiotics for Prevention of Severe Necrotizing Enterocolitis: Experience of New Zealand Neonatal Intensive Care Units. Front Pediatr 2020; 8:119.

144. Karbalaei M, Keikha M. Probiotics and intestinal decolonization of antibiotic-resistant microorganisms; A reality or fantasy? Ann Med Surg (Lond) 2022; 80:104269.

145. Bonnet M, Eckert C, Tournebize R. Decolonization of asymptomatic carriage of multi-drug resistant bacteria by bacteriophages? Front Microbiol 2023; 14:1266416.

146. Fang Q, Yin X, He Y, Feng Y, Zhang L, Luo H, et al. Safety and efficacy of phage application in bacterial decolonisation: a systematic review. The Lancet Microbe.

147. Tacconelli E, Mazzaferri F, de Smet AM, Bragantini D, Eggimann P, Huttner BD, et al. ESCMID-EUCIC clinical guidelines on decolonization of multidrug-resistant Gram-negative bacteria carriers. Clin Microbiol Infect 2019; 25:807-17.

148. Brink AJ, Coetzee J, Corcoran C, Clay CG, Hari-Makkan D, Jacobson RK, et al. Emergence of OXA-48 and OXA-181 carbapenemases among Enterobacteriaceae in South

Africa and evidence of in vivo selection of colistin resistance as a consequence of selective decontamination of the gastrointestinal tract. J Clin Microbiol 2013; 51:369-72.
149. Mackenzie JS, Jeggo M. The One Health Approach-Why Is It So Important? Trop Med Infect Dis 2019; 4.
150. Barbier E, Rodrigues C, Depret G, Passet V, Gal L, Piveteau P, et al. The ZKIR Assay, a novel Real-Time PCR Method for the Detection of Klebsiella pneumoniae and Closely Related Species in Environmental Samples. Appl Environ Microbiol 2020.
151. Hakonsholm F, Hetland MAK, Svanevik CS, Lunestad BT, Lohr 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. Int J Hyg Environ Health 2022; 242:113967.
152. Davies YM, Cunha MP, Oliveira MG, Oliveira MC, Philadelpho N, Romero DC, et al. Virulence and antimicrobial resistance of Klebsiella pneumoniae isolated from passerine and psittacine birds. Avian Pathol 2016; 45:194-201.
153. Chen CM, Tang HL, Chiou CS, Tung KC, Lu MC, Lai YC. Colonization dynamics of Klebsiella pneumoniae in the pet animals and human owners in a single household. Vet Microbiol 2021; 256:109050.
154. Blin C, Passet V, Touchon M, Rocha EPC, Brisse S. Metabolic diversity of the emerging pathogenic lineages of Klebsiella pneumoniae. Environ Microbiol 2017; 19:1881-98.
155. Jacoby GA, Medeiros AA. More extended-spectrum beta-lactamases. Antimicrob Agents Chemother 1991; 35:1697-704.
156. Martinez-Martinez L, Pascual A, Jacoby GA. Quinolone resistance from a transferable plasmid. Lancet 1998; 351:797-9.
157. Jacoby GA, Walsh KE, Mills DM, Walker VJ, Oh H, Robicsek A, et al. qnrB, another plasmid-mediated gene for quinolone resistance. Antimicrob Agents Chemother 2006; 50:1178-82.
158. Smith Moland E, Hanson ND, Herrera VL, Black JA, Lockhart TJ, Hossain A, et al. Plasmid-mediated, carbapenem-hydrolysing beta-lactamase, KPC-2, in Klebsiella pneumoniae isolates. J Antimicrob Chemother 2003; 51:711-4.
159. Yong D, Toleman MA, Giske CG, Cho HS, Sundman K, Lee K, et al. Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. Antimicrob Agents Chemother 2009; 53:5046-54.

160. Poirel L, Heritier C, Tolun V, Nordmann P. Emergence of oxacillinase-mediated resistance to imipenem in Klebsiella pneumoniae. Antimicrob Agents Chemother 2004; 48:15-22.
161. Villegas MV, Lolans K, Correa A, Kattan JN, Lopez JA, Quinn JP, et al. First identification of Pseudomonas aeruginosa isolates producing a KPC-type carbapenem-hydrolyzing beta-lactamase. Antimicrob Agents Chemother 2007; 51:1553-5.
162. Robledo IE, Aquino EE, Sante MI, Santana JL, Otero DM, Leon CF, et al. Detection of KPC in Acinetobacter spp. in Puerto Rico. Antimicrob Agents Chemother 2010; 54:1354-7.
163. Canton R, Gonzalez-Alba JM, Galan JC. CTX-M Enzymes: Origin and Diffusion. Front Microbiol 2012; 3:110.
164. Poirel L, Rodriguez-Martinez JM, Mammeri H, Liard A, Nordmann P. Origin of plasmid-mediated quinolone resistance determinant QnrA. Antimicrob Agents Chemother 2005; 49:3523-5.
165. Tacao M, Araujo S, Vendas M, Alves A, Henriques I. Shewanella species as the origin of bla(OXA-48) genes: insights into gene diversity, associated phenotypes and possible transfer mechanisms. Int J Antimicrob Agents 2018; 51:340-8.
166. Ludden C, Moradigaravand D, Jamrozy D, Gouliouris T, Blane B, Naydenova P, et al. A One Health Study of the Genetic Relatedness of Klebsiella pneumoniae and Their Mobile Elements in the East of England. Clin Infect Dis 2020; 70:219-26.
167. Thorpe HA, Booton R, Kallonen T, Gibbon MJ, Couto N, Passet V, et al. A large-scale genomic snapshot of Klebsiella spp. isolates in Northern Italy reveals limited transmission between clinical and non-clinical settings. Nat Microbiol 2022; 7:2054-67.
168. Calland JK, Haukka K, Kpordze SW, Brusah A, Corbella M, Merla C, et al. Population structure and antimicrobial resistance among Klebsiella isolates sampled from human, animal, and environmental sources in Ghana: a cross-sectional genomic One Health study. Lancet Microbe 2023; 4:e943-e52.
169. Dereeper A, Gruel G, Pot M, Couvin D, Barbier E, Bastian S, et al. Limited Transmission of Klebsiella pneumoniae among Humans, Animals, and the Environment in a Caribbean Island, Guadeloupe (French West Indies). Microbiol Spectr 2022; 10:e0124222.
170. Morgado S, Fonseca E, Vicente AC. Genomics of Klebsiella pneumoniae Species Complex Reveals the Circulation of High-Risk Multidrug-Resistant Pandemic Clones in Human, Animal, and Environmental Sources. Microorganisms 2022; 10.
171. Cocker D, Chidziwisano K, Mphasa M, Mwapasa T, Lewis JM, Rowlingson B, et al. Investigating One Health risks for human colonisation with extended spectrum beta-lactamase-

producing Escherichia coli and Klebsiella pneumoniae in Malawian households: a longitudinal cohort study. Lancet Microbe 2023; 4:e534-e43.
172. Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA, Robinson TP, et al. Global trends in antimicrobial use in food animals. Proc Natl Acad Sci U S A 2015; 112:5649-54.
173. Wang Y, Xu C, Zhang R, Chen Y, Shen Y, Hu F, et al. Changes in colistin resistance and mcr-1 abundance in Escherichia coli of animal and human origins following the ban of colistin-positive additives in China: an epidemiological comparative study. Lancet Infect Dis 2020; 20:1161-71.
174. Liu MC, Jian Z, Liu W, Li J, Pei N. One Health Analysis of mcr-Carrying Plasmids and Emergence of mcr-10.1 in Three Species of Klebsiella Recovered from Humans in China. Microbiol Spectr 2022; 10:e0230622.
175. Pfaller MA, Herwaldt LA. The clinical microbiology laboratory and infection control: emerging pathogens, antimicrobial resistance, and new technology. Clin Infect Dis 1997; 25:858-70.
176. Fournier PE, Drancourt M, Colson P, Rolain JM, La Scola B, Raoult D. Modern clinical microbiology: new challenges and solutions. Nat Rev Microbiol 2013; 11:574-85.
177. van Belkum A, Durand G, Peyret M, Chatellier S, Zambardi G, Schrenzel J, et al. Rapid clinical bacteriology and its future impact. Ann Lab Med 2013; 33:14-27.
178. Gajic I, Kabic J, Kekic D, Jovicevic M, Milenkovic M, Mitic Culafic D, et al. Antimicrobial Susceptibility Testing: A Comprehensive Review of Currently Used Methods. Antibiotics (Basel) 2022; 11.
179. Lagier JC, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 2015; 28:208-36.
180. Forbes JD, Knox NC, Ronholm J, Pagotto F, Reimer A. Metagenomics: The Next Culture-Independent Game Changer. Front Microbiol 2017; 8:1069.
181. Langley G, Besser J, Iwamoto M, Lessa FC, Cronquist A, Skoff TH, et al. Effect of Culture-Independent Diagnostic Tests on Future Emerging Infections Program Surveillance. Emerg Infect Dis 2015; 21:1582-8.
182. Bonnet M, Lagier JC, Raoult D, Khelaifia S. Bacterial culture through selective and non-selective conditions: the evolution of culture media in clinical microbiology. New Microbes New Infect 2020; 34:100622.

183. Perry JD. A Decade of Development of Chromogenic Culture Media for Clinical Microbiology in an Era of Molecular Diagnostics. Clin Microbiol Rev 2017; 30:449-79.
184. Giske CG, Gezelius L, Samuelsen O, Warner M, Sundsfjord A, Woodford N. A sensitive and specific phenotypic assay for detection of metallo-beta-lactamases and KPC in Klebsiella pneumoniae with the use of meropenem disks supplemented with aminophenylboronic acid, dipicolinic acid and cloxacillin. Clin Microbiol Infect 2011; 17:552-6.
185. Dortet L, Poirel L, Nordmann P. Rapid identification of carbapenemase types in Enterobacteriaceae and Pseudomonas spp. by using a biochemical test. Antimicrob Agents Chemother 2012; 56:6437-40.
186. Papagiannitsis CC, Studentova V, Izdebski R, Oikonomou O, Pfeifer Y, Petinaki E, et al. Matrix-assisted laser desorption ionization-time of flight mass spectrometry meropenem hydrolysis assay with NH4HCO3, a reliable tool for direct detection of carbapenemase activity. J Clin Microbiol 2015; 53:1731-5.
187. Wareham DW, Shah R, Betts JW, Phee LM, Momin MH. Evaluation of an Immunochromatographic Lateral Flow Assay (OXA-48 K-SeT) for Rapid Detection of OXA-48-Like Carbapenemases in Enterobacteriaceae. J Clin Microbiol 2016; 54:471-3.
188. Giske CG, Turnidge J, Canton R, Kahlmeter G, Committee ES. Update from the European Committee on Antimicrobial Susceptibility Testing (EUCAST). J Clin Microbiol 2022; 60:e0027621.
189. Humphries RM, Abbott AN, Hindler JA. Understanding and Addressing CLSI Breakpoint Revisions: a Primer for Clinical Laboratories. J Clin Microbiol 2019; 57.
190. Tabak YP, Vankeepuram L, Ye G, Jeffers K, Gupta V, Murray PR. Blood Culture Turnaround Time in U.S. Acute Care Hospitals and Implications for Laboratory Process Optimization. J Clin Microbiol 2018; 56.
191. Doern GV. Detection of selected fastidious bacteria. Clin Infect Dis 2000; 30:166-73.
192. Vartoukian SR. Cultivation strategies for growth of uncultivated bacteria. J Oral Biosci 2016; 58:142-9.
193. Peker N, Couto N, Sinha B, Rossen JW. Diagnosis of bloodstream infections from positive blood cultures and directly from blood samples: recent developments in molecular approaches. Clin Microbiol Infect 2018; 24:944-55.
194. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med 2006; 34:1589-96.

195. Cummings LA, Hoogestraat DR, Rassoulian-Barrett SL, Rosenthal CA, Salipante SJ, Cookson BT, et al. Comprehensive evaluation of complex polymicrobial specimens using next generation sequencing and standard microbiological culture. Sci Rep 2020; 10:5446.
196. Rhoads DD, Wolcott RD, Sun Y, Dowd SE. Comparison of culture and molecular identification of bacteria in chronic wounds. Int J Mol Sci 2012; 13:2535-50.
197. Kirkup BC. Culture-independence for surveillance and epidemiology. Pathogens 2013; 2:556-70.
198. Singh K, Mangold KA, Wyant K, Schora DM, Voss B, Kaul KL, et al. Rectal screening for Klebsiella pneumoniae carbapenemases: comparison of real-time PCR and culture using two selective screening agar plates. J Clin Microbiol 2012; 50:2596-600.
199. Liu Q, Jin X, Cheng J, Zhou H, Zhang Y, Dai Y. Advances in the application of molecular diagnostic techniques for the detection of infectious disease pathogens (Review). Mol Med Rep 2023; 27.
200. Gerace E, Mancuso G, Midiri A, Poidomani S, Zummo S, Biondo C. Recent Advances in the Use of Molecular Methods for the Diagnosis of Bacterial Infections. Pathogens 2022; 11.
201. Yang S, Rothman RE. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. Lancet Infect Dis 2004; 4:337-48.
202. Schoonbroodt S, Ichante JL, Boffe S, Devos N, Devaster JM, Taddei L, et al. Real-time PCR has advantages over culture-based methods in identifying major airway bacterial pathogens in chronic obstructive pulmonary disease: Results from three clinical studies in Europe and North America. Front Microbiol 2022; 13:1098133.
203. Mitton B, Rule R, Said M. Laboratory evaluation of the BioFire FilmArray Pneumonia plus panel compared to conventional methods for the identification of bacteria in lower respiratory tract specimens: a prospective cross-sectional study from South Africa. Diagn Microbiol Infect Dis 2021; 99:115236.
204. Rule R, Paruk F, Becker P, Neuhoff M, Chausse J, Said M. Clinical utility of the BioFire FilmArray Blood Culture Identification panel in the adjustment of empiric antimicrobial therapy in the critically ill septic patient. PLoS One 2021; 16:e0254389.
205. Domingues RB, Santos MVD, Leite F, Senne C. FilmArray Meningitis/Encephalitis (ME) panel in the diagnosis of bacterial meningitis. Braz J Infect Dis 2019; 23:468-70.
206. Espy MJ, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Vetter EA, et al. Real-time PCR in clinical microbiology: applications for routine laboratory testing. Clin Microbiol Rev 2006; 19:165-256.

207. Kralik P, Ricchi M. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. Front Microbiol 2017; 8:108.
208. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. Nat Protoc 2006; 1:1559-82.
209. Borst A, Box AT, Fluit AC. False-positive results and contamination in nucleic acid amplification assays: suggestions for a prevent and destroy strategy. Eur J Clin Microbiol Infect Dis 2004; 23:289-99.
210. Charnock C, Samuelsen O, Nordlie AL, Hjeltnes B. Use of a Commercially Available Microarray to Characterize Antibiotic-Resistant Clinical Isolates of Klebsiella pneumoniae. Curr Microbiol 2018; 75:163-72.
211. Ma X, Li Y, Liang Y, Liu Y, Yu L, Li C, et al. Development of a DNA microarray assay for rapid detection of fifteen bacterial pathogens in pneumonia. BMC Microbiol 2020; 20:177.
212. Martinez MA, Soto-Del Rio Mde L, Gutierrez RM, Chiu CY, Greninger AL, Contreras JF, et al. DNA microarray for detection of gastrointestinal viruses. J Clin Microbiol 2015; 53:136-45.
213. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Hum Immunol 2021; 82:801-11.
214. Hilt EE, Ferrieri P. Next Generation and Other Sequencing Technologies in Diagnostic Microbiology and Infectious Diseases. Genes (Basel) 2022; 13.
215. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 2012; 13:341.
216. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 2020; 21:30.
217. Tang P, Croxen MA, Hasan MR, Hsiao WWL, Hoang LM. Infection control in the new age of genomic epidemiology. Am J Infect Control 2017; 45:170-9.
218. Jeon SA, Park JL, Park SJ, Kim JH, Goh SH, Han JY, et al. Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. Genes Genomics 2021; 43:713-24.
219. Kumar KR, Cowley MJ, Davis RL. Next-Generation Sequencing and Emerging Technologies. Semin Thromb Hemost 2019; 45:661-73.
220. Anslan S, Mikryukov V, Armolaitis K, Ankuda J, Lazdina D, Makovskis K, et al. Highly comparable metabarcoding results from MGI-Tech and Illumina sequencing platforms. PeerJ 2021; 9:e12254.

221. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 2010; 327:78-81.
222. Xu Y, Lin Z, Tang C, Tang Y, Cai Y, Zhong H, et al. A new massively parallel nanoball sequencing platform for whole exome research. BMC Bioinformatics 2019; 20:153.
223. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. Hum Mol Genet 2018; 27:R234-R41.
224. Taylor TL, Volkening JD, DeJesus E, Simmons M, Dimitrov KM, Tillman GE, et al. Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology. Sci Rep 2019; 9:16350.
225. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature 2016; 530:228-32.
226. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet 2018; 19:9-20.
227. Zhao W, Zeng W, Pang B, Luo M, Peng Y, Xu J, et al. Oxford nanopore long-read sequencing enables the generation of complete bacterial and plasmid genomes without short-read sequencing. Front Microbiol 2023; 14:1179966.
228. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sorensen EA, Wollenberg RD, et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. Nat Methods 2022; 19:823-6.
229. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. Pathology 2015; 47:199-210.
230. Ladner JT, Sahl JW. Towards a post-pandemic future for global pathogen genome sequencing. PLoS Biol 2023; 21:e3002225.
231. Lo SW, Jamrozy D. Genomics and epidemiological surveillance. Nat Rev Microbiol 2020; 18:478.
232. AMR NGHRUoGSo. Whole-genome sequencing as part of national and international surveillance programmes for antimicrobial resistance: a roadmap. BMJ Glob Health 2020; 5.
233. Croucher NJ, Didelot X. The application of genomics to tracing bacterial pathogen transmission. Curr Opin Microbiol 2015; 23:62-7.
234. Carlos CC, Masim MAL, Lagrada ML, Gayeta JM, Macaranas PKV, Sia SB, et al. Genome Sequencing Identifies Previously Unrecognized Klebsiella pneumoniae Outbreaks in Neonatal Intensive Care Units in the Philippines. Clin Infect Dis 2021; 73:S316-S24.

235. Benulic K, Pirs M, Couto N, Chlebowicz M, Rossen JWA, Zorec TM, et al. Whole genome sequencing characterization of Slovenian carbapenem-resistant Klebsiella pneumoniae, including OXA-48 and NDM-1 producing outbreak isolates. PLoS One 2020; 15:e0231503.

236. Zhou K, Lokate M, Deurenberg RH, Tepper M, Arends JP, Raangs EG, et al. Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum beta-lactamase producing ST15 Klebsiella pneumoniae. Sci Rep 2016; 6:20840.

237. Ko KKK, Chng KR, Nagarajan N. Metagenomics-enabled microbial surveillance. Nat Microbiol 2022; 7:486-96.

238. Batool M, Galloway-Pena J. Clinical metagenomics-challenges and future prospects. Front Microbiol 2023; 14:1186424.

239. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 2017; 35:833-44.

240. van Dijk LR, Walker BJ, Straub TJ, Worby CJ, Grote A, Schreiber HLt, et al. StrainGE: a toolkit to track and characterize low-abundance strains in complex microbial communities. Genome Biol 2022; 23:74.

241. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G, Professional Practice C, et al. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. Arch Pathol Lab Med 2017; 141:776-86.

242. Wang Y, Chen T, Zhang S, Zhang L, Li Q, Lv Q, et al. Clinical evaluation of metagenomic next-generation sequencing in unbiased pathogen diagnosis of urinary tract infection. J Transl Med 2023; 21:762.

243. Wilson MR, Shanbhag NM, Reid MJ, Singhal NS, Gelfand JM, Sample HA, et al. Diagnosing Balamuthia mandrillaris Encephalitis With Metagenomic Deep Sequencing. Ann Neurol 2015; 78:722-30.

244. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med 2014; 370:2408-17.

245. Sun Y, Li H, Chen J, Ma Z, Han P, Liu Y, et al. Case Report: Metagenomics Next-Generation Sequencing Can Be Performed for the Diagnosis of Disseminated Mucormycosis. Front Med (Lausanne) 2021; 8:675030.

