Citation for published version:
Gibbon, MJ, Couto, N, David, S, Barden, R, Standerwick, R, Jagadeesan, K, Birkwood, H, Dulyayangkul, P, Avison, MB, Kannan, A, Kibbey, D, Craft, T, Habib, S, Thorpe, HA, Corander, J, Kasprzyk-Hordern, B & Feil, EJ 2021, 'A high prevalence of blaOXA-48 in Klebsiella (Raoultella) ornithinolytica and related species in hospital wastewater in South West England', *Microbial Genomics*, vol. 7, no. 3. https://doi.org/10.1099/mgen.0.000509

DOI:
10.1099/mgen.0.000509

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Publisher Rights
CC BY-NC

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 16. Jul. 2024

Microbial Genomics

A high prevalence of blaOXA-48 in Klebsiella (Raoultella) ornithinolytica and related species in hospital wastewater in South West England.

--Manuscript Draft--

Manuscript Number:	MGEN-D-20-00245R1
Full Title:	A high prevalence of blaOXA-48 in Klebsiella (Raoultella) ornithinolytica and related species in hospital wastewater in South West England.
Article Type:	Research Article
Section/Category:	Pathogens and Epidemiology
Keywords:	carbapenemase-producing Enterobacteriaceae (CPE); Klebsiella; blaOXA-48; Raoultella ornithinolytica; one health; wastewater
Corresponding Author:	Edward J. Feil, Ph.D University of Bath Bath, BANES, UNITED KINGDOM
First Author:	Marjorie J Gibbon
Order of Authors:	Marjorie J Gibbon
	Natacha Couto
	Sophia David
	Ruth Barden
	Richard Standerwick
	Kishore Jagadeesan
	Hollie Birkwood
	Punyawee Dulyayangkul
	Matthew B Avison
	Andrew Kannan
	Dan Kibbey
	Tim Craft
	Samia Habib
	Harry A. Thorpe
	Jukka Corander
	Barbara Kasprzyk-Hordern
	Edward J. Feil, Ph.D
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Klebsiella species occupy a wide range of environmental and animal niches, and occasionally cause opportunistic infections that are resistant to multiple antibiotics. In particular, Klebsiella pneumoniae (Kpne) has gained notoriety as a major nosocomial pathogen, due principally to the rise in non-susceptibility to carbapenems and other beta-lactam antibiotics. Whilst it has been proposed that the urban water cycle facilitates transmission of pathogens between clinical settings and the environment, the level of risk posed by resistant Klebsiella strains in hospital wastewater remains unclear. We used whole genome sequencing (WGS) to compare Klebsiella species in contemporaneous samples of wastewater from an English hospital and influent to the associated wastewater treatment plant (WWTP). As we aimed to characterise representative samples of Klebsiella communities, we did not actively select for antibiotic resistance (other than for ampicillin), nor for specific Klebsiella species. Two species, Kpne and K. (Raoultella) ornithinolytica (Korn

CONFIDENTIAL

	), were of equal dominance in the hospital wastewater, and four other Klebsiella species were present in low abundance in this sample. In contrast, despite being the species most closely associated with health-care settings, Kpne was dominant species within the WWTP influent. 29% of all isolates harboured the bla OXA-48 gene on a pOXA-48-like plasmid, and these isolates were almost exclusively recovered from the hospital wastewater. This gene was far more common in Korn (68% of isolates) than in Kpne (3.4% of isolates). In general plasmid-borne, but not chromosomal, resistance genes were significantly enriched in the hospital wastewater sample. These data implicate hospital wastewater as an important reservoir for antibiotic resistant Klebsiella , and point to an unsuspected role of species within the Raoultella group in the maintenance and dissemination of plasmid-borne bla OXA-48 .
Author Comments:	NA
Suggested Reviewers:	Kat Holt Monash University kathryn.holt@monash.edu Expert in Klebsiella genomics
	Sandra Reuter University Medical Center Freiburg: Universitatsklinikum Freiburg sandra.reuter@uniklinik-freiburg.de Expert in Klebsiella genomics and AMR
	Will Gaze University of Exeter w.h.gaze@exeter.ac.uk Expert in environmental AMR

CONFIDENTIAL

Click here to access/download
Response to Reviewer
Responses_final.docx

CONFIDENTIAL

Manuscript (Word document)
Click here to access/download;Manuscript (Word document);revised revised text final.docx

A high prevalence of *bla*$_{OXA-48}$ in *Klebsiella (Raoultella) ornithinolytica* and related species in hospital wastewater in South West England.

Marjorie J. Gibbon[1], Natacha Couto[1], Sophia David[2], Ruth Barden[3], Richard Standerwick[3], Kishore Jagadeesan[4], Hollie Birkwood[1], Punyawee Dulyayangkul[6], Matthew B. Avison[6], Andrew Kannan[4], Dan Kibbey[4], Tim Craft[5], Samia Habib[1], Harry A. Thorpe[8], Jukka Corander[2,7,8], Barbara Kasprzyk-Hordern[4], Edward J. Feil[1*]

1. The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, UK
2. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK.
3. Wessex Water, Bath, BA2 7WW, UK
4. Department of Chemistry, University of Bath, Bath, BA2 7AY, United Kingdom
5. Department of R&D, Royal United Hospitals Bath, NHS Foundation Trust, Bath, BA1 3NG, United Kingdom
6. University of Bristol, School of Cellular and Molecular Medicine, University Walk, Bristol, BS8 1TD, UK
7. Helsinki Institute for Information Technology, Department of Mathematics and Statistics, University of Helsinki, FIN-00014 Helsinki, Finland.
8. Department of Biostatistics, University of Oslo, N-0317, Oslo, Norway.

*Corresponding author: Edward J. Feil, email address: e.feil@bath.ac.uk

Keywords: carbapenemase-producing Enterobacteriaceae (CPE); *Raoultella ornithinolytica*; one health; wastewater

CONFIDENTIAL

Abstract

Klebsiella species occupy a wide range of environmental and animal niches, and occasionally cause opportunistic infections that are resistant to multiple antibiotics. In particular, *Klebsiella pneumoniae* (*Kpne*) has gained notoriety as a major nosocomial pathogen, due principally to the rise in non-susceptibility to carbapenems and other beta-lactam antibiotics. Whilst it has been proposed that the urban water cycle facilitates transmission of pathogens between clinical settings and the environment, the level of risk posed by resistant *Klebsiella* strains in hospital wastewater remains unclear. We used whole genome sequencing (WGS) to compare *Klebsiella* species in contemporaneous samples of wastewater from an English hospital and influent to the associated wastewater treatment plant (WWTP). As we aimed to characterise representative samples of *Klebsiella* communities, we did not actively select for antibiotic resistance (other than for ampicillin), nor for specific *Klebsiella* species. Two species, *Kpne* and *K.* (*Raoultella*) *ornithinolytica* (*Korn*), were of equal dominance in the hospital wastewater, and four other *Klebsiella* species were present in low abundance in this sample. In contrast, despite being the species most closely associated with health-care settings, *Kpne* was dominant species within the WWTP influent. 29% of all isolates harboured the $bla_{\text{OXA-48}}$ gene on a pOXA-48-like plasmid, and these isolates were almost exclusively recovered from the hospital wastewater. This gene was far more common in *Korn* (68% of isolates) than in *Kpne* (3.4% of isolates). In general plasmid-borne, but not chromosomal, resistance genes were significantly enriched in the hospital wastewater sample. These data implicate hospital wastewater as an important reservoir for antibiotic resistant *Klebsiella*, and point to an unsuspected role of species within the *Raoultella* group in the maintenance and dissemination of plasmid-borne $bla_{\text{OXA-48}}$.

Impact statement

Klebsiella pneumoniae is recognised as a high priority health-care associated pathogen due largely to the rapid emergence and global spread of resistance genes encoding extended-spectrum beta-lactamases (ESBLs) and, more recently, carbapenemases. These plasmid-borne resistance genes are readily transferred between other *Klebsiella* species and the *Enterobacteriacae* family that inhabit multiple animal and environmental niches. Effective management of antimicrobial resistance in the *Klebsiella* spp. therefore needs to incorporate both a broad epidemiological perspective (the "One-Health" framework) and a broad phylogenetic perspective, as environmental *Klebsiella* species can act as important reservoirs of resistance, as well as directly causing serious infections. We used WGS to characterise isolates of *Klebsiella* spp. from wastewater of a UK hospital and from the influent to the WWTP processing this wastewater. Although we did not select for carbapenem non-susceptibility, the hospital wastewater yielded a high prevalence of the carbapenemase gene $bla_{\text{OXA-48}}$. This gene was carried on a stable and highly transferable pOXA-48-like plasmid within multiple species, and in particular *Klebsiella* (*Raoultella*) *ornithinolytica*. This observation warrants increased surveillance of hospital wastewater, and a greater recognition of the potential public health impact of resistant strains belonging to the *Raoultella* group, both in the environment and in the clinic.

Data Summary

Sequencing reads are available from the European Nucleotide Archive (ENA) under the study accession number, PRJEB39942 (ERP123516). Individual accession numbers for raw sequence data are available in Table S1. All supporting metadata and tree (Newick) files are available to explore and download via the Microreact project at https://microreact.org/project/Wastewater. Sixteen supplementary figures, two supplementary tables and three supplementary notes are available with

CONFIDENTIAL

the online version of this article. The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

Abbreviations

CPE, carbapenemase producing *Enterobacteriaceae*; ESBL, extended-spectrum beta-lactamase; GTR, general time reversible; ICE, integrative conjugative element; *Kmic*, *K. michiganensis*; *Korn*, *K.* (*Raoultella*) *ornithinolytica*; *Kpla, K. (Raoultella) planticola*; *Kpne*, *Klebsiella pneumoniae*; *Kqps, K. quasipneumoniae* subspecies *similipneumoniae*; *Kvar, K. variicola*; SCAI, Simmons citrate agar with 1% myo-inositol; SNP, single nucleotide polymorphism; ST, sequence type; WGS, whole genome sequencing; WWTP, wastewater treatment plant.

CONFIDENTIAL

Introduction

Klebsiella pneumoniae (*Kpne*) is recognised by the World Health Organisation as one of the most high priority bacterial pathogens due to its ability to cause life-threatening conditions that are increasingly difficult to treat with antibiotics [1]. Healthcare-associated clones of *Kpne* that are non-susceptible to carbapenems are of particular concern. Genes encoding carbapenemases, that confer non-susceptibility to carbapenems, are typically plasmid-borne and are widely disseminated among multiple *Klebsiella* species, as well as within the broader *Enterobacteriaceae* family. The resultant public health burden is significant; the mortality of patients infected with carbapenamase-producing *Enterobacteriacae* (CPE) is 26–44% higher than in patients infected with carbapenem-susceptible isolates [2].

Five major types of carbapenemase genes have been described; KPC, OXA-48-like, NDM, VIM and IMP, and these genes primarily disseminate through the healthcare network via the movement of healthcare workers or patient referral [3]. However, many *Klebsiella* species, including *Kpne*, are commonly carried asymptomatically in the human gut, and are also recovered from animals, plants, water or the rhizosphere. Environmental or animal reservoirs may play a role in the circulation of resistance strains or genes, and the urban water cycle in particular is implicated as a major "One Health" driver of resistance. However, direct transmission across different settings is very difficult to detect and monitor[4-6].

In this study, we use whole-genome sequencing to compare isolates of *Klebsiella spp.* recovered from the wastewater of a hospital in South West England, with those isolated from a contemporaneous sample of the influent to the wastewater treatment plant (WWTP) serving the hospital and the local population. By profiling the *Klebsiella* communities at these two sites we sought to gauge the extent to which the bacterial strains or resistance genes present in hospital wastewater penetrate through the wastewater network. We used a selective culturing approach incorporating ampicillin to sample representative isolates belonging to the *Klebsiella* or *Raoultella* groups, but otherwise did not enrich for antibiotic resistance. Nevertheless, we observed a high prevalence of *Klebsiella* isolates harbouring plasmid-borne $bla_{\text{OXA-48}}$ within the hospital wastewater sample (28/95; 29.5%). The pOXA-48-like plasmid is associated with multiple species, but most notably *Klebsiella* (*Raoultella*) *ornithinolytica* (*Korn*). In contrast, the WWTP influent sample contained very low levels of both *Korn* and the $bla_{\text{OXA-48}}$ bearing plasmids, consistent with a moderate level of environmental impact from the hospital wastewater.

Methods

Sampling

Wastewater samples were collected over 5 consecutive days (5-9 August 2019) from two main sites: a hospital outlet accounting for 60% of total wastewater originating from a large hospital in SW England (> 700 beds and a catchment of 500,000 people), and a WWTP serving 105,000 people, including the hospital. The WWTP influent is mostly of domestic origin, the contribution from industrial sources being only about 1%. There is approximately 10 km of pipe connecting the hospital to the WWTP. WWTP influent was collected as flow proportional 24 h composites with average sample collection frequencies of approximately 15 min using an ISCO 3700 autosampler. Wastewater from the hospital outlet was collected as time proportional composites, 24 h composites with 50 ml collected every 15 min using an ISCO 3700 autosampler. In addition, a single sample was taken from the local river at a depth of 30 cm, approximately 8 km upstream of the WWTP in May 2019. All samples were packed on ice to maintain 4˚C and were transported to the laboratory within one hour.

CONFIDENTIAL

Recovery of *Klebsiella* isolates, DNA extraction and sequencing

An aliquot (100 µl) of the samples from the hospital drains and WWTP influent were spread on Simmons citrate agar (Fisher) with 1% myo-inositol (Sigma), a culture medium that selectively favours the growth of *Klebsiella* spp without the use of antibiotics (SCAI media; [7]). The SCAI media was supplemented with 10 µg ml^{-1} ampicillin (Oxoid) in order to further select against Gram-positive bacteria and other susceptible species, as *Klebsiella* species are typically intrinsically resistant to this antibiotic owing to a chromosomally encoded SHV beta-lactamase. Plates were incubated at 37 °C for 24 h. *Klebsiella / Raoultella* colonies were identified as yellow/orange and shiny, sometimes mucoid. Discrete colonies were picked and streaked on fresh SCAI amp_{10} and incubated at 37 °C for 48 h. Ninety-four isolates with large, yellow and shiny colonies were selected for DNA sequencing; 49 from the hospital site (at least 7 from each of 5 days) and 45 from WWTP (at least 9 from each of four days). All of these colonies were confirmed as *Klebsiella* or *Raoultella* species by whole genome sequencing (WGS) without any other prior molecular characterisation. The single river water sample (250 ml) was filtered through a 0.45 µm pore membrane (Millipore) then transferred to LB broth (Miller; Fisher) with ampicillin (10 µg ml^{-1}) and incubated at 37 °C overnight with shaking before spreading on SCAI amp_{10} and incubating at 37 °C. A single yellow colony was selected for DNA sequencing, which was also subsequently confirmed as a *Klebsiella* species.

A single colony of each isolate was picked from a fresh SCAI amp_{10} plate into LB broth (Miller) with ampicillin (10 µg ml^{-1}) and incubated at 37 °C overnight with shaking. DNA was extracted using QIAamp DNA Mini Kit (Qiagen). Isolates were sequenced using the Illumina HiSeq platform (HiSeq 10 X, 150 bp Paired End). Reads were trimmed with Trimmomatic v0.33 [8] and the trimmed reads were used to generate *de novo* assemblies using SPAdes 3.10.0 [9] using k-mer sizes of 41, 49, 57, 65, 77, 85 and 93 and with the –cov-cutoff flag set to 'auto' as part of an in-house pipeline at the Wellcome Sanger Institute, Cambridge, UK. Genomes were annotated using Prokka v1.14.5 ([10]; https://github.com/tseemann/prokka).

Phylogenetic analysis

Short reads of all isolates were mapped to the genome of *Kpne* isolate 5Sd using Snippy v4.3.6 (https://github.com/tseemann/snippy). This reference was chosen as a high quality short-read assembly consisting of 15 contigs. A core alignment of 5,100,718 nucleotides with 184,671 polymorphisms was used to generate an approximate maximum-likelihood phylogenetic tree based on a general time reversible (GTR) model using FastTree v2.1.11 [11, 12]. We constructed an additional tree that also included the publicly available *Kpne* ST983 genome (ED01500733, accession number NZ_POWS00000000.1 [13], using the same reference, that contained 184,119 single nucleotide polymorphism (SNP) variants. We also constructed a larger *Korn* tree using *Korn* genomes generated as part of other studies from Italy (n=25), Pakistan (n=5) and two isolates from WWTP in the East of England described previously [14] by mapping to the *Korn* isolate SPARK_1635_C1 from Italy. This generated a core alignment of 6,004,523 nucleotides with 24,659 polymorphisms. These trees were combined with metadata and output from Kleborate and visualised using Microreact v23.0.0 [15]. Distance matrices of isolates in each tree were generated using snp-dists (https://github.com/tseemann/snp-dists). Phylogenetic analysis of the bla_{OXA-48} gene and *ybt* locus are described below.

Identification of resistance and virulence determinants

Genome assemblies produced using SPAdes 3.10.0 were assigned species and multilocus sequence type (in those species where schemes are available), as well as being screened for virulence and

CONFIDENTIAL

resistance genes using Kleborate v0.4.0-beta (https://github.com/katholt/Kleborate). Recent phylogenetic analysis based on WGS data has revealed that the *Raoultella* species are unequivocally nested within the *Klebsiella* genus [16]; an observation further supported by the current data. We therefore follow Wyres *et al* [16] and refer to the '*Raoultella*' species *R. ornithinolytica* and *R. planticola* as '*Klebsiella*' *ornithinolytica (Korn)* and *K. planticola (Kpla*).

Abricate v0.9.8 (https://github.com/tseemann/abricate) was used for further screening for resistance genes in the ResFinder database (downloaded 29 April 2020) and virulence factors in the virulence factors database VFDB (downloaded 19 April 2020). We scored the presence or absence of genes in our genomes according to a threshold of >80% nucleotide identity and coverage. In addition, we searched for virulence factors with lower thresholds (>40% nucleotide identity and coverage). In the very small number of cases where 2 copies of the same gene were noted in a single genome we simply scored the gene as present.

In order to confirm the presence of the $bla_{\text{OXA-48}}$ gene and the associated plasmid, assemblies identified as harbouring this gene were screened using BLAST v2.9.0+ to identify homologous sequences carried on plasmid pOXA-48 ([17]; accession number, JN626286). The contigs were aligned with pOXA-48 and Tn*1999* ([18]; accession no AY236073.2) using Clustal Omega v1.2.3 as implemented in Geneious Prime® v2020.1.1 (Biomatters, Ltd., Auckland, New Zealand). The alignment of representative contigs was visualised with Easyfig v2.2.2 [19]. Sequence reads of isolates harbouring $bla_{\text{OXA-48}}$ were mapped to pOXA-48 using Burrows−Wheeler Aligner v0.7.12 [20]. and visualised in Tablet [21]. Read sets from the isolates and publicly available sequences of 21 pOXA-48-like plasmids (described in Table S2) were mapped to pOXA-48 using Snippy v4.3.6 (https://github.com/tseemann/snippy), giving an alignment of 61,881 nucleotides and 28 core single nucleotide variants. Where reads were available, the variants were confirmed by mapping to the reference (as above). An approximate maximum-likelihood phylogenetic tree using a GTR model was generated with FastTree v2.1.11 [11, 12].

The *ybt* locus encodes the siderophore and virulence factor yersiniabactin and is typically associated with an ICE (integrative conjugative element). The *ybt*-positive annotated genomes (n=33) were manually inspected to determine the location of the tRNA-*Asn* occupied site using Geneious Prime®. The contig containing the *ybt* locus was re-annotated with an in-house ICE*Kp*-*ybt* database using Geneious Prime®. The sequences of the genes in the *ybt*-locus (*ybtS-ybtX-ybtQ-ybtP-ybtA-irp2-irp1-ybtu-ybtT-ybtE-fyuA*) were extracted and concatenated. An alignment of the concatenated *ybt* locus was produced using MAFFT v7.450 [22] and a tree was generated using FastTree v2.1.11 [11, 12] with the GTR model. The *ybt* locus of *Kpne* NCTC11697 was used as reference to re-root the tree. Recombination events were identified using Gubbins v2.4.1 [23]. A Perl-based local version of ICEfinder [24] was used to detect potential ICEs. A tanglegram linking the core genome and *ybt* locus were generated using the "tanglegram" function from the "dendextend" package v1.13.4 in R v3.6.3 (https://www.r-project.org).

Plasmid detection and characterisation

Plasmid replicons were identified using Abricate v0.9.8 with the PlasmidFinder database (downloaded 13 Jan 2020) based on a threshold of >80% nucleotide identity and coverage [25]. To determine the plasmid content of the isolates, we used MOB-suite [26] and mlplasmids v1.0.0 [27] to classify contigs as plasmid-borne or chromosomal. MOB-suite uses Mash distances to assign contigs to plasmids in a closed reference database. We used the default parameters that are already

CONFIDENTIAL

optimised for Enterobacteriaceae plasmids. This approach identifies the accession number of the plasmid with the shortest Mash distance to a given set of contigs but, depending on the database, we recognise that substantial size or structural variation may still be present between the query contigs and the returned plasmid. In contrast, mlplasmids uses a machine learning tool to assign contigs as plasmid-borne or chromosomal and was run specifying *Kpne* as the species model, a minimum contig length of 1,000 bp, and a posterior probability threshold of 0.7. Only contigs that were consistently assigned as plasmid-borne or chromosomal by both MOB-suite and mlplasmids were accepted, cases where the results from these two approaches were discordant were assigned as ambiguous. We also carried out a clustering analysis to detect linkage between resistance genes, plasmids and sequence type (ST) using the pheatmap package in R v3.6.3. The NbClust function (method "ward.D2" and index "silhouette") available in the RNbClust package (version 3.0) was used to evaluate an optimal number of clusters. The hierarchical clustering using the method "ward.D2" was computed by cutting the resulting trees specifying 15 clusters.

Statistical analysis

Box plots and Wilcoxon Rank Sum tests were carried out using R v4.0.2 (https://cran.r-project.org/).

Results

The phylogenetic tree, all metadata, as well as combined and parsed outputs from Kleborate, ResFinder, PlasmidFinder, mlplasmids, MOB-suite are freely available to explore and download via the Microreact project at https://microreact.org/project/Wastewater. A brief explanation to the metadata fields is given in supplementary note 1, and full instructions on how to use Microreact are available at https://microreact.org/instructions. The metadata is also available as an excel file (Table S1).

Species diversity and distribution

High quality genome assemblies were obtained for 95 isolates, 49 from the hospital wastewater, 45 from the WWTP and 1 from the local river. A total of six species were isolated, but *Kpne* and *Korn* dominated the samples, together accounting for 88% of the isolates (Table 1). All six species were present in the hospital sample, but only three were present in the WWTP sample (*Kpne*, *Korn* and a single isolate of *K. variicola*; *Kvar*). Within the hospital sample, 19 were *Kpne* (39%) and 20 isolates were *Korn* (41%). In contrast, for the WWTP sample, 39 (87%) were *Kpne* and only 5 isolates (11%) were *Korn*. This was unexpected as *Kpne* is considered a major healthcare pathogen whilst *Korn* is considered a predominantly environmental species. Studies of hospital-acquired infection by Klebsiella species do not typically target *Korn*, hence the prevalence of this species in health-care settings may have been under-estimated. However, although a Fisher's Exact Test confirmed that the WWTP sample was significantly enriched for *Kpne* over *Korn* compared to the hospital sample ($P < 0.0001$), the number of isolates sequenced is too small to draw firm conclusions on species distributions. Moreover, the samples were taken over a small temporal range (5 days), which renders the data vulnerable to transient clonal expansion.