246. Li X, Liang S, Zhang D, He M, Zhang H. The clinical application of metagenomic next-generation sequencing in sepsis of immunocompromised patients. Front Cell Infect Microbiol 2023; 13:1170687.
247. Liu Q, Liu X, Hu B, Xu H, Sun R, Li P, et al. Diagnostic performance and clinical impact of blood metagenomic next-generation sequencing in ICU patients suspected monomicrobial and polymicrobial bloodstream infections. Front Cell Infect Microbiol 2023; 13:1192931.
248. Leggett RM, Alcon-Giner C, Heavens D, Caim S, Brook TC, Kujawska M, et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. Nat Microbiol 2020; 5:430-42.
249. Fresia P, Antelo V, Salazar C, Gimenez M, D'Alessandro B, Afshinnekoo E, et al. Urban metagenomics uncover antibiotic resistance reservoirs in coastal beach and sewage waters. Microbiome 2019; 7:35.
250. Zhao Y, Su JQ, An XL, Huang FY, Rensing C, Brandt KK, et al. Feed additives shift gut microbiota and enrich antibiotic resistance in swine gut. Sci Total Environ 2018; 621:1224-32.
251. Yang X, Noyes NR, Doster E, Martin JN, Linke LM, Magnuson RJ, et al. Use of Metagenomic Shotgun Sequencing Technology To Detect Foodborne Pathogens within the Microbiome of the Beef Production Chain. Appl Environ Microbiol 2016; 82:2433-43.
252. Rowe WPM, Baker-Austin C, Verner-Jeffreys DW, Ryan JJ, Micallef C, Maskell DJ, et al. Overexpression of antibiotic resistance genes in hospital effluents over time. J Antimicrob Chemother 2017; 72:1617-23.
253. Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, Lukjancenko O, et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat Commun 2019; 10:1124.
254. Tang J, Bu Y, Zhang XX, Huang K, He X, Ye L, et al. Metagenomic analysis of bacterial community composition and antibiotic resistance genes in a wastewater treatment plant and its receiving surface water. Ecotoxicol Environ Saf 2016; 132:260-9.
255. Li X, Wu Z, Dang C, Zhang M, Zhao B, Cheng Z, et al. A metagenomic-based method to study hospital air dust resistome. Chem Eng J 2021; 406:126854.
256. Lax S, Sangwan N, Smith D, Larsen P, Handley KM, Richardson M, et al. Bacterial colonization and succession in a newly opened hospital. Sci Transl Med 2017; 9.

257. Brooks B, Olm MR, Firek BA, Baker R, Thomas BC, Morowitz MJ, et al. Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. Nat Commun 2017; 8:1814.
258. Afzaal M, Saeed F, Shah YA, Hussain M, Rabail R, Socol CT, et al. Human gut microbiota in health and disease: Unveiling the relationship. Front Microbiol 2022; 13:999001.
259. Khan I, Bai Y, Zha L, Ullah N, Ullah H, Shah SRH, et al. Mechanism of the Gut Microbiota Colonization Resistance and Enteric Pathogen Infection. Front Cell Infect Microbiol 2021; 11:716299.
260. Ebinger A, Fischer S, Hoper D. A theoretical and generalized approach for the assessment of the sample-specific limit of detection for clinical metagenomics. Comput Struct Biotechnol J 2021; 19:732-42.
261. Gweon HS, Shaw LP, Swann J, De Maio N, AbuOun M, Niehus R, et al. The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. Environ Microbiome 2019; 14:7.
262. Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, et al. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Rep 2018; 8:5890.
263. Guitor AK, Raphenya AR, Klunk J, Kuch M, Alcock B, Surette MG, et al. Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes. Antimicrob Agents Chemother 2019; 64.
264. Kok CR, Mulakken N, Thissen JB, Grey SF, Avila-Herrera A, Upadhyay MM, et al. Targeted metagenomic assessment reflects critical colonization in battlefield injuries. Microbiol Spectr 2023; 11:e0252023.
265. Kerkvliet JJ, Bossers A, Kers JG, Meneses R, Willems R, Schurch AC. Metagenomic assembly is the main bottleneck in the identification of mobile genetic elements. PeerJ 2024; 12:e16695.
266. Martinez JL, Coque TM, Baquero F. What is a resistance gene? Ranking risk in resistomes. Nat Rev Microbiol 2015; 13:116-23.
267. Stalder T, Press MO, Sullivan S, Liachko I, Top EM. Linking the resistome and plasmidome to the microbiome. ISME J 2019; 13:2437-46.
268. Eggen AE, Mathiesen EB, Wilsgaard T, Jacobsen BK, Njolstad I. The sixth survey of the Tromso Study (Tromso 6) in 2007-08: collaborative research in the interface between clinical medicine and epidemiology: study objectives, design, data collection procedures, and

attendance in a multipurpose population-based health survey. Scand J Public Health 2013; 41:65-80.

269. Ravi A, Halstead FD, Bamford A, Casey A, Thomson NM, van Schaik W, et al. Loss of microbial diversity and pathogen domination of the gut microbiota in critically ill patients. Microb Genom 2019; 5.

270. Mukhopadhya I, Martin JC, Shaw S, McKinley AJ, Gratz SW, Scott KP. Comparison of microbial signatures between paired faecal and rectal biopsy samples from healthy volunteers using next-generation sequencing and culturomics. Microbiome 2022; 10:171.

271. Sun S, Zhu X, Huang X, Murff HJ, Ness RM, Seidner DL, et al. On the robustness of inference of association with the gut microbiota in stool, rectal swab and mucosal tissue samples. Sci Rep 2021; 11:14828.

272. Tang Q, Jin G, Wang G, Liu T, Liu X, Wang B, et al. Current Sampling Methods for Gut Microbiota: A Call for More Precise Devices. Front Cell Infect Microbiol 2020; 10:151.

273. Van Horn KG, Audette CD, Sebeck D, Tucker KA. Comparison of the Copan ESwab system with two Amies agar swab transport systems for maintenance of microorganism viability. J Clin Microbiol 2008; 46:1655-8.

274. Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MCW, et al. Impact of Preservation Method and 16S rRNA Hypervariable Region on Gut Microbiota Profiling. mSystems 2019; 4.

275. 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; 20:936-41.

276. Poussin C, Sierro N, Boue S, Battey J, Scotti E, Belcastro V, et al. Interrogating the microbiome: experimental and computational considerations in support of study reproducibility. Drug Discov Today 2018; 23:1644-57.

277. Videnska P, Smerkova K, Zwinsova B, Popovici V, Micenkova L, Sedlar K, et al. Stool sampling and DNA isolation kits affect DNA quality and bacterial composition following 16S rRNA gene sequencing using MiSeq Illumina platform. Sci Rep 2019; 9:13837.

278. Fiedorova K, Radvansky M, Nemcova E, Grombirikova H, Bosak J, Cernochova M, et al. The Impact of DNA Extraction Methods on Stool Bacterial and Fungal Microbiota Community Recovery. Front Microbiol 2019; 10:821.

279. Rodrigues C, Hauser K, Cahill N, Ligowska-Marzeta M, Centorotola G, Cornacchia A, et al. High Prevalence of Klebsiella pneumoniae in European Food Products: a Multicentric Study Comparing Culture and Molecular Detection Methods. Microbiol Spectr 2022; 10:e0237621.

280. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 2009; 55:611-22.

281. Klymus KE MC, Allison MJ, Goldberg CS, Helbing CC, Hunter ME, Jackson CA, Lance RF, Mangan AM, Monroe EM, Piaggio AJ, Stokdyk JP, Wilson CC, Richter CA. Reporting the limits of detection and quantification for environmental DNA assays. Environmental DNA 2019; 00:1-12.

282. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, et al. Evaluating the Information Content of Shallow Shotgun Metagenomics. mSystems 2018; 3.

283. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. Nature 2017; 550:61-6.

284. Tomofuji Y, Sonehara K, Kishikawa T, Maeda Y, Ogawa K, Kawabata S, et al. Reconstruction of the personal information from human genome reads in gut metagenome sequencing data. Nat Microbiol 2023; 8:1079-94.

285. Shi Y, Wang G, Lau HC, Yu J. Metagenomic Sequencing for Microbial DNA in Human Samples: Emerging Technological Advances. Int J Mol Sci 2022; 23.

286. Bush SJ, Connor TR, Peto TEA, Crook DW, Walker AS. Evaluation of methods for detecting human reads in microbial sequencing datasets. Microb Genom 2020; 6.

287. Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 2011; 6:e17288.

288. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res 2018; 7:1338.

289. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012; 9:357-9.

290. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25:1754-60.

291. Whiteford N, Skelly T, Curtis C, Ritchie ME, Lohr A, Zaranek AW, et al. Swift: primary data analysis for the Illumina Solexa sequencing platform. Bioinformatics 2009; 25:2194-9.

292. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. LBNL-7065E. Berkeley, CA: Lawrence Berkeley National Laboratory. 2014.

293. Kulkarni P, Frommolt P. Challenges in the Setup of Large-scale Next-Generation Sequencing Analysis Workflows. Comput Struct Biotechnol J 2017; 15:471-7.

294. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. 2010.

295. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30:2114-20.
296. Krueger F. Trimgalore. GitHub repository, https://github.com/FelixKrueger/TrimGalore. 2021.
297. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 2010.
298. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 2016; 9:88.
299. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018; 34:i884-i90.
300. Blanco-Miguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat Biotechnol 2023; 41:1633-44.
301. Ounit R, Wanamaker S, Close TJ, Lonardi S. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. BMC Genomics 2015; 16:236.
302. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RG, Getz G, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. Nat Biotechnol 2011; 29:393-6.
303. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res 2016; 26:1721-9.
304. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. Peerj Comput Sci 2017.
305. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol 2019; 20:257.
306. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking Metagenomics Tools for Taxonomic Classification. Cell 2019; 178:779-94.
307. Anyansi C, Straub TJ, Manson AL, Earl AM, Abeel T. Computational Methods for Strain-Level Microbial Detection in Colony and Metagenome Sequencing Data. Front Microbiol 2020; 11:1925.
308. Maklin T, Kallonen T, David S, Boinett CJ, Pascoe B, Meric G, et al. High-resolution sweep metagenomics using fast probabilistic inference. Wellcome Open Res 2020; 5:14.

309. Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, et al. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 2018; 6:e4794.
310. Lever J, Martin K, Naomi A. Model selection and overfitting. Nature Methods 2016; 13:703-4.
311. Jafari M, Ansari-Pour N. Why, When and How to Adjust Your P Values? Cell J 2019; 20:604-7.
312. Noble WS. How does multiple testing correction work? Nat Biotechnol 2009; 27:1135-7.
313. Liu J, Zhang Q, Dong YQ, Yin J, Qiu YQ. Diagnostic accuracy of metagenomic next-generation sequencing in diagnosing infectious diseases: a meta-analysis. Sci Rep 2022; 12:21032.
314. Chen Y, Fan LC, Chai YH, Xu JF. Advantages and challenges of metagenomic sequencing for the diagnosis of pulmonary infectious diseases. Clin Respir J 2022; 16:646-56.
315. Baquero F. Threats of antibiotic resistance: an obliged reappraisal. Int Microbiol 2021; 24:499-506.
316. Elias R, Spadar A, Hendrickx APA, Bonnin RA, Dortet L, Pinto M, et al. Emergence of KPC-3- and OXA-181-producing ST13 and ST17 Klebsiella pneumoniae in Portugal: genomic insights on national and international dissemination. J Antimicrob Chemother 2023; 78:1300-8.
317. Lopez-Camacho E, Pano-Pardo JR, Ruiz-Carrascoso G, Wesselink JJ, Lusa-Bernal S, Ramos-Ruiz R, et al. Population structure of OXA-48-producing Klebsiella pneumoniae ST405 isolates during a hospital outbreak characterised by genomic typing. J Glob Antimicrob Resist 2018; 15:48-54.
318. Moubareck CA, Mouftah SF, Pal T, Ghazawi A, Halat DH, Nabi A, et al. Clonal emergence of Klebsiella pneumoniae ST14 co-producing OXA-48-type and NDM carbapenemases with high rate of colistin resistance in Dubai, United Arab Emirates. Int J Antimicrob Agents 2018; 52:90-5.
319. Oliveira RA, Ng KM, Correia MB, Cabral V, Shi H, Sonnenburg JL, et al. Klebsiella michiganensis transmission enhances resistance to Enterobacteriaceae gut invasion by nutrition competition. Nat Microbiol 2020; 5:630-41.
320. Velazquez EM, Nguyen H, Heasley KT, Saechao CH, Gil LM, Rogers AWL, et al. Endogenous Enterobacteriaceae underlie variation in susceptibility to Salmonella infection. Nat Microbiol 2019; 4:1057-64.

321. Barlow GM, Mathur R. Type 2 Diabetes and the Microbiome. J Endocr Soc 2022; 7:bvac184.
322. Salles BIM, Cioffi D, Ferreira SRG. Probiotics supplementation and insulin resistance: a systematic review. Diabetol Metab Syndr 2020; 12:98.
323. Maseda E, Salgado P, Anillo V, Ruiz-Carrascoso G, Gomez-Gil R, Martin-Funke C, et al. Risk factors for colonization by carbapenemase-producing enterobacteria at admission to a Surgical ICU: A retrospective study. Enferm Infecc Microbiol Clin 2017; 35:333-7.
324. Zhu WM, Yuan Z, Zhou HY. Risk factors for carbapenem-resistant Klebsiella pneumoniae infection relative to two types of control patients: a systematic review and meta-analysis. Antimicrob Resist Infect Control 2020; 9:23.
325. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. Nat Microbiol 2018; 3:1255-65.
326. Yip AYG, King OG, Omelchenko O, Kurkimat S, Horrocks V, Mostyn P, et al. Antibiotics promote intestinal growth of carbapenem-resistant Enterobacteriaceae by enriching nutrients and depleting microbial metabolites. Nat Commun 2023; 14:5094.
327. Yoon MY, Yoon SS. Disruption of the Gut Ecosystem by Antibiotics. Yonsei Med J 2018; 59:4-12.
328. Watts GS, Thornton JE, Jr., Youens-Clark K, Ponsero AJ, Slepian MJ, Menashi E, et al. Identification and quantitation of clinically relevant microbes in patient samples: Comparison of three k-mer based classifiers for speed, accuracy, and sensitivity. PLoS Comput Biol 2019; 15:e1006863.
329. Vezina B, Watts SC, Hawkey J, Cooper HB, Judd LM, Jenney AWJ, et al. Bactabolize is a tool for high-throughput generation of bacterial strain-specific metabolic models. Elife 2023; 12.

Paper I

GUT MICROBES
2022, VOL. 14, NO. 1, e2118500 (19 pages)
https://doi.org/10.1080/19490976.2022.2118500

Taylor & Francis
Taylor & Francis Group

METHODS

Detection of *Klebsiella pneumoniae* human gut carriage: a comparison of culture, qPCR, and whole metagenomic sequencing methods

Kenneth Lindstedt[a], Dorota Buczek[a], Torunn Pedersen[b], Erik Hjerde[c], Niclas Raffelsberger[a,d], Yutaka Suzuki[e], Sylvain Brisse[f], Kathryn Holt[g,h], Ørjan Samuelsen[b,i], and Arnfinn Sundsfjord[a,b]

[a]Department of Medical Biology, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway; [b]Norwegian National Advisory Unit on Detection of Antimicrobial Resistance, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway; [c]Department of Chemistry, UiT the Arctic University of Norway, Tromsø, Norway; [d]Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway; [e]Department of Computational Biology and Medical Sciences, The University of Tokyo, Tokyo, Japan; [f]Biodiversity and Epidemiology of Bacterial Pathogens Unit, Institut Pasteur, Université Paris Cité, Paris, France; [g]Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Australia; [h]Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, UK; [i]Department of Pharmacy, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway

ABSTRACT

Klebsiella pneumoniae is an important opportunistic healthcare-associated pathogen and major contributor to the global spread of antimicrobial resistance. Gastrointestinal colonization with *K. pneumoniae* is a major predisposing risk factor for infection and forms an important hub for the dispersal of resistance. Current culture-based detection methods are time consuming, give limited intra-sample abundance and strain diversity information, and have uncertain sensitivity. Here we investigated the presence and abundance of *K. pneumoniae* at the species and strain level within fecal samples from 103 community-based adults by qPCR and whole metagenomic sequencing (WMS) compared to culture-based detection. qPCR demonstrated the highest sensitivity, detecting *K. pneumoniae* in 61.2% and 75.8% of direct-fecal and culture-enriched sweep samples, respectively, including 52/52 culture-positive samples. WMS displayed lower sensitivity, detecting *K. pneumoniae* in 71.2% of culture-positive fecal samples at a 0.01% abundance cutoff, and was inclined to false positives in proportion to the relative abundance of other Enterobacterales present. qPCR accurately quantified *K. pneumoniae* to 16 genome copies/reaction while WMS could estimate relative abundance to at least 0.01%. Quantification by both methods correlated strongly with each other (Spearman's rho = 0.91). WMS also supported accurate intra-sample *K. pneumoniae* sequence type (ST)-level diversity detection from fecal microbiomes to 0.1% relative abundance, agreeing with the culture-based detected ST in 16/19 samples. Our results show that qPCR and WMS are sensitive and reliable tools for detection, quantification, and strain analysis of *K. pneumoniae* from fecal samples with potential to support infection control and enhance insights in *K. pneumoniae* gastrointestinal ecology.

PLAIN LANGUAGE SUMMARY

What is the context?

- *Klebsiella pneumoniae* is a major cause of healthcare-associated infections and a key contributor to the spread of resistance to last-line antimicrobials.
- Gastrointestinal colonization by *K. pneumoniae* is a risk factor for developing infection and can facilitate the spread of antimicrobial resistance.
- Culture-based detection may lack sensitivity and is ill-suited to detecting within-sample *K. pneumoniae* abundance and diversity.
- Developing molecular methods to improve *K. pneumoniae* abundance and strain diversity detection are essential in understanding human gut colonization and ecology.

What is new?

- We compared culture-based detection of *K. pneumoniae* within human fecal samples to two molecular-based techniques: 1) qPCR, which amplifies *K. pneumoniae* species complex-specific DNA targets with high sensitivity, and 2) whole metagenomic sequencing (WMS), which sequences the entire DNA content of complex microbial communities.

ARTICLE HISTORY
Received 24 May 2022
Revised 17 August 2022
Accepted 24 August 2022

KEYWORDS
Klebsiella pneumoniae; qPCR; whole metagenomic sequencing; gastrointestinal carriage; detection; quantification; sequence type

CONTACT Kenneth Lindstedt kenneth.w.lindstedt@uit.no; Arnfinn Sundsfjord arnfinn.sundsfjord@uit.no Department of Medical Biology, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, 9038, Norway

Supplemental data for this article can be accessed online at https://doi.org/10.1080/19490976.2022.2118500

- Our findings show:
- qPCR had the highest sensitivity, detecting *K. pneumoniae* in all (52/52) culture-positive samples and 11/51 and 23/47 culture-negative samples, using a direct-fecal and culture-sweep method, respectively. qPCR could accurately quantify *K. pneumoniae* to 16 genome copies/reaction.
- WMS had lower sensitivity, positive in 37/52 culture-positive samples, and demonstrated false positives arising from closely related bacterial species. Relative abundance estimates could be performed to 0.01%.
- WMS performed accurate strain-level detection of *K. pneumoniae* to 0.1% relative abundance and could detect within-sample strain diversity.

What is the impact?

qPCR and WMS are valid methods for the detection and characterization of colonizing *K. pneumoniae* with potential to enhance our understanding of the gastrointestinal ecology of this important pathogen.

Introduction

Klebsiella pneumoniae (Kp) is a critical priority pathogen that has become a major contributor in the spread of antimicrobial resistance (AMR) within and between sectors.[1–4] Multidrug-resistant (MDR) Kp clones have disseminated globally and are a leading cause of opportunistic healthcare-associated infections, with limited treatment options and high morbidity and mortality rates.[5–8] In parallel, 'hypervirulent' (Hv) Kp clones have emerged which are typically non-MDR, however, can cause invasive community-acquired infections in otherwise healthy individuals.[9]

Kp is part of the phylogenetically broader *Klebsiella pneumoniae* species complex (KpSC), consisting of the seven closely related taxa (or phylogroups): *K. pneumoniae sensu strictu* (Kp1), *K. quasipneumoniae* subsp. *quasipneumoniae* (Kp2) and subsp. *similipneumoniae* (Kp4), *K. variicola* subsp. *variicola* (Kp3) and subsp. *tropica* (Kp5), '*K. quasivariicola*' (Kp6), and *K. africana* (Kp7).[8,10–13] Kp1 (referred to here as Kp) is of most clinical relevance and consists of a highly diverse population structure corresponding to thousands of unique sequence types (STs) as defined by the seven-gene multi-locus sequence typing (MLST) scheme; https://bigsdb.pasteur.fr/klebsiella/.[8,14]

Kp is a common colonizer of the human gastrointestinal tract. The prevalence of gastrointestinal colonization in the community can range from 4% and 6% in the USA and Australia, 40% and 65% in Senegal and Madagascar, and up to 75% and 87% in Taiwan and Malaysia.[15–18] We recently described a KpSC carriage rate of 16.3% among 2975 adults in a general urban population in Northern Norway using the KpSC selective Simmon's citrate agar with inositol (SCAI).[19,20] Gastrointestinal colonization itself is a major risk factor for invasive infection in hospitalized patients and an increased relative abundance corresponds to a higher infection risk.[15,21–23] The gut is also an important reservoir for the spread of AMR through clonal dissemination and horizontal gene transfer (HGT).[24,25]

Despite the clinical and epidemiological importance of gastrointestinal carriage, significant knowledge gaps regarding the prevalence, abundance, and diversity of Kp in human gut colonization remain. Kp detection is generally performed by culture-based screening of fecal samples or rectal swabs, which is time-consuming and gives limited information regarding abundance and intra-sample strain diversity. Culture-based detection has also been shown to lack sensitivity in detection of Gram-negative pathogens from fecal samples.[26,27] Molecular methods such as quantitative PCR (qPCR) and shotgun whole metagenomic sequencing (WMS) offer a potential to compensate for these shortcomings. Both qPCR- and metagenomics-based methods have demonstrated equivalent or improved detection sensitivity for pathogenic bacteria and AMR genes in clinical and environmental samples compared to culture.[26–30]

The aim of this study was to evaluate and compare WMS, qPCR, and culture for the detection and quantification of Kp from human fecal samples at both the species and strain level. Using the extensive culture and whole-genome sequencing (WGS) data gathered during our previous culture-based

Kp carriage study, we analyzed a representative selection of Kp culture-positive and negative fecal samples by both qPCR and WMS. Results were compared to culture for Kp detection sensitivity and analyzed for Kp relative abundance and intra-sample strain diversity.

Results

Efficiency and sensitivity of the ZKIR-qPCR in human fecal samples

We employed the recently developed ZKIR-qPCR for Kp detection in this study due to its high sensitivity and specificity for KpSC detection in environmental and food samples.[30,31] BLAST analysis of the 78 bp ZKIR-qPCR target sequence revealed high sequence conservation in all 484 KpSC genomes from our previous cross-sectional carriage study, with 98.6% (477/484) having three or less bp mismatches in the forward primer, and a single conserved A to G substitution at the 5' end of the reverse primer region.[20] Importantly the 3' ends of both forward and reverse primer regions were perfectly conserved, except for a single Kp2 isolate with an A to G substitution at the 3' terminal end of the forward primer (Suppl. Figure 1). Calculated melting temperatures (T_M) of the PCR product of each sequence variant ranged from 78.8 to 79.9°C.

The ZKIR-qPCR had an amplification efficiency > 90% and $R^2 > 0.99$ in a linear dynamic range from 250,000 to 3 genome copies per reaction, both in the presence and absence of 25 ng KpSC-negative fecal microbiome DNA, when assessed against representative strains of each of the four human-associated KpSC subspecies (Kp1-4, Suppl. Table 1, Suppl. Figure 2). Each selected KpSC strain had the most prevalent number of forward primer mismatches seen for that subspecies in the BLAST analysis (Kp1 = 1, Kp3 = 3, Kp2 = 2, Kp4 = 2, Suppl. Figure 1). In-line with Poisson distribution, dilutions below three genome copies per reaction only intermittently detected the ZKIR amplicon.[32]

Limit of Detection (LOD) was three genome copies per reaction for all four KpSC subspecies, both in the presence and absence of 25 ng KpSC-negative fecal microbiome DNA (Suppl. Table 2). At 16 genome copies, copy number could be quantified to a coefficient of variation (CV) ≤ 35%, a previously reported Limit of Quantification (LOQ).[33]

Detection of Kp in human fecal samples by the ZKIR-qPCR

To determine the Kp detection sensitivity of the ZKIR-qPCR in human fecal samples, 52 Kp culture-positive and 51 KpSC culture-negative human fecal samples were selected from our previous study.[20] DNA was prepared as a direct fecal microbiome extraction (Direct samples), as well as from a plate-sweep of each sample re-grown on SCAI media (Sweep samples). Four culture-negative samples failed to grow on SCAI media. A total of 61.2% (63/103) Direct and 75.8% (75/99) Sweep samples were positive by the ZKIR-qPCR (Table 1). All culture-positive samples (52/52) were positive by the ZKIR-qPCR in both Direct and Sweep sample preparations. Of the ZKIR-qPCR positive Direct samples, 6.4% (4/63) were not detected after Sweep enrichment. Sanger sequencing performed on seven Direct sample amplicons confirmed the correct ZKIR sequence.

Quantification of genome copy number in Direct samples demonstrated that culture-positive samples had a significantly higher KpSC abundance than culture-negative samples (median 33.72 and 0.17 genome copies/ng DNA, respectively, $p < .001$), (Figure 1a, Suppl. Table 3). This difference was amplified by Sweep enrichment (culture-positive median: 40,865 genome copies/ng DNA, culture-negative median: 0.15 genome copies/ng DNA, $p < .001$), (Figure 1b, Suppl. Table 3).

Table 1. Comparison of KpSC detection by SCAI culture compared to the ZKIR-qPCR in 103 fecal samples using the Direct fecal microbiome (Direct) and SCAI sweep enrichment (Sweep) DNA extraction methods.

	No. (%) of samples positive by ZKIR-qPCR	
Culture result[a]	*Direct*	*Sweep*
positive n = 52	52 (100)	52 (100)
negative n = 51	11 (22)	23 (49)[b]

Direct, Direct fecal microbiome DNA extraction; Sweep, DNA extraction from a plate sweep of fecal samples cultured 48 hours on SCAI media.
[a]SCAI culture detection result as per our previous culture-based Kp gut carriage study[20]
[b]Four culture negative samples failed to grow on SCAI media, thus % Sweep positive was calculated using n = 47 culture negative samples.

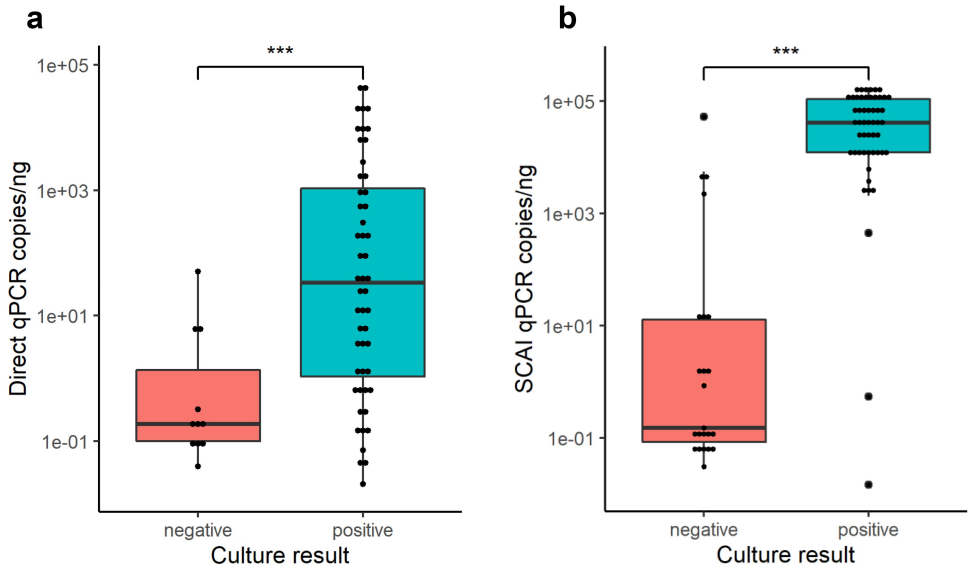


Figure 1. Comparison of KpSC abundance between culture-positive (teal) and culture-negative (red) samples detected by the ZKIR-qPCR using the Direct fecal method (a) and SCAI sweep enrichment (b). *** = $p < .001$ (Mann Whitney U test).

Profiling WMS samples and effect of sample collection kit

Taxonomic profiling of whole metagenomic sequenced Direct samples demonstrated the presence of the major human gut associated phyla (Suppl. Table 4A).[34,35] Enterobacterales, which dominated the Proteobacteria phyla, had a mean of relative abundance of 17.2% (Suppl. Table 4B). This was considerably higher than previously reported mean of ≤ 2% relative abundance for Enterobacterales within gut microbiomes from healthy adults.[34,35] As expected, Enterobacterales abundance increased to almost complete domination of metagenomes following SCAI sweep enrichment, (mean 95.6%, SD 9%) (Supplementary Table 4C).

To investigate whether our sampling method had caused an artificially increased Enterobacterales abundance, we performed repeat sampling, WMS, and taxonomy profiling of ten participants from our previous culture-based Kp carriage study using the original collection method (ESwabs) and compared these to a validated preservative microbiome sample collection kit (Norgen) (Suppl. Table 5A-D).[36] Taxonomic comparison revealed significant expansion of Enterobacterales in ESwab samples compared to Norgen samples, (ESwab median 38.2% vs Norgen median 0.62%, p = .002, Suppl. Table 5C and D). This was predominantly driven by *Escherichia coli* overgrowth, (median 26.2% vs 0.16% in ESwabs and Norgen samples, respectively, p = .002), which constituted a median of 86.8% of total Enterobacterales in ESwabs, compared to 39.8% in Norgen samples. While Kp abundance also underwent a significant increase in ESwabs compared to Norgen samples (median 0.42% vs 0.03%, respectively, p = .006), its total relative abundance within Enterobacterales reduced (median 0.99% vs 4.2% in ESwabs and Norgen samples, respectively). The biased microbiome profile caused by collection in ESwabs prevented any Kp-microbiome association analyses to be included as part of this study.

Detection of Kp by WMS

To determine the sensitivity and specificity of Kp detection by WMS, samples were analyzed using the taxonomic profiler Centrifuge and compared to the ZKIR-qPCR and culture. Reads were assigned to Kp in all except four Direct samples (n = 99) and all Sweep samples that had grown (n = 99). Two additional Direct samples had < 10^{-5}% Kp relative abundance, so were considered as negative. Kp relative abundance in qPCR positive samples (median 0.027% and 6.05% in Direct and Sweep samples,

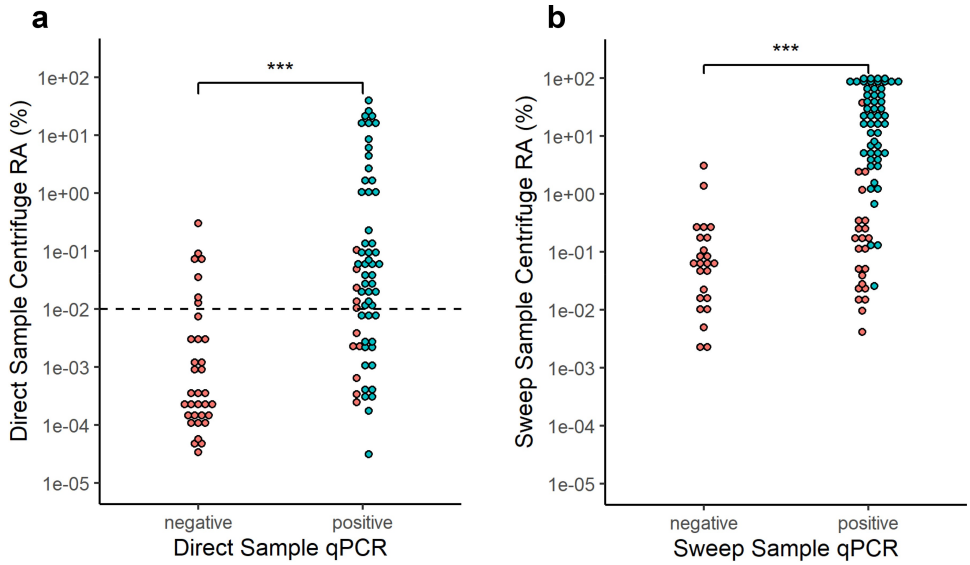


Figure 2. Kp abundance detected by Centrifuge compared to ZKIR-qPCR and SCAI culture detection in (a) Direct and (b) Sweep samples. Teal = SCAI culture positive, Red = SCAI culture negative. RA = relative abundance. Dotted-line line in Figure 2a represents 0.01% abundance detection cutoff used in Direct samples. *** = $p < .001$ (Mann Whitney U test).

respectively), was significantly higher than qPCR negative samples, (median 0.00035% and 0.063% for Direct and Sweep samples, respectively, both had $p < .001$) (Figure 2, Suppl. Table 3). Despite this, considerable overlap in abundances was observed between qPCR- and culture-positive and negative groups, precluding easy distinction of Kp presence or absence by WMS from either the Direct or Sweep preparations.

We hypothesized many of the qPCR and culture-negative samples with high Kp relative abundance by WMS were false positives misassigned from closely related Enterobacterales species. To investigate this further, two non-Kp containing *in silico* binary species mixes were constructed and Kp abundance measured. Each species mix consisted of reads from *Bacteroides fragilis* and either *E. coli* or *Klebsiella aerogenes* in increasing proportions from 0.1% to 100% relative abundance. An increasing rate of Kp false positives were observed as the abundance of both Enterobacterales species increased, with a higher effect observed for *K. aerogenes* (Suppl. Figure 3), which is the species most closely related to the KpSC. Kp false positives exceeded 0.01% between 10–15% and 30–35% *K. aerogenes* and *E. coli* relative abundance, respectively. At 100% *K. aerogenes* or *E. coli* relative abundance, Kp false positives were 0.20% and 0.056%, respectively.

E. coli was the most abundant Enterobacterales species in our Direct samples, with mean relative abundance of 12.1% (Suppl. Table 3). Since Kp false positives over 0.01% did not to appear in our *in silico* species mix until *E. coli* relative abundance was greater than 30%, 0.01% relative abundance was used as a cutoff to report Kp presence to optimize detection sensitivity and specificity. Using this cutoff, Kp was detected in 66.7% (42/63) and 71.2% (37/52) of ZKIR-qPCR and culture-positive Direct samples, respectively (Figure 2a, Suppl. Table 3). Of the 49 Direct samples with Kp WMS abundance above 0.01%, 85.7% (42/49) and 75.5% (37/49) were positive by the ZKIR-qPCR and SCAI culture, respectively. Of the seven Direct samples negative by both qPCR and culture with a Centrifuge Kp abundance above 0.01%, six had Enterobacterales abundances above 15%, suggestive of false positives. Due to the high Enterobacterales abundance in Sweep samples (mean 95.6%), which would be expected to generate high Kp false positives, no such detection cutoff was applied (Suppl. Table 3).

Screening assembled metagenomes for Kp-specific sequences

To use a more specific approach for WMS-based detection, assembled metagenomes were screened for the seven Kp MLST genes and the

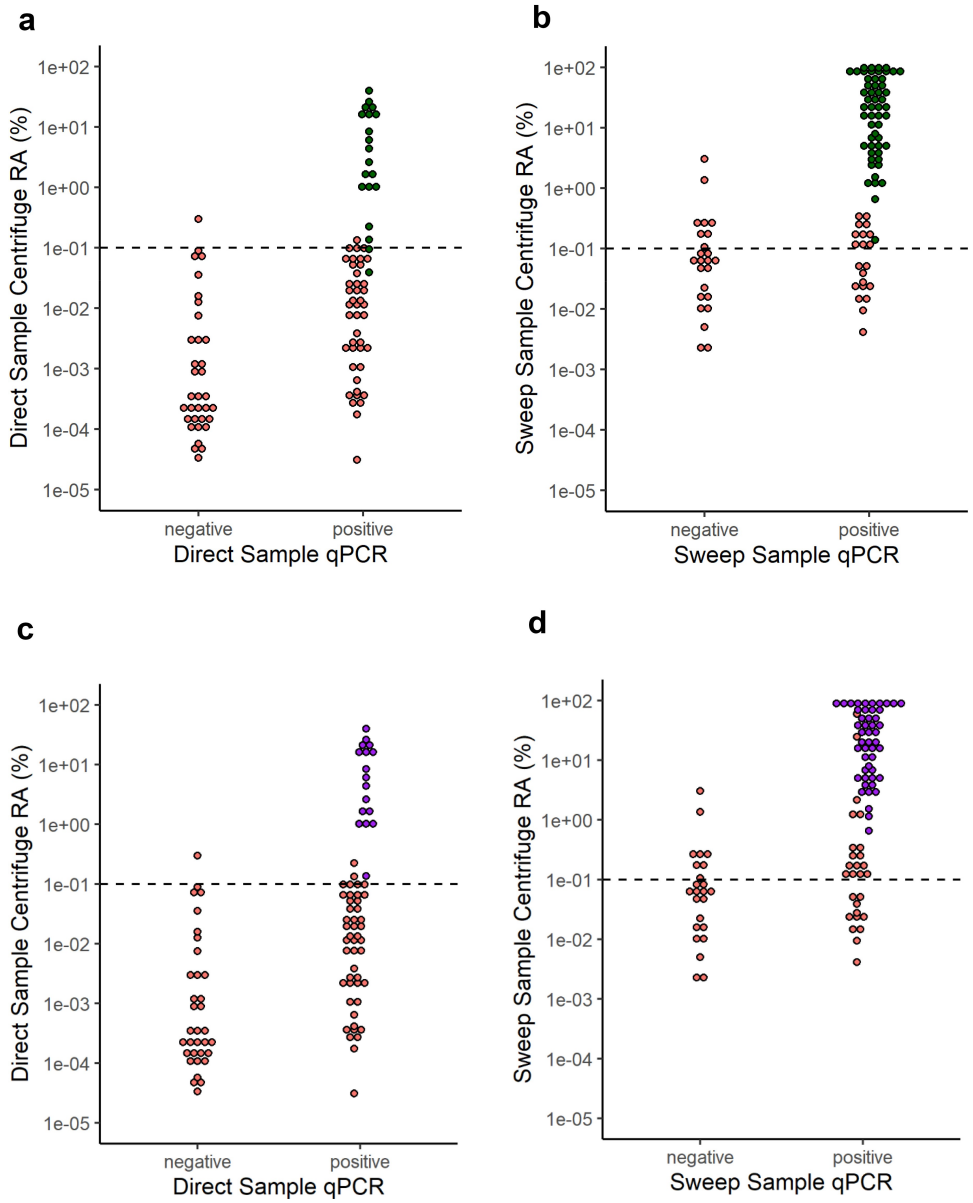


Figure 3. Kp detection from assembled metagenomes using KpSC specific 78bp ZKIR sequence (green) and 4/7 Kp MLST genes (purple) in Direct and Sweep samples. (a) and (b): ZKIR sequence detected in Direct and Sweep samples, respectively. (c) and (d): 4/7 Kp MLST genes detected in Direct and Sweep samples, respectively. Red = ZKIR/MLST sequences not detected. RA = relative abundance. Dotted line at 0.1% Centrifuge abundance represents approximate threshold for detection by these methods.

ZKIR sequence. The ZKIR sequence was detected in 54.5% (54/99) of Sweep samples, all of which were positive by qPCR and 92.6% (50/54) were positive by culture (Figure 3b, Suppl. Table 3). Similarly, using a detection cutoff of 4/7 MLST genes, 52.5% (52/99) Sweep samples were positive, all of which were also positive by qPCR and 94.2% (49/52) were positive by culture (Figure 3d, Suppl. Table 3). Detection sensitivity was considerably lower in Direct samples, with 19.4% (20/103) and 16.5% (17/103) positive by ZKIR sequence and 4/7 MLST gene detection, respectively. All Direct ZKIR and MLST positive samples, however, were positive by both qPCR and culture (Figure 3 a and c, Suppl. Table 3). All Sweep and Direct samples with at least 4/7 MLST genes detected were also positive for the ZKIR sequence. Detection by these methods had clear dependence on Kp abundance, with the ZKIR sequence detected in only two samples below approximately 400 genome copies/ng DNA by qPCR and 0.1% relative abundance by

WMS, while 4/7 MLST genes were not detected in any samples below this threshold.