The phylogenetic relationships between the species are shown in Figure 1. This confirms the close relatedness of the species corresponding to the *Kpne* species complex: *Kpne*, *Kvar* and *K. quasipneumoniae* subspecies *similipneumoniae* (*Kqps*). The two species previously assigned as '*Raoultella*' (*Korn* and *Kpla*) are also related. The single isolate of *K. michigenensis (Kmic)* is more closely related to the '*Raoultella*' group than to the *Kpne* species complex, which justifies the re-

CONFIDENTIAL

inclusion of the '*Raoultella'* species within the *Klebsiella* genus as previously observed [16]. *Kmic* belongs to a third species complex with *K. grimontii* and *K. oxytoca*, which were not recovered in this study. The enrichment of *Kpne* within WWTP isolates (shown in blue) is also evident in Figure 1. Discussion of within species diversity and prevalent clonal lineages is given below and in supplementary material.

The prevalence and distribution of resistance genes

We used both ARG-Annot (via Kleborate) and ResFinder (via Abricate) in order to detect resistance genes in our data. The outputs from these tools were consistent, and available to explore via the Microreact project. The analysis below is primarily based on the more inclusive ResFinder data. We identified 58 resistance genes or gene variants, predicted to encode resistance to antibacterial drugs from the following 8 classes, where the number in brackets refer to gene variants rather than the number of isolates: Aminoglycoside (n=6), Beta-lactam (n=30), Chloramphenicol (n=1), Fosfomycin (n=7), Quinolones (n=5), Sulfonamide (n=3), Tetracycline (n=3), Trimethoprim (n=3). A minor caveat is that some of the genes assigned by this analysis as resistance genes are in fact intrinsic chromosomal genes present in the vast majority of *Klebsiella* strains, and may only confer resistance when highly expressed or when additional plasmid-borne duplicates are present. These genes include intrinsic *oqxA* and *oqxB* genes which encode efflux pumps and (unless highly expressed) only confer very low level resistance to fluoroquinolones [28, 29]. Other examples include *fosA*, which typically confers only low levels of resistance to fosfomycin, and $bla_{\text{SHV-187}}$ which confers intrinsic resistance to ampicillin and amoxicillin across all *Kpne* strains (Figure S1).

The median number of resistance genes per isolate (combining chromosomal and plasmid-borne) from the hospital wastewater is 9, whereas for the WWTP sample it is 4, and this difference is statistically significant by a Wilcoxon Rank Sum test ($P<0.001$; Figure 2A). The significant enrichment of resistance genes within the hospital wastewater holds true when only considering *Kpne* isolates, and therefore does not simply reflect the non-random distribution of species between sites ($P<0.001$; Figure 2B). As there is only one *Kpne* ST that is found in both sites (ST983), this difference is more likely due to differences in strain composition rather than to individual strains losing or gaining resistance genes according to which site they are located. This single species analysis was only possible for *Kpne* due to the low prevalence of *Korn* isolates from the WWTP influent.

Although there is no significant difference between *Kpne* and *Korn* in terms of the number of resistance genes identified by ResFinder per isolate ($P = 0.059$; Figure S2), there is a surprising difference between these two species with respect to the carbapenemase gene $bla_{\text{OXA-48}}$. Twenty-eight isolates over all species (29.5%) harboured this gene, and all of these except a single *Korn* isolate were recovered from the hospital wastewater. A total of 17 of the *Korn* isolates carry the $bla_{\text{OXA-48}}$ gene, accounting for 68% of the isolates from this species. In contrast, of the 59 *Kpne* isolates, only 2 (3.4%) carried this gene (7Rg and 8Rg), and these are clonally related thus have co-inherited this gene (these isolates differ by only 33 core SNPs and are marked with a red asterisk in Figure 3). Of the 11 isolates from species other than *Kpne* and *Korn*, 9 harboured $bla_{\text{OXA-48}}$, the exceptions being one *Kvar* isolate from the WWTP and the single *Kmic* isolate (Figure 1). We emphasise that our culturing procedure did not select for non-susceptibility to carbapenems, thus the high frequency of $bla_{\text{OXA-48}}$ reflects a high abundance of this gene within the underlying community at the time of sampling.

Characterisation of a pOXA-48-like plasmid

CONFIDENTIAL

We characterised the genetic context of *bla*$_{OXA-48}$ within the assemblies of the 28 isolates harbouring this gene. In 17/28 assemblies this gene is present on a conserved contig of approximately 62 kb. Alignments between these contigs revealed they share >99% sequence identity with each other and with the 63 kb plasmid pOXA-48 ([17]; accession number JN626286; Figure 4). This IncL conjugative plasmid (and minor variants) is the predominant source of *bla*$_{OXA-48}$ in Enterobacteriales worldwide [30]. pOXA-48 carries a composite transposon Tn*1999*, made up of *bla*$_{OXA-48}$, *lysR* (encoding a helix-turn-helix type transcriptional regulatory protein) and two copies of IS*1999*. A common variant is Tn*1999.2*, which harbours IS*1R* between bla$_{OXA-48}$ and IS*1999*. Our 62 kb contigs included three of the four elements making up Tn*1999* (Figure 4), but our assemblies failed to resolve the expected second copy of IS*1999*. Alignments with pOXA-48-like plasmids confirm that 8/17 of our contigs harbour Tn*1999* (Figure 4 groups 1 and 3). The assemblies of the remaining 9/17 contigs do not extend far enough beyond *bla*$_{OXA-48}$ to distinguish between Tn*1999* and Tn*1999.2* (Figure 4 groups 2 and 4), although we can be sure they are not Tn1999.3, .4 or .5 in which additional mobile elements are present between *bla*$_{OXA-48}$ and *lys*R, or within the *lys*R sequence [30]. In the absence of any evidence that our isolates do harbour Tn*1999.2* we shall refer to them all as harbouring Tn*1999*. The transposon is inserted in the plasmid backbone in the same orientation as in pOXA-48 in 9/17 of these contigs (Figure 4, groups 1 and 4). Within the limitations of the short-read assemblies, the Tn*1999* in our contigs are identical to each other at the nucleotide level, and differ from the reference sequence in pOXA-48 by 10 SNPs. The plasmid backbone (approx. 58 kb) is also highly conserved. Our contigs have two IS*1* family transposase genes, not present in pOXA-48 (Figure 4), otherwise the identity is >99% throughout. Allowing for the presence of the additional IS*1999*, the size of our contigs (~62 kb) would be highly consistent with other pOXA-48-like plasmids [30].

In the remaining 11/28 *bla*$_{OXA-48}$ positive isolates, the gene is harboured on a short contig, up to 3.7 kb. Mapping of reads from these isolates to pOXA-48 revealed that more than 99.8% of the plasmid is represented, although in all but one case the short-read assembled contig only carries *bla*$_{OXA-48}$ and *lysR* so we cannot determine the wider genetic context of the gene. In two of these isolates the *lysR* sequence is interrupted by an insertion sequence. *Korn* isolate 8Rl harbours a fragment with homology to IS*Kpn26*, an IS*5* family transposase, whereas *Kpla* 6Rh contains a region with homology to a Tn*3*-like element Tn*5403* family transposase. Multiple copies of these sequences are present in many of our isolates, as well as in many publicly available *Klebsiella* genomes and other Enterobacteriales, but they have not, to our knowledge, previously been shown to be associated with *bla*$_{OXA-48}$.

Despite the potential for mobility, we found no evidence for Tn*1999* or *bla*$_{OXA-48}$ transferring independently of a pOXA-48-like plasmid. Reads from all of our *bla*$_{OXA-48}$ positive isolates mapped to at least 99.8% of pOXA-48, and BLAST searches with the *bla*$_{OXA-48}$, Tn*1999* or the plasmid backbone only identified significant matches in the *bla*$_{OXA-48}$ positive isolates. An analysis of the *bla*$_{OXA-48}$ containing contigs with MOB-suite and mlplasmids v1.0.0 further supported the view that these contigs were of plasmid origin. MOB-recon assigned the contigs to plasmid cluster 681 and more specifically to plasmid accession number CP015075, a 63.5 kb pOXA-48-like plasmid pEC745_OXA48 isolated from *E. coli* ST131 [31]. Our contigs are more similar to this plasmid than to pOXA-48. The presence/absence of this plasmid was 100% consistent with the presence/absence of *bla*$_{OXA-48}$. Hierarchical clustering analysis of predicted plasmids, resistance genes and replicon types further confirmed the association between *bla*$_{OXA-48}$, replicon type IncL/M and a pOXA-48-like plasmid (Figures 5, S3).

Our data thus suggest that *bla*$_{OXA-48}$ has spread within and between species in the hospital wastewater community via the transfer of the pOXA-48 plasmid. Whereas most of the ~62 kb *bla*$_{OXA-}$

CONFIDENTIAL

$_{48}$ contigs differ by one or two SNPs, there are two pairs of isolates from different species in which the contigs are 100% identical, consistent with recent interspecies transfer (*Kpne* 8Rg / *Kvar* 9Rb, and *Korn* 9Ri / *Kpne* 6Rf). To examine whether there were single or multiple introductions of the pOXA-48 plasmid into this population, we constructed a phylogenetic tree to compare our isolates with 21 publicly available sequences of pOXA-48-like plasmids (Figure S4; Table S2). These publicly available plasmid sequences were from diverse geographic origins, multiple species and harbour $bla_{\text{OXA-48}}$ on either Tn*1999* or the variant Tn*1999.2*. They include two pOXA-48-like plasmids associated with *Korn* isolates previously recovered from wastewater in the UK [14]. This analysis confirmed that the plasmid bearing $bla_{\text{OXA-48}}$ in our isolates is highly similar to those plasmids previously reported from global origins. Most plasmids on the tree are separated by only a single core SNP, confirming that this plasmid is very widespread and highly conserved. The plasmids harbouring $bla_{\text{OXA-48}}$ in our isolates are no more closely related to each other than they are to plasmids from global sources, hence it is not possible to deduce how many times this plasmid was introduced into our study community. A minor exception to the high degree of plasmid conservation is found in the two isolates 6Rbi (*Korn*, hospital wastewater) and 7Sd (*Korn*, WWTP); the pOXA-48 plasmids in these strains are identical but separated from the other plasmids by a single core SNP (indicated by the red circle in Figure S4). As the host chromosomes of these two isolates differ by only 107 core SNPs (Figure 6), this congruence between plasmid and chromosomal phylogeny is likely to reflect common inheritance rather than horizontal transfer.

Summary of plasmid distribution across species and sites

Having characterised the pOXA-48-like plasmid in our data, we used a combination of the outputs from MOB-suite, mlplasmids, PlasmidFinder and ResFinder to determine the plasmid and replicon profiles of all our isolates and, as far as possible, assign each resistance gene either to a specific plasmid or as chromosomal. The distribution of plasmids, replicon types and resistance genes over the whole dataset are shown in Figures S5, S6 and S1, respectively, and are available to explore on the Microreact project at https://microreact.org/project/Wastewater (see supplementary note 1). MOB-suite assigned the plasmid with accession number CP015075 as the closest hit to the $bla_{\text{OXA-48}}$ carrying pOXA-48 like plasmid in our data. Our data show this plasmid to be highly promiscuous, being present in 5 of the 6 species, the exception being the single isolate of *Kmic*. Out of a total of 80 plasmids identified by MOB-suite from our data (representing 26 replicon types; Figure S5), 20 carried at least one resistance gene (25%), only one other showed the same degree of cross-species distribution as the CP015075-like plasmid. This was a CP011607-like plasmid with replicon type Col440I, and this small (~5 kb) plasmid does not harbour any resistance genes in our data. Two other plasmids were detected in 5 species, three in four, three in three, 17 in two and 55 (68.8%) were restricted to a single species. An important caveat to these figures is that there is likely to be variation between plasmids that are placed in the same cluster by MOB-suite, and these differences may in turn impact on the ability of the plasmid to transfer between species.

Wilcoxon Rank Sum tests confirmed that the isolates from the hospital wastewater contain significantly more plasmids per isolate ($P<0.001$), and significantly more replicon types per isolate ($P<0.001$), than the isolates from the WWTP (Figure S7). As noted earlier, the hospital wastewater sample also contains significantly more resistance genes than the WWTP sample (Figure 2). Considering all 58 resistance gene variants across all 96 isolates, there were a total 665 resistance gene assignments attempted. In 374 cases (56.2%) the resistance gene was assigned as chromosomal, 205 assignments (30.8%) were plasmid-borne by both methods, 73 (11%) were ambiguous (ie assigned as plasmid-borne by only one of the two methods, and hence excluded), and 13 (1.9%) were assigned as plasmid by both methods, but with no clear match in the MOB-suite

CONFIDENTIAL

database ('unknown'). The resistance genes that were assigned plasmid-borne by both MOB-suite and mlplasmids were significantly enriched in the hospital wastewater sample (P<0.001), but there was no difference between the two sites with regards to chromosomal resistance genes (Figure 7).

To explore plasmid and resistance gene distribution within a finer sub-species phylogenetic context, we then considered each species in turn.

Plasmids, resistance genes and phylogeny of *K. ornithinolytica (Korn)*

The distribution, according to source and phylogeny, of all 29 plasmids detected by MOB-suite (representing 16 replicon types) within the *Korn* isolates is given in Figure 6. Only 4 of these plasmids are associated with at least one resistance gene (13.8%). The distribution of replicon types is given in Figure S8. Nineteen of the 25 *Korn* isolates resolve into two clades, represented by 12 and 7 isolates. All of the 19 isolates belonging to these clades were recovered from the hospital wastewater, whereas of the remaining 6 isolates only one was isolated from the hospital wastewater. Isolates within the larger clade differ between 42 and 136 SNPs, and within the smaller clade between 25-95 SNPs. These levels of diversity are within the maximum distance typically observed within single STs [32]. In order to place our *Korn* genomes within a wider phylogenetic context we rebuilt the tree with additional *Korn* genomes from Italy and Pakistan that were recovered from diverse sources and sequenced as part of other studies. We also included two *Korn* genomes recovered previously from UK wastewater and which harbour a pOXA-48-like plasmid ([14]; Figure S9). This revealed that the *Korn* isolates from the current study did not represent a monophyletic group, but are scattered across the tree indicating multiple introductions. However, none of these additional isolates clustered with either of the major two *Korn* lineages, which in turn points to subsequent local clonal expansion consistent with the low diversity within each of these clones.

The two major clones show distinct repertoires of plasmids and resistance genes (Figure 6), although this is less clear for replicon types (Figure S8). The larger clone is associated with a total of 13 plasmids across all isolates, of which 5 are specific to this clone and only two of which carry resistance genes; the $bla_{\text{OXA-48}}$-carrying CP015075-like plasmid, and the KT896499(pKPSH169)-like plasmid which carries the *qnrS2* gene encoding reduced susceptibility to quinolones. The original pKPSH169 plasmid was harboured by a *Kpne* isolate recovered from wastewater biosolids in Israel, and belongs to the widespread pKPN3-like family of small plasmids known to carry *qnrS* genes [33]. Ten plasmids are detected among all isolates within the smaller clone, but again only two of these contain resistance genes; the $bla_{\text{OXA-48}}$-carrying CP015075-like plasmid, and a KJ541070(pG5A4Y413)-like plasmid harbouring the ß-lactamase $bla_{\text{MOX-3}}$ and $bla_{\text{OXA-10}}$ genes. The fourth resistance gene-bearing plasmid found in the *Korn* isolates is a KY798506-like plasmid that carries the *aac(3')-IIa*, *aac(6')-IIb*, $bla_{\text{OXA-1}}$ and $bla_{\text{SHV-12}}$ resistance genes. This plasmid is present in the pair of closely related *Korn* isolates (6Rbi and 7Sd) that also harbour a single SNP variant of the pOXA-48-like plasmid CP015075 (Figure S4). $bla_{\text{SHV-12}}$ is an important extended-spectrum beta-lactamase (ESBL) variant gene that is assigned as associated with the KY798506-like plasmid in strain 7Sd, but is not unambiguously assigned to this plasmid in isolate 6Rbi. The KY798506-like plasmid is also present within three isolates of *Kvar* ST454 (see below), and belongs to the pKpQIL-like family of plasmids that commonly carry $bla_{\text{KPC-2}}$ [34, 35]. An alignment of the KY798506 reference plasmid with the corresponding contigs in our data confirmed regions of high homology (Figure S10).

Plasmids, resistance genes and phylogeny of *K. pneumoniae (Kpne)*

CONFIDENTIAL

A total of 60 different plasmids and 19 replicon types were detected in 59 *Kpne* isolates using MOB-suite. The distribution of resistance genes and their associated plasmids is given in Figure 3, and the distribution of all plasmids and replicon types is given in Figures S11 and S12 respectively. Although the average number of plasmids per isolate is lower for *Kpne* (3.3) than for *Korn* (6), a higher proportion of the *Kpne* plasmids harbour at least one resistance gene (15/60; 25%) than the *Korn* plasmids (4/29; 13.8%). A higher proportion of the *Kpne* plasmids were present in only one isolate (26/60; 43%) than the *Korn* plasmids (7/29; 24.1%), consistent with a more diverse sample.

Kleborate was used to assign multilocus STs to the *Kpne* data. The most common ST is ST983, which is represented by 9 isolates, and these differ from each other by between 15 and 47 core SNPs, suggesting local clonal expansion. Eight of the 9 *Kpne* ST983 isolates are from the hospital wastewater, and one from the WWTP influent. A total of 10 different plasmids are observed within this clone, five of which carry resistance genes (Figures 3, S13). It is noteworthy that these five different resistance plasmids carry the same set, or a subset, of 9 resistance genes: the aminoglycoside resistance genes *aac(3)-IIa*, *aac(6')-Ib-cr*, *aph(3'')-Ib*, *aph(6')-Id*, in addition to *dfrA14*, $bla_{\text{CTX-M-15}}$, $bla_{\text{OXA-1}}$, $bla_{\text{TEM-1}}$ and *sul2*. The CP021953(AR_0148)-like plasmid carries all these genes in three ST983 isolates (5Rk, 8Rc and 8Rhi), plasmid JX424423(pKD01)-like carries all these genes in 5Rg (this plasmid is also present in a *Kpla* isolate where it harbours 11 resistance genes), and plasmid CP016925(pCTXM15_DHQP1400954)-like contains all the genes except for *aac(3)-IIa* in 5Rd (Figure S13).

The data therefore suggest that a suite of resistance genes are linked on a mobile element and have been co-transferred between different plasmids within the ST983 clone. This is supported by hierarchical clustering analysis, which confirmed the linkage of the resistance genes $bla_{\text{CTX-M-15}}$, $bla_{\text{TEM-1B}}$, *aph(3'')-Ib*, *sul2*, *dfrA14* and *aph(6)-Id* within *Kpne* ST983, but with multiple plasmids (Figure 5; cluster 2). Furthermore, comparison of the ST983 contigs with published data from South Africa [13] provides additional support that these genes are linked, mobile and have a global distribution (Figure S14) and more detailed analyses of the relevant contigs also point to the mobility of these genes (supplementary note 2). There are four *Kpne* STs (STs 35, 13, 17 and 584) that are each represented by 3 isolates (Figure 3). The plasmid and resistance gene profiles of these lineages are described in supplementary note 3.

Plasmids and resistance genes in other species

In addition to the 59 *Kpne* isolates, we also recovered two other species corresponding to the *Kpne* species complex, *Kqps* (n=4) and *K. variicola* (*Kvar*) (n=4). All of these 8 isolates were recovered from hospital wastewater and harbour $bla_{\text{OXA-48}}$, except one divergent *Kvar* isolate which was isolated from the WWTP and does not harbour $bla_{\text{OXA-48}}$. The four isolates of *Kqps* all correspond to ST3590 and harbour identical plasmid and resistance profiles. Three of the four *Kvar* isolates corresponded to ST454. Although this is not a clone of recognised clinical importance, all three ST454 isolates harbour a KY798506-like plasmid that was also found in *Korn* and belongs to the pKpQil-like family of plasmids as discussed above. Similar to the *Korn* isolates, this plasmid is also associated with the bla_{SHV12} in *Kvar* ST454 isolates 8Rji and 5Rj, whereas this gene is assigned as chromosomal in *Kvar* ST454 isolate 9Rb. Finally, *Kpla* isolate 6Rf harbours plasmid a JX424423(pKDO1)-like plasmid with 11 resistance genes, more than any other plasmid in the dataset. These are *aac(3)-IIa*, *aac(6')-Ib-cr*, $bla_{\text{OXA-1}}$, $bla_{\text{CTX-M-15}}$, *dfrA14*, *aph(3'')-Ib*, *aph(6)-Id*, bla_{TEM}, *qnrB1*, *sul2*, *tet*(A). This plasmid is also present in two *Kpne* isolates, including one ST983 isolate as previously noted, where it harbours subsets of nine and six of these genes.

Characterisation of virulence factors

CONFIDENTIAL

Kleborate revealed that 33 of the isolates contained the major virulence factor *ybt* that encodes the siderophore yersiniabactin. This locus was detected in all 25 *Korn* isolates and in each case was assigned an "unknown" type. We note that the *ybt* locus in *Korn* is chromosomally located close to an tRNA-*Asn* site, with no evidence for an associated ICE, and is phylogenetically distinct from the *ybt* locus in *Kpne* (Figure S15). A tanglegram linking the *ybt*-based and core genome-based phylogenies shows that different *ybt* variants match the different core genome lineages in *Korn*, suggesting a single acquisition of the *ybt* locus into this species (Figure S16), however, the *Korn* phylogeny lacks the resolving power to show this conclusively.

A total of 8 *Kpne* isolates harbour a yersiniabactin locus, 4 from the hospital wastewater, 3 from the WWTP sample and the single isolate from the river, and these correspond to 4 unique STs. The three *Kpne* ST13 isolates are all ybST21 (*ybt* 10; ICE*Kp4*). The two highly related ST35 isolates (River_C and 7Sl) harboured a novel 2LV of ybST183 (*ybt* 9; ICE*Kp3*), whilst the third ST35 isolate (8Sd) harboured ybST10 (*ybt* 5; ICE*Kp6*). A single locus variant of *Kpne* ST25 from the hospital wastewater (5Ra) harboured a 1LV of ybST63 (*ybt* 6; ICE*Kp5*) and finally an isolate of *Kpne* ST1536 was isolated (7Sc) from the WWTP sample harbouring a 2LV of ybST 209 (*ybt* 9; ICE*Kp3*). We confirmed the presence of the *ybt* locus using the vfdb database through ABRicate.

Kleborate and ABRicate also revealed the absence of type I or III *Klebsiella* fimbriae (although *fim* genes are present), and the presence of the *astA* gene in 3 *Korn* isolates (6Rbi, 7Sa, 7Sd) that encodes a heat-stable enterotoxin 1. We note that the *astA* gene is embedded in an IS*256*-family transposon, and a BLASTP search suggested that the closest *astA*-encoded protein (from strain 6Rbi), named EAST1, is from *Edwardsiella ictalurid* (100% coverage, 77.14% identity; data not shown).

Discussion

The acquisition of multiple antibiotic resistance genes by *Klebsiella* spp. accounts for a significant public health burden, particularly in health care settings. Here we compared *Klebsiella* spp. isolates from the wastewater of a large hospital in South West England with those from the influent to a WWTP serving the hospital. We targeted *Klebsiella* spp. isolates through selective culturing, but we did not actively enrich for antibiotic resistance (other than for ampicillin), nor for specific *Klebsiella* species. We note marked differences in resistance profile and species composition between the two sites which suggests that the signal from hospital wastewater is markedly diluted once it has reached the WWTP [36]. However, it remains likely that bacteria from the hospital are still present in the WWTP influent, and possibly also the effluent, albeit at a low frequency [14]. In support of this, the WWTP sample contained a single *Korn* isolate (7Sd) that carries both a pOXA-48-like plasmid (harbouring bla_{oxa48}) and a pKpQIL-D2-like plasmid (harbouring bla_{SHV12}). Although we cannot rule out the possibility that this isolate derived from the community rather than the hospital, isolate 7Sd is clonally related to the hospital wastewater *Korn* isolate 6Rbi, differing by only 107 core SNPs, and both 7Sd and 6Rbi possess a variant pOXA-48 plasmid containing a unique SNP. This exception aside, the prevalence of plasmid-borne resistance genes is significantly greater within the hospital wastewater than the WWTP influent, which again is consistent with previous studies examining a range of different hosts and including hospital associated isolates [37]. In contrast, we find no evidence for a difference in the prevalence of chromosomal resistance genes, indicating that these are commonly core genes in *Klebsiella* spp. [38].

Whereas both *Kpne* and *Korn* are present in the hospital wastewater at equal abundances, and four other species are also present at this site, the WWTP sample is overwhelmingly dominated by a

CONFIDENTIAL

single species, *Kpne*. As *Kpne* is health-care associated, and *Korn* is considered an environmental species, their relative abundances at either site is the opposite of that expected. Moreover, the WWTP sample also represents a much more heterogeneous mixture of inputs from the surrounding community, thus would be expected to contain a higher species diversity. The bulk of the WWTP influent derives from domestic, rather than industrial, sources. Thus the high prevalence of *Kpne* in this sample presumably reflects, at least in part, a high rate of asymptomatic carriage of *Kpne* in the local community. We note that nursing homes in particular have been implicated as a major reservoir of ESBL-producing *E. coli* and *Kpne* [39].

Our data reveal a high abundance of a 63 kb IncL/M pOXA-48-like plasmid harbouring the carbapenemase gene bla_{OXA-48}. This gene encodes the OXA-48 class D β-lactamase that hydrolyses penicillins at a high level and carbapenems at a low level. This enzyme confers a high level of resistance to imipenem [40] but is ineffective against cephalosporins [41]. The pOXA-48 plasmid is almost exclusively present in the hospital wastewater, with the *Korn* WWTP isolate 7Sd being the only exception. The pOXA-48 plasmid has disseminated worldwide, and has previously been associated with *Korn* isolates from sewage in the UK [14]. Although minor variants have been described [30], it appears to be highly stable [42]. This plasmid has also been observed to freely transfer between species within the *Enterobacteriaceae* family [14], and has a very low fitness burden [43]. These characteristics are evident in our data. Even without selective culturing, the pOXA-48-like plasmid is present in the majority of isolates of 4 species within the hospital wastewater (*Korn*, *Kpla*, *Kvar* and *Kqps*), with the notable exception being that only 2 clonally related isolates of *Kpne* carry this plasmid. The pOXA-48-like plasmid present in our data is highly related to publicly available sequences from diverse geographic origins, meaning it is not possible to infer how many times the local community has acquired this plasmid. Caution should thus be exercised before inferring local epidemiological spread of a pOXA-48-like plasmid from a single source.

In addition to the high abundance of plasmid-borne bla_{OXA-48}, we note multiple plasmids that harbour clinically important ESBL genes. Cases where these plasmids co-reside with the pOXA-48-like plasmid (eg *Kpla* isolate 6Rf and *Korn* isolate 7Sd) are likely to present particular therapeutic challenges due to resistance to both cefpodoxime and meropenem. Two ESBL-bearing plasmids are present in multiple species; first, a bla_{SHV-12} carrying KY798506(pKpQIL-D2)-like plasmid is present in the WWTP *Korn* isolate 7Sd and two hospital wastewater *Kvar* ST454 isolates 8Rji and 5Rj. This pKpiQIL-like plasmid commonly carries the key carbapenemase gene bla_{KPC-2}, and is facilitating the spread of this gene in the UK [34] and globally [43]. Although we did not detect the bla_{KPC-2} gene in our data, we did observe a plasmid with high homology to the KY798506 reference plasmid that is known to carry this carbapenemase gene. Whilst the presence of this plasmid might potentially increase the risk of the acquisition and spread of bla_{KPC-2} within this population it is currently not possible to speculate further on the public health relevance of this finding. The second ESBL-plasmid in multiple species is a JX424423(pKDO1)-like plasmid, which carries a $bla_{CTX-M-15}$ gene in a *Kpne* ST983 isolates and *Kpla* isolate 6Rf. In the latter case, this plasmid also harbours an additional 10 resistance genes and co-resides in isolate 6Rf with the pOXA-48-like plasmid. Thus both *Korn* and *Kpla* isolates can harbour multidrug resistance plasmids in the environment, and improved diagnostics are required to improve the reporting of serious infections caused by these species [44, 45].

The $bla_{CTX-M-15}$ gene is found associated with other plasmids in different *Kpne* isolates, including two highly related isolates of *Kpne* ST35 (7Sl and River_C) isolated from the WWTP and river respectively (supplementary note 3). This finding suggests that this clinically relevant clone, which also harbours a *yersiniabactin* virulence locus in our data, is circulating within the wider environment. The $bla_{CTX-M-15}$ gene is also associated with all nine isolates of *Kpne* ST983. This clone is associated with a suite of

CONFIDENTIAL

linked resistance genes, including *bla*$_{CTX\text{-}M\text{-}15}$, that co-mobilise between the different plasmids associated with this clone. Comparison with published data from an ST983 isolate from South Africa [13] suggests this association is stable on a global scale (Figure S14; supplementary note 2). This observation highlights that epidemiological surveillance should not only incorporate strains and plasmids, but should also consider mobile transposable elements that can independently transfer between them. Plasmid-borne transposons harbouring resistance genes have been noted to exhibit varying degrees of autonomy in *Kpne* [42], suggesting that the epidemiologically most relevant unit of mobility will vary between different species, strains, plasmids and resistance genes [46].

Although plasmids can be maintained even in the absence of the relevant antibiotics[47], the key observation that plasmid-borne resistance genes are enriched in the hospital wastewater sample points to selection pressures resulting from antibiotic exposure. However, it is unclear as to where this selection is predominantly operating. There have been a number of studies comparing wastewater and local clinical samples, and it is reasonably assumed that the microbial composition of hospital wastewater closely mirrors that of the hospital itself [14, 48-50]. However, an alternative is that resistance is both acquired and maintained directly within drain-associated biofilms, possibly driven by sub-lethal concentrations of antibiotics in these environments. In the absence of comparative data from the hospital itself, or on the concentration of carbapenems and other beta-lactam antibiotics within the wastewater, it is currently not possible to argue strongly in favour of one scenario over the other. Parallel sampling of hospital wastewater and patients over time would help to resolve this question.

This study has important practical implications regarding the public-health impact of *Klebsiella* species other than *Kpne*, and in particular *Korn* and other species corresponding to the *Raoultella* group. These species pose a risk both as emerging pathogens [45], but also as a potential reservoir for resistance determinants that can be readily transferred between species [51]. There are numerous reports of species of the *Raoultella* group harbouring carbapenemase genes, including those encoding OXA-48-like, VIM, KPC and NDM type enzymes in health-care, community and environmental settings [52-54]. Moreover, *bla*$_{OXA\text{-}48}$ specifically has been recognised as an increasing problem in the UK [55] and a major cause of hospital outbreaks [56, 57]. *Korn* isolates harbouring a pOXA-48 like plasmid have been previously reported from wastewater in the UK [14], although it is not possible to show that these plasmids are epidemiologically linked to those in the current study.

The pathogenic potential of *Korn*, and the putative virulence factors in this species, have not been widely studied. Our data confirm the presence of a chromosomally encoded yersiniabactin (*ybt*) locus, which is a major virulence factor for pulmonary infection in *Kpne* [58]. In *Kpne*, the *ybt* locus is typically located within an integrative conjugative element (ICE*Kp*). In contrast, in *Korn* the *ybt* locus is located in the chromosome next to a tRNA-*Asn* site but with no identifiable integrase gene [59]. We show that the *Korn ybt* variant is phylogenetically distinct, but its role in the virulence of this species is yet to be determined. Additionally, we also found a homologue of the *astA* gene embedded in an IS*256*-family transposon in 3 *Korn* isolates, two of which also contain the pOXA-48-like plasmid and the pKpQIL-like plasmid that commonly carries *bla*$_{KPC\text{-}2}$. The *astA* gene encodes for a heat-stable enterotoxin 1 (EAST1) in *E. coli* (usually EAEC or ETEC; [60]) and it has been described in enteropathogenic *Kpne* [61]. However, to the best of our knowledge, it has never been found in *Korn* (supplementary note 3). The relevance of this gene for the pathogenicity of *Korn* also remains to be elucidated, but it further raises its pathogenic potential.

In conclusion, our data reveal a high abundance of a pOXA-48-like plasmid in hospital wastewater in multiple *Klebsiella* species, and in particular within *Korn*. This plasmid was detected at a much lower

CONFIDENTIAL

frequency in the influent to the WWTP serving the hospital. These data warrant close surveillance both of this plasmid and of *Korn* and related species.

Authors and contributors

EJF and BKH conceived the study. MJG and NC carried out the bioinformatics analyses, with input from HAT, JC and SD. Sampling was carried out by KJ, AK and DK with logistical support from RB, RS and TC. The microbiology was carried out by MJG, and the sequencing was managed by MJG, with additional data supplied by SH. The paper was written by EJF, MJG, NC, MBA, BKH, SD, HAT and JC.

Conflicts of interest

No conflicts of interest to declare.

Funding information

EJF, MJG, NC, PD and MA are funded by MRC project awarded to MA (OH-DART MR/S004769/1). EJF, HAT and JC are funded by JPI-AMR project SpARK awarded to EJF (MR/R00241X/1). SH is funded by the Schlumberger Foundation. BKH is funded by Wessex Water and EPSRC IAA (grant no. EP/R51164X/1) and NERC (grant no. 2109393).

Ethical approval

NA

Table 1

	Hospital drain	WWTP influent	River	Total
K. pneumoniae (*Kpne*)	19	39	1	59
K. ornithinolytica (*Korn*)	20	5	0	25
K. quasipneumoniae subsp. *similipneumoniae* (*Kqps*)	4	0	0	4
K. variicola subsp. *variicola* (*Kvar*)	3	1	0	4
K. planticola (*Kpla*)	2	0	0	2
K. michiganensis (*Kmic*)	1	0	0	1
Total	49	45	1	95

CONFIDENTIAL

Table legends

Table 1. Number of isolates corresponding to each species isolated from hospital drain, influent at WWTP serving the hospital and local community and river approximately 8 km upstream.

Figure legends

Figure 1. Approximate maximum likelihood phylogenetic tree of the 95 isolates analysed in this study constructed using an alignment of 184,671 core SNPs. Species, source of each isolate and presence of carbapenemase resistance gene $bla_{\text{OXA-48}}$ are indicated. The tree and all metadata discussed can be accessed at https://microreact.org/project/Wastewater

Figure 2. Comparison of number of resistance genes identified using Abricate with the Resfinder database from the hospital wastewater and WWTP influent in (A) all isolates and (B) *Kpne* only.

Figure 3. Approximate maximum likelihood phylogenetic tree of the 59 *Kpne* isolates in this study isolated from the hospital wastewater, the WWTP influent and the river. The presence of plasmids identified by MOB-suite and listed by accession number, and of resistance genes identified using Abricate with the ResFinder database, are shown. STs are indicated. Only plasmids associated with resistance genes are included. The *Kpne* isolates 7Rg and 8Rg carry the pOXA-48 plasmid and are marked with a red asterisk.

Figure 4. Alignment of pOXA-48 ([17]; accession JN626286) and representative contigs from our isolates harbouring $bla_{\text{OXA-48}}$. Arrows represent ORFs; $bla_{\text{OXA-48}}$ is shown in green and insertion sequences in red.

Figure 5. Simplified hierarchical clustering of predicted plasmid, resistance genes and replicon types in 95 isolates. Green refers to absence and pink to presence. For the complete hierarchical clustering see Figure S3.

Figure 6. Approximate maximum likelihood phylogenetic tree of the 25 *Korn* isolates in this study isolated from the hospital wastewater and WWTP influent. The presence of plasmids identified by MOB-suite and listed by accession number, and of resistance genes identified using Abricate with the ResFinder database, are shown. Plasmids associated with resistance genes are marked with an asterisk.

Figure 7. Frequency of resistance genes identified as (A) plasmid-borne and (B) chromosomal in isolates from the hospital wastewater and WWTP influent. Contigs harbouring resistance genes were characterised as plasmid or chromosomal using MOB-suite and mlplasmids; plasmid numbers assigned by MOB-suite were used to determine number of plasmids per isolate.

References

1. Tacconelli, E., 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): p. 318-327 doi: 10.1016/S1473-3099(17)30753-3.

CONFIDENTIAL

2. Falagas, M.E., et al., *Deaths attributable to carbapenem-resistant Enterobacteriaceae infections.* Emerg Infect Dis, 2014. **20**(7): p. 1170-5 doi: 10.3201/eid2007.121004.
3. David, S., et al., *Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread.* Nat Microbiol, 2019. **4**(11): p. 1919-1929 doi: 10.1038/s41564-019-0492-8.
4. Amador, P.P., et al., *Antibiotic resistance in wastewater: occurrence and fate of Enterobacteriaceae producers of class A and class C beta-lactamases.* J Environ Sci Health A Tox Hazard Subst Environ Eng, 2015. **50**(1): p. 26-39 doi: 10.1080/10934529.2015.964602.
5. Korzeniewska, E., A. Korzeniewska, and M. Harnisz, *Antibiotic resistant Escherichia coli in hospital and municipal sewage and their emission to the environment.* Ecotoxicol Environ Saf, 2013. **91**: p. 96-102 doi: 10.1016/j.ecoenv.2013.01.014.
6. Pazda, M., et al., *Antibiotic resistance genes identified in wastewater treatment plant systems - A review.* Sci Total Environ, 2019. **697**: p. 134023 doi: 10.1016/j.scitotenv.2019.134023.
7. Van Kregten, E., N.A. Westerdaal, and J.M. Willers, *New, simple medium for selective recovery of Klebsiella pneumoniae and Klebsiella oxytoca from human feces.* J Clin Microbiol, 1984. **20**(5): p. 936-41 doi: 10.1128/JCM.20.5.936-941.1984.
8. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20 doi: 10.1093/bioinformatics/btu170.
9. Nurk, S., et al., *Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads*, in *RECOMB 2013*, M. Deng, et al., Editors. 2013, Springer Berlin Heidelberg: Beijing, China. p. 158-170.
10. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9 doi: 10.1093/bioinformatics/btu153.
11. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree: computing large minimum evolution trees with profiles instead of a distance matrix.* Mol Biol Evol, 2009. **26**(7): p. 1641-50 doi: 10.1093/molbev/msp077.
12. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree 2--approximately maximum-likelihood trees for large alignments.* PLoS One, 2010. **5**(3): p. e9490 doi: 10.1371/journal.pone.0009490.
13. Founou, R.C., et al., *Whole Genome Sequencing of Extended Spectrum beta-lactamase (ESBL)-producing Klebsiella pneumoniae Isolated from Hospitalized Patients in KwaZulu-Natal, South Africa.* Sci Rep, 2019. **9**(1): p. 6266 doi: 10.1038/s41598-019-42672-2.
14. Ludden, C., et al., *Sharing of carbapenemase-encoding plasmids between Enterobacteriaceae in UK sewage uncovered by MinION sequencing.* Microb Genom, 2017. **3**(7): p. e000114 doi: 10.1099/mgen.0.000114.
15. Argimon, S., et al., *Microreact: visualizing and sharing data for genomic epidemiology and phylogeography.* Microb Genom, 2016. **2**(11): p. e000093 doi: 10.1099/mgen.0.000093.
16. Wyres, K.L., M.M.C. Lam, and K.E. Holt, *Population genomics of Klebsiella pneumoniae.* Nat Rev Microbiol, 2020. **18**(6): p. 344-359 doi: 10.1038/s41579-019-0315-1.
17. Poirel, L., R.A. Bonnin, and P. Nordmann, *Genetic features of the widespread plasmid coding for the carbapenemase OXA-48.* Antimicrob Agents Chemother, 2012. **56**(1): p. 559-62 doi: 10.1128/AAC.05289-11.
18. Aubert, D., et al., *Functional characterization of IS1999, an IS4 family element involved in mobilization and expression of beta-lactam resistance genes.* J Bacteriol, 2006. **188**(18): p. 6506-14 doi: 10.1128/JB.00375-06.
19. Sullivan, M.J., N.K. Petty, and S.A. Beatson, *Easyfig: a genome comparison visualizer.* Bioinformatics, 2011. **27**(7): p. 1009-10 doi: 10.1093/bioinformatics/btr039.
20. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60 doi: 10.1093/bioinformatics/btp324.
21. Milne, I., et al., *Using Tablet for visual exploration of second-generation sequencing data.* BRIEFINGS IN BIOINFORMATICS, 2012. **14**: p. 193-202 doi: 10.1093/bib/bbs012.

CONFIDENTIAL

22. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* Mol Biol Evol, 2013. **30**(4): p. 772-80 doi: 10.1093/molbev/mst010.
23. Croucher, N.J., et al., *Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.* Nucleic Acids Res, 2015. **43**(3): p. e15 doi: 10.1093/nar/gku1196.
24. Liu, M., et al., *ICEberg 2.0: an updated database of bacterial integrative and conjugative elements.* Nucleic Acids Res, 2019. **47**(D1): p. D660-D665 doi: 10.1093/nar/gky1123.
25. Carattoli, A., et al., *In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing.* Antimicrob Agents Chemother, 2014. **58**(7): p. 3895-903 doi: 10.1128/AAC.02412-14.
26. Robertson, J. and J.H.E. Nash, *MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies.* Microb Genom, 2018. **4**(8) doi: 10.1099/mgen.0.000206.
27. Arredondo-Alonso, S., et al., *mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species.* Microb Genom, 2018. **4**(11) doi: 10.1099/mgen.0.000224.
28. Rodriguez-Martinez, J.M., et al., *Contribution of OqxAB efflux pumps to quinolone resistance in extended-spectrum-beta-lactamase-producing Klebsiella pneumoniae.* J Antimicrob Chemother, 2013. **68**(1): p. 68-73 doi: 10.1093/jac/dks377.
29. Wan Nur Ismah, W.A.K., et al., *Impact of OqxR loss of function on the envelope proteome of Klebsiella pneumoniae and susceptibility to antimicrobials.* J Antimicrob Chemother, 2018. **73**(11): p. 2990-2996 doi: 10.1093/jac/dky293.
30. Pitout, J.D.D., et al., *The Global Ascendency of OXA-48-Type Carbapenemases.* Clin Microbiol Rev, 2019. **33**(1) doi: 10.1128/CMR.00102-19.
31. Stoesser, N., et al., *Complete Sequencing of Plasmids Containing blaOXA-163 and blaOXA-48 in Escherichia coli Sequence Type 131.* Antimicrob Agents Chemother, 2016. **60**(11): p. 6948-6951 doi: 10.1128/AAC.01130-16.
32. Holt, K.E., et al., *Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health.* Proc Natl Acad Sci U S A, 2015. **112**(27): p. E3574-81 doi: 10.1073/pnas.1501049112.
33. Kaplan, E., et al., *Genomic and Functional Characterization of qnr-Encoding Plasmids from Municipal Wastewater Biosolid Klebsiella pneumoniae Isolates.* Front Microbiol, 2015. **6**: p. 1354 doi: 10.3389/fmicb.2015.01354.
34. Doumith, M., et al., *Major role of pKpQIL-like plasmids in the early dissemination of KPC-type carbapenemases in the UK.* J Antimicrob Chemother, 2017. **72**(8): p. 2241-2248 doi: 10.1093/jac/dkx141.
35. Martin, J., et al., *Covert dissemination of carbapenemase-producing Klebsiella pneumoniae (KPC) in a successfully controlled outbreak: long- and short-read whole-genome sequencing demonstrate multiple genetic modes of transmission.* J Antimicrob Chemother, 2017. **72**(11): p. 3025-3034 doi: 10.1093/jac/dkx264.
36. Verburg, I., et al., *Abundance and Antimicrobial Resistance of Three Bacterial Species along a Complete Wastewater Pathway.* Microorganisms, 2019. **7**(9) doi: 10.3390/microorganisms7090312.
37. Arredondo-Alonso, S., et al., *Plasmids Shaped the Recent Emergence of the Major Nosocomial Pathogen Enterococcus faecium.* mBio, 2020. **11**(1) doi: 10.1128/mBio.03284-19.
38. Ford, P.J. and M.B. Avison, *Evolutionary mapping of the SHV beta-lactamase and evidence for two separate IS26-dependent blaSHV mobilization events from the Klebsiella pneumoniae chromosome.* J Antimicrob Chemother, 2004. **54**(1): p. 69-75 doi: 10.1093/jac/dkh251.
39. Wiener, J., et al., *Multiple antibiotic-resistant Klebsiella and Escherichia coli in nursing homes.* JAMA, 1999. **281**(6): p. 517-23 doi: 10.1001/jama.281.6.517.