Kp relative abundance estimation by WMS and correlation with qPCR

KpSC abundance measured by qPCR had a strong correlation with Kp relative abundance measured by Centrifuge in both ZKIR-qPCR positive Direct and Sweep samples (Spearman's rho = 0.91, and 0.96 respectively, both with $p < .001$), (Figure 4). Sweep samples also demonstrated clear separation between culture-positive and negative samples by both qPCR and Centrifuge (Figure 4b). In these samples, below a limit of approximately 2000 copies/ng by qPCR and 0.6% by Centrifuge, only 9.5% (2/21) of samples were detected as positive by culture, compared to 92.5% (50/54) above this threshold.

The accuracy of Kp abundance quantification by WMS was investigated using two *in vitro* Kp-spiked microbiomes that consisted of: 1) a mock microbiome of six bacterial species representing the major gut taxa, including 1% *E. coli* to represent total Enterobacterales, and 2) a KpSC-negative human fecal microbiome with 0.52% total Enterobacterales abundance. Both microbiomes were spiked with Kp at 1%, 0.1%, 0.01%, and 0.001%, and 0% relative abundance. The mock microbiome was spiked with a single strain (ST11), while the fecal microbiome was spiked with a combination of three Kp strains (ST11, ST23, and ST101), at a ratio of 60:30:10, respectively (Suppl. Table 1).

Centrifuge achieved close estimations of all Kp relative abundances, differing from spiked abundances by a factor of 0.46 to 0.86 and 3.52 to 4.79 in the mock and fecal microbiomes, respectively (Table 2 and Table 3). Background Kp abundances measured in the non-spiked mock and fecal microbiomes were 0.0004% and 0.00068%, respectively, which were close to the observed Kp false positives in the *in silico* binary species mixes containing 0.5% and 1% total Enterobacterales abundances (Suppl. Figure 3). These findings suggest in 'healthy' microbiomes, which typically contain <1% total Enterobacterales abundance, Kp quantification can be performed to as low as 0.01% without substantial influence from false positives.[34,35]

Table 2. *K. pneumoniae* abundance in the spiked mock microbiome measured by WMS (Centrifuge) and the ZKIR-qPCR.

Sample	Kp spike (%)	WMS measured abundance (%)	Ratio spiked/ measured	ZKIR-qPCR ave copies/reaction
M1	0.00	0.0004	NA	0
M2	0.001	0.00086	0.86	41
M3	0.01	0.0046	0.46	412
M4	0.1	0.050	0.50	4776
M5	1.0	0.59	0.59	61395

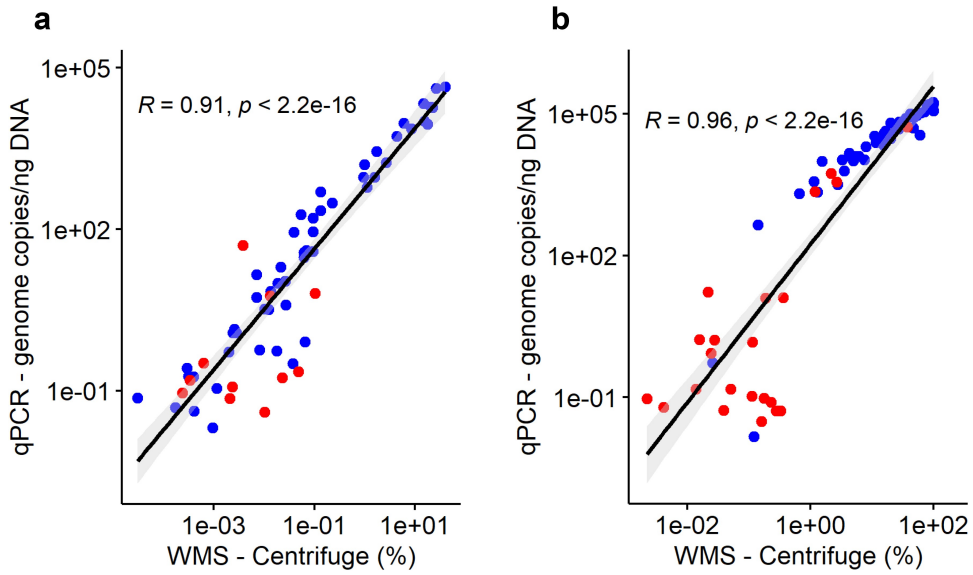


Figure 4. Correlation of Kp abundances quantified by qPCR vs WMS (Centrifuge) in Direct samples (a) and Sweep samples (b). R = Spearman's rho. Blue = SCAI culture positive, Red = SCAI culture negative.

Table 3. *K. pneumoniae* abundance in the spiked fecal microbiome measured by WMS (Centrifuge) and the ZKIR-qPCR.

Sample	Kp spike (%)	WMS measured abundance (%)	Ratio spiked/ measured	ZKIR-qPCR ave copies/reaction
F1	0.00	0.00068	NA	0
F2	0.001	0.0035	3.52	16
F3	0.01	0.038	3.78	235
F4	0.1	0.43	4.34	2905
F5	1.0	4.79	4.79	31925

NA = not applicable

The average genome copy numbers measured by the ZKIR-qPCR in the 0.001% spiked mock and fecal microbiomes of 41 and 16 genome copies/reaction, respectively, were close to the measured LOQ for this assay. This suggests at 0.001% relative abundance the ZKIR-qPCR is approaching the lower limit at which it can accurately quantify Kp abundance.

Kp strain-level detection from metagenomic data

Metagenomic Kp strain-level analysis of Direct samples was performed with StrainGST, part of the StrainGE toolkit.[37] To examine the tool's accuracy, we explored our Kp-spiked mock and fecal microbiomes described above (Table 4 and Table 5). The spiked Kp isolate ST11 was correctly identified in the mock microbiome at 1% and 0.1% abundance. In the fecal microbiome, the three spiked isolates, ST11, ST23, and ST101, were all correctly identified at 1%, while ST11 and ST23 only were identified at 0.1%. No strains were detected in either spiked microbiome at abundances below 0.1%, in line with the reported lower detection limit for this tool.[37] A spiked mock microbiome sample containing 1% Kp (ST11) and 1% Kp3 (ST697) was also tested, in which both strains were identified correctly. Estimated abundances of each spiked strain by StrainGST were close approximations of the true abundances. No false positives were reported in the spiked fecal microbiome samples. In the Kp spiked mock microbiomes ST12 and ST340 were reported in the samples spiked with 1% and 0.1% ST11, respectively. Comparison of the MLST profiles revealed ST12 and ST340 are single and double locus variants of ST11, respectively.

Since no spiked Kp strains were detected below 0.1%, all Direct samples with Centrifuge Kp relative abundance ≥ 0.1%, including two Kp culture-negative samples, were selected for StrainGST analysis, (n = 21 samples, median Kp relative abundance: 2.64%, range: 0.1% – 39.55%, Suppl. Table 3). Kp strains were detected in all culture-positive samples (n = 19), and matched culture detected strains in 84.2% (16/19) (Table 6). Kp strains detected which did not match culture were either four or five locus variants of their culture detected counterparts, suggesting these were

Table 4. Metagenomic Kp strain-level detection performed using StrainGST in a Kp-spiked mock microbiome sample.

Spiked Microbiome			StrainGST result			
Mock	Spiked strain(s)	Spiked abundance (%)	Strain(s) detected	Strain est. abundance (%)	Total est. abundance (%)	Confidence Score
M1	none	NA	ND	0	0	NA
M2	ST11	0.001	ND	0	0	NA
M3	ST11	0.01	ND	0	0	NA
M4	ST11	0.1	ST12[a], ST11	0.051, 0.166	0.217	0.207, 0.044
M5	ST11	1.0	ST11, ST340[a]	0.429, 0.762	1.191	0.760, 0.027
M6	ST11, ST697[b]	1:1	ST11, ST697, ST340[a]	0.798, 0.377, 0.756	1.93	0.705, 0.591, 0.03

[a]ST340 and ST12 are single and double locus variants of ST11, respectively
[b]*Klebsiella variicola*
NA = not applicable
ND = not detected.

Table 5. Metagenomic Kp strain-level detection performed using StrainGST in a Kp-spiked fecal microbiome sample.

Spiked Microbiome			StrainGST result			
Fecal	Spiked strains (ratio 60:30:10)	Spiked abundance (%)	Strain(s) detected	Strain est. abundance (%)	Total est. abundance (%)	Confidence Score
F1	none	NA	ND	0	0	NA
F2	ST11, ST23, ST101	0.001	ND	0	0	NA
F3	ST11, ST23, ST101	0.01	ND	0	0	NA
F4	ST11, ST23, ST101	0.1	ST11, ST23	0.056, 0.028	0.084	0.245, 0.029
F5	ST11, ST23, ST101	1.0	ST11, ST23, ST101	0.398, 0.171, 0.2	0.769	0.747, 0.230, 0.082

NA = not applicable

Table 6. Metagenomic Kp strain-level detection performed using StrainGST in 21 adult fecal samples with Kp relative abundance ≥ 0.1%.

T7 Sample Number	SCAI culture result[a]		Kp Abundance		StrainGST result		
	Kp detected?	Strain	WMS (%)[b]	qPCR (copies/ng DNA)	Strain(s) detected	est. abundance (%)	Confidence Score
89	yes	**ST14**	39.55	45900	**ST14**	30.639	0.94
101	yes	ST2042	26.04	39390	ST2039	27.521	0.94
75	yes	**ST485**	22.09	18240	**ST485**, ST35[c]	6.146, 3.875	0.86, 0.66
92	yes	**ST35**	20.55	18660	**ST35**	12.651	0.96
74	yes	**ST27**	17.81	9120	**ST27**	4.061	0.98
18	yes	**ST4039**	15.29	9940	**ST4039**	5.27	0.85
91	yes	**ST1496**	14.73	21240	**ST1496**	14.655	0.98
45	yes	**ST25**	8.39	7400	**ST25**, ST2549, ST4039[c]	0.37, 2.75, 1.16	0.76, 0.1, 0.02
97	yes	**ST704**	6.02	9410	**ST704**	7.38	0.74
12	yes	**ST70**	4.37	5290	**ST70**	6.922	0.65
44	yes	**ST23**	2.64	1716	**ST23**	1.524	0.67
100	yes	ST375	1.73	2785	ST2042, ST1660	2.07, 0.6	0.63, 0.03
46	yes	**ST25**	1.56	920	**ST25**, ST461	0.29, 0.09	0.58, 0.16
2	yes	**ST3043**	1.12	604	**ST3043**	0.32	0.84
90	yes	**ST1106**	0.98	1612	**ST1106**	1.84	0.28
84	yes	**ST200**	0.95	926	**ST200**	0.98	0.67
80	no	NA	0.30	0	ND	NA	NA
72	yes	**ST20**	0.22	298	**ST20**	0.59	0.18
21	yes	**ST151**	0.13	492	**ST151**	1.85	0.22
35	yes	ST25	0.13	217	ST10	0.13	0.33
62	no	NA	0.10	7	ND	NA	NA

Samples in bold represent ST matches between culture detection and strainGST detection
[a]SCAI culture detection result as per our previous culture-based Kp gut carriage study[20]
[b]WMS relative abundance measured by Centrifuge
[c]STs detected in these samples were double (sample 45) and triple (sample 75) locus variants of each other which may have resulted from assignment a single strain to multiple closely related reference genomes in the database rather than true intra-sample strain diversity.
NA = not applicable

not closely related. No strains were detected in either of the two culture-negative samples. Both of these samples had high Enterobacterales abundance (59.6% and 63.9%), while one was qPCR negative and the other had only seven genome copies/ng DNA detected, suggestive of falsely high relative abundance estimated by Centrifuge. Multiple strains were reported in four samples, three with two strains and one had three strains. Three of these samples had strains matching culture-detection, which were the highest confidence and abundance strain in each case. The three strains detected in sample 45 and the two strains in sample 75 were double and triple locus variants of each other, respectively, and considered possibly to be one strain that had been assigned to multiple reference genomes in the database. The two strains detected in both samples 46 and 100 shared only a single MLST locus each, likely representing true microbiome Kp strain diversity.

NA = not applicable

Discussion

The purpose of this study was to investigate WMS and the ZKIR-qPCR as methods for the detection and analysis of Kp from human fecal samples and compare these to culture-based detection. Overall, the ZKIR-qPCR demonstrated the highest Kp detection sensitivity which was reflected in the very low LOD of this assay of just three genome copies per reaction. This corresponds to the lowest possible limit for qPCR according to Poisson distribution, indicating the very high efficiency of this assay.[32] The fact that no culture-positive samples were negative by the ZKIR-qPCR indicates a low false negative rate. Our findings suggest SCAI-based detection may underestimate the true prevalence of KpSC gastrointestinal carriage. This may be related to technical challenges identifying KpSC given other common Enterobacterales, including *Enterobacter, Citrobacter, Serratia*, and other non-KpSC *Klebsiella* species, are capable of growth on SCAI media often with similar morphologies to KpSC.[19,38,39] In-line with our findings, the ZKIR-qPCR has previously demonstrated a higher KpSC detection sensitivity compared to SCAI culture in plant, soil, chicken meat, and salad samples.[30,31] Similar to the findings by Barbier *et al.*[30] we found a culture enrichment step prior to qPCR significantly enhanced detection sensitivity. Together

these results demonstrate the ZKIR-qPCR is a rapid and highly accurate tool for KpSC detection in a range of sample types, thereby facilitating targeted selection of samples for further culture- or metagenomics-based analysis.

It is noteworthy that the high detection sensitivity of the ZKIR-qPCR was achieved despite our tested KpSC strains having one (Kp1), two (Kp2 and Kp4), and three (Kp3) bp mismatches within 15 nucleotides of the 5' end of the forward ZKIR primer and one mismatch at the penultimate position of the 5' end of the reverse primer. In accordance with these findings, it has been shown up to five bp mismatches within primers can be well tolerated provided the 3' primer region is well conserved.[40–42]

Accurate species-level profiling is essential for high quality shotgun WMS analyses. Using the taxonomic profiling tool Centrifuge with a 0.01% relative abundance cutoff, Kp was detected at the species level in 66.7% and 71.2% of qPCR and culture-positive Direct samples, respectively. Despite the comparatively high reported sensitivity and specificity of Centrifuge, however, like other metagenomic classifiers it is known to generate false positives at lower species abundances.[43–45] Using *in silico* binary species mixes, we demonstrated the rate of Kp false positives is proportional to the abundance and relatedness of other Enterobaterales species present in the sample, surpassing 0.01% Kp false positives at 10–15% *K. aerogenes* and 30–35% *E. coli* relative abundance, respectively. Within a 'healthy' gut microbiome, in which the average Enterobacterales abundance is ≤2%, this may not significantly impact Kp detection specificity.[34,35] Much higher Enterobacterales abundance can occur in dysbiotic states including inflammatory bowel disease, type 2 diabetes mellitus, and following antimicrobial therapy.[46–48] Detection of low abundance Kp by taxonomic classifiers in these settings would therefore need careful interpretation.

Gastrointestinal microbiomes with an increased Kp relative abundance are associated with an increased risk of Kp bacteremia, nosocomial transmission, and may predispose to prolonged colonization.[22,23,49,50] Accurate measurement of Kp abundance could therefore provide important clinical information relevant for infection risk stratification and infection control purposes. In our Kp spiked microbiomes, which contained ≤1% total Enterobacterales, we found WMS gave close estimations of Kp relative abundance to 0.01%, below which false positives began to have a substantial influence. The ZKIR-qPCR, however, accurately quantified Kp to as low as 16 genome copies/reaction, corresponding to approximately 0.001% relative abundance in the Kp-spiked microbiomes, with the additional advantage of providing quantification information in a clinically relevant timeframe.

The spread of AMR by Kp is predominantly driven by the expansion of MDR high-risk clones disseminating between hospitals and across borders.[1,8,51–55] The utility of WMS in Kp infection control thus requires timely and sensitive Kp strain-level detection. Using StrainGST, part of the StrainGE toolkit, we demonstrated fast and accurate strain-level detection can be achieved from fecal metagenomes to Kp abundances as low as 0.1%, matching culture-detected strains in 16/19 samples.[37] Interestingly, only one ST type was detected in most samples, suggesting gut microbiomes may be largely dominated by a single Kp strain. This contrasts with recent small sample sized culture-based studies in which multiple carriage strains were found with much higher frequency.[56–58] More robust longitudinal studies are needed to determine whether these culture-detected strains represent true gastrointestinal colonizers versus low abundance transitory passengers that are being enriched by culture. Alternatively, the partial enrichment of Enterobacterales, including Kp, as shown in ESwab compared to DNA preserved Norgen samples, may have led to a single ST-type selection overwhelming strain diversity in our Direct samples.

Although we demonstrated high accuracy of Kp strain detection by WMS in our samples, two false positive STs, ST12 and ST340, were detected in our mock microbiome in addition to the spiked strain, both of which were closely related to the spiked ST11. These may have arisen from the stricter database clustering we used to increase resolution between closely related ST types, e.g., ST11 and ST258, resulting in assignment of a single strain to multiple closely related reference genomes.

Mismatches between the Kp strain detected by culture and WMS also occurred on three occasions. Whether these differences were the result of misassignment by StrainGE, or alternatively, overgrowth of a low abundance non-dominant strain induced by culture, requires further study.

Our findings suggest a lower limit of 0.1% relative abundance for reliable retrieval of Kp-specific strain and allele-level information from metagenomes, as metagenomic detection of the ZKIR sequence, the Kp MLST alleles, and Kp ST-level detection occurred very seldomly below this level. While this is a considerable level of sensitivity, it nevertheless represented less than half of our Direct samples, which had a median Kp relative abundance of 0.007%, suggesting Kp gastrointestinal carriage typically occurs at lower abundances than this threshold. Our target sequencing depth of 20 million paired-end reads per sample, while shown to be sufficient for species-level detection, may have limited the amount of information recoverable from our samples at the subspecies-level.[59,60] Performing deeper sequencing in large-scale metagenomic studies, however, is challenging due to high costs and data storage and processing requirements. Strain-level detection performed on culture-enriched samples, such as our SCAI Sweep samples, using tools such as StrainGST or the recently described mSweep/mGEMS pipeline, or through targeted enrichment of metagenomes by RNA-probe hybridization-capture, may provide the most sensitive and cost-effective method for high-resolution strain analysis from metagenomes.[61,62] Studies are currently underway to explore these important possibilities.

Samples used in this study were initially collected for culture-based purposes, thus ESwab collection devices were used to maintain bacterial viability. The extensive KpSC culture and associated single colony WGS data gathered from these samples made them ideal for the purpose of this study. This collection method was also a major limitation, as it resulted in significant overgrowth of Enterobacterales, particularly *E. coli*, as shown when compared to the validated Norgen collection method. The resultant biased ESwab microbiome meant no Kp-microbiome association analyses could be performed from this data, nor any strong conclusions drawn regarding the normal relative abundance range of Kp gastrointestinal carriage. Further studies are currently underway utilizing the methods described here using validated microbiome collection devices to address these important questions.

In conclusion, we have shown the ZKIR-qPCR and WMS are reliable tools for detection and quantification of Kp within human gastrointestinal samples. Both methods exhibited differing and complementary strengths and weaknesses. This is evidenced by the speed, high sensitivity, and low cost of the ZKIR-qPCR, allowing targeted selection of samples for WMS, which, although less sensitive and more time and resource intensive, can provide in-depth microbiome and strain-level Kp analysis. Future studies using the methods evaluated herein therefore have great potential to enhance our understanding of Kp gastrointestinal ecology. Placed into a One Health context, these approaches will help in elucidating the role of the gastrointestinal tract of humans and animals in the spread of Kp and associated AMR genes between niches.