CONFIDENTIAL

40. Poirel, L., et al., *Emergence of oxacillinase-mediated resistance to imipenem in Klebsiella pneumoniae.* Antimicrob Agents Chemother, 2004. **48**(1): p. 15-22 doi: 10.1128/aac.48.1.15-22.2004.
41. Poirel, L., A. Potron, and P. Nordmann, *OXA-48-like carbapenemases: the phantom menace.* J Antimicrob Chemother, 2012. **67**(7): p. 1597-606 doi: 10.1093/jac/dks121.
42. David, S., 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**(40): p. 25043-25054 doi: 10.1073/pnas.2003407117.
43. Hamprecht, A., et al., *Pathogenicity of Clinical OXA-48 Isolates and Impact of the OXA-48 IncL Plasmid on Virulence and Bacterial Fitness.* Front Microbiol, 2019. **10**: p. 2509 doi: 10.3389/fmicb.2019.02509.
44. Demiray, T., et al., *A rare cause of infection, Raoultella planticola: emerging threat and new reservoir for carbapenem resistance.* Infection, 2016. **44**(6): p. 713-717 doi: 10.1007/s15010-016-0900-4.
45. Hajjar, R., et al., *Raoultella ornithinolytica: Emergence and Resistance.* Infect Drug Resist, 2020. **13**: p. 1091-1104 doi: 10.2147/IDR.S191387.
46. Sheppard, A.E., et al., *Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene blaKPC.* Antimicrob Agents Chemother, 2016. **60**(6): p. 3767-78 doi: 10.1128/AAC.00464-16.
47. Rahube, T.O., et al., *Characterization and comparative analysis of antibiotic resistance plasmids isolated from a wastewater treatment plant.* Front Microbiol, 2014. **5**: p. 558 doi: 10.3389/fmicb.2014.00558.
48. Brechet, C., et al., *Wastewater treatment plants release large amounts of extended-spectrum beta-lactamase-producing Escherichia coli into the environment.* Clin Infect Dis, 2014. **58**(12): p. 1658-65 doi: 10.1093/cid/ciu190.
49. Raven, K.E., et al., *Genomic surveillance of Escherichia coli in municipal wastewater treatment plants as an indicator of clinically relevant pathogens and their resistance genes.* Microb Genom, 2019. **5**(5) doi: 10.1099/mgen.0.000267.
50. Surleac, M., et al., *Whole genome sequencing snapshot of multi-drug resistant Klebsiella pneumoniae strains from hospitals and receiving wastewater treatment plants in Southern Romania.* PLoS One, 2020. **15**(1): p. e0228079 doi: 10.1371/journal.pone.0228079.
51. Baquero, F., et al., *Gene Transmission in the One Health Microbiosphere and the Channels of Antimicrobial Resistance.* Front Microbiol, 2019. **10**: p. 2892 doi: 10.3389/fmicb.2019.02892.
52. Wang, S., et al., *Emergence of NDM-1- and CTX-M-3-Producing Raoultella ornithinolytica in Human Gut Microbiota.* Front Microbiol, 2019. **10**: p. 2678 doi: 10.3389/fmicb.2019.02678.
53. Piedra-Carrasco, N., et al., *Carbapenemase-producing enterobacteriaceae recovered from a Spanish river ecosystem.* PLoS One, 2017. **12**(4): p. e0175246 doi: 10.1371/journal.pone.0175246.
54. Zou, H., et al., *Genetic characterization and virulence of a carbapenem-resistant Raoultella ornithinolytica isolated from well water carrying a novel megaplasmid containing blaNDM-1.* Environ Pollut, 2020. **260**: p. 114041 doi: 10.1016/j.envpol.2020.114041.
55. Findlay, J., et al., *OXA-48-like carbapenemases in the UK: an analysis of isolates and cases from 2007 to 2014.* J Antimicrob Chemother, 2017. **72**(5): p. 1340-1349 doi: 10.1093/jac/dkx012.
56. Ledda, A., et al., *Hospital outbreak of carbapenem-resistant Enterobacteriales associated with an OXA-48 plasmid carried mostly by Escherichia coli ST399.* 2020 doi: 10.1101/2020.06.15.148189.
57. Sivaramakrishnan, A., et al., *Epidemiology and control measures of an OXA-48-producing Enterobacteriaceae hospital outbreak.* Infection Prevention in Practice, 2020. **2**(3) doi: 10.1016/j.infpip.2019.100021.

CONFIDENTIAL

58. Lawlor, M.S., C. O'Connor, and V.L. Miller, *Yersiniabactin is a virulence factor for Klebsiella pneumoniae during pulmonary infection.* Infect Immun, 2007. **75**(3): p. 1463-72 doi: 10.1128/IAI.00372-06.
59. Lam, M.M.C., et al., *Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICEKp in Klebsiella pneumoniae populations.* Microb Genom, 2018. **4**(9) doi: 10.1099/mgen.0.000196.
60. McVeigh, A., et al., *IS1414, an Escherichia coli insertion sequence with a heat-stable enterotoxin gene embedded in a transposase-like gene.* Infect Immun, 2000. **68**(10): p. 5710-5 doi: 10.1128/iai.68.10.5710-5715.2000.
61. Nguyen Thi, P.L., et al., *Enteropathogenic Klebsiella pneumoniae HIV-infected adults, Africa.* Emerg Infect Dis, 2003. **9**(1): p. 135-7 doi: 10.3201/eid0901.020138.

CONFIDENTIAL

Click here to access/download;Figure;figure 1.pdf

Figure 1.

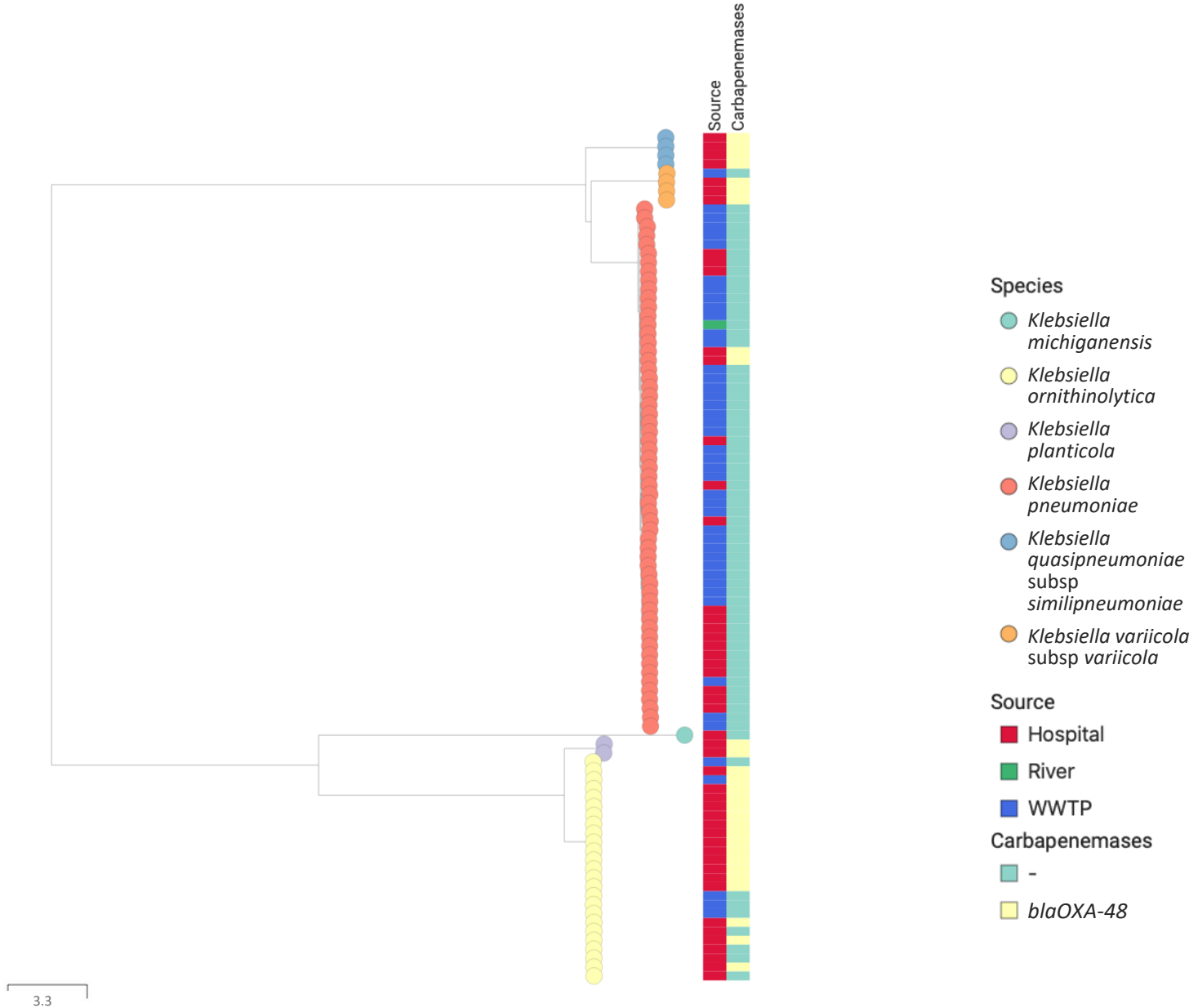
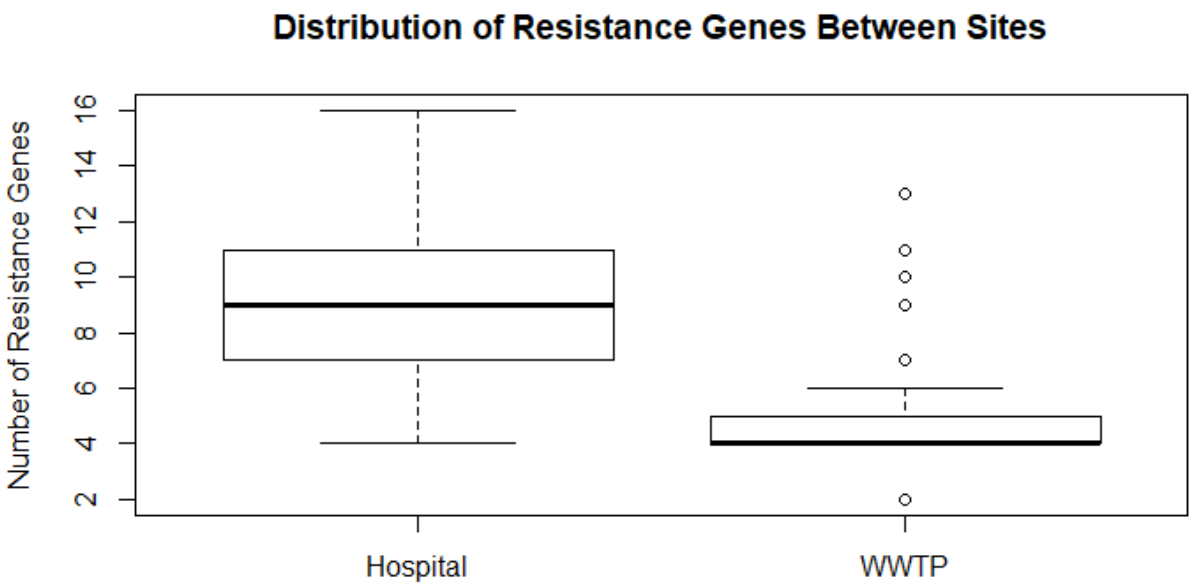


Figure 2.

A


Distribution of Resistance Genes Between Sites
Number of Resistance Genes
2
4
6
8
10
12
14
16
Hospital
WWTP

B

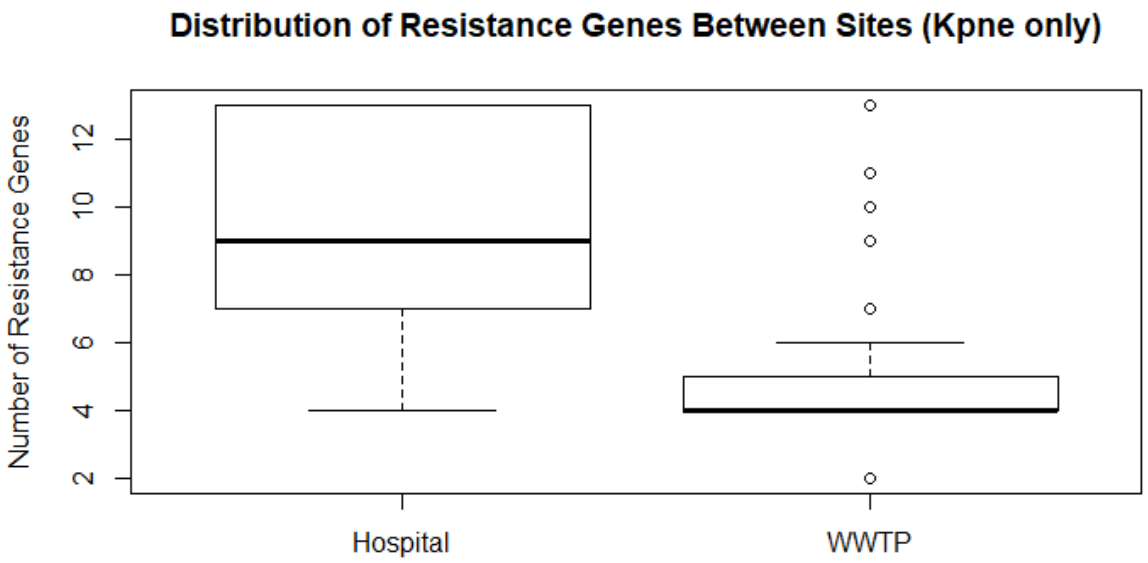

Distribution of Resistance Genes Between Sites (Kpne only)
Number of Resistance Genes
2
4
6
8
10
12
Hospital
WWTP

Figure 3.

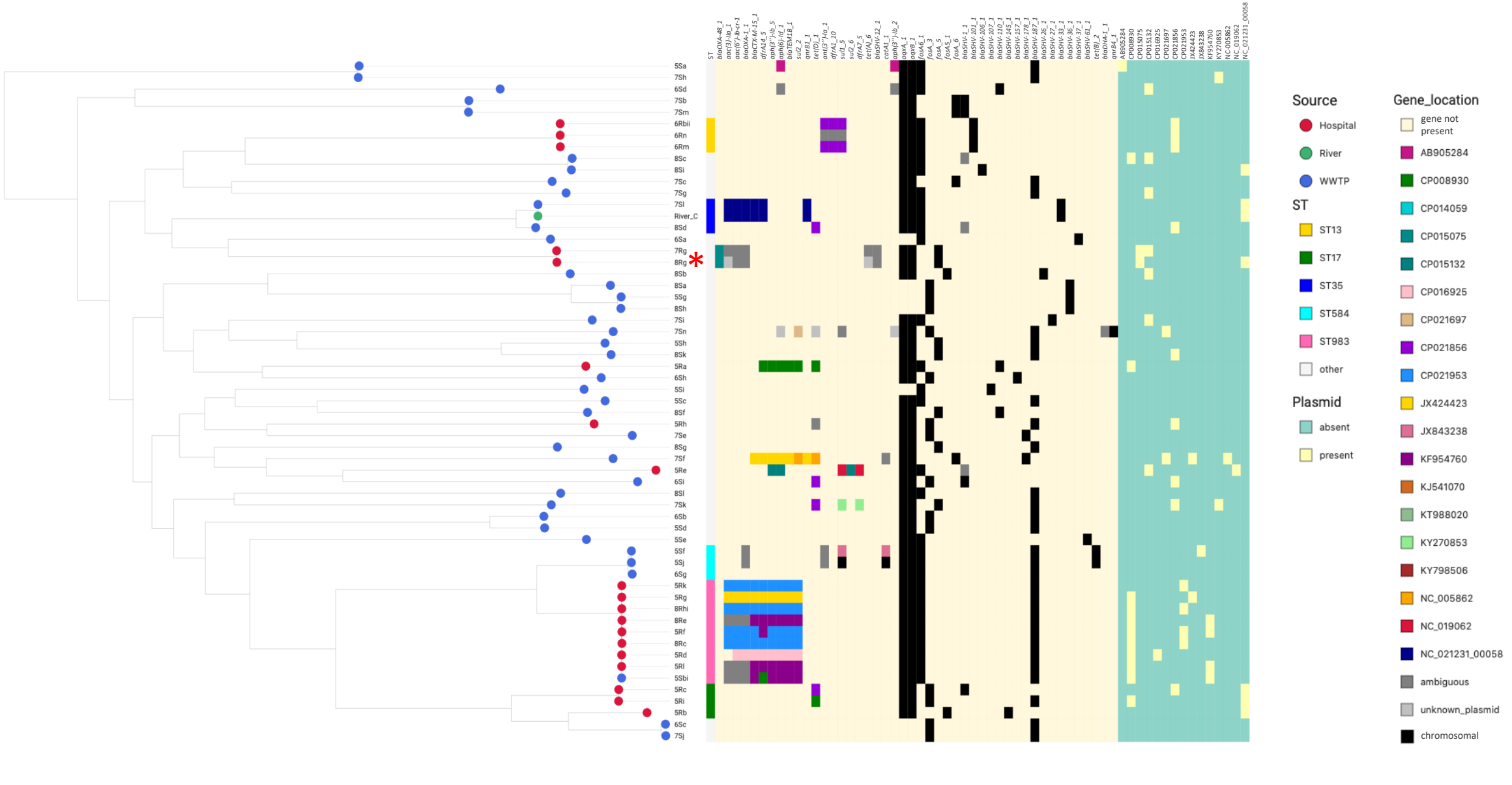

ST
blaOXA-48_1
aac(3)-IIe_1
aac(6')-Ib-cr-1
blaOXA-1_1
blaCTX-M-15_1
dfrA14_5
aph(3'')-Ib_5
aph(6)-Id_1
blaTEM-1B_1
sul2_2
qnrB1_1
tet(D)_1
ant(3'')-Ia_1
dfrA1_10
sul1_5
sul2_6
dfrA7_5
tet(A)_6
blaSHV-12_1
catA1_1
aph(3'')-Ib_2
aadA1_1
aadA8_1
fosA6_1
fosA_3
fosA_5
fosA5_1
fosA_6
blaSHV-1_1
blaSHV-101_1
blaSHV-106_1
blaSHV-107_1
blaSHV-110_1
blaSHV-145_1
blaSHV-157_1
blaSHV-178_1
blaSHV-187_1
blaSHV-26_1
blaSHV-27_1
blaSHV-33_1
blaSHV-36_1
blaSHV-37_1
blaSHV-61_1
tet(B)_2
blaDHA-1_1
qnrB4_1
AB905284
CP008930
CP015075
CP015132
CP016925
CP021697
CP021856
CP021953
JX424423
JX843238
KF954760
KY270853
NC_005862
NC_019062
NC_021231_00058
5Sa
7Sh
6Sd
7Sb
7Sm
6Rbii
6Rn
6Rm
8Sc
8Si
7Sc
7Sg
7Sl
River_C
8Sd
6Sa
7Rg
8Rg
8Sb
8Sa
5Sg
8Sh
7Si
7Sn
5Sh
8Sk
5Ra
6Sh
5Si
5Sc
8Sf
5Rh
7Se
8Sg
7Sf
5Re
6Si
8Sl
7Sk
6Sb
5Sd
5Se
5Sf
5Sj
6Sg
5Rk
5Rg
8Rhi
8Re
5Rf
8Rc
5Rd
5Rl
5Sbi
5Rc
5Ri
5Rb
6Sc
7Sj
0.0015
Source
Hospital
River
WWTP
ST
ST13
ST17
ST35
ST584
ST983
other
Plasmid
absent
present
Gene_location
gene not present
AB905284
CP008930
CP014059
CP015075
CP015132
CP016925
CP021697
CP021856
CP021953
JX424423
JX843238
KF954760
KJ541070
KT988020
KY270853
KY798506
NC_005862
NC_019062
NC_021231_00058
ambiguous
unknown_plasmid
chromosomal

Figure 4.

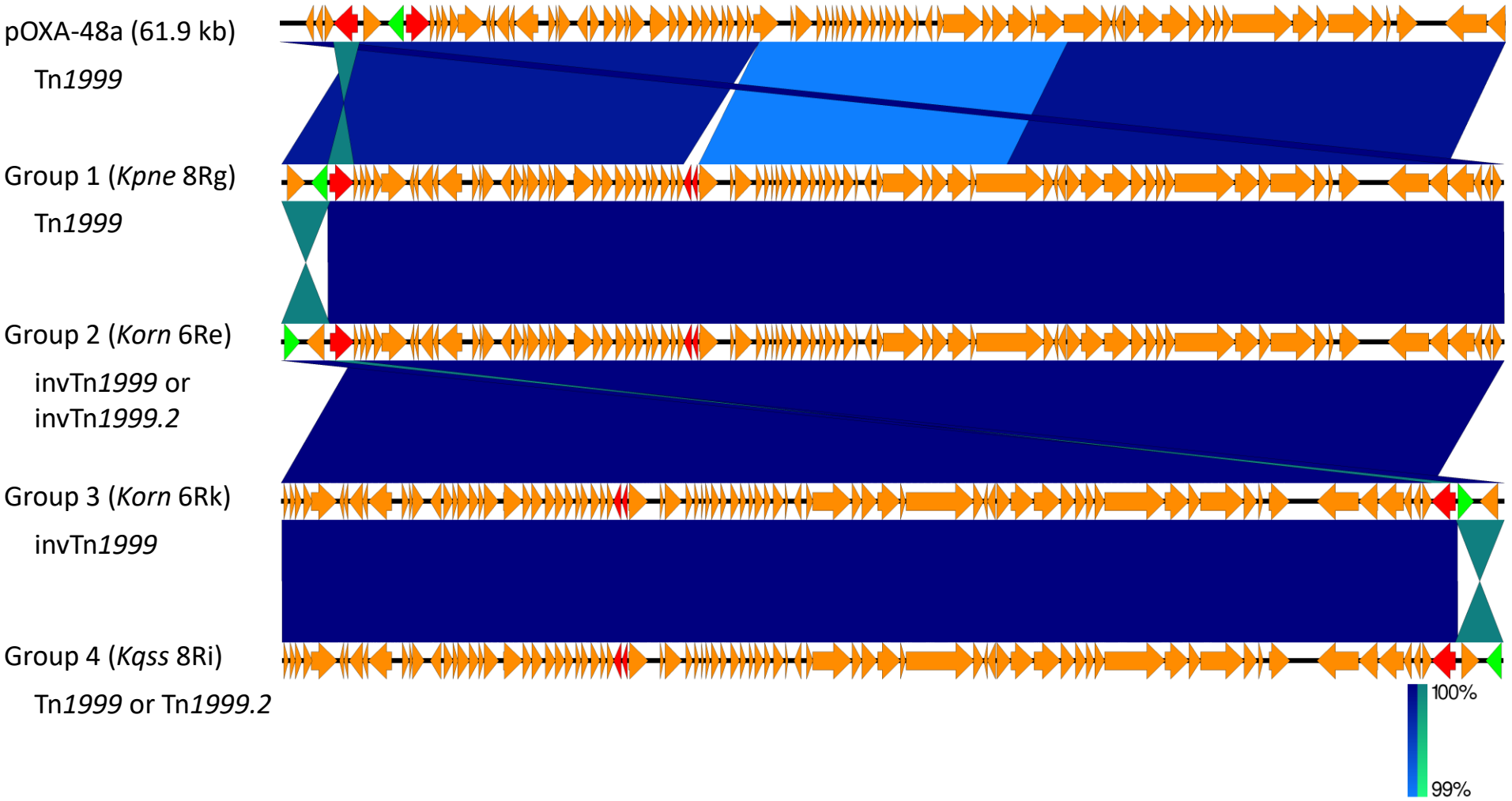

pOXA-48a (61.9 kb)
Tn*1999*
Group 1 (*Kpne* 8Rg)
Tn*1999*
Group 2 (*Korn* 6Re)
invTn*1999* or
invTn*1999.2*
Group 3 (*Korn* 6Rk)
invTn*1999*
Group 4 (*Kqss* 8Ri)
Tn*1999* or Tn*1999.2*
100%
99%

Figure 5.