Materials and Methods

Human fecal samples

Fecal samples were drawn from a collection of 2975 KpSC culture-screened samples obtained during our cross-sectional KpSC carriage study and the seventh survey of the Tromsø Study, (The Tromsø Study: Tromsø7).[20] Briefly, samples were self-collected from community-based adult participants using a nylon-flocked ESwab 490CE.A (Copan, Brescia, Italy). Upon arrival to the laboratory, 200 µL of 85% glycerol was added and samples were stored at −80°C. Samples were screened for KpSC on SCAI media and suspect colonies underwent KpSC identification by MALDI-TOF.[19] Confirmed KpSC isolates underwent WGS and MLST-typing by Kleborate.[63] 103 samples were selected for the current study based on i) the presence/absence of Kp as confirmed by WGS, (n = 52 and 51, respectively) and ii) less than 2 days transit time from initial collection to freezing at −80°C. Prior to this study samples had undergone one freeze-thaw cycle.

Sample preparation and DNA isolation

After thawing on ice, 50 µL of each fecal sample was plated on SCAI media (Sigma-Aldrich, cat # 85462–500 G and I5125-500 G) and incubated at 37°C for 48 hours. Remaining sample was centrifuged (4000 x *g* for 10 minutes at 4°C) and pellet used for a whole microbiome DNA extraction (Direct sample). All growth on SCAI plates was scraped using a 10 µL inoculation loop, and approximately 50 µL (one loaded inoculation loop) used for a SCAI culture sweep DNA extraction (Sweep sample). DNA extractions were performed using the Purelink Microbiome DNA Purification kit (Thermo Fisher Scientific, cat# A29790), according to the manufacturer's instructions, with the following minor modifications: Step 1a) samples resuspended in 800 µL S1 Lysis Buffer plus 20 mg/mL lysozyme (Thermo Fisher Scientific, cat# 89833) and incubated at 37°C for 10 minutes. Step 1c) following addition of S2 Lysis Enhancer, samples were incubated at 95°C for 10 minutes. Step 1e) samples were homogenized in lysing matrix E tubes (MP Bio, cat# 6914050) using a Precellys Evolution tissue homogenizer (6500 rpm for 2 × 23 s at 4°C) (Bertin Technologies, Montigny Le Bretonneux, France), followed by 2 rounds of centrifugation at 14000 x *g* for 5 min. Step 1 h) prior to addition of S3 Cleanup Buffer, 2 µL of 10 mg/mL RNase A (Thermo Fisher Scientific, cat# EN0531) was added and samples were incubated at room temperature for 5 minutes. Quality control of purified DNA was performed using Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, USA) and concentration determined with Qubit 3.0 fluorometer (Thermo Fisher Scientific).

In silico analysis of ZKIR target region

BLAST analysis of the 78 bp ZKIR region was performed against the 484 assembled KpSC genomes from our previous carriage study (BioProject: PRJEB42350) using default nucleotide-nucleotide BLAST parameters (NCBI-blast v2.10.0+).[20,64] Melting temperature (T_M) for each ZKIR variant was calculated using the oligo analysis tool available at: https://eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/. Final amplicon sequences for T_M calculation consisted of forward and reverse ZKIR primers plus the 30 bp intervening region from each ZKIR sequence variant found by BLAST analysis.

ZKIR-qPCR assays

Reaction mixture, primers, and cycling conditions

PCR mixture, ZKIR primers, and cycling conditions were as described by Barbier *et al.*[30] All qPCR reactions were performed on an Applied Biosystems 7500 Real-Time PCR System (Thermo Fisher Scientific).

Standard curve

ZKIR-qPCR standard curves were prepared using whole genome sequenced representatives of each of the four human-associated KpSC members: Kp1, Kp2, Kp3 and Kp4 (Suppl. Table 1). Strains were grown overnight on tryptose blood agar with lactose/bromothymolblue (Thermo Fisher Scientific cat# CM0233, Thermo Fisher Scientific cat# LP0070, VWR cat# 1.03026.0025) at 37°C, and DNA extraction performed as described. Seven five-fold dilutions of genomic DNA (gDNA) were made for each isolate at $2.5x10^5$, $5x10^4$, 10^4, $2x10^3$, 400, 80, 16 and 3 genome copies per qPCR reaction, according to the equation: *genome copy number = [(mass of input DNA in ng) * (6.0221*10^23 molecules/mole)]/(length of genome in bp * 660 g/mol * 10^9ng/g)*, where length of Kp genome = $5.5x10^6$ bp.[65] Each dilution point was performed in technical triplicate. Reactions were performed both with and without addition of 25 ng of human fecal microbiome DNA from a healthy donor which was KpSC-negative by the ZKIR-qPCR. Slope, reaction efficiency, R^2, Y-intercept, and melting temperatures (T_M) were calculated using 7500 Real-Time PCR Analysis Software v2.3 (Applied Biosystems, Life technologies, Waltham, USA).

Limit of Detection (LOD) and Limit of Quantification (LOQ)

Limit of Detection (LOD) is defined by the Minimum Information for the Publication of Quantitative Real-Time PCR Experiments as the lowest concentration of target detectable with reasonable certainty.[66] LOD was therefore taken as the lowest number of genome copies detectable in ten out of ten technical replicates. LOQ was estimated

as the lowest dilution at which the coefficient of variation (CV) of genome copy number of ten technical replicates was ≤ 35%, where genome copy number = $[1+(efficiency/100)]^{y\text{-}Cq}$, and CV = *[(standard deviation of genome copy number)/(average of genome copy number)] * 100*.[33] LOD and LOQ were performed using gDNA from each of the four KpSC strains above with and without addition of 25 ng KpSC negative human fecal microbiome DNA. ZKIR-qPCR assays were performed as described for Standard Curve, with an additional dilution point at eight genome copies/reaction, and dilution points 16, 8 and 3 genome copies/reaction were performed in 10 technical replicates.

Detection of Kp by the ZKIR-qPCR

All Direct and Sweep DNA samples diluted to 10 ng/μL and 2.5 μL (25 ng) was used as input for each qPCR reaction. Reaction mixture and cycling conditions were as described previously.[30] Samples were assayed in technical triplicate and considered positive if amplicons were produced in at least two with a T_M between 78.3°C and 80.4°C and $C_q < 40$. T_M range was based on values from the *in silico* analysis of KpSC isolates (described above) ± 0.5°C for inter-assay variability between predicted and measured values. Although microbial detection by qPCR requires amplicons to be present in only a single technical replicate, we increased this threshold to two positive replicates to minimize false positives.[32,33] Non-template controls were used in all qPCR experiments. Additionally, *E. coli* underwent all processing steps from DNA extraction to ZKIR-qPCR assay in parallel with Direct samples, and *Klebsiella oxytoca* underwent all steps from culture on SCAI media, DNA extraction and ZKIR-qPCR in parallel with Sweep samples (Suppl. Table 1). As both these species do not contain the ZKIR sequence, this controlled for cross-contamination of KpSC DNA between samples at any of the sample processing steps.

WMS sample processing and analysis

Library preparations and sequencing

DNA was fragmented using the Focused-ultrasonicator M220 (Covaris, Woburn, USA). 100ng of fragmented DNA underwent library preparation using TrueSeq Nano DNA Library Prep Kit (Illumina, cat# 20015965) and Swift Turbo 2S flex DNA Library Prep Kit (Swift Biosciences, cat# 45096) in accordance with the manufacturer's instructions. Sequencing was performed on the NovaSeq 6000 platform (Illumina, San Diego, USA) to a target depth of 20×10^6 pair-end reads at 150 bp.

Data processing

FASTQ files underwent removal of optical duplicates using Clumpify (version 38.82), a part of the BBmap package (version 38.79), removal of adapters and poor-quality sequences by fastp (version 0.20.1), and removal of human DNA residues by FastQ Screen (version 0.14.0) against the GRCh38 reference assembly (accession number GCF_000001405.39).[67–69] Unpaired reads were synchronized by the Repair tool of BBmap package (version 38.79).[67]

WMS assembly and Taxonomic profiling

Paired-end and singleton reads were assembled into contigs using MetaSPAdes (v3.13.0) with default parameters.[70] Kp detection and estimation of abundance was performed using the taxonomy profiler Centrifuge (version 1.0.4) with the default database, p_compressed+h + v.[44] Centrifuge uses a Burrows-Wheeler transform (BWT) and Ferragina-Manzini (FM) index to create a comparatively small reference database by concatenating and compressing multiple genomes of the same species for rapid and accurate species identification.[44] For other taxonomic profiling, Kraken 2 (version 2.1.2) and Bracken (version 2.6.1) with the MiniKraken DB_8GB v202003 were used.[71,72]

Screening WMS assemblies for Kp-specific sequences

The seven Kp MLST alleles, downloaded from the PasteurMLST database, and the 78bp ZKIR sequence were used as reference databases for identification of Kp in the WMS-assembled contigs.[14] To screen the contigs, nucleotide BLAST (v2.10.1) was used with DNA identity and coverage parameters set to ≥ 95% (MLST allele detection) or default parameters (ZKIR sequence detection).[64]

Computational resources

All computational analyses were performed on the Norwegian academic high-performance computing and storage services maintained by the Sigma2 Norwegian Research Infrastructure Service (NRIS).[73] Data was stored and shared in the Norwegian e-infrastructure for Life Science (NeLS) maintained by ELIXIR Norway.

Validation of ESwab versus Norgen collected fecal samples

Ten previous Tromsø7 participants were re-recruited as part of an ongoing longitudinal Kp gut carriage study. Participants were sampled using ESwabs, under the same conditions as the original collection including less than 2 days from sample collection to arrival at the laboratory, and compared to collection taken at the same time using the Norgen Nucleic Acid Preservation system (Norgen Biotek, cat# 53700). All samples underwent library preparation using MGIEasy FS DNA Library Prep Set v2.1 (MGI Tech Co, cat# 1000005254) on the 7-MGISP-960 automated library preparation system (software version: V1.2.0.163, automation version: V1.0), as per manufacturer's instructions (MGI Tech Co, Shenzen, China). Sequencing was performed on the G400 platform (MGI Tech Co). Processing of sequenced reads and taxonomic profiling was performed as described.

In silico binary species mixes

FASTQ sequence reads from *B. fragilis, K. aerogenes*, and *E. coli* (Suppl. Table 1) were retrieved from the Sequence Read Archive (NCBI) and subsampled using SEQTK (https://github.com/lh3/seqtk/blob/master/README.md). Subsampled reads were combined to create two binary species mixes containing reads from *B. fragilis* and either *K. aerogenes* or *E. coli* in the following ratios: 99.99/0.01, 99.95/0.05, 99/1, 95/5, 90/10, 75/25, 50/50, 25/75, 0/100. Binary species mixes underwent processing and taxonomy profiling as described samples above.

Kp-spiked microbiomes

Kp-spiked mock microbiome

The mock microbiome was constructed from six bacterial strains: *Bacteroides vulgatus, Clostridium septicum, Bifidobacterium longum, Helicobacter pylori, Aeromonas hydrophila*, and *E. coli* (Suppl. Table 1). gDNA was extracted from each strain and combined in the following relative abundance calculated on genome copy numbers: 40% *B. vulgatus*, 40% *C. septicum*, 10% *B. longum*, 5% *H. pylori*, 4% *A. hydrophila*, and 1% *E. coli*. Abundances represented typical relative abundance of major phyla found in a healthy adult gut microbiome.[34,35] Kp ST11 gDNA was spiked into six mock microbiome aliquots at relative abundance: 0%, 0.001%, 0.01%, 0.1%, and 1%, as well as 1% Kp ST11 plus 1% Kp3 ST697 (Suppl. Table 1).

Kp-spiked fecal microbiome

Whole microbiome DNA was extracted from a fecal sample collected from a healthy adult donor using the Norgen Stool Nucleic Acid Preservation system (Norgen Biotek, cat# 53700) and confirmed KpSC negative by the ZKIR-qPCR. Total bacterial abundance was estimated by qPCR quantification of the bacterial 16S gene using the universal 16S primers described by Clifford *et al.*[65] qPCR reaction mixture, cycling conditions were as described by Barbier *et al.*[30] 0.25ng microbiome DNA was used as input, and standard curve was set up as for the ZKIR-qPCR above except the following five-fold dilution series was used: $1.25x10^6$, $2.5x10^5$, $5x10^4$, 10^4, $2x10^3$, 400, 80, and 16 genome copies per qPCR reaction. gDNA from Kp ST11, ST23, and ST101 (Suppl. Table 1) were combined in the ratio 60:30:10, respectively, and spiked into aliquots of the donor microbiome DNA at 0%, 0.001%, 0.01%, 0.1%, and 1%. All Kp-spiked mock and fecal microbiome samples underwent WMS sequencing, sample processing, and taxonomic analysis as described.

Kp strain analysis

Kp strain analysis was performed using StrainGST, part of the Strain Genome Explorer (StrainGE) toolkit.[37] A custom database of KpSC genomes (n = 3604) was constructed with default k-mer size 23. The database consisted of i) all Kp genomes from refseq (NCBI) (n = 1010), downloaded on the 02/02/2022 using NCBI Genome Downloading Scripts, (https://github.com/kblin/ncbi-genome-download), ii) 484 KpSC genomes from our KpSC carriage study

(303 Kp1, 134 Kp3, 31 Kp2, and 16 Kp4 genomes), and iii) 2109 KpSC genomes from the recent SPARK study (1705 Kp1, 279 Kp3, 76 Kp2, and 49 Kp4 genomes).[20,74] The default lower limit for database clustering of 0.90 k-mer similarity resulted in closely related ST types co-clustered (e.g., ST11 and ST258), thus, a lower limit of 0.95 was used for final database clustering.

Statistical analysis

Statistical differences between sample groups were determined using a one-tailed Mann Whitney U test (independent samples) or one-sided Wilcoxon signed-rank test (paired samples) using Jasp version 0.14.1 (University of Amsterdam, Amsterdam, Netherlands) (https://jasp-stats.org/download/). Correlation analysis of qPCR and Centrifuge Kp abundances was performed using R Studio version 3.6.1. p-values <0.05 were regarded as statistically significant.

Acknowledgments

We thank Jukka Corander for constructive feedback during the bioinformatic analysis, and thanks to Cyril Muluh and Ellen Haldis Josefsen for their technical assistance. We thank Sigma2, the national provider of e-infrastructure, for access to High-Performance Computing and large-scale data storage (Project ID: NS9794K).

Author contributions

AS conceptualized the study and acquired funding in collaboration with ØS, KH, and SB. KL was responsible for selecting samples, culture, and DNA extraction with support and expertise from NR, TP, ØS, and AS. KL was responsible for qPCR with support and expertise from TP, SB, and AS. KL and DB arranged samples for WMS which was performed by YS. Bioinformatic analysis was performed by DB and KL supported with expertise from EH and KH. DB was responsible for data storage and public availability of data supported by EH. Statistical analysis was performed by KL. KL prepared the first draft of the manuscript in collaboration with DB and AS. All authors contributed to review and editing of the manuscript and approved the final version.

Data availability

Metagenomic data (raw Illumina/MGI reads) are publicly available in ENA under BioProject: PRJEB52877.

Disclosure statement

All authors report no conflicts of interest.

Ethics

Collection and analysis of samples for this study was approved by the Regional Committee for Medical and Health Research Ethics, North Norway (REC North reference: 2016/1788, 2014/940, and 137064). All participants signed informed consent forms and the study was performed in compliance with the Declaration of Helsinki.

Funding

This study was supported by the Trond Mohn Foundation, contract TMF2019TMT03 to AS, and Northern Norway Regional Health Authority, contracts HNF1589-21 to AS and HNF1415-18 to ØS. KL was funded by a PhD-fellowship from UiT-The Arctic University of Norway. The funders of the study played no role in study design, data collection, data analysis, data interpretation, or writing of the repor; Trond Mohn stiftelse

ORCID

Kenneth Lindstedt http://orcid.org/0000-0002-5919-2479
Dorota Buczek http://orcid.org/0000-0002-2282-6298
Torunn Pedersen http://orcid.org/0000-0002-9593-8119
Erik Hjerde http://orcid.org/0000-0002-6014-1249
Niclas Raffelsberger http://orcid.org/0000-0002-9463-8915
Yutaka Suzuki http://orcid.org/0000-0003-4852-1879
Sylvain Brisse http://orcid.org/0000-0002-2516-2108
Kathryn Holt http://orcid.org/0000-0003-3949-2471
Ørjan Samuelsen http://orcid.org/0000-0002-5525-2614
Arnfinn Sundsfjord http://orcid.org/0000-0002-3728-2270

References

1. Navon-Venezia S, Kondratyeva K, Carattoli A. *Klebsiella pneumoniae*: a major worldwide source and shuttle for antibiotic resistance. Fems Microbiol Rev. 2017;41:252–275. doi:10.1093/femsre/fux013
2. Wyres KL, Holt KE. *Klebsiella pneumoniae* as a key trafficker of drug resistance genes from environmental to clinically important bacteria. Curr Opin Microbiol. 2018;45:131–139. doi:10.1016/j.mib.2018.04.004
3. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, Pulcini C, Kahlmeter G, Kluytmans J, Carmeli Y, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect Dis. 2018;18(3):318–327. doi:10.1016/S1473-3099(17)30753-3

4. Antimicrobial Resistance C. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399:629–655. doi:10.1016/S0140-6736(21)02724-0
5. Scheuerman O, Schechner V, Carmeli Y, Gutierrez-Gutierrez B, et al. Comparison of predictors and mortality between bloodstream infections caused by ESBL-Producing *Escherichia coli* and ESBL-Producing *Klebsiella pneumoniae*. Infect Control Hosp Epidemiol. 2018;39:660–667. doi:10.1017/ice.2018.63
6. Xu L, Sun X, Ma X. Systematic review and meta-analysis of mortality of patients infected with carbapenem-resistant *Klebsiella pneumoniae*. Ann Clin Microbiol Antimicrob. 2017;16:18. doi:10.1186/s12941-017-0191-3
7. Podschun R, Ullmann U. Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clin Microbiol Rev. 1998;11:589–603. doi:10.1128/CMR.11.4.589
8. Wyres KL, Lam MMC, Holt KE. Population genomics of *Klebsiella pneumoniae*. Nat Rev Microbiol. 2020;18:344–359. doi:10.1038/s41579-019-0315-1
9. Russo TA, Marr CM. Hypervirulent *Klebsiella pneumoniae*. Clin Microbiol Rev. 2019;32:e00001–19.
10. Rosenblueth M, Martinez L, Silva J, Martinez-Romero E. Klebsiella variicola, a novel species with clinical and plant-associated isolates. Syst Appl Microbiol. 2004;27:27–35. doi:10.1078/0723-2020-00261
11. Brisse S, Passet V, Grimont PAD. Description of Klebsiella quasipneumoniae sp. nov., isolated from human infections, with two subspecies, Klebsiella quasipneumoniae subsp. quasipneumoniae subsp. nov. and Klebsiella quasipneumoniae subsp. similipneumoniae subsp. nov., and demonstration that Klebsiella singaporensis is a junior heterotypic synonym of Klebsiella variicola. Int J Syst Evol Microbiol. 2014;64:3146–3152. doi:10.1099/ijs.0.062737-0
12. Long SW, Linson SE, Ojeda Saavedra M, Cantu C, et al. Whole-Genome sequencing of a human clinical isolate of the novel species Klebsiella quasivariicola sp. nov. Genome Announc. 2017;5. doi:10.1128/genomeA.01057-17
13. Rodrigues C, Passet V, Rakotondrasoa A, Diallo TA, Criscuolo A, Brisse S. Description of Klebsiella africanensis sp. nov., Klebsiella variicola subsp. tropicalensis subsp. nov. and Klebsiella variicola subsp. variicola subsp. nov. Res Microbiol. 2019;170:165–170. doi:10.1016/j.resmic.2019.02.003
14. Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S. Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. J Clin Microbiol. 2005;43:4178–4182. doi:10.1128/JCM.43.8.4178-4182.2005
15. Gorrie CL, Mirceta M, Wick RR, et al. Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* Infection in intensive care patients. Clin Infect Dis. 2017;65:208–215. doi:10.1093/cid/cix270
16. Conlan S, Kong HH, Segre JA, Highlander SK. Species-level analysis of DNA sequence data from the NIH Human Microbiome Project. PLoS One. 2012;7:e47075. doi:10.1371/journal.pone.0047075
17. Lin YT, Siu LK, Lin JC, Chen TL, Tseng CP, Yeh KM, Chang F-Y, Fung C-P. Seroepidemiology of *Klebsiella pneumoniae* colonizing the intestinal tract of healthy Chinese and overseas Chinese adults in Asian countries. BMC Microbiol. 2012;12:13. doi:10.1186/1471-2180-12-13
18. Huynh BT, Passet V, Rakotondrasoa A, Diallo T, Kerleguer A, Hennart M, Lauzanne AD, 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;11:1287–1299. doi:10.1080/19490976.2020.1748257
19. 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;20:936–941. doi:10.1128/jcm.20.5.936-941.1984
20. Raffelsberger N, Hetland MAK, Svendsen K, Smabrekke L, Lohr IH, Andreassen LLE, Brisse S, Holt KE, Sundsfjord A, Samuelsen Ø, et al. Gastrointestinal carriage of *Klebsiella pneumoniae* in a general adult population: a cross-sectional study of risk factors and bacterial genomic diversity. Gut Microbes. 2021;13:1939599. doi:10.1080/19490976.2021.1939599
21. Martin RM, Cao J, Brisse S, et al. Molecular Epidemiology of Colonizing and Infecting Isolates of *Klebsiella pneumoniae*. mSphere. 2016;1:e00261–16.
22. Shimasaki T, Seekatz A, Bassis C, Rhee Y, Yelin RD, Fogg L, Dangana T, Cisneros EC, Weinstein RA, Okamoto K, et al. 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. Clin Infect Dis. 2019;68(12):2053–2059. doi:10.1093/cid/ciy796
23. Sun Y, Patel A, SantaLucia J, Roberts E, Zhao L, Kaye K, Rao K, Bachman MA. Measurement of Klebsiella intestinal colonization density to assess infection risk. mSphere. 2021;6(3):e0050021. doi:10.1128/mSphere.00500-21
24. Conlan S, Park M, Deming C, Thomas PJ, Young AC, Coleman H, Sison C, Weingarten RA, Lau AF, Dekker JP, et al. Plasmid Dynamics in KPC-Positive *Klebsiella pneumoniae* during Long-Term Patient Colonization. mBio. 2016;7(3). doi:10.1128/mBio.00742-16
25. Leon-Sampedro R, DelaFuente J, Diaz-Agero C, Crellen T, Musicha P, Rodriguez-Beltran J, de la Vega C, Hernández-García M, López-Fresneña N, Ruiz-Garbajosa P, et al. Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota