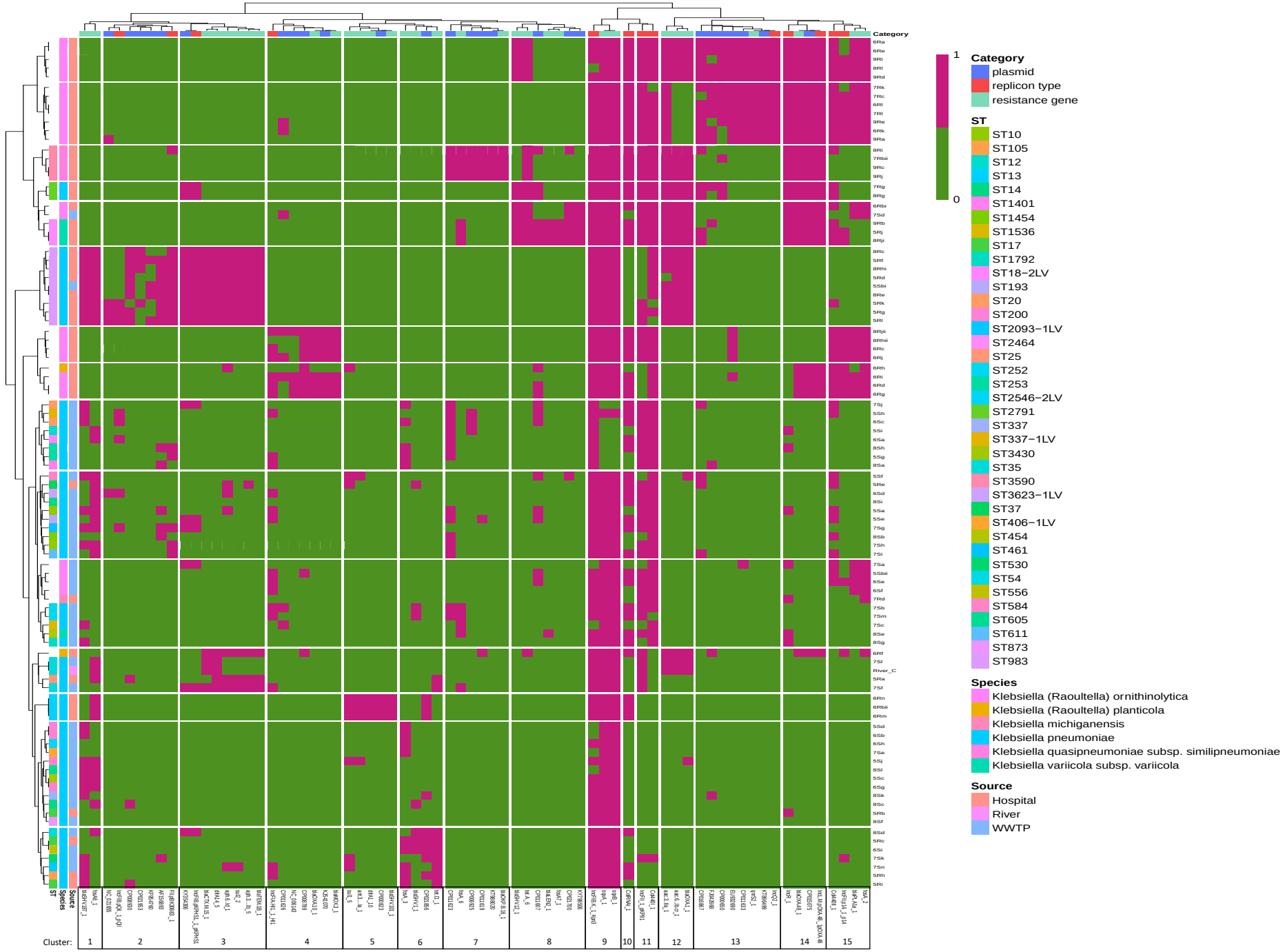

Category
plasmid
replicon type
resistance gene
ST
ST10
ST105
ST12
ST13
ST14
ST1401
ST1454
ST1536
ST17
ST1792
ST18−2LV
ST193
ST20
ST200
ST2093−1LV
ST2464
ST25
ST252
ST253
ST2546−2LV
ST2791
ST337
ST337−1LV
ST3430
ST35
ST3590
ST3623−1LV
ST37
ST406−1LV
ST454
ST461
ST530
ST54
ST556
ST584
ST605
ST611
ST873
ST983
Species
Klebsiella (Raoultella) ornithinolytica
Klebsiella (Raoultella) planticola
Klebsiella michiganensis
Klebsiella pneumoniae
Klebsiella quasipneumoniae subsp. similipneumoniae
Klebsiella variicola subsp. variicola
Source
Hospital
River
WWTP
1
0
ST
Species
Source
Cluster:
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Figure 6.

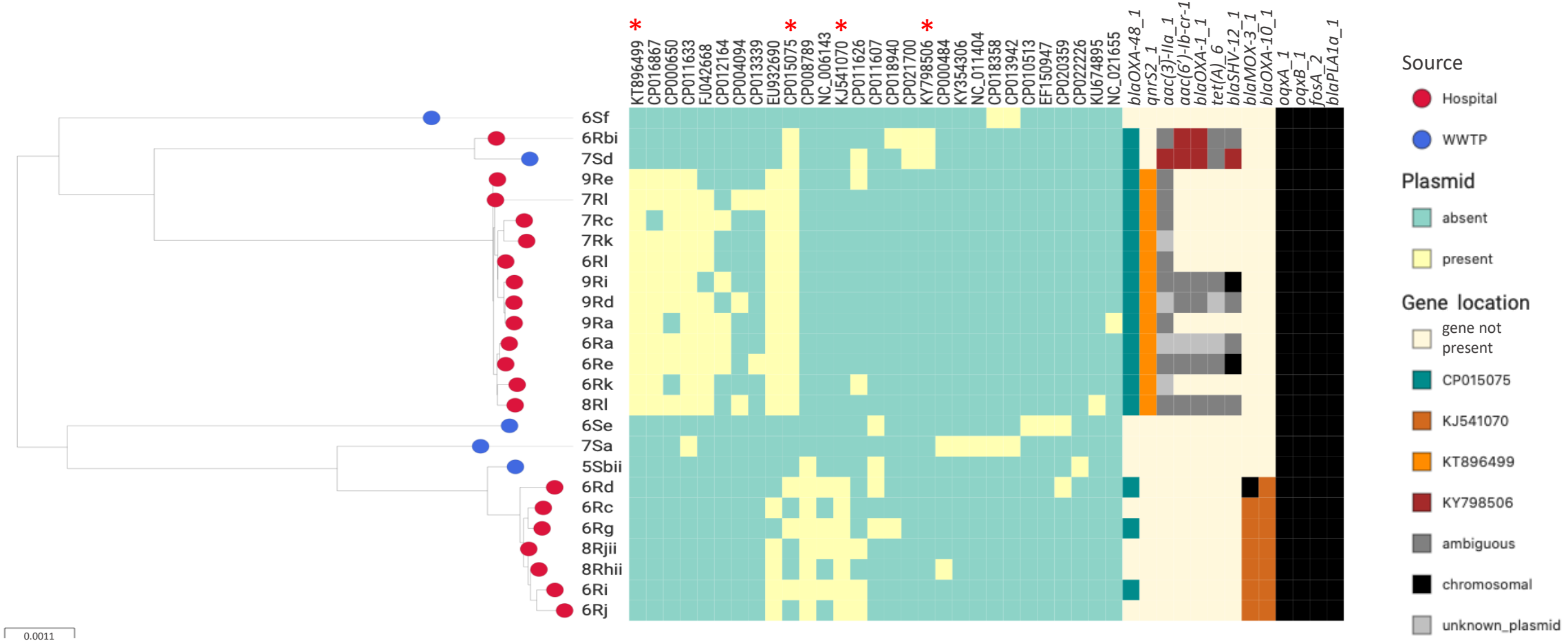
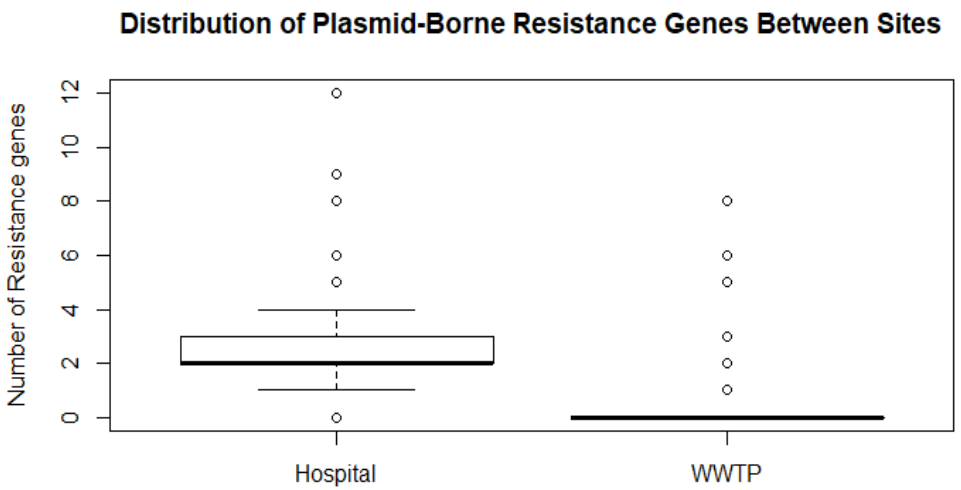
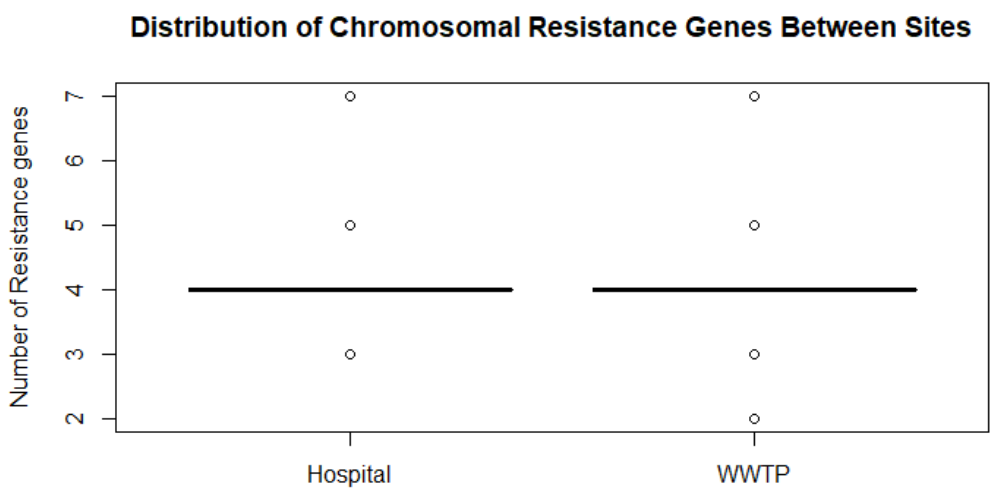

KT896499
CP016867
CP000650
CP011633
FJ042668
CP012164
CP004094
CP013339
EU932690
CP015075
CP008789
NC_006143
KJ541070
CP011626
CP011607
CP018940
CP021700
KY798506
CP000484
KY354306
NC_011404
CP018358
CP013942
CP010513
EF150947
CP020359
CP022226
KU674895
NC_021655
blaOXA-48_1
qnrS2_1
aac(3)-IIa_1
aac(6')-Ib-cr-1
blaOXA-1_1
tet(A)_6
blaSHV-12_1
blaMOX-3_1
blaOXA-10_1
oqxA_1
oqxB_1
fosA_2
blaPLA1a_1
6Sf
6Rbi
7Sd
9Re
7Rl
7Rc
7Rk
6Rl
9Ri
9Rd
9Ra
6Ra
6Re
6Rk
8Rl
6Se
7Sa
5Sbii
6Rd
6Rc
6Rg
8Rjii
8Rhii
6Ri
6Rj
0.0011
Source
Hospital
WWTP
Plasmid
absent
present
Gene location
gene not present
CP015075
KJ541070
KT896499
KY798506
ambiguous
chromosomal
unknown_plasmid

Figure 7.

A


Distribution of Plasmid-Borne Resistance Genes Between Sites
Number of Resistance genes
0
2
4
6
8
10
12
Hospital
WWTP

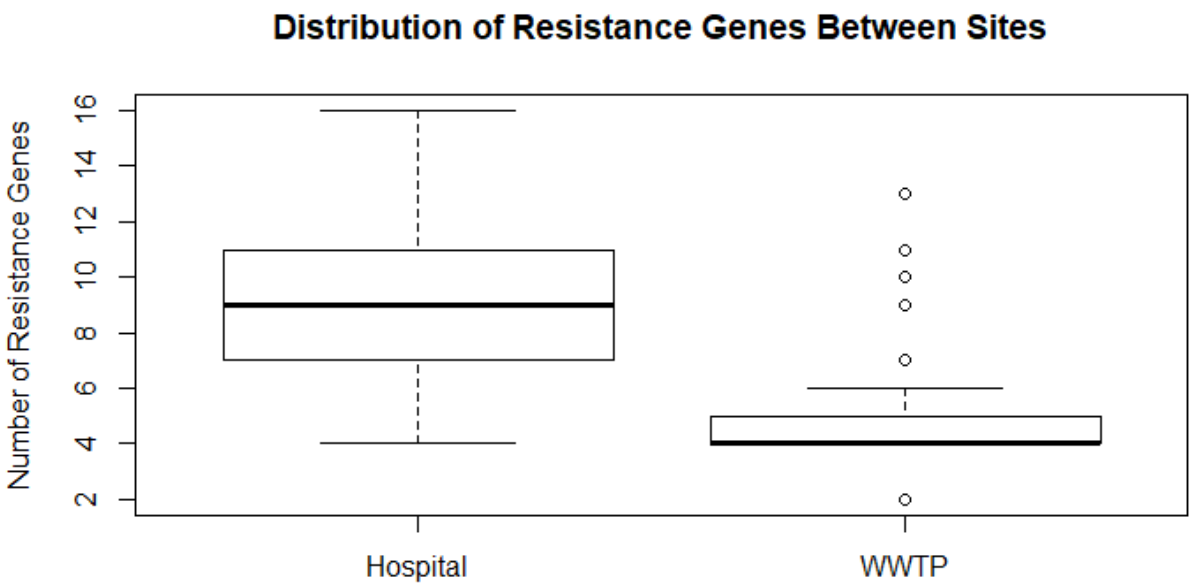
B


Distribution of Chromosomal Resistance Genes Between Sites
Number of Resistance genes
2
3
4
5
6
7
Hospital
WWTP

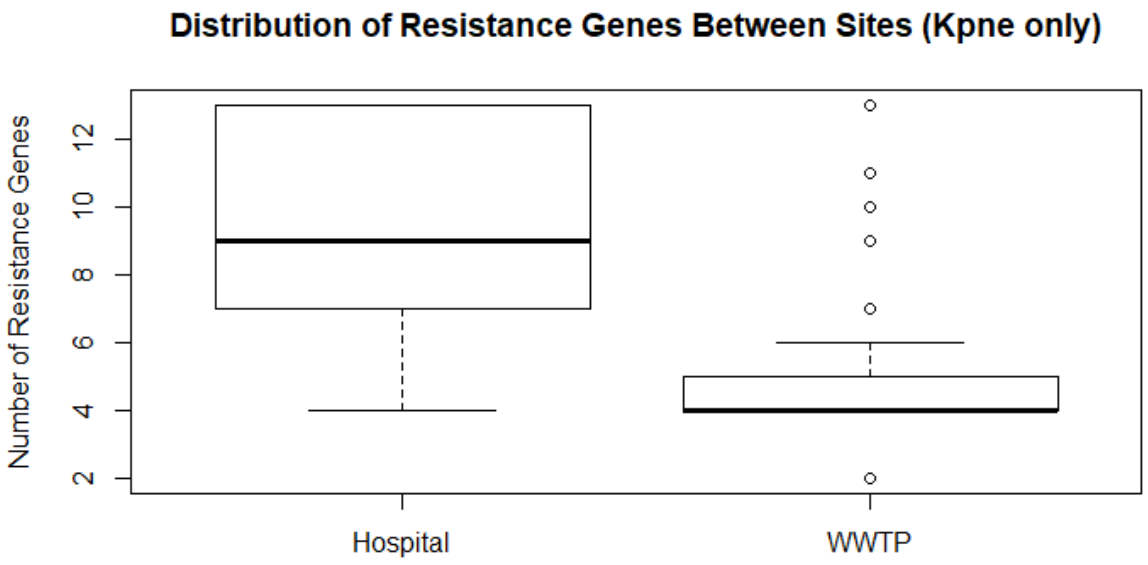
Click here to access/download;Figure;Figure 2.pdf

Figure 2.

A

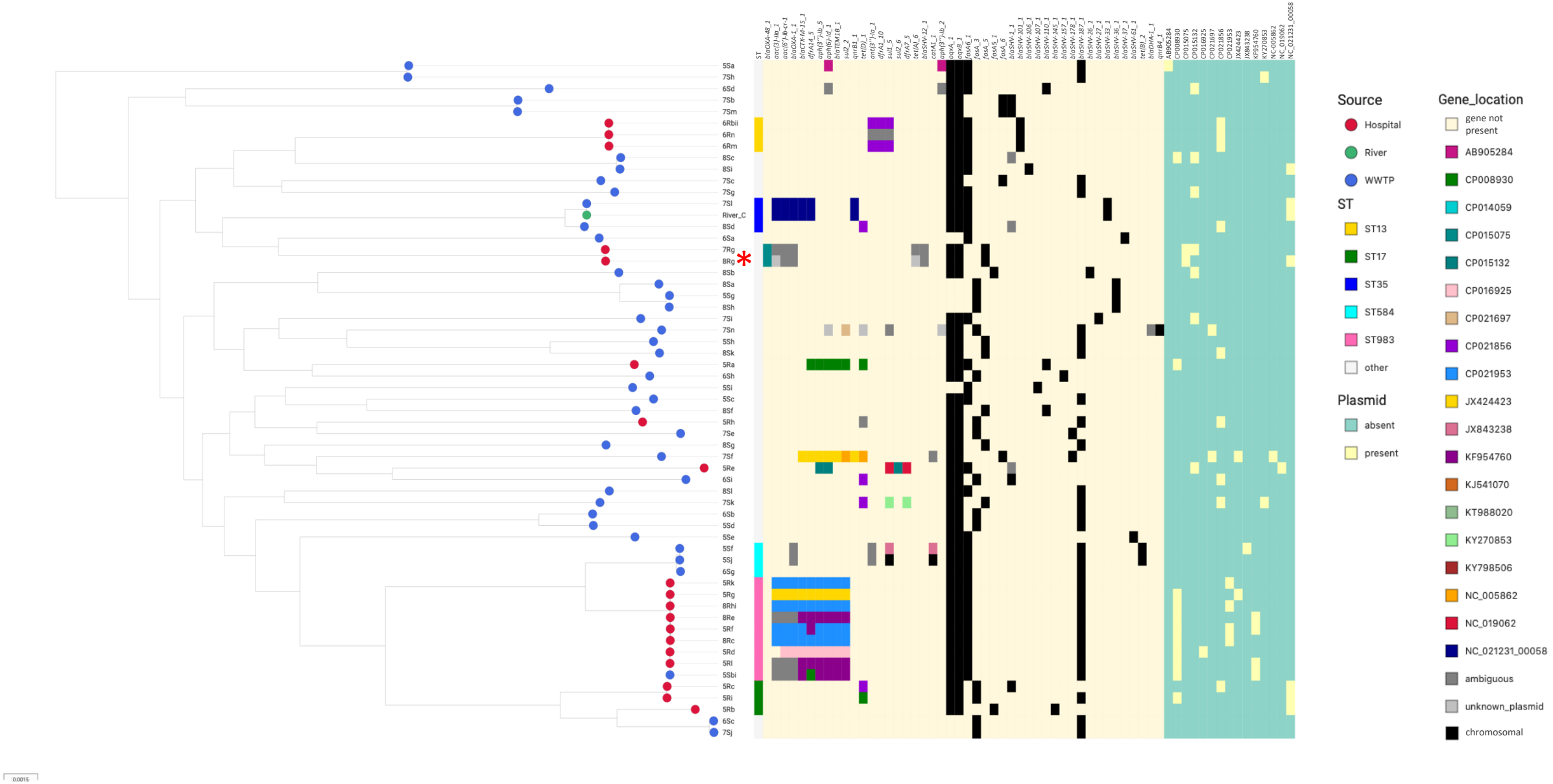


B



Click here to access/download;Figure;Figure 3.pdf

Figure 3.



Click here to access/download;Figure;Figure 4.pdf

Figure 4.

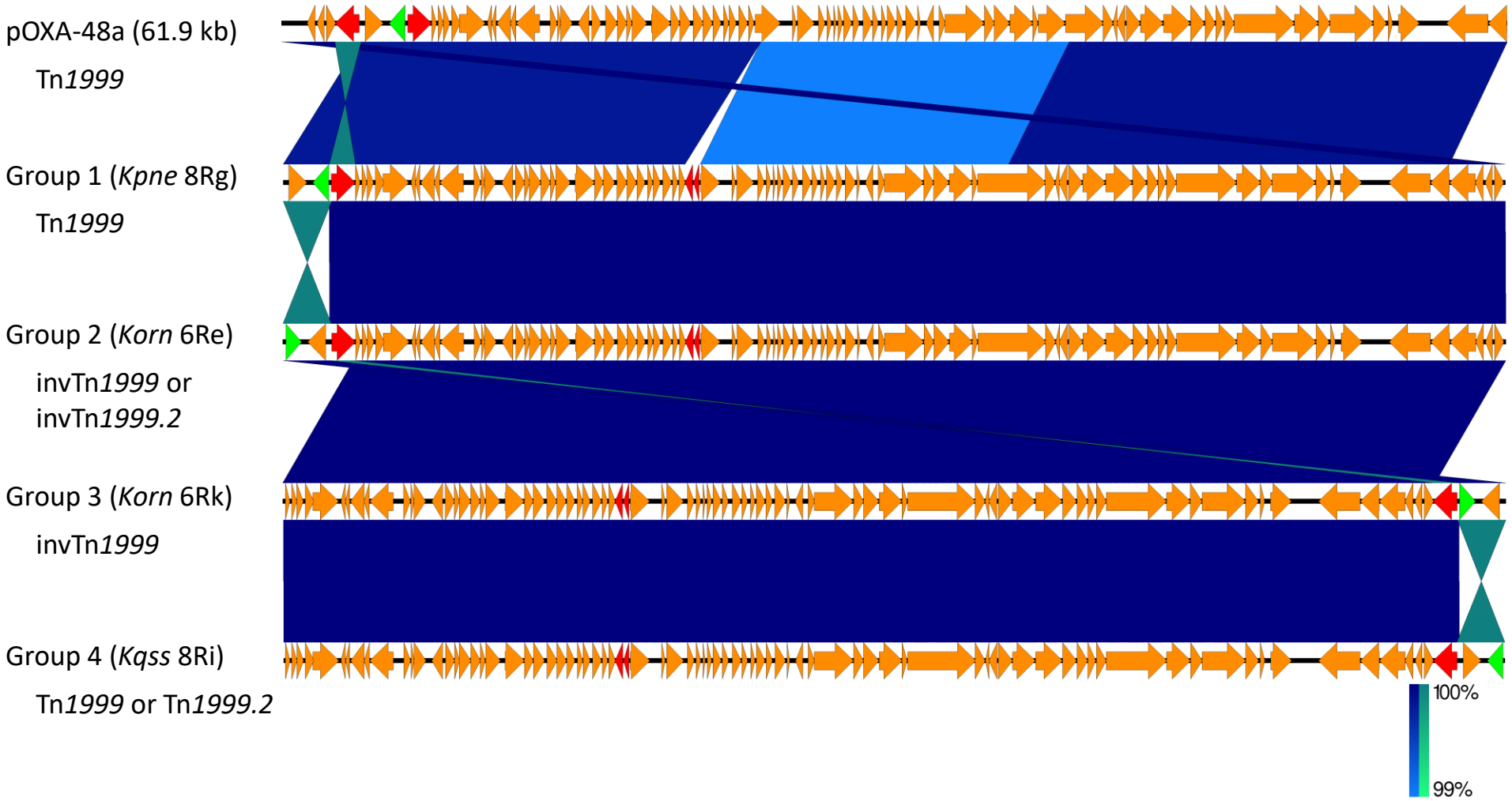


Figure 5

Click here to access/download;Figure;Figure 5.pdf

Figure 5.