of hospitalized patients. Nat Microbiol. 2021;6 (5):606–616. doi:10.1038/s41564-021-00879-y
26. de Boer RF, Ott A, Kesztyus B, Kooistra-Smid AM. Improved detection of five major gastrointestinal pathogens by use of a molecular screening approach. J Clin Microbiol. 2010;48(11):4140–4146. doi:10.1128/JCM.01124-10
27. Buchan BW, Olson WJ, Pezewski M, Marcon MJ, Novicki T, Uphoff TS, Chandramohan L, Revell P, Ledeboer NA. Clinical evaluation of a real-time PCR assay for identification of Salmonella, Shigella, Campylobacter (Campylobacter jejuni and C. coli), and shiga toxin-producing *Escherichia coli* isolates in stool specimens. J Clin Microbiol. 2013;51:4001–4007. doi:10.1128/JCM.02056-13
28. Munk P, Andersen VD, de Knegt L, Jensen MS, Knudsen BE, Lukjancenko O, Mordhorst H, Clasen J, Agersø Y, Folkesson A, et al. A sampling and metagenomic sequencing-based methodology for monitoring antimicrobial resistance in swine herds. J Antimicrob Chemother. 2017;72:385–392. doi:10.1093/jac/dkw415
29. Abayasekara LM, Perera J, Chandrasekharan V, Gnanam VS, Udunuwara NA, Liyanage DS, Bulathsinhala NE, Adikary S, Aluthmuhandiram JVS, Thanaseelan CS, et al. Detection of bacterial pathogens from clinical specimens using conventional microbial culture and 16S metagenomics: a comparative study. BMC Infect Dis. 2017;17(1):631. doi:10.1186/s12879-017-2727-8
30. 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*. Appl Environ Microbiol. 2020;86(7). doi:10.1128/AEM.02711-19
31. 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. doi:10.1128/spectrum.02376-21
32. Kralik P, Ricchi M. A Basic Guide to Real Time PCR in Microbial Diagnostics: definitions, Parameters, and Everything. Front Microbiol. 2017;8:108. doi:10.3389/fmicb.2017.00108
33. MC KKE, Allison MJ, Goldberg CS, Helbing CC, Hunter ME, Jackson CA, Lance RF, Mangan AM, Monroe EM, Piaggio AJ, et al. Reporting the limits of detection and quantification for environmental DNA assays. Environmental DNA. 2019;00:1–12.
34. 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:e0206484. doi:10.1371/journal.pone.0206484
35. Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, Huttenhower C, Izard J. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biol. 2012;13:R42. doi:10.1186/gb-2012-13-6-r42
36. Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MCW, Wong MCS, Ng SC, Chan FKL, Chan PKS. Impact of Preservation Method and 16S rRNA Hypervariable Region on Gut Microbiota Profiling. mSystems. 2019;4(1). doi:10.1128/mSystems.00271-18
37. van Dijk LR, Walker BJ, Straub TJ, Worby CJ, Grote A, SchreiberIV HL, Anyansi C, Pickering AJ, Hultgren SJ, Manson AL, et al. StrainGE: a toolkit to track and characterize low-abundance strains in complex microbial communities. Genome Biol. 2022;23:74. doi:10.1186/s13059-022-02630-0
38. Passet V, Brisse S. Association of tellurite resistance with hypervirulent clonal groups of *Klebsiella pneumoniae*. J Clin Microbiol. 2015;53:1380–1382. doi:10.1128/JCM.03053-14
39. Ohtomo R, Saito M. A New Selective Medium for Detection of Klebsiella from Dairy Environments. Microbes and Environments. 2003;18:138–144. doi:10.1264/jsme2.18.138
40. Suss B, Flekna G, Wagner M, Hein I. Studying the effect of single mismatches in primer and probe binding regions on amplification curves and quantification in real-time PCR. J Microbiol Methods. 2009;76:316–319. doi:10.1016/j.mimet.2008.12.003
41. Whiley DM, Sloots TP. Sequence variation in primer targets affects the accuracy of viral quantitative PCR. J Clin Virol. 2005;34:104–107. doi:10.1016/j.jcv.2005.02.010
42. Christopherson C, Sninsky J, Kwok S. The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. Nucleic Acids Res. 1997;25:654–658. doi:10.1093/nar/25.3.654
43. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking Metagenomics Tools for Taxonomic Classification. Cell. 2019;178:779–794. doi:10.1016/j.cell.2019.07.010
44. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26:1721–1729. doi:10.1101/gr.210641.116
45. Watts GS, Thornton JEsJr., Youens-Clark K, Ponsero AJ, Slepian MJ, Menashi E, Hu C, Deng W, Armstrong DG, Reed S, et al. Identification and quantitation of clinically relevant microbes in patient samples: comparison of three k-mer based classifiers for speed, accuracy, and sensitivity. PLoS Comput Biol. 2019;15:e1006863. doi:10.1371/journal.pcbi.1006863
46. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13:R79. doi:10.1186/gb-2012-13-9-r79

47. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, Hansen TH, Liang S, Feng Q, Zhang C, et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. Nat Microbiol. 2018;3:1255–1265. doi:10.1038/s41564-018-0257-9
48. Frost F, Kacprowski T, Ruhlemann M, Pietzner M, Bang C, Franke A, Nauck M, Völker U, Völzke H, Dörr M, et al. Long-term instability of the intestinal microbiome is associated with metabolic liver disease, low microbiota diversity, diabetes mellitus and impaired exocrine pancreatic function. Gut. 2021;70:522–530. doi:10.1136/gutjnl-2020-322753
49. Nordberg V, Jonsson K, Giske CG, Iversen A, Aspevall O, Jonsson B, Camporeale A, Norman M, Navér L. Neonatal intestinal colonization with extended-spectrum beta-lactamase-producing Enterobacteriaceae-a 5-year follow-up study. Clin Microbiol Infect. 2018;24:1004–1009. doi:10.1016/j.cmi.2017.12.028
50. Lerner A, Adler A, Abu-Hanna J, Cohen Percia S, Kazma Matalon M, Carmeli Y. Spread of KPC-producing carbapenem-resistant Enterobacteriaceae: the importance of super-spreaders and rectal KPC concentration. Clin Microbiol Infect. 2015;21:470 e1–7. doi:10.1016/j.cmi.2014.12.015
51. Wyres KL, Holt KE. *Klebsiella pneumoniae* Population Genomics and Antimicrobial-Resistant Clones. Trends Microbiol. 2016;24:944–956. doi:10.1016/j.tim.2016.09.007
52. Rettedal S, Lohr IH, Natas O, Giske CG, Sundsfjord A, Oymar K. First outbreak of extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* in a Norwegian neonatal intensive care unit; associated with contaminated breast milk and resolved by strict cohorting. APMIS. 2012;120:612–621. doi:10.1111/j.1600-0463.2012.02879.x
53. Giani T, Arena F, Vaggelli G, Conte V, Chiarelli A, Henrici De Angelis L, Fornaini R, Grazzini M, Niccolini F, Pecile P, et al. Large Nosocomial Outbreak of Colistin-Resistant, Carbapenemase-Producing *Klebsiella pneumoniae* Traced to Clonal Expansion of an mgrB Deletion Mutant. J Clin Microbiol. 2015;53:3341–3344. doi:10.1128/JCM.01017-15
54. Politi L, Gartzonika K, Spanakis N, Zarkotou O, Poulou A, Skoura L, Vrioni G, Tsakris A. Emergence of NDM-1-producing *Klebsiella pneumoniae* in Greece: evidence of a widespread clonal outbreak. J Antimicrob Chemother. 2019;74:2197–2202. doi:10.1093/jac/dkz176
55. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, Abudahab K, Goater R, Giani T, Errico G, et al. Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. Nat Microbiol. 2019;4:1919–1929. doi:10.1038/s41564-019-0492-8
56. Lepuschitz S, Hauser K, Schriebl A, Schlagenhaufen C, Stoger A, Chakeri A, Vötsch K, Pekard-Amenitsch S, Springer B, Allerberger F, et al. Fecal *Klebsiella pneumoniae* Carriage Is Intermittent and of High Clonal Diversity. Front Microbiol. 2020;11:581081. doi:10.3389/fmicb.2020.581081
57. Chen CM, Tang HL, Chiou CS, Tung KC, Lu MC, Lai YC. Colonization dynamics of *Klebsiella pneumoniae* in the pet animals and human owners in a single household. Vet Microbiol. 2021;256:109050. doi:10.1016/j.vetmic.2021.109050
58. 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. J Clin Microbiol. 2019;57. doi:10.1128/JCM.01537-18
59. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. Evaluating the Information Content of Shallow Shotgun Metagenomics. mSystems. 2018;3(6). doi:10.1128/mSystems.00069-18
60. Gweon HS, Shaw LP, Swann J, De Maio N, AbuOun M, Niehus R, Hubbard ATM, Bowes MJ, Bailey MJ, Peto TEA, et al. The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. Environ Microbiome. 2019;14:7. doi:10.1186/s40793-019-0347-1
61. Maklin T, Kallonen T, David S, Boinett CJ, Pascoe B, Meric G, Aanensen DM, Feil EJ, Baker S, Parkhill J, et al. High-resolution sweep metagenomics using fast probabilistic inference. Wellcome Open Res. 2020;5:14. doi:10.12688/wellcomeopenres.15639.1
62. Guitor AK, Raphenya AR, Klunk J, Kuch M, Alcock B, Surette MG, McArthur AG, Poinar HN, Wright GD. Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes. Antimicrob Agents Chemother. 2019;64(1). doi:10.1128/AAC.01324-19
63. 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;12:4188. doi:10.1038/s41467-021-24448-3
64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–410. doi:10.1016/S0022-2836(05)80360-2
65. 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:e48558. doi:10.1371/journal.pone.0048558

66. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55:611–622. doi:10.1373/clinchem.2008.112797
67. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. LBNL-7065E. 2014. Berkeley, CA: Lawrence Berkeley National Laboratory.
68. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–i90. doi:10.1093/bioinformatics/bty560
69. Wingett SW, Andrews S. FastQ Screen: a tool for multi-genome mapping and quality control. F1000Res. 2018;7:1338. doi:10.12688/f1000research.15931.1
70. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–834. doi:10.1101/gr.213959.116
71. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. Peerj Comput Sci. 2017. doi:10.7717/peerj-cs.142
72. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:257. doi:10.1186/s13059-019-1891-0
73. HPC. The Norwegian academic high-performance computing and storage services. 2019.
74. Thorpe H, Booton R, Kallonen T, Gibbon MJ, Couto N, Passet V, Fernandez JSL, Rodrigues C, Matthews L, Mitchell S, et al. One Health or Three? Transmission modelling of Klebsiella isolates reveals ecological barriers to transmission between humans, animals and the environment. bioRxiv. 2021. doi:10.1101/2021.08.05.455249

Supplementary Material

Detection of *Klebsiella pneumoniae* human gut carriage: a comparison of culture, qPCR, and whole metagenomic sequencing methods

Kenneth Lindstedt (ORCID **0000-0002-5919-2479**)[a*], Dorota Buczek (ORCID 0000-0002-2282-6298)[a], Torunn Pedersen (ORCID 0000-0002-9593-8119)[b], Erik Hjerde (ORCID 0000-0002-6014-1249)[c], Niclas Raffelsberger (ORCID 0000-0002-9463-8915)[a,d], Yutaka Suzuki (0000-0003-4852-1879)[e], Sylvain Brisse (ORCID 0000-0002-2516-2108)[f], Kathryn Holt (ORCID 0000-0003-3949-2471)[g,h], Ørjan Samuelsen (ORCID 0000-0002-5525-2614)[b,i], Arnfinn Sundsfjord (ORCID **0000-0002-3728-2270)**[a,b*]

[a]*Department of Medical Biology, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway;* [b]*Norwegian National Advisory Unit on Detection of Antimicrobial Resistance, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway;* [c]*Department of Chemistry, UiT The Arctic University of Norway, Tromsø, Norway;* [d]*Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway;* [e]*Department of Computational Biology and Medical Sciences, The University of Tokyo, Tokyo, Japan;* [f]*Biodiversity and Epidemiology of Bacterial Pathogens Unit, Institut Pasteur, Université Paris Cité, Paris, France;* [g]*Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Australia;* [h]*Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, UK;* [i]*Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway*

*Corresponding authors: Kenneth Lindstedt, email: kenneth.w.lindstedt@uit.no, and Arnfinn Sundsfjord, arnfinn.sundsfjord@uit.no. Department of Medical Biology, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, 9038, Norway

Table of Contents

Supplementary Table 1. Bacterial strains used in this study

Strain	Species	Use	Reference
K47-25 (ST258)	Kp1	ZKIR qPCR reaction efficiency/standard curve	[1]
T7-004 (ST681)	Kp3	ZKIR qPCR reaction efficiency	[2]
T7-071 (ST4653)	Kp2	ZKIR qPCR reaction efficiency	[2]
T7-021 (ST4625)	Kp4	ZKIR qPCR reaction efficiency	[2]
ATCC 25922	*E. coli*	ZKIR qPCR negative control (Direct samples)	NA
ATCC 43863	*K. oxytoca*	ZKIR qPCR negative control (Sweep samples)	NA
DSM 2151	*B. fragilis*	*In silico* binary species mix	SRR16258999
DSM 30053	*K. aerogenes*	*In silico* binary species mix	SRR15076202
ETEC H10407	*E. coli*	*In silico* binary species mix	ERR2910163
ATCC 8482	*B. vulgatus*	Mock microbiome	NA
ATCC 12464	*C. septicum*	Mock microbiome	NA
UNN S1	*B. longum*	Mock microbiome	ERS1507214[3]
ATCC 43504	*H. pylori*	Mock microbiome	NA
ATCC 7966	*A. hydrophila*	Mock microbiome	NA
ATCC 35218	*E. coli*	Mock microbiome	NA
K66-45 (ST11)	Kp1	Kp-spike strain	[4]
T7-263 (ST697)	Kp3	Kp3-spike strain	[2]
T7-442 (ST23)	Kp1	Kp-spike strain	[2]
P19-10 (ST101)	Kp1	Kp spike strain	[5]

Supplementary Table 2. Comparison of Limit of Detection (LOD) and Limit of Quantification (LOQ) of four human-associated KpSC members (Kp1-4) by the ZKIR-qPCR.

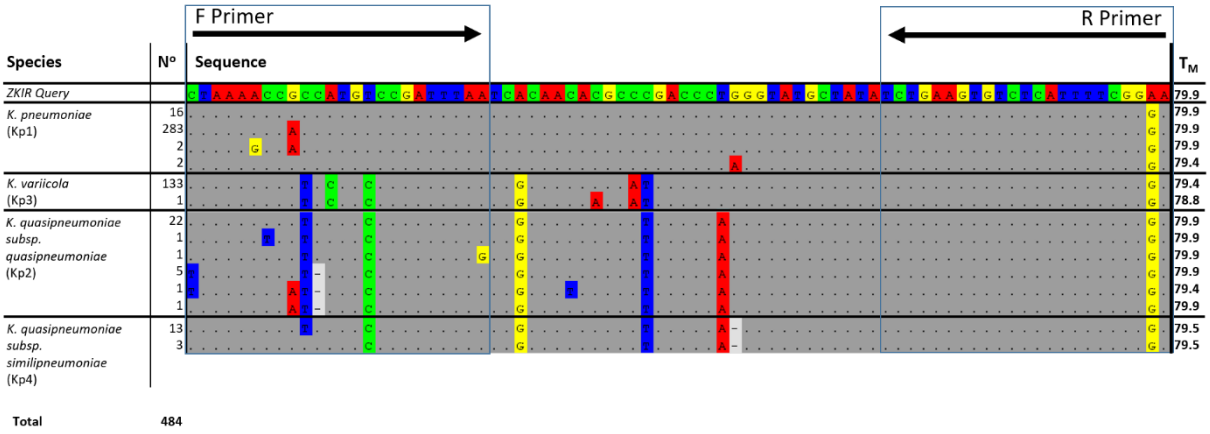
Species	Limit of Detection (LOD)[a]		Limit of Quantification (LOQ)[a]	
	gDNA only	gDNA with microbiome DNA[b]	gDNA only	gDNA with microbiome DNA[b]
K. pneumoniae (Kp1)	3 genomes	3 genomes	16 genomes **CV = 23.5%**	16 genomes **CV = 16.7%**
K. variicola (Kp3)	3 genomes	3 genomes	16 genomes **CV = 28.2%**	16 genomes **CV = 28.51%**
K. quasipneumoniae subsp. quasipneumoniae (Kp2)	3 genomes	3 genomes	16 genomes **CV = 35.2%**	16 genomes **CV = 27.5%**
K. quasipneumoniae subsp. similipneumoniae (Kp4)	3 genomes	3 genomes	16 genomes **CV = 23.5%**	16 genomes **CV = 21.78%**

[a]Limits were defined as lowest genome copy number which could be detected (LOD) or quantified (LOQ) in 10/10 technical replicates.

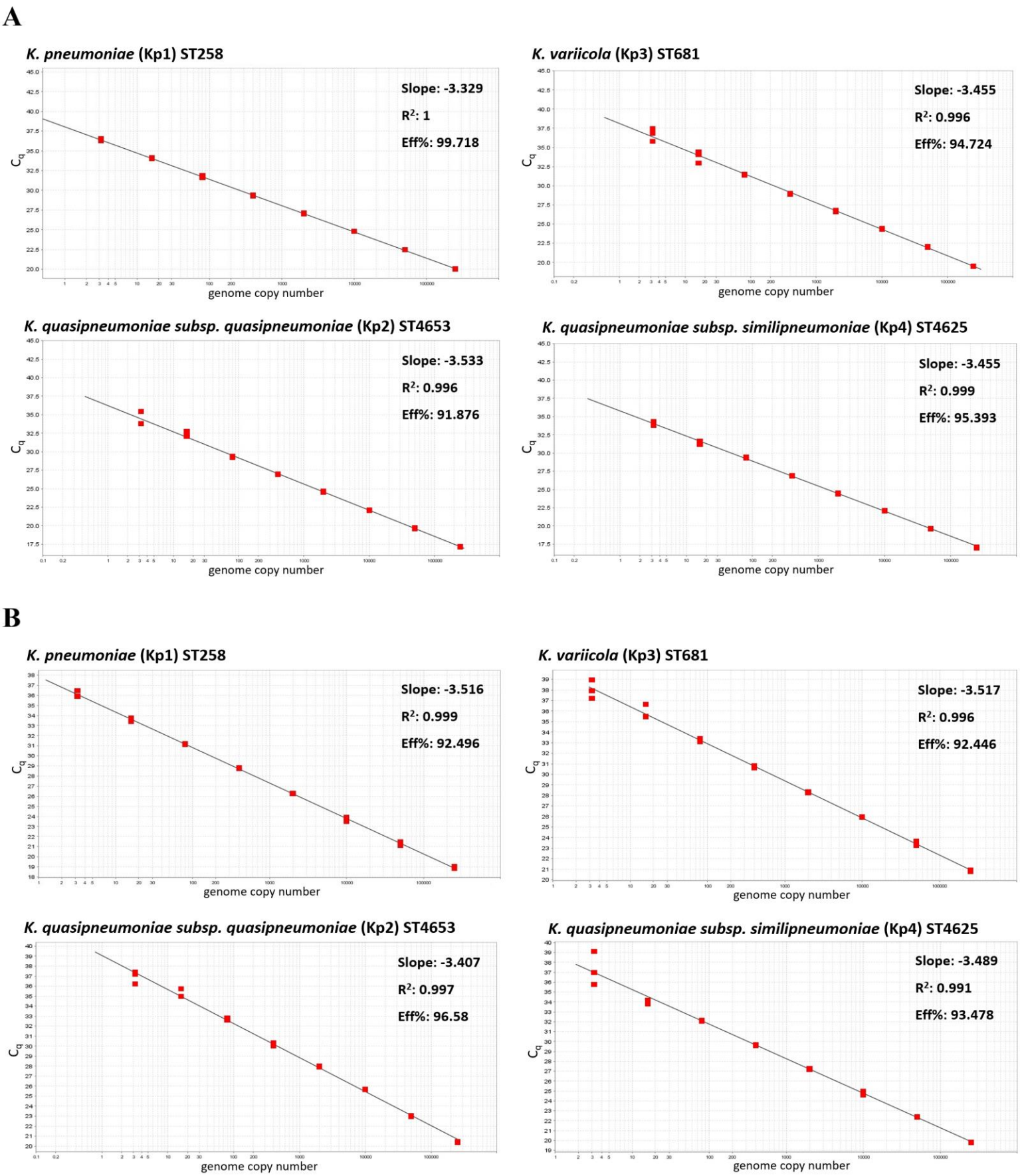
[b]Assays were performed as isolate genomic DNA (gDNA) only as well as challenged by the presence of 25 ng KpSC negative faecal microbiome DNA.

CV = coefficient of variation.

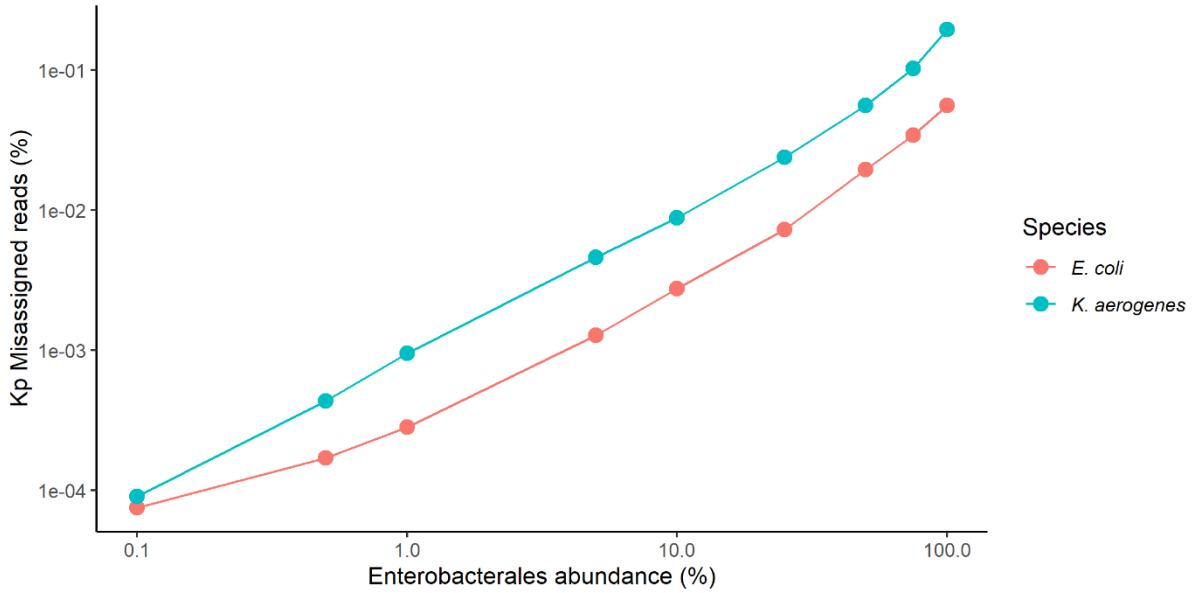
Supplementary 3, 4, and 5. Provided as Excel table



Supplementary Figure 1: Blast analysis of the 78bp ZKIR-qPCR target sequence against the 484 KpSC isolates detected during our previous Kp population carriage study.[2] Base-pair mismatches are highlighted. Calculated amplicon melting temperatures (T_M) (ºC) for each sequence variant are shown. Results visualised using AliView Alignment Viewer and Editor (v1.26).



Supplementary Fig 2: ZKIR-qPCR standard curves generated from representative isolates of each of the four human associated KpSC species. Assays were performed on isolate gDNA only (A) and in the presence of 25 ng KpSC negative faecal microbiome DNA (B) in technical triplicates from 250,000 genomes/reaction to 3 genomes per reaction. C_q = Quantification Cycle



Supplementary Figure 3: Proportion of reads misassigned to the *Kp* genome by Centrifuge from related Enterobacterales in *in silico* binary species mixes containing *B. fragilis* and increasing abundance of either *E. coli* (teal) or *K. aerogenes* (orange).

References

1. Samuelsen O, Naseer U, Tofteland S, Skutlaberg DH, Onken A, Hjetland R, et al. Emergence of clonally related Klebsiella pneumoniae isolates of sequence type 258 producing plasmid-mediated KPC carbapenemase in Norway and Sweden. J Antimicrob Chemother 2009; 63:654-8.
2. Raffelsberger N, Hetland MAK, Svendsen K, Smabrekke L, Lohr IH, Andreassen LLE, et al. Gastrointestinal carriage of Klebsiella pneumoniae in a general adult population: a cross-sectional study of risk factors and bacterial genomic diversity. Gut Microbes 2021; 13:1939599.
3. Esaiassen E, Hjerde E, Cavanagh JP, Simonsen GS, Klingenberg C, Norwegian Study Group on Invasive Bifidobacterial I. Bifidobacterium Bacteremia: Clinical Characteristics and a Genomic Approach To Assess Pathogenicity. J Clin Microbiol 2017; 55:2234-48.
4. Heikal A, Samuelsen O, Kristensen T, Okstad OA. Complete Genome Sequence of a Multidrug-Resistant, blaNDM-1-Expressing Klebsiella pneumoniae K66-45 Clinical Isolate from Norway. Genome Announc 2017; 5.
5. Pedersen T, Sekyere JO, Govinden U, Moodley K, Sivertsen A, Samuelsen O, et al. Spread of Plasmid-Encoded NDM-1 and GES-5 Carbapenemases among Extensively Drug-Resistant and Pandrug-Resistant Clinical Enterobacteriaceae in Durban, South Africa. Antimicrob Agents Chemother 2018; 62.

Supplementary Table 3

Supplementary Table 3 - Comparison of Kp detection by SCAI culture to the ZKIR-qPCR and WMS

seq number	SCAI-sweep qPCR (genome copies/ng DNA)	Direct-fecalqPCR (genome copies/ng DNA)	Culture (1 = positive, 0 = negative)	SCAI-sweep Centrifuge (%)	Direct-fecal Centrifuge (%)	SCAI-sweep ZKIR_blast (1 = positive, 0 = negative)	Direct-fecal ZKIR_blast (1 = positive, 0 = negative)	SCAI-sweep Kp MLST gene match	Direct-fecal Kp MLST gene match	SCAI-sweep Enterobacterales abundance (%)	Direct-fecal Enterobacterales abundance (%)	Direct-fecal E. coli abundance (%)
1	47466.67	0.17	1	27.1671000	0.0004	1	0	6/7	0/7	97.69	0.42	0.04
2	132400.00	605.33	1	91.0528000	1.1157	1	1	7/7	6/7	97.28	1.15	0.22
3	43240.00	3.88	1	16.6954000	0.0271	1	0	6/7	1/7	98.33	15.04	14.55
4	0.06	0.00	0	0.0041341	0.0000	0	0	1/7	0/7	97.94	0.36	0.07
5	0.05	0.00	0	0.2779390	0.0001	0	0	1/7	0/7	97.93	0.51	0.02
6	no growth	0.00	0	no growth	0.0001	no growth	0	no growth	0/7	no growth	0.34	0.02
7	30920.00	36.78	1	15.6781000	0.0618	1	0	5/7	2/7	98.1	28.75	27.26
8	9760.00	0.12	1	1.5321500	0.0012	1	0	5/7	0/7	98.17	0.35	0.24
9	0.06	0.00	0	0.0094919	0.0008	0	0	1/7	1/7	98.86	7.17	6.93
10	0.15	0.00	0	0.0501797	0.0000	0	0	2/7	0/0	97.65	0.35	0.04
11	0.01	0.04	1	0.1202310	0.0004	0	0	1/7	0/7	98	0.22	0.09
12	93000.00	5306.67	1	53.4967000	4.3727	1	1	7/7	5/7	98.9	57.07	36
13	49600.00	39.04	1	46.6466000	0.0944	1	0	7/7	0/7	50.65	0.17	0.03
14	145500.00	7.08	1	95.2601000	0.0135	1	0	7/7	0/7	97.09	0.06	0
15	0.89	0.00	0	0.0242805	0.0003	0	0	1/7	1/7	97.85	1.8	1.73
16	0.10	0.00	0	0.1751710	0.0030	0	0	2/7*	2/7	98.43	12.17	11.67
17	0.00	0.00	0	0.0021864	0.0001	0	0	1/7	0/7	97.97	6.36	6.09
18	38490.00	10120.00	1	15.6161000	15.2881	1	1	6/7	6/7	97.6	42.15	16.05
19	36900.00	10.89	1	19.6783000	0.0267	1	0	6/7	2/7	98.3	1.31	1.19
20	10640.00	0.33	1	3.2757000	0.0375	1	0	5/7	2/7	97.84	78.99	70.18
21	63300.00	493.33	1	19.8564000	0.1347	1	1	7/7	6/7	98.51	73.49	24.1
22	3175.00	1.18	1	2.7578900	0.0029	1	0	5/7	0/7	97.42	0.15	0.02
23	14880.00	29.93	1	4.2790400	0.0618	1	0	6/7	2/7	97.48	1.19	1.05
24	0.00	0.00	0	0.1053350	0.0711	0	0	2/7	2/7	97.95	53.18	52.38
25	0.00	0.00	0	0.0569806	0.0351	0	0	1/7	0/7	81.16	24.16	0.68
26	2241.00	5.79	0	1.1713500	0.0134	1	0	4/7	1/7	98.27	24.74	24.22
27	0.00	0.00	0	0.0701593	0.0001	0	0	1/7	1/7	99.04	1.06	0.95
28	no growth	0.19	0	no growth	0.0003	no growth	0	no growth	0/7	no growth	0.09	0
29	0.00	0.00	0	0.2354680	0.0013	0	0	2/7	0/7	97.45	0.12	0.09
30	0.00	0.00	0	0.0639786	0.0011	0	0	1/7	1/7	98.84	2.84	2.73
31	16.22	0.09	0	0.0217370	0.0002	0	0	2/7	1/7	31.59	1.08	0.77
32	0.08	0.04	0	0.2270570	0.0103	0	0	2/7	1/7	98.77	64.84	63.34
33	0.11	0.00	0	0.1114000	0.0031	0	0	2/7	2/7	96.82	1.39	1.09
34	80100.00	159.87	1	35.5232000	0.0938	1	1	7/7	1/7	98.85	8.9	8.55
35	68700.00	220.40	1	31.5129000	0.1344	1	0	7/7	3/7	97.61	0.82	0.46
36	106700.00	89.60	1	61.9730000	0.0934	1	0	7/7	1/7	97.8	41.86	41.03
37	0.00	0.11	0	0.0108182	0.0024	0	0	2/7	2/7	98.03	24.25	23.87
38	0.14	0.00	0	0.0138394	0.0010	0	0	2/7	2/7	97.37	10.56	10.12
39	0.05	0.00	0	0.3266130	0.0002	0	0	1/7	0/7	98.52	0.63	0.42
40	10310.00	0.52	1	4.9277100	0.0082	1	0	6/7	2/7	98.13	91.94	86.93
41	183200.00	189.20	1	97.2029000	0.0548	1	0	7/7	0/7	97.81	0.17	0.01
42	34420.00	14.35	1	10.8363000	0.0072	1	0	5/7	2/7	97.85	1.81	1.67
43	35990.00	0.18	1	59.6509000	0.0003	1	0	4/7	0/7	98.07	0.13	0.01
44	131700.00	1729.33	1	77.0745000	2.6353	1	1	7/7	5/7	98.46	10.64	8.08
45	99500.00	7400.00	1	41.8037000	8.3946	1	1	6/7	7/7	96.92	19.6	8.42
46	157700.00	924.00	1	85.9353000	1.5632	1	1	7/7	7/7	98.01	1.67	0.4
47	12850.00	3.32	1	5.3145300	0.0125	1	0	5/7	2/7	96.39	9.16	8.5
48	5510.00	0.23	0	2.1702000	0.0487	1	0	3/7	1/7	96.62	77.58	76.48
49	0.03	0.07	0	0.1607700	0.0021	0	0	2/7	2/7	97.93	1.06	0.79
50	0.00	0.00	0	0.0096351	0.0028	0	0	1/7	1/7	98.67	24.86	24.45
51	0.00	0.00	0	0.0023640	0.0000	0	0	2/7	1/7	97.1	0.14	0.03
52	0.00	0.00	0	0.2963090	0.0896	0	0	2/7	2/7	97.49	16.54	16.02

Supplementary Table 3 (Continued)

52	0.00	0.00	0	0.2963090	0.0896	0	0	2/7	2/7	97.49	16.54	16.02
53	0.00	0.32	0	0.2545720	0.0006	0	0	2/7	0/7	97.27	0.39	0.17
54	no growth	0.00	0	no growth	0.0001	no growth	0	no growth	0/7	no growth	0.24	0.02
55	0.00	0.00	0	0.0159129	0.0001	0	0	1/7	0/7	91.85	0.31	0.16
56	0.00	0.00	0	0.0500097	0.0002	0	0	1/7	0/7	97.39	0.43	0.26
57	0.00	0.00	0	0.0439509	0.0002	0	0	2/7	2/7	96.96	0.75	0.38
58	0.00	0.00	0	0.0766105	0.0003	0	0	1/7	1/7	98.09	0.68	0.54
59	0.00	0.00	0	0.0891014	0.0002	0	0	2/7	0/7	92.11	0.56	0.34
60	0.00	0.00	0	0.1559930	0.0001	0	0	2/7	0/7	97.4	0.28	0.02
61	0.00	0.00	0	0.1914110	0.0074	0	0	1/7	2/7	97.51	5.09	4.83
62	13.67	6.47	0	0.1845210	0.1040	0	0	2/7*	2/7*	98.11	63.9	4.16
63	1.46	0.00	0	0.1129590	0.0000	0	0	1/7	0/7	78.98	0.16	0.02
64	1.47	0.00	0	0.0158004	0.0159	0	0	1/7	1/7	96.43	60.55	59.5
65	52400.00	50.40	0	37.2196000	0.0038	1	0	7/7	2/7	97.66	43.99	38.09
66	12.69	0.00	0	0.3624190	0.0000	0	0	2/7	0/7	97.27	0.16	0.02
67	1.65	0.00	0	0.0275184	0.0126	0	0	2/7	2/7	98.17	36.13	26.97
68	0.00	0.00	0	0.0223370	0.0001	0	0	2/7	0/7	79.46	1.38	0.48
69	0.00	0.00	0	0.0049777	0.0000	0	0	1/7	0/7	98.27	0.19	0.02
70	6120.00	9.97	1	3.4939000	0.0187	1	0	6/7	0/7	76.2	0.95	0.73
71	444.00	0.05	1	0.1391000	0.0002	1	0	3/7	0/7	97.42	0.44	0.09
72	29960.00	304.53	1	20.0863000	0.2246	1	1	5/7	2/7	82.4	79.84	77.58
73	84000.00	40.12	1	47.8308000	0.0692	1	0	7/7	1/7	94.09	4.84	4.04
74	132400.00	8946.67	1	90.9906000	17.8094	1	1	7/7	7/7	97.7	14.27	0.13
75	151200.00	18133.33	1	97.9701000	22.0899	1	1	7/7	7/7	98.07	20.87	0.16
76	0.00	0.00	0	0.0617131	0.0002	0	0	2/7	2/7	98.81	1.03	0.88
77	0.00	0.00	0	3.0317900	0.0744	0	0	1/7	1/7	97.12	4.93	0.07
78	0.00	0.00	0	0.0156912	0.0000	0	0	2/7	0/7	97.55	0.64	0.24
79	0.00	0.00	0	1.3577400	0.0003	0	0	2/7	0/7	95.87	27.32	0.13
80	3576.00	0.00	0	2.6566400	0.2975	1	0	6/7	0/7	97.33	59.62	0.52
81	0.04	0.16	0	0.0512507	0.0231	0	0	1/7	1/7	98.79	38.7	38.14
82	0.05	0.00	0	0.0391432	0.0002	0	0	2/7	0/7	96.59	0.45	0.16
83	no growth	0.00	0	no growth	0.0000	no growth	0	no growth	0/7	no growth	0.21	0.02
84	2069.00	925.33	1	0.6618850	0.9487	1	1	5/7	6/7	95.21	10.1	8.89
85	12590.00	0.54	1	6.0515800	0.0178	1	0	6/7	2/7	97.3	7.22	6.8
86	168600.00	1.39	1	98.6748000	0.0026	1	0	7/7	0/7	99	0.13	0
87	2182.00	0.07	1	1.3073800	0.0000	1	0	4/7	0/7	97.84	0.34	0.04
88	10550.00	3.23	1	7.5532800	0.0103	1	0	5/7	0/7	95.92	0.4	0.07
89	111600.00	44533.33	1	70.2734000	39.5527	1	1	7/7	7/7	98.55	62.25	17.34
90	20820.00	1605.33	1	15.3346000	0.9814	1	1	5/7	5/7	98.61	67.42	62.86
91	115900.00	21920.00	1	98.5909000	14.7314	1	1	7/7	7/7	98.31	23.57	0.06
92	116900.00	18893.33	1	96.7098000	20.5545	1	1	7/7	7/7	98.73	22.43	0.72
93	0.54	19.63	1	0.0256789	0.0217	0	0	2/7	0/7	97.18	0.46	0.32
94	11020.00	0.80	1	4.8983200	0.0648	1	0	6/7	1/7	98.55	47.09	46.39
95	24340.00	0.26	1	11.6076000	0.0003	1	0	5/7	2/7	98.26	41.56	39.55
96	84600.00	1.18	1	39.4113000	0.0024	1	0	7/7	0/7	98.89	39.67	39
97	144000.00	9440.00	1	77.8163000	6.0247	1	1	7/7	6/7	98.36	43.21	35.31
98	12380.00	5.46	1	4.7337100	0.0072	1	0	6/7	0/7	95.79	0.33	0.24
99	3670.00	0.02	1	1.1409100	0.0010	1	0	5/7	1/7	98.46	1.2	1.09
100	53800.00	2810.67	1	24.6969000	1.7325	1	1	4/7	6/7	97.92	18.01	15.49
101	20300.00	41333.33	1	7.9513300	26.0352	1	1	5/7	6/7	98.16	65.29	30.68
102	154800.00	87.87	1	80.3980000	0.0392	1	1	6/7	2/7	99.31	1.77	1.63
103	65200.00	0.57	1	27.8006000	0.0020	1	0	7/7	0/7	98.65	0.4	0.27
mean	32764.27	1907.44	0.50	19.32	1.81	0.55	0.19	NA	NA	95.61	17.16	12.09
median	3175.00	0.18	1.00	1.53	0.007	1.00	0.00	NA	NA	97.84	1.80	0.79
SD	50204.93	6898.20	0.50	30.95	6.04	0.50	0.40	NA	NA	9.00	24.23	20.47

Supplementary Table 4

Supplementary Table 4: Taxonomic comparison of samples to two large microbiome profiling reference studies; Segata *et al.* (2012)[1] and King *et al* . (2019)[2]. (**A)** Abundance of major gut phyla in T7 Direct samples (n=103) compared to references. (B) Abundance of Enterobacterales (order) and Escherichia coli (species) in T7 Direct samples compared to references. (C) Abundance of Enterobacterales (order) in T7 Sweep samples compared to references.