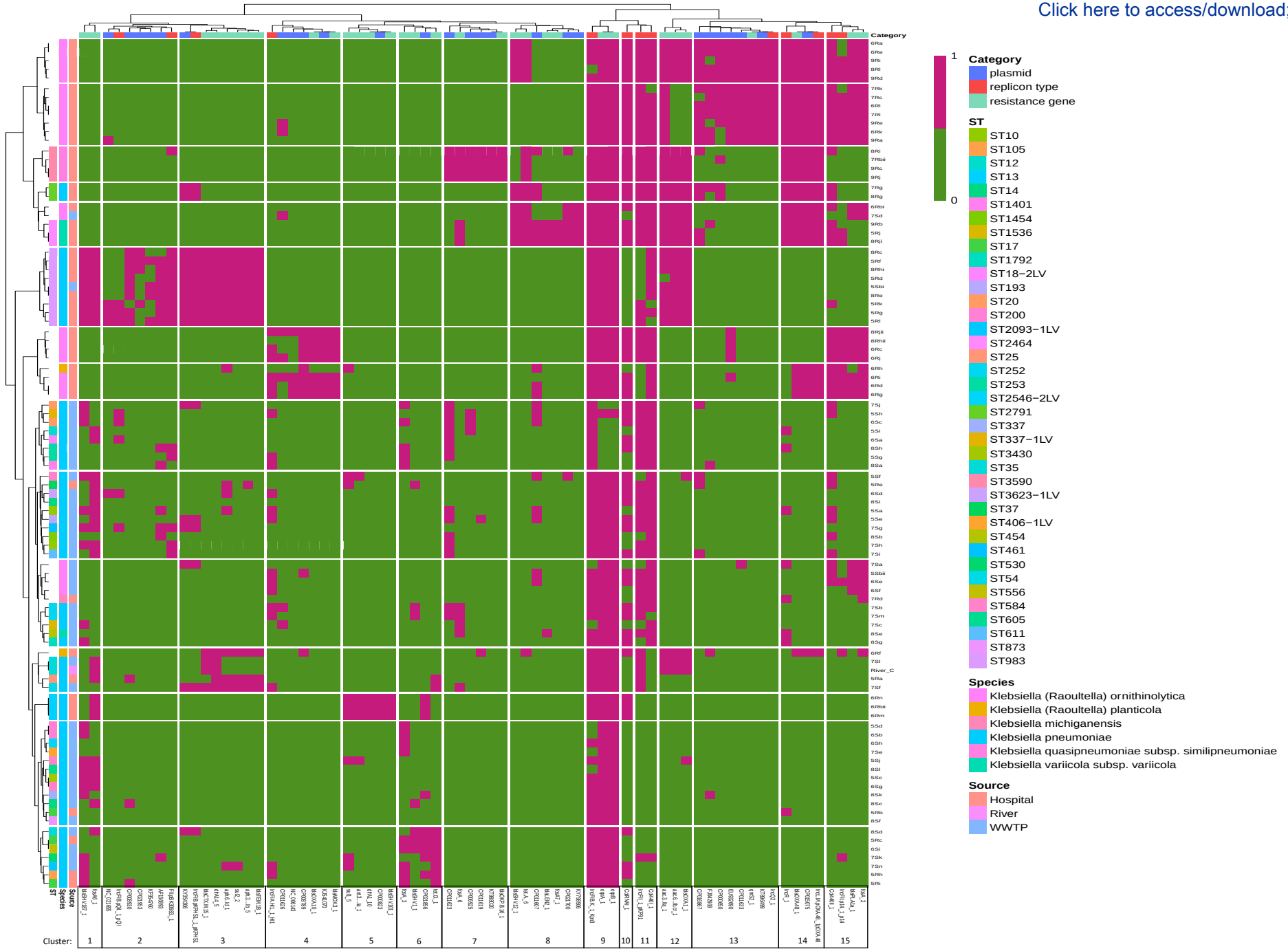
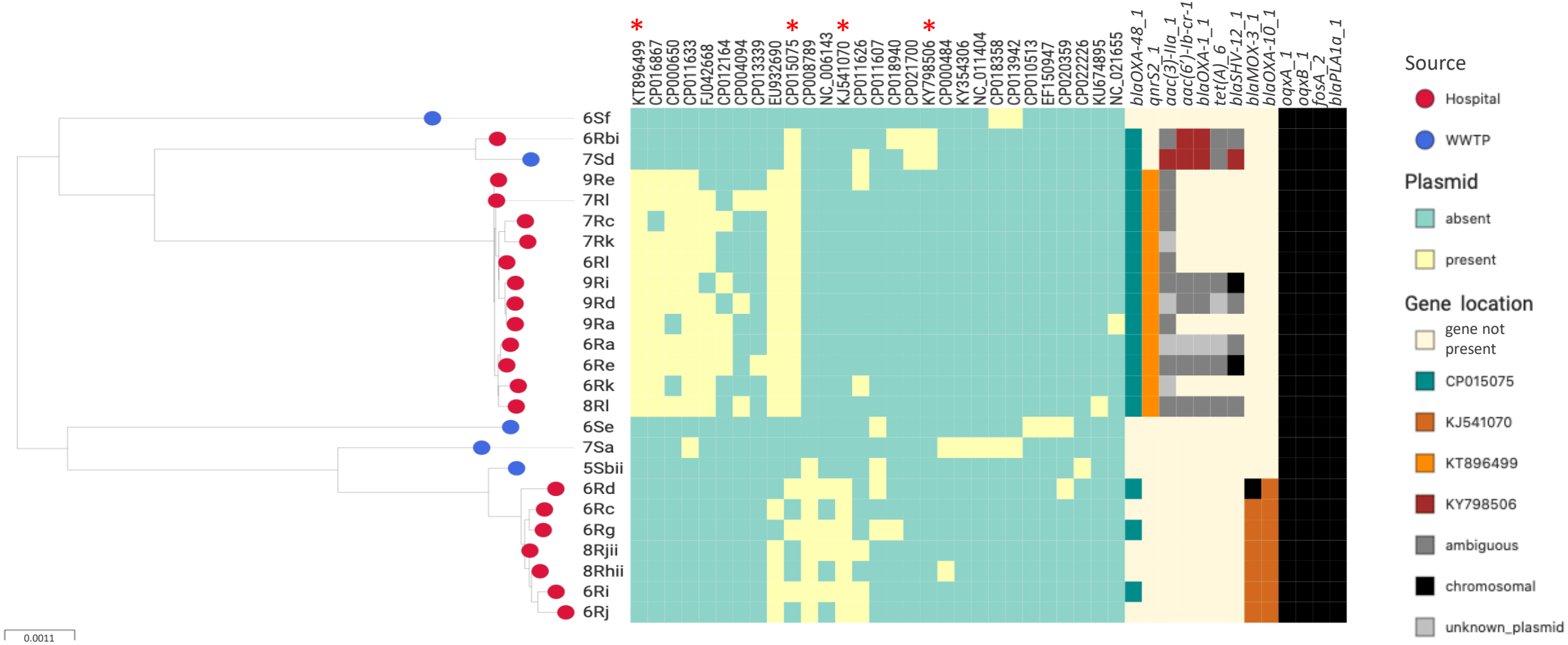


Figure 6

Click here to access/download;Figure;Figure 6.pdf

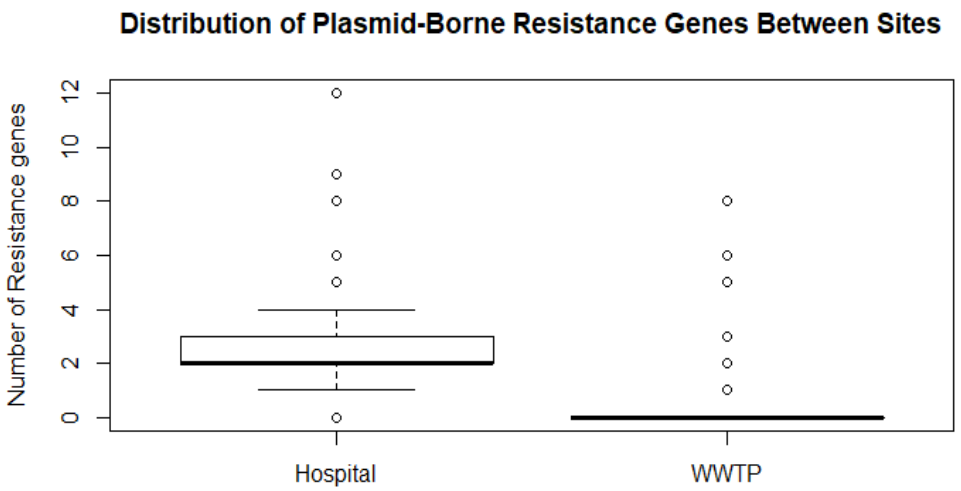
Figure 6.



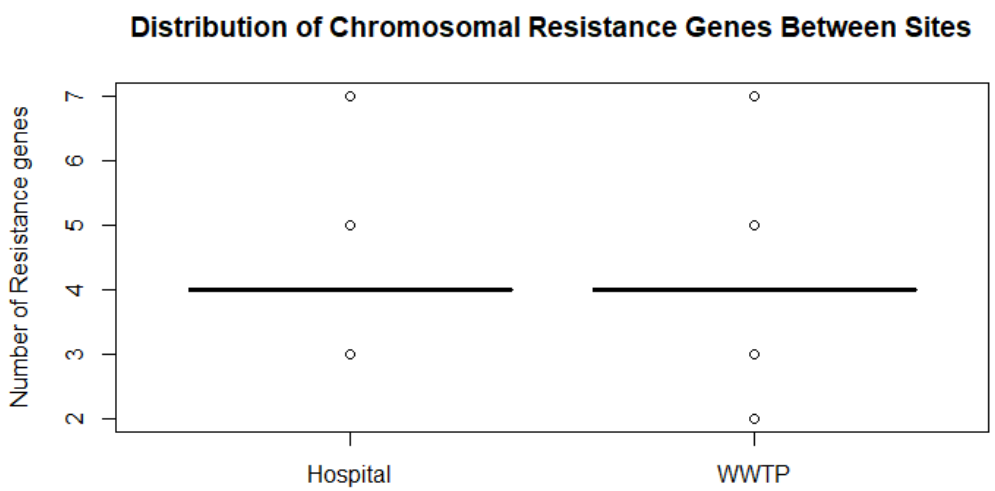
Click here to access/download;Figure;Figure 7.pdf

Figure 7.

A



B



Supplementary Material
Click here to access/download;Supplementary Material;supplementary material.pdf

Supplementary note 1 - **Analysing the data via the microreact project at**:

https://microreact.org/project/Wastewater

Columns B - CX include output from Kleborate v0.4.0-beta, including species, MLST, virulence loci and antimicrobial resistance genes. Column CY lists the number of resistance genes identified in the isolates using Abricate v0.9.8 with the ResFinder database (downloaded 29 April 2020) using the threshold of >80 % nucleotide identity and coverage. The locations of these genes are detailed in columns CZ - FE, using output from MOB-suite and mlplasmids v1.0.0; 'chromosomal', 'ambiguous' (MOB-suite and mlplasmids predictions were discordant) or plasmid (when MOB-suite and mlplasmids both called 'plasmid', the accession number returned by MOB-suite is shown, or unknown in the case of novel plasmids). Columns FF - GE show replicon types identified using Abricate with the PlasmidFinder database (downloaded 13 Jan 2020) with a threshold of >80 % nucleotide identity and coverage. Columns GF - JG are presence ('1')/ absence ('-') of plasmids identified by MOB-suite identified by accession number (as above). Columns JH and JI show the number of virulence factors identified using Abricate with the virulence factors database (downloaded 19 April 2020) with thresholds of both >80 and >40 % nucleotide identity and coverage respectively. Columns JJ - NG show the presence of these virulence factors in the isolates as identified using thresholds of >40 % ('40') or >80 % ('80') nucleotide identity and coverage. Detailed instructions on how to use microreact are available at https://microreact.org/instructions.

Supplementary note 2 - linked resistance genes on multiple plasmids in *Kpne* ST983

In all nine ST983 isolates, the resistance genes *bla*$_{\text{CTX-M-15}}$, *bla*$_{\text{TEM-1B}}$, *aph(3'')-Ib*, *sul2* and *aph(6)-Id* genes are on homologous contigs of 9 - 14 kb, which also harbour genes for IS*1380* family transposase IS*Ecp1* and transposon Tn*3* resolvase. In the longer contigs these are flanked by an IS*110* family transposase IS*5075* and Tn*3* family transposase Tn*2*. The resistance genes *aac(6')-Ib-cr* and *bla*$_{\text{OXA-1}}$ are on another homologous contig with a gene encoding chloramphenicol acetyltransferase in all isolates, and also carries *dfrA14* in one isolate (5Rd). Resistance gene *aac(3)-IIa* is on a homologous contig in eight of the isolates, with an IS*3* family transposase IS*Kpn11* and *tmrB* which encodes a tunicamycin resistance protein. Similarly, *dfrA14* is on a homologous contig in these eight isolates, alongside an IS*6* family transposase IS*6100* and flanked in the longer contigs by IS*6* family transposase IS*26*. When aligned to CP021953 and JX424423, two of the resistance plasmids predicted by MOB-suite to harbour all of these resistance genes in four of these isolates, all of the contigs from all of the ST983 isolates align to the plasmid over an approximately 30 kb region, interspersed with insertion sequences (data not shown). These arrangements likely explain the mobility of these resistance genes. The ST983 isolates in a previous study from South Africa [1] show a similar profile of resistance genes with those in the present study (*aph(6')-Id*, *aph(3'')-Ib*, *bla*$_{\text{TEM-1B}}$, *bla*$_{\text{SHV-38}}$, *bla*$_{\text{SHV-168}}$, *bla*$_{\text{CTX-M-15}}$, *qnrB6*, *oqxA*, *oqxB*, *fosA*, *sul2*, *tet(A)*, *tet(C)*, *dfrA14*, *qnrB1*). Phylogenetic analysis confirmed the close relatedness between the ST983 reported from South Africa and those in the current study (Figure S14).

1. Founou RC, Founou LL, Allam M, Ismail A, Essack SY. Whole Genome Sequencing of Extended Spectrum beta-lactamase (ESBL)-producing Klebsiella pneumoniae Isolated from Hospitalized Patients in KwaZulu-Natal, South Africa. Sci Rep. 2019;9(1):6266. doi: 10.1038/s41598-019-42672-2

CONFIDENTIAL

Supplementary note 3 – Other *Kpne* lineages

Kpne ST35 is a global multidrug-resistant clone [1]. Two isolates of ST35 were recovered from the WWTP (7Sl, 8Sd) and one from the river (River_C). These two isolates have identical plasmid and resistance gene profiles, and differ by 80 core SNPs over the alignment of 5,100,720 nucleotides. Both isolates harbour a single NC_021231_00058-like plasmid, which contains 6 resistance genes (*aac(3')-IIa*, *aac(6')-Ib-cr*, bla_{OXA-1}, $bla_{CTX-M-15}$, *dfrA14* and *qnrB1*; Figure 3), possess the same two replicon types (IncFII_1_pKP91 and IncFIB(K)_Kpn3; Figure S12), the same four chromosomal resistance genes *oqxA*, *oqxB*, *fosA6* and bla_{SHV-33} (Figure 3) and the virulence factor yersiniabactin (type ybt9; ICEKp3; as discussed below). The high identity between these two isolates is striking, given that one isolate was from the WWTP influent, and the other sampled directly from the river approximately 8 km upstream from the WWTP and 3 months earlier. This suggests that this multidrug-resistant clone is stable and relatively abundant in the local aquatic environment. The third ST35 isolate, 8Sd from the WWTP, is slightly divergent on the tree and exhibits a different plasmid and resistance gene profile.

Kpne ST13 is also a globally disseminated clone of clinical importance [2, 3, 4, 5]. This ST was represented by 3 hospital wastewater isolates, each of which contains two plasmids, a CP000823-like plasmid (which does not carry any resistance genes) and a CP021856(tig00000001_pilon)-like plasmid, which harbours *ant(3")-Ia*, *dfrA1* and *sul1*. All three ST13 isolates possess the chromosomal resistance genes *oqxA*, *oqxB*, *fosA6* and $bla_{SHV-101}$. Similarly, ST17 has been detected from global sources and is known to harbour multiple resistance genes [6, 7]. ST17 was also represented by 3 hospital isolates which show more heterogeneous plasmid and resistance profiles; a total of 5 plasmids were detected within these 3 isolates, but 4 of these were only present in only one isolate. Only a single plasmid-borne resistance gene was detected in ST17, *tet(D)*, but 7 chromosomal genes were noted. Finally, ST584 was assigned to three isolates from the WWTP sample. One of these isolates, 5Sf, harbours 5 plasmids including a JX843238(pTOR_02)-like plasmid [8] that carries *catA1* encoding resistance to chloramphenicol (the only unambiguous example of the presence of this gene on a plasmid in our data) and *sul1* genes. ST584 is not recognised as clinically significant, and is most notable for being recovered at high prevalence from wild boar and barbary macaques in Algeria [9].

CONFIDENTIAL

1. Shen Z, Gao Q, Qin J, Liu Y, Li M. Emergence of an NDM-5-Producing Hypervirulent Klebsiella pneumoniae Sequence Type 35 Strain with Chromosomal Integration of an Integrative and Conjugative Element, ICEKp1. Antimicrob Agents Chemother. 2019;64(1). doi: 10.1128/AAC.01675-19

2. Marcade G, Brisse S, Bialek S, Marcon E, Leflon-Guibout V, Passet V, et al. The emergence of multidrug-resistant Klebsiella pneumoniae of international clones ST13, ST16, ST35, ST48 and ST101 in a teaching hospital in the Paris region. Epidemiol Infect. 2013;141(8):1705-12. doi: 10.1017/S0950268812002099

3. Cejas D, Elena A, Guevara Nunez D, Sevillano Platero P, De Paulis A, Magarinos F, et al. Changing epidemiology of KPC-producing Klebsiella pneumoniae in Argentina: Emergence of hypermucoviscous ST25 and high-risk clone ST307. J Glob Antimicrob Resist. 2019;18:238-42. doi: 10.1016/j.jgar.2019.06.005

4. Mairi A, Pantel A, Ousalem F, Sotto A, Touati A, Lavigne JP. OXA-48-producing Enterobacterales in different ecological niches in Algeria: clonal expansion, plasmid characteristics and virulence traits. J Antimicrob Chemother. 2019;74(7):1848-55. doi: 10.1093/jac/dkz146

5. Fursova NK, Astashkin EI, Gabrielyan NI, Novikova TS, Fedyukina GN, Kubanova MK, et al. Emergence of Five Genetic Lines ST395(NDM-1), ST13(OXA-48), ST3346(OXA-48), ST39(CTX-M-14), and Novel ST3551(OXA-48) of Multidrug-Resistant Clinical Klebsiella pneumoniae in Russia. Microb Drug Resist. 2020;26(8):924-33. doi: 10.1089/mdr.2019.0289

6. Ding Y, Wang Y, Hsia Y, Sharland M, Heath PT. Systematic review of carbapenem-resistant Enterobacteriaceae causing neonatal sepsis in China. Ann Clin Microbiol Antimicrob. 2019;18(1):36. doi: 10.1186/s12941-019-0334-9

7. Strydom KA, Chen L, Kock MM, Stoltz AC, Peirano G, Nobrega DB, et al. Klebsiella pneumoniae ST307 with OXA-181: threat of a high-risk clone and promiscuous plasmid in a resource-constrained healthcare setting. J Antimicrob Chemother. 2020;75(4):896-902. doi: 10.1093/jac/dkz550

8. Rahube TO, Viana LS, Koraimann G, Yost CK. Characterization and comparative analysis of antibiotic resistance plasmids isolated from a wastewater treatment plant. Front Microbiol. 2014;5:558. doi: 10.3389/fmicb.2014.00558

9. Bachiri T, Bakour S, Ladjouzi R, Thongpan L, Rolain JM, Touati A. High rates of CTX-M-15-producing Escherichia coli and Klebsiella pneumoniae in wild boars and Barbary macaques in Algeria. J Glob Antimicrob Resist. 2017;8:35-40. doi: 10.1016/j.jgar.2016.10.005

CONFIDENTIAL

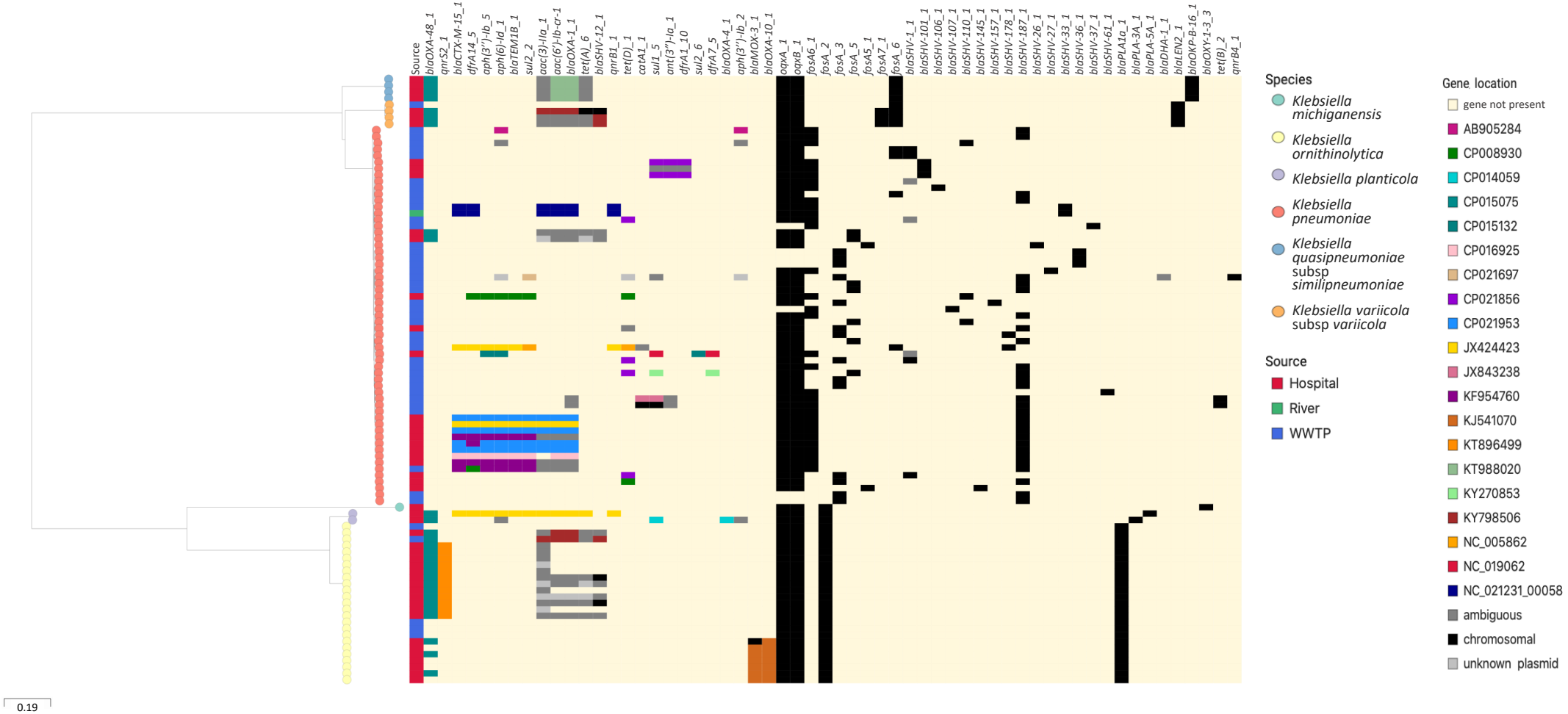


Figure S1. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the 95 isolates analysed in this study constructed using an alignment of 184,671 core SNPs. Species and source of each isolate are indicated, with the location of resistance genes identified using Abricate with the Resfinder database. Contigs harbouring the resistance genes were classified as chromosomal or plasmid using MOB-suite and mlplasmids; contigs classified as plasmid by one method and chromosomal by another are marked as 'ambiguous'. Where both methods assigned a contig to plasmid origin, the accession number of the plasmid as reported by MOB-suite is shown. Contigs assigned as plasmids but without a match to the database are assigned as "unknown".

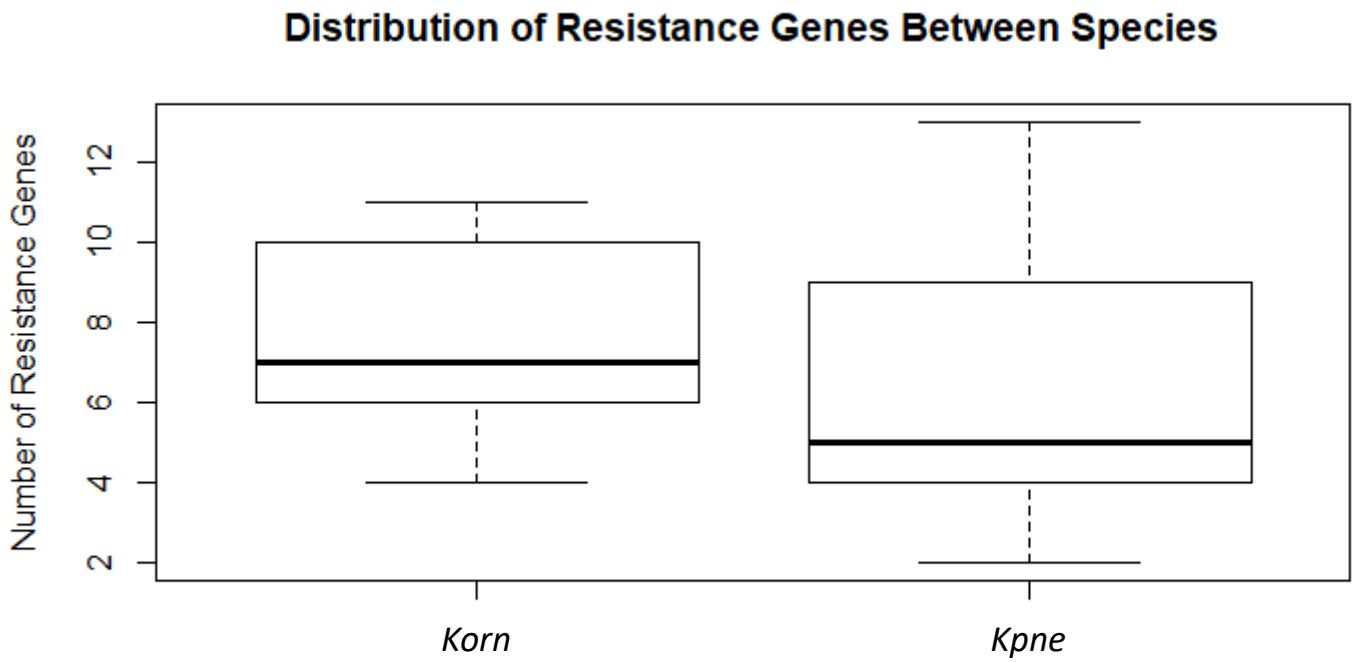
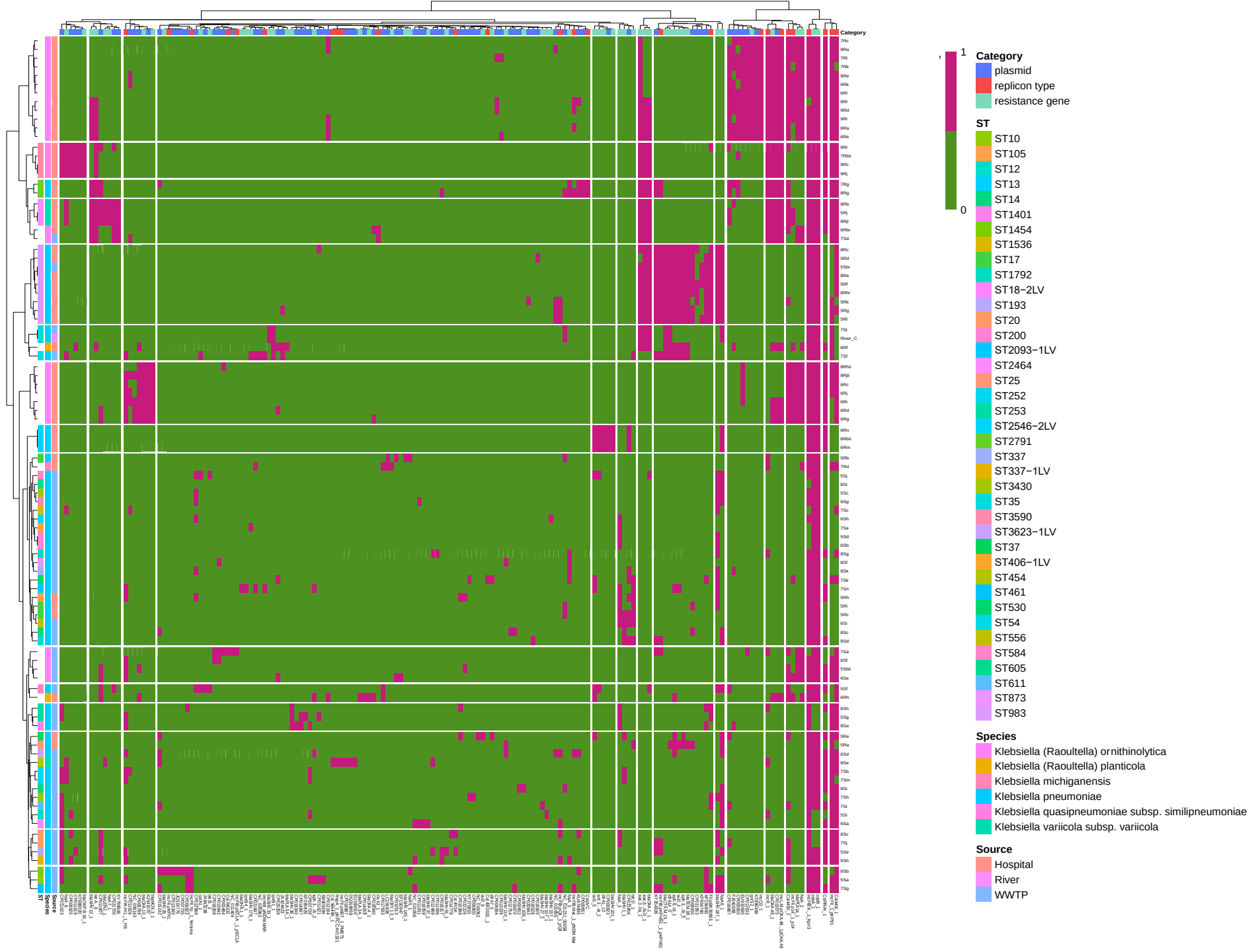
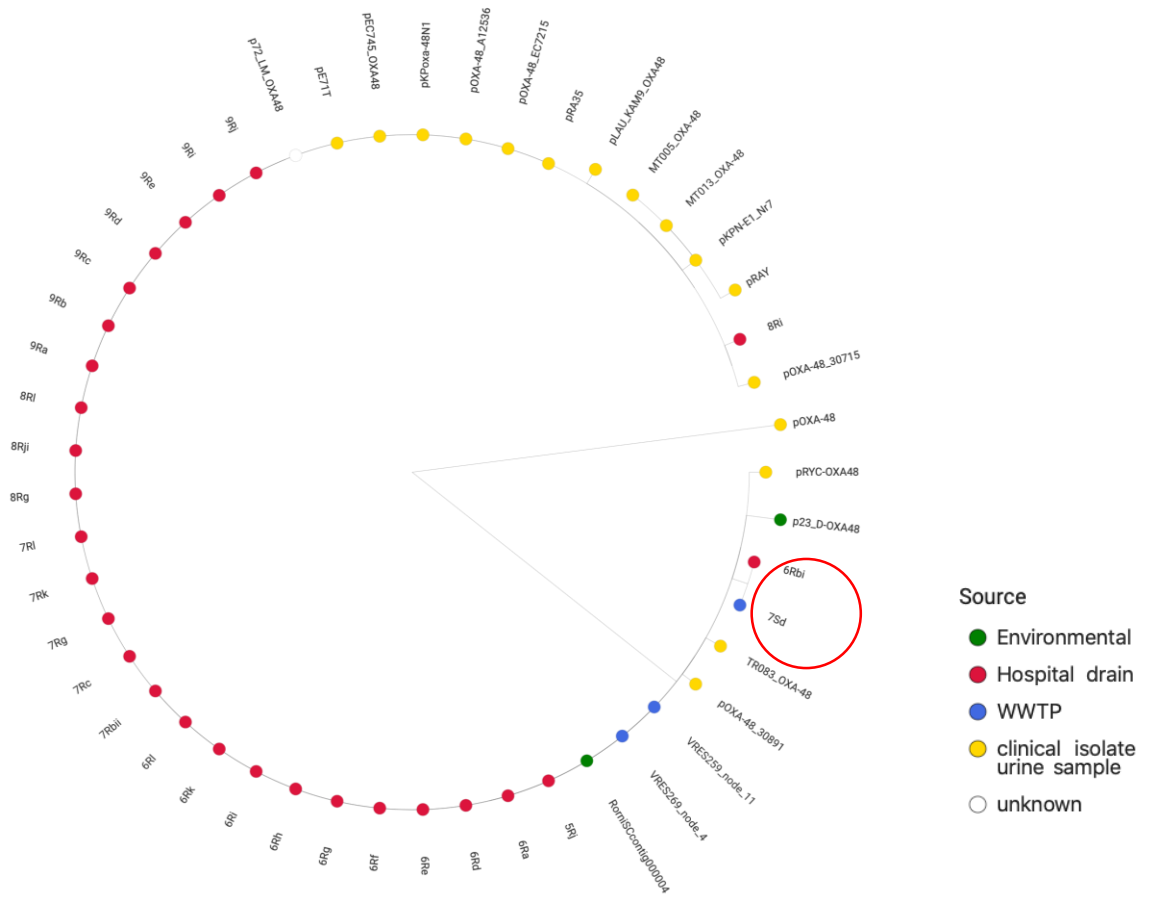


Figure S2. Distribution of resistance genes identified using Abricate with the Resfinder database between *Korn* and *Kpne*; not significantly different by a Wilcoxon test (p = 0.059).

Figure S3. Complete hierarchical clustering analysis (ward.D2) of predicted plasmids, resistance genes and replicon types in 95 isolates. Green refers to absence and pink to presence.





Source
- Environmental
- Hospital drain
- WWTP
- clinical isolate urine sample
- unknown

Figure S4. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the pOXA-48-like plasmids recovered in our study and 21 publicly available sequences (listed in Table S2). Two *Korn* isolates (6Rbi and 7Sd) separated by a single SNP indicated with a red circle.

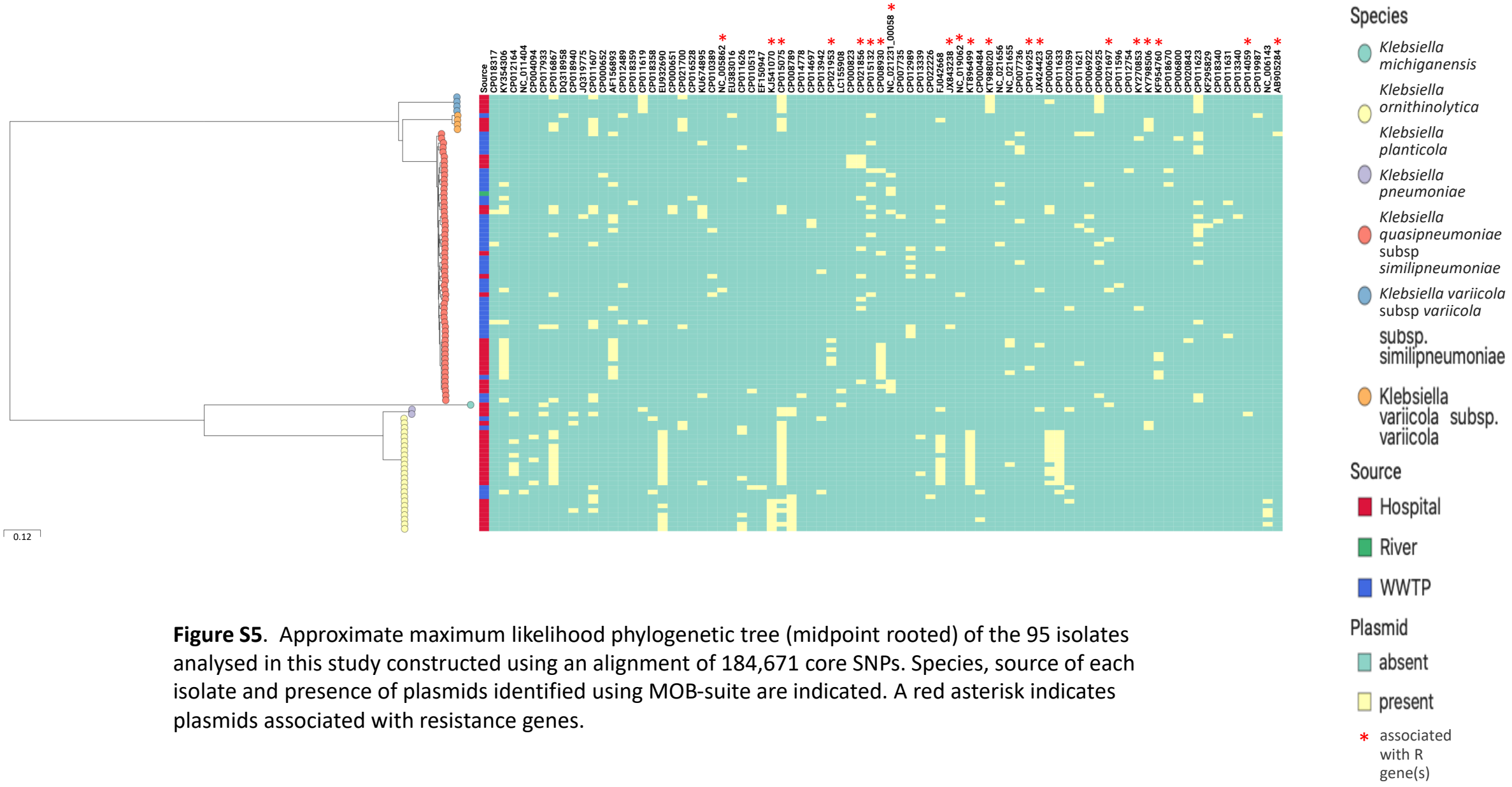


Figure S5. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the 95 isolates analysed in this study constructed using an alignment of 184,671 core SNPs. Species, source of each isolate and presence of plasmids identified using MOB-suite are indicated. A red asterisk indicates plasmids associated with resistance genes.

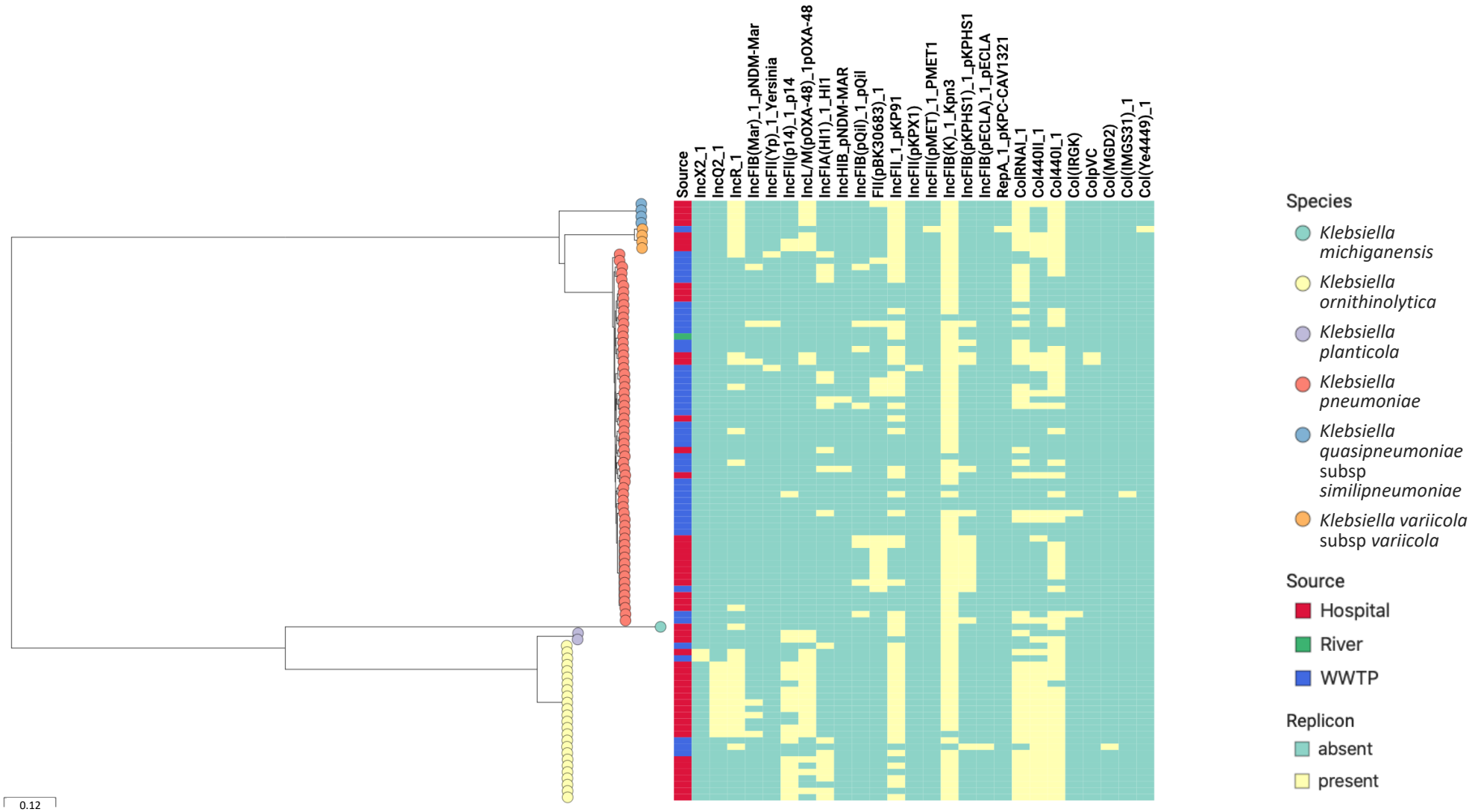
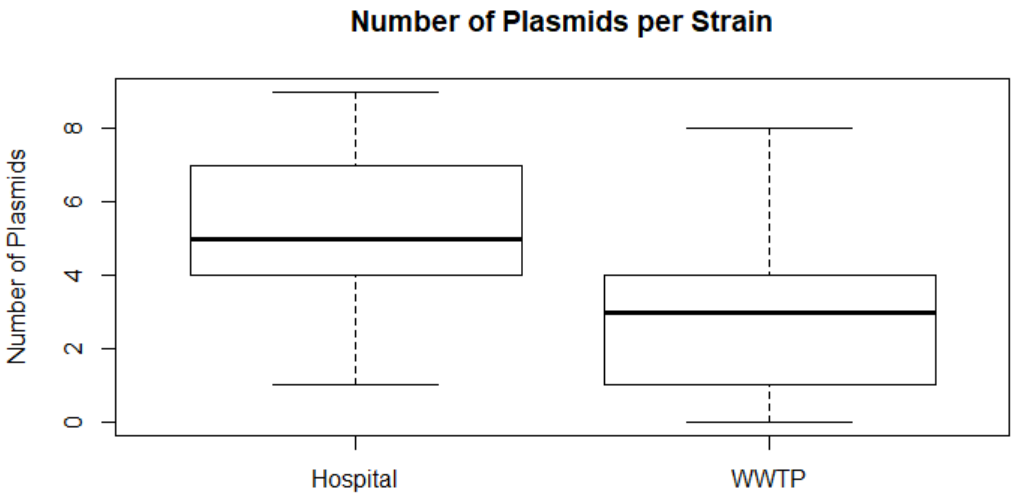


Figure S6. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the 95 isolates analysed in this study constructed using an alignment of 184,671 core SNPs. Species, source of each isolate and presence of replicon types identified using Abricate with the PlasmidFinder database are indicated.

A



B

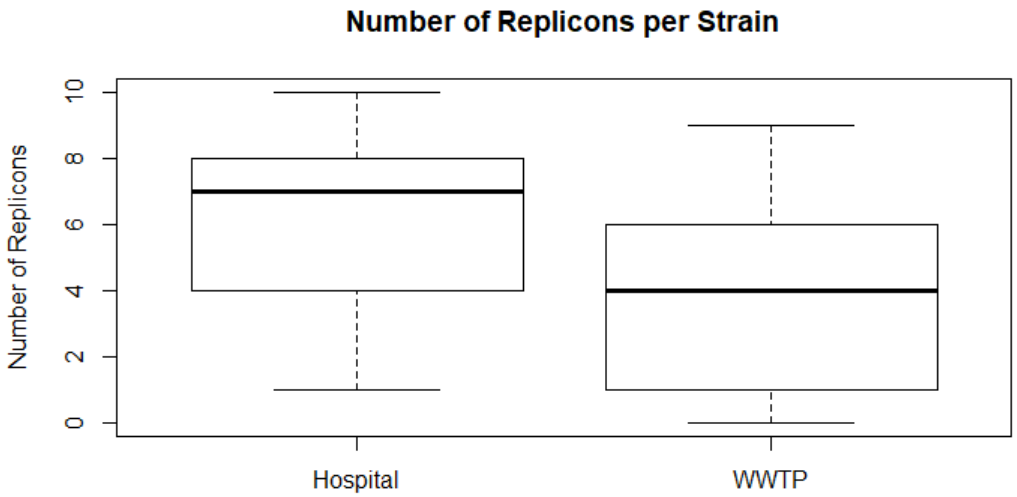


Figure S7. Comparison of isolates from hospital drain and WWTP influent. (A) Number of plasmids as determined using MOB-suite; those that also not confirmed as plasmid by mlplasmids were classed as ambiguous, (B) number of replicon types identified by Abricate using the PlasmidFinder database. Wilcoxon test shows a significant difference in both cases.