Supplementary Table 4A

Phyla	Segata et al. (2012)[1]	King et al. (2019)[2]	T7 Direct Samples -%																				
	mean % +/- SD	mean % +/- SD	mean % +/- SD	DIR01	DIR02	DIR03	DIR04	DIR05	DIR06	DIR07	DIR08	DIR09	DIR10	DIR11	DIR12	DIR13	DIR14	DIR15	DIR16	DIR17	DIR18	DIR19	DIR20
Bacteroidetes	65.201 +/- 20.5	73.13 +/- 22.16	42.00 +/- 19.65	40.93	44.12	57.31	47.03	31.62	39.54	33.12	74.84	45.23	55.12	23.69	33.23	36.86	39.41	47.35	53.19	49.5	32.27	77.49	7.93
Firmicutes	29.637 +/- 19.260	22.20 +/- 18.66	32.52 +/- 17.45	49.83	46.65	20.78	43.97	56.43	52.69	31.32	22.51	37.13	33.77	69.37	7.5	53.12	55.26	45.88	30.87	40.13	20.94	19.93	9.44
Proteobacteria	2.910 +/- 4.735	2.15 +/- 10.39	18.94 +/- 24.79	3.61	3.07	16.67	2.35	5.87	2.58	30.77	1.44	8.75	4.67	1.74	57.87	2.37	0.38	2.21	12.53	6.52	43.25	1.48	79.13
Actinobacteria	0.507 +/- 2.819	1.82 +/- 3	4.22 +/- 4.09	3.53	3.97	4.09	4.75	4.22	3.82	3.63	0.79	7.84	4.43	4.35	1.33	7	1.13	4.48	2.37	0.89	2.68	0.66	1.91
Verrucomicrobia	0.442 +/- 1.736	0.70 +/- 1.68	1.78 +/- 3.11	0.05	1.46	0.1	1.13	0.13	0.59	0.36	0.02	0.1	0.24	0.03	0.04	0.15	3.75	0.02	0.02	2.91	0.38	0.38	0.8
Fusobacteria	0.072 +/- 0.5	ND	0.03 +/- 0.10	0.08	0.03	0.04	0.05	0.07	0.04	0.04	0.09	0.03	0.05	0.01	0	0.01	0	0	0.01	0	0.01	0	0

Supplementary Table 4B

Order/species	Segata et al. (2012)[1]	King et al. (2019)[2]	T7 Direct Samples - %																				
	mean % +/- SD	mean % +/- SD	mean +/- SD	DIR01	DIR02	DIR03	DIR04	DIR05	DIR06	DIR07	DIR08	DIR09	DIR10	DIR11	DIR12	DIR13	DIR14	DIR15	DIR16	DIR17	DIR18	DIR19	DIR20
Enterobacterales	0.159 +/- 0.957	2.02 +/- 10.33	17.16 +/- 24.23	0.42	1.15	15.04	0.36	0.51	0.34	28.75	0.35	7.17	0.35	0.22	57.07	0.17	0.06	1.8	12.17	6.36	42.15	1.31	78.99
Escherichia coli	NA	NA	12.09 +/- 20.47	0.04	0.22	14.55	0.07	0.02	0.02	27.26	0.24	6.93	0.04	0.09	36	0.03	0	1.73	11.67	6.09	16.05	1.19	70.18

Supplementary Table 4C

Order	Segata et al. (2012)[1]	King et al. (2019)[2]	T7 Sweep Samples - %																				
	mean % +/- SD	mean % +/- SD	mean +/- SD	SWE01	SWE02	SWE03	SWE04	SWE05	SWE06	SWE07	SWE08	SWE09	SWE10	SWE11	SWE12	SWE13	SWE14	SWE15	SWE16	SWE17	SWE18	SWE19	SWE20
Enterobacterales	0.025 +/- 0.097	2.02 +/- 10.33	95.6 +/- 9.0	97.81	99	98.01	99.31	98.07	ND	97.09	98.36	97.28	97.7	98.46	98.73	98.31	98.55	97.8	96.92	98.9	98.89	94.09	98.85

[1]Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biol 2012; 13:R42.

[2]King CH, Desai H, Sylvetsky AC, LoTempio J, Ayanyan S, Carrie J, et al. Baseline human gut microbiota profile in healthy people and standard reporting template. PLoS One 2019; 14:e0206484.

Supplementary Table 4 (continued)

DIR21	DIR22	DIR23	DIR24	DIR25	DIR26	DIR27	DIR28	DIR29	DIR30	DIR31	DIR32	DIR33	DIR34	DIR35	DIR36	DIR37	DIR38	DIR39	DIR40	DIR41	DIR42	DIR43	DIR44	DIR45	DIR46	DIR47	DIR48
5.77	27.1	19.07	17.8	30.61	42.92	47.01	39.3	70.22	38.3	20.89	11.62	81.7	79.99	75.49	32.42	19.17	69.34	58.26	2.88	75.31	81.58	59.58	54.32	51.92	20.95	50.7	8.62
17.34	60.55	71.15	13.97	34.52	29.11	33.88	48.43	26.93	47.05	48.91	6.29	16.36	5.85	20.81	11.15	23.75	15.97	39.55	2.17	18.97	13.4	38.57	25.81	21.59	72.22	27.17	8.43
73.53	1.86	1.68	54.18	24.67	25.06	1.4	0.6	0.41	3.9	2.09	69.84	1.54	10.11	1.32	50.25	27.59	11.3	0.98	93.3	0.33	2.13	0.47	11.04	23.89	3.66	12.29	81.74
2.53	9.23	7.96	5.03	9.64	2.71	13.36	11.5	0.86	7.79	10.69	5.17	0.32	3.04	0.96	5.96	26.2	3.36	0.55	0.27	2.94	2.86	1.26	0.7	1.23	2.03	5.42	1.12
0.13	0.59	0.02	8.19	0.12	0.01	3.79	0.01	0.89	2.25	13.64	6.16	0.04	0.41	1.3	0.03	2.19	0	0.4	1.28	2.42	0	0.01	8.08	0.03	0.09	0.84	0
0	0.01	0.01	0	0.01	0	0.01	0.01	0	0.01	0.02	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0.01	0.02	0.01	0

DIR21	DIR22	DIR23	DIR24	DIR25	DIR26	DIR27	DIR28	DIR29	DIR30	DIR31	DIR32	DIR33	DIR34	DIR35	DIR36	DIR37	DIR38	DIR39	DIR40	DIR41	DIR42	DIR43	DIR44	DIR45	DIR46	DIR47	DIR48
73.49	0.15	1.19	53.18	24.16	24.74	1.06	0.09	0.12	2.84	1.08	64.84	1.39	8.9	0.82	41.86	24.25	10.56	0.63	91.94	0.17	1.81	0.13	10.64	19.6	1.67	9.16	77.58
24.1	0.02	1.05	52.38	0.68	24.22	0.95	0	0.09	2.73	0.77	63.34	1.09	8.55	0.46	41.03	23.87	10.12	0.42	86.93	0.01	1.67	0.01	8.08	8.42	0.4	8.5	76.48

SWE21	SWE22	SWE23	SWE24	SWE25	SWE26	SWE27	SWE28	SWE29	SWE30	SWE31	SWE32	SWE33	SWE34	SWE35	SWE36	SWE37	SWE38	SWE39	SWE40	SWE41	SWE42	SWE43	SWE44	SWE45	SWE46	SWE47	SWE48
97.61	98.65	98.51	97.92	97.66	50.65	97.69	ND	98.33	97.6	98.3	98.07	97.85	98.1	82.4	98.26	98.61	98.16	97.48	96.39	97.3	95.79	98.55	97.84	95.92	98.13	98.17	76.2

Supplementary Table 4 (continued)

DIR49	DIR50	DIR51	DIR52	DIR53	DIR54	DIR55	DIR56	DIR57	DIR58	DIR59	DIR60	DIR61	DIR62	DIR63	DIR64	DIR65	DIR66	DIR67	DIR68	DIR69	DIR70	DIR71	DIR72	DIR73	DIR74	DIR75	DIR76
53.79	54.31	59.76	13.8	35.64	74.44	54.39	56.76	51.48	33.08	46.67	41.51	32.33	14.54	59.91	9.29	22.32	53.78	47.41	28.51	43.14	45.96	17.25	6.26	20.47	28.88	35.63	55.25
34.13	19.06	37.21	49.26	58.8	22.02	37.17	33.43	43.46	54.59	35.43	47.66	57.89	8.46	37.86	22.94	14.51	44.55	7.38	51.38	49.54	43.39	62.42	8.69	43.33	38.16	33.66	21.19
1.45	26.58	0.46	25.69	0.69	0.61	0.64	0.76	0.96	1.33	2.88	1.39	3.78	75.68	1.44	65.21	50.47	0.63	41.86	3.49	0.52	1.39	0.94	83.82	5.84	21.08	25.29	3.71
8.27	0.04	0.61	5.11	4.08	1.36	4.62	7.53	3.72	6.85	5.21	4.16	8.41	1.1	0.6	2.43	11.42	0.64	3.03	6.26	5.16	7.66	18.47	0.59	12.51	2.12	0.92	3.59
1.89	0	1.89	3.72	0.18	1.41	3.1	0.09	0.26	3.92	8.82	4.62	0.11	0.02	0.02	0	0.48	0.31	0.01	10.19	0.82	0.68	0.08	0.03	13.43	8.62	4.41	15.96
0.01	0	0	0.02	0.01	0.01	0.01	0.01	0.01	0	0.11	0.01	0.01	0	0	0.05	0	0.01	0	0.01	0.01	0.02	0.01	0	0.02	0.01	0	0.02

DIR49	DIR50	DIR51	DIR52	DIR53	DIR54	DIR55	DIR56	DIR57	DIR58	DIR59	DIR60	DIR61	DIR62	DIR63	DIR64	DIR65	DIR66	DIR67	DIR68	DIR69	DIR70	DIR71	DIR72	DIR73	DIR74	DIR75	DIR76
1.06	24.86	0.14	16.54	0.39	0.31	0.43	0.75	0.68	0.56	0.28	5.09	63.9	0.16	60.55	43.99	0.16	36.13	1.38	0.19	0.95	0.44	79.84	4.84	14.27	20.87	1.03	4.93
0.79	24.45	0.03	16.02	0.17	0.02	0.16	0.26	0.38	0.54	0.34	0.02	4.83	4.16	0.02	59.5	38.09	0.02	26.97	0.48	0.02	0.73	0.09	77.58	4.04	0.13	0.16	0.88

SWE49	SWE50	SWE51	SWE52	SWE53	SWE54	SWE55	SWE56	SWE57	SWE58	SWE59	SWE60	SWE61	SWE62	SWE63	SWE64	SWE65	SWE66	SWE67	SWE68	SWE69	SWE70	SWE71	SWE72	SWE73	SWE74	SWE75	SWE76
96.62	98.46	97.33	97.42	98.27	ND	97.84	95.21	97.42	31.59	98.11	97.27	98.17	96.43	78.98	97.85	97.18	97.65	97.37	96.82	98.43	98.77	97.94	98.86	96.59	97.93	98.52	98.79

Supplementary Table 4 (continued)

DIR77	DIR78	DIR79	DIR80	DIR81	DIR82	DIR83	DIR84	DIR85	DIR86	DIR87	DIR88	DIR89	DIR90	DIR91	DIR92	DIR93	DIR94	DIR95	DIR96	DIR97	DIR98	DIR99	DIR100	DIR101	DIR102	DIR103
43.94	34.43	55.55	25.19	49.68	37.03	35.34	67.69	64.56	70.19	25.22	34.18	21.68	16.52	44.33	62.07	53.01	23.03	32.57	25.73	44.13	45.04	66.28	66.38	8.04	72.52	42.12
49.65	50.09	14.2	12.95	10.54	53.7	56.89	19.8	21.8	25.38	62.12	53.98	13.11	10.27	29.59	12.27	32.17	22.08	21.16	29.91	10.67	48.05	25.87	11.91	21.88	14.91	49.71
5.53	4.27	28.37	60.21	39.23	3.52	2.52	11.22	8.26	2.36	2.96	2.91	62.76	68.16	24.6	23.66	1.42	48.1	42.44	40.71	44.03	2.86	2.75	18.36	65.88	3.24	3.74
0.7	6.05	1.3	1.1	0.49	3.32	1.22	1.08	2.17	1.67	5.46	3.07	2.26	1.2	1.34	1.17	13.23	6.59	1.6	3.55	0.94	2.98	4.39	3.3	3.06	9.29	3.47
0.01	4.43	0.43	0.4	0.01	1.94	3.7	0.13	3.01	0.22	3.76	5.26	0.06	3.72	0.02	0.69	0.02	0.02	2.05	0.01	0.13	0	0.49	0	1.04	0.01	0.72
0.02	0.08	0.01	0.01	0.01	0.07	0.04	0.01	0.02	0.01	0.06	0.07	0.02	0.01	0.01	0.02	0.02	0.01	0.02	0.01	0.01	0.96	0.03	0	0.01	0	0.03

DIR77	DIR78	DIR79	DIR80	DIR81	DIR82	DIR83	DIR84	DIR85	DIR86	DIR87	DIR88	DIR89	DIR90	DIR91	DIR92	DIR93	DIR94	DIR95	DIR96	DIR97	DIR98	DIR99	DIR100	DIR101	DIR102	DIR103
0.64	27.32	59.62	38.7	0.45	0.21	10.1	7.22	0.13	0.34	0.4	62.25	67.42	23.57	22.43	0.46	47.09	41.56	39.67	43.21	0.33	1.2	18.01	65.29	1.77	0.4	0.24
0.07	0.24	0.13	0.52	38.14	0.16	0.02	8.89	6.8	0	0.04	0.07	17.34	62.86	0.06	0.72	0.32	46.39	39.55	39	35.31	0.24	1.09	15.49	30.68	1.63	0.27

SWE77	SWE78	SWE79	SWE80	SWE81	SWE82	SWE83	SWE84	SWE85	SWE86	SWE87	SWE88	SWE89	SWE90	SWE91	SWE92	SWE93	SWE94	SWE95	SWE96	SWE97	SWE98	SWE99	SWE100	SWE101	SWE102	SWE103
97.93	98	97.97	97.95	81.16	99.04	ND	97.45	98.84	98.03	98.67	97.1	97.49	97.27	91.85	97.39	96.96	98.09	92.11	97.4	97.51	79.46	98.27	98.81	97.12	97.55	95.87

Supplementary Table 5

Supplementary Table 5: Taxonomic comparison of Norgen and Eswab samples to two large microbiome profiling reference studies; Segata et al. (2012)[1] and King et al. (2019)[2] for major gut phyla abundance in Norgen (**A**) and Eswab (**B**) samples, and Enterobacterales (order), *Escherichia coli* (species) and *Klebsiella pneumoniae* (species) in Norgen (**C**) and Eswab (**D**) samples

Supplementary Table 5A

Phyla	Segata et al. (2012)[1]	King et al. (2019)[2]	Norgen Samples - %											
	mean % +/- SD	mean % +/- SD	mean % +/- SD	median	1A	2A	3A	4A	5A	6A	7A	8A	9A	10A
Bacteroidetes	65.201 +/- 20.5	73.13 +/- 22.16	29.79 +/- 16.42	26.695	16.63	41.14	23.95	4.18	34.76	17.46	68.08	15.9	29.44	53.09
Firmicutes	29.637 +/- 19.260	22.20 +/- 18.66	51.41 +/- 15.38	53.8	61.15	53.67	56.09	75.07	53.93	42.74	29.22	78.91	37.46	25.62
Proteobacteria	2.910 +/- 4.735	2.15 +/- 10.39	5.24 +/- 8.00	1.79	0.51	0.52	12.2	0.35	3.97	2.43	1.15	0.79	29.58	13.77
Actinobacteria	0.507 +/- 2.819	1.82 +/- 3	12.02 +/- 12.51	5.695	21.62	4.29	7.1	19.76	2.66	36.91	1.41	4.24	2.74	7.43
Verrucomicrobia	0.442 +/- 1.736	0.70 +/- 1.68	1.25 +/- 2.07	0.08	0.02	0.33	0.02	0.57	2.84	0.08	0.08	0.03	0.68	0.03
Fusobacteria	0.072 +/- 0.5	ND	0.01 +/- 0.01	0	0	0	0.01	0	0.03	0.01	0	0.01	0	0

Supplementary Table 5B

Phyla	Segata et al. (2012)[1]	King et al. (2019)[2]	Eswab Samples - %											
	mean % +/- SD	mean % +/- SD	mean % +/- SD	median	1B	2B	3B	4B	5B	6B	7B	8B	9B	10B
Bacteroidetes	65.201 +/- 20.5	73.13 +/- 22.16	25.98 +/- 19.06	17.645	15.56	69.57	11.68	7.14	3.82	11.15	58.13	19.73	39.42	37.87
Firmicutes	29.637 +/- 19.260	22.20 +/- 18.66	16.31 +/- 11.97	14.34	46.48	20.85	14.71	21.48	3.37	12.96	12.6	39.23	13.97	4.68
Proteobacteria	2.910 +/- 4.735	2.15 +/- 10.39	45.44 +/- 24.78	38.805	0.58	3.28	65.16	57.62	91.82	35.12	27.36	28.08	42.49	50.94
Actinobacteria	0.507 +/- 2.819	1.82 +/- 3	11.60 +/- 12.45	7.44	37.27	6	8.39	13.27	0.57	40.59	1.82	12.86	2.97	6.49
Verrucomicrobia	0.442 +/- 1.736	0.70 +/- 1.68	0.60 +/- 0.82	0.06	0.02	0.26	0.01	0.44	0.33	0.06	0.06	0.02	1.11	0.01
Fusobacteria	0.072 +/- 0.5	ND	0.00 +/- 0.00	0	0	0	0	0	0	0	0	0	0	0

Supplementary Table 5C

Order/species	Segata et al. (2012)[1]	King et al. (2019)[2]	Norgen Samples - %											
	mean % +/- SD	mean % +/- SD	mean +/- SD	median	1A	2A	3A	4A	5A	6A	7A	8A	9A	10A
Enterobacterales	0.159 +/- 0.957	2.02 +/- 10.33	4.55 +/- 7.99	0.62	0.14	0.29	10.64	0.06	2.37	0.2	0.94	0.25	29	13.45
E. coli	NA	NA	2.33 +/- 5.16	0.16	0.04	0.15	0.17	0.01	1.97	0.02	0.41	0.09	17.42	11.94
K. pneumoniae	NA	NA	0.64 +/- 1.87	0.025	0.02	0.01	0.29	0	0.03	0.01	0.46	0.02	1.41	0.07

Supplementary Table 5D

Order/species	Segata et al. (2012)[1]	King et al. (2019)[2]	Eswab Samples - %											
	mean % +/- SD	mean % +/- SD	mean +/- SD	median	1B	2B	3B	4B	5B	6B	7B	8B	9B	10B
Enterobacterales	0.159 +/- 0.957	2.02 +/- 10.33	43.58 +/- 24.56	38.21	0.12	3.07	64.88	57.4	91.54	34.52	27.21	27.58	41.9	50.81
E. coli	NA	NA	30.83 +/- 28.33	26.16	0.01	2.95	1.18	55.12	88.49	29.16	18.61	25.61	26.7	45.25
K. pneumoniae	NA	NA	3.72 +/- 9.89	0.42	0.01	0.01	0.59	0.62	0.25	3.18	7.63	0.06	1.85	0.2

[1]Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biol 2012; 13:R42.

[2]King CH, Desai H, Sylvetsky AC, LoTempio J, Ayanyan S, Carrie J, et al. Baseline human gut microbiota profile in healthy people and standard reporting template. PLoS One 2019; 14:e0206484.

Paper II

Paper III