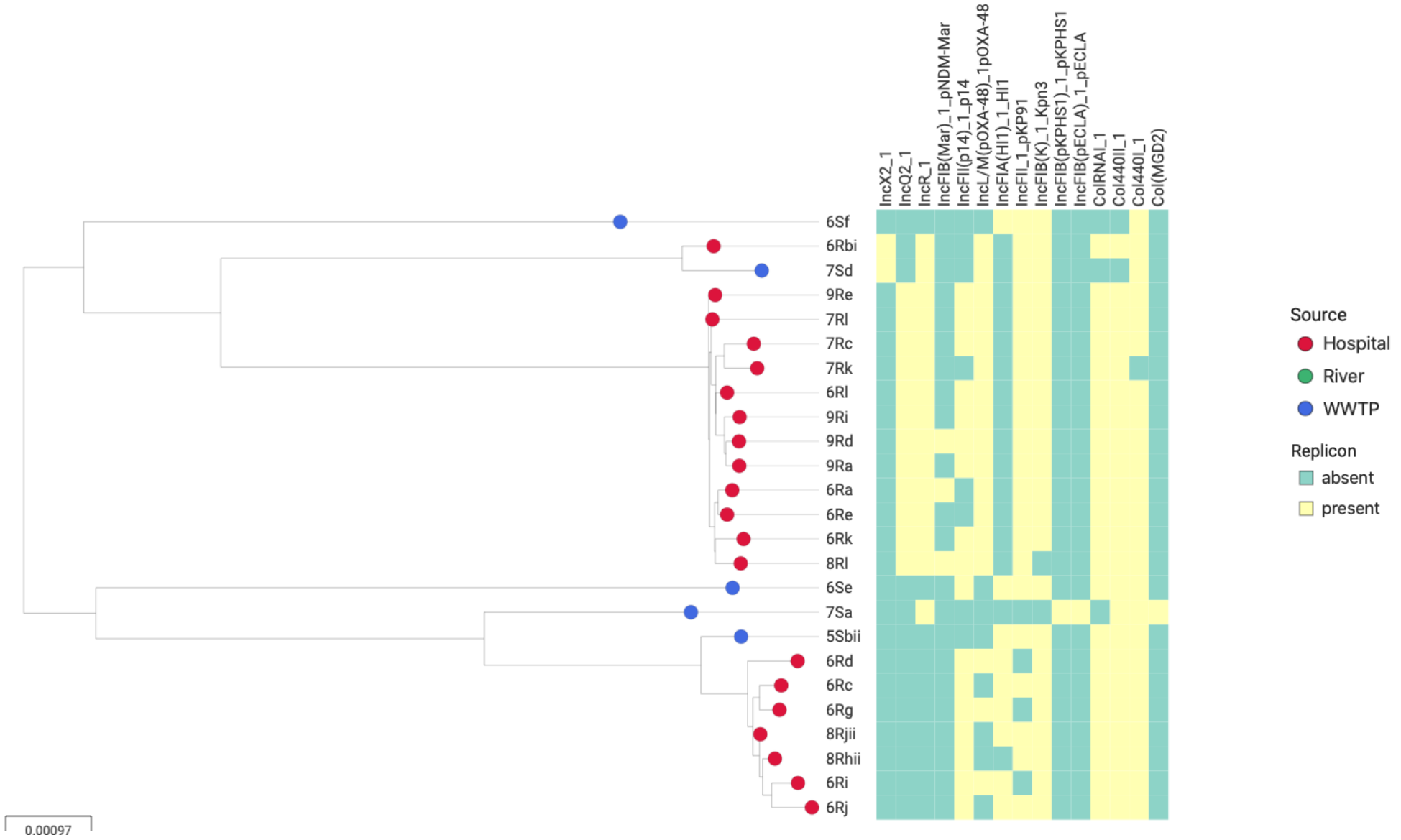


Figure S8. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the 25 *Korn* isolates in this study isolated from the hospital drain and WWTP influent showing the presence of replicon types identified using Abricate with the PlasmidFinder database.

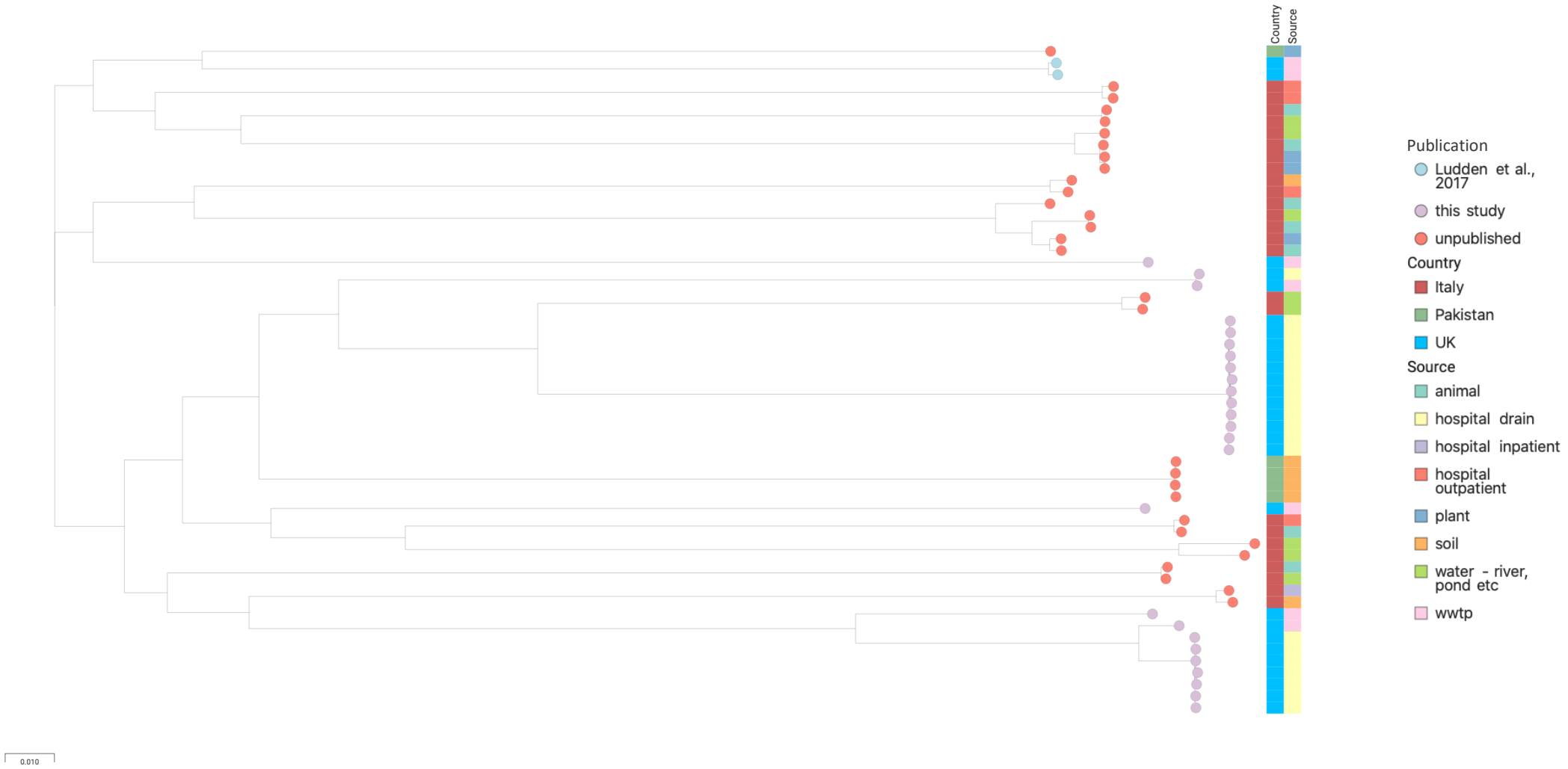


Figure S9. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the 25 *Korn* isolates from this study with isolates from WWTP in the East of England (n=2; Ludden *et al.*, 2017) and unpublished sequences of isolates from Italy (n=25) and Pakistan (n=5). Isolates were aligned to SPARK_1625_C1 from Italy; the tree was constructed using an alignment of 6,004,523 core SNPs. The source of each isolate is indicated.

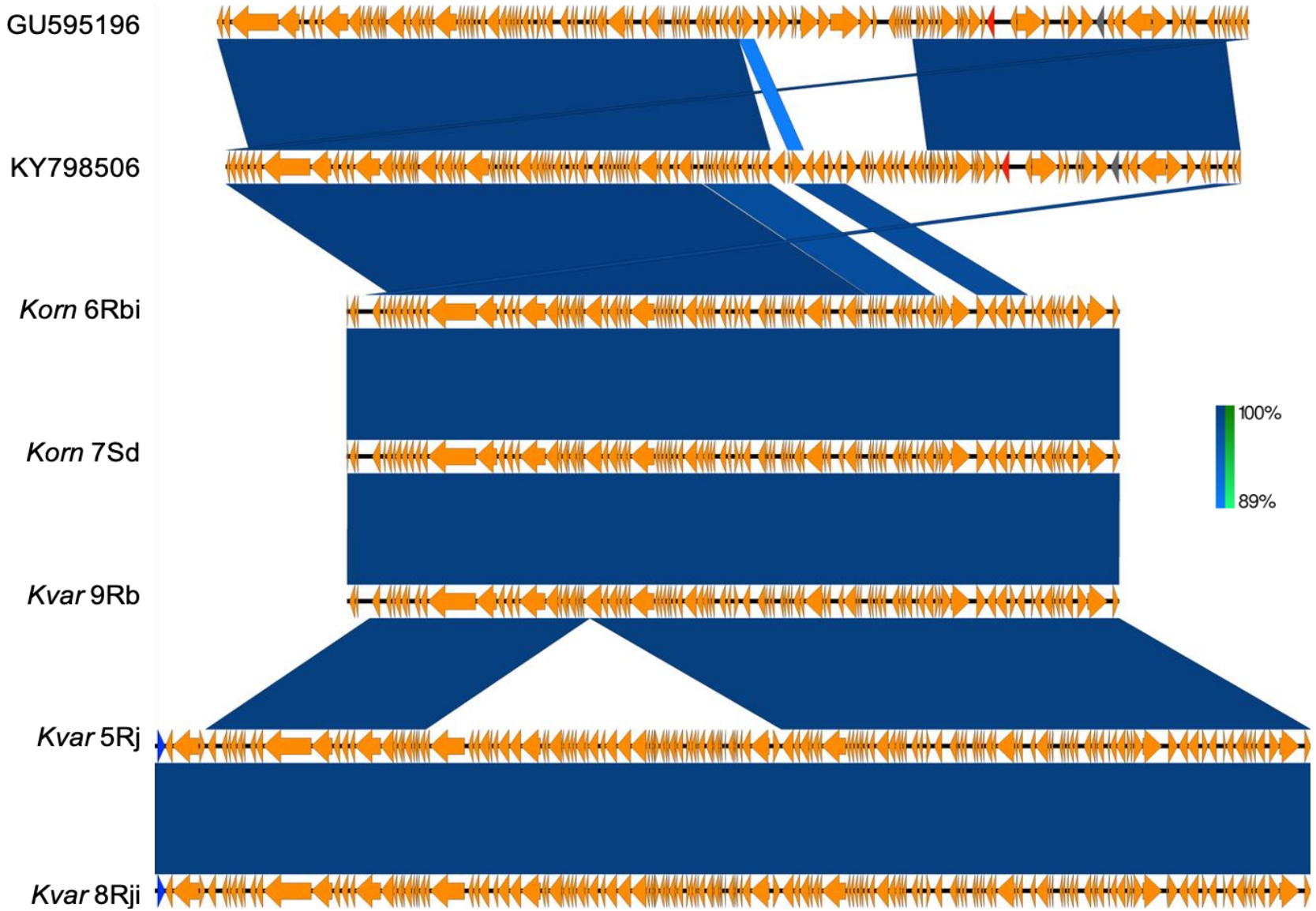


Figure S10. Alignment of the GU595196 and KY798506 reference plasmids with the contigs identified in our data as KY798506-like. In red are the bla_{KPC} genes, in grey the bla_{TEM} and in blue the bla_{SHV}.

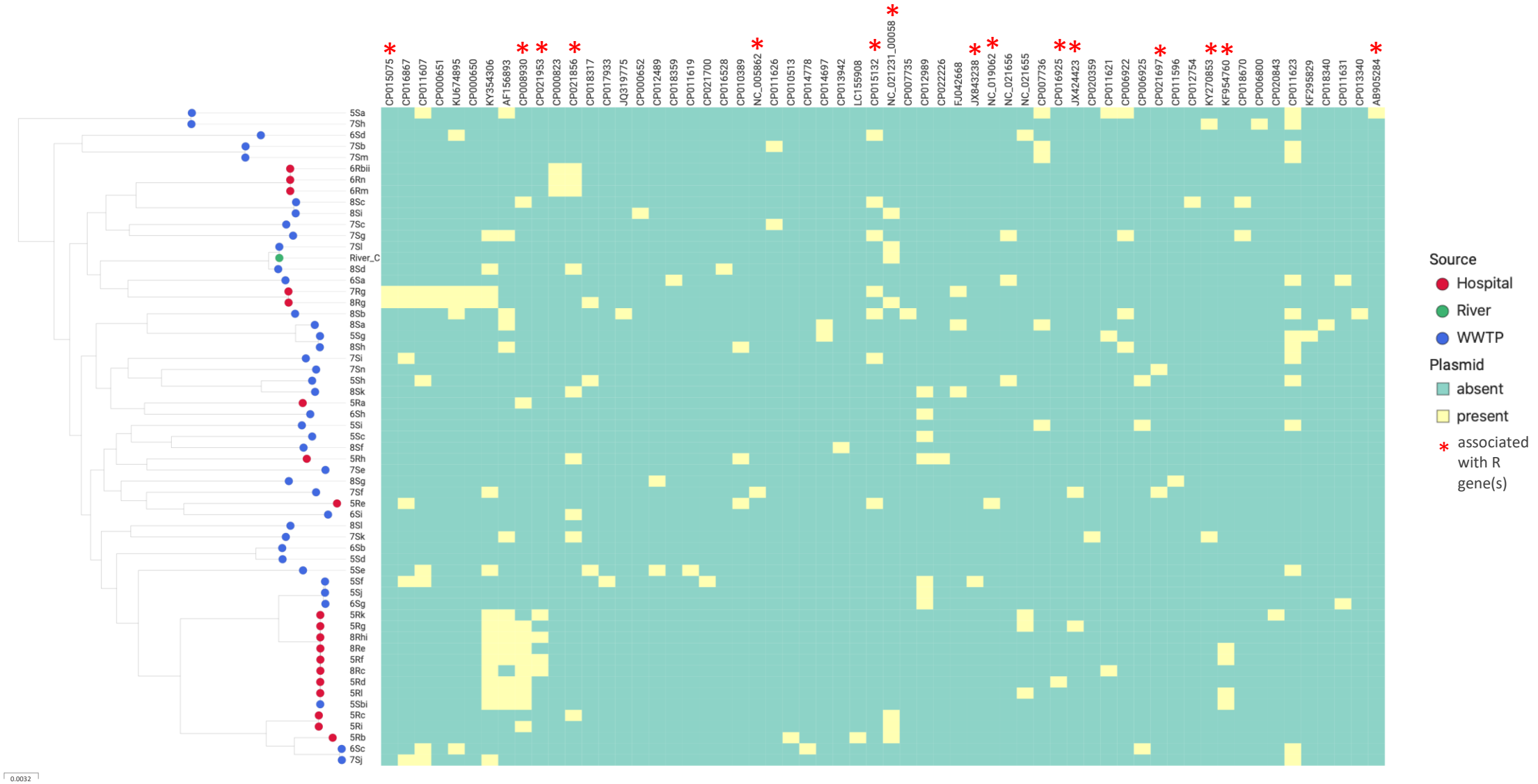


Figure S11. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the 59 *Kpne* isolates in this study isolated from the hospital wastewater, WWTP influent and river showing ST and the presence of plasmids as identified by MOB-suite. Contigs were also analysed using mlplasmids and any that were not identified as plasmid by both methods were omitted from this data. Plasmids associated with resistance genes are marked with a red asterisk.

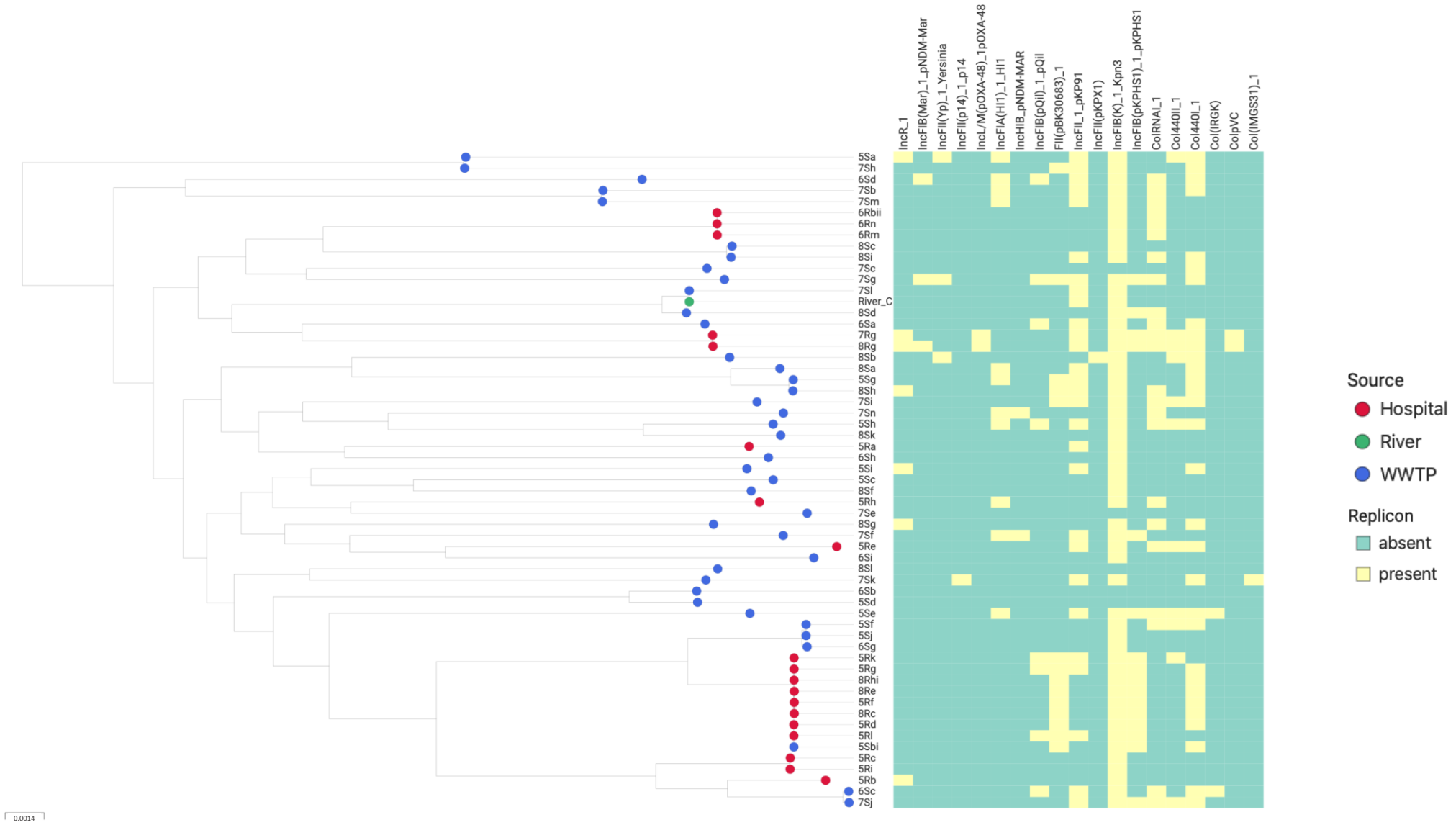


Figure S12. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the 59 *Kpne* isolates in this study isolated from the hospital drain, WWTP influent and river showing ST and the presence of replicon types identified using Abricate with the PlasmidFinder database.

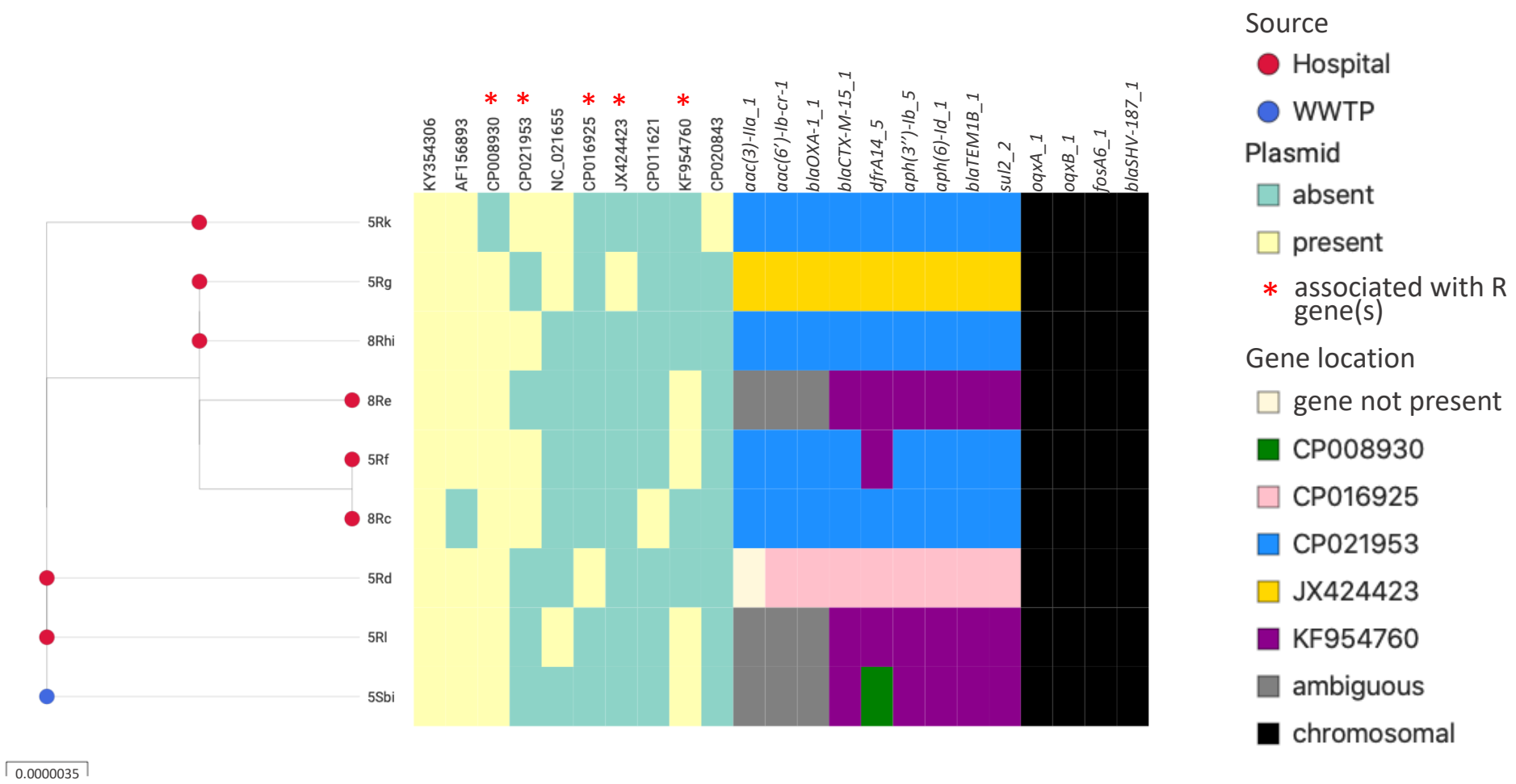


Figure S13. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the 9 *Kpne* ST983 isolates in this study isolated from the hospital drain and WWTP influent. The presence of plasmids identified using MOB-suite and mlplasmids, listed by accession number, is shown. The location of resistance genes (plasmid accession number / chromosomal / ambiguous) is shown; ambiguous was assigned to contigs when results from MOB-suite and mlplasmids did not agree. Plasmids associated with resistance genes are marked with an asterisk.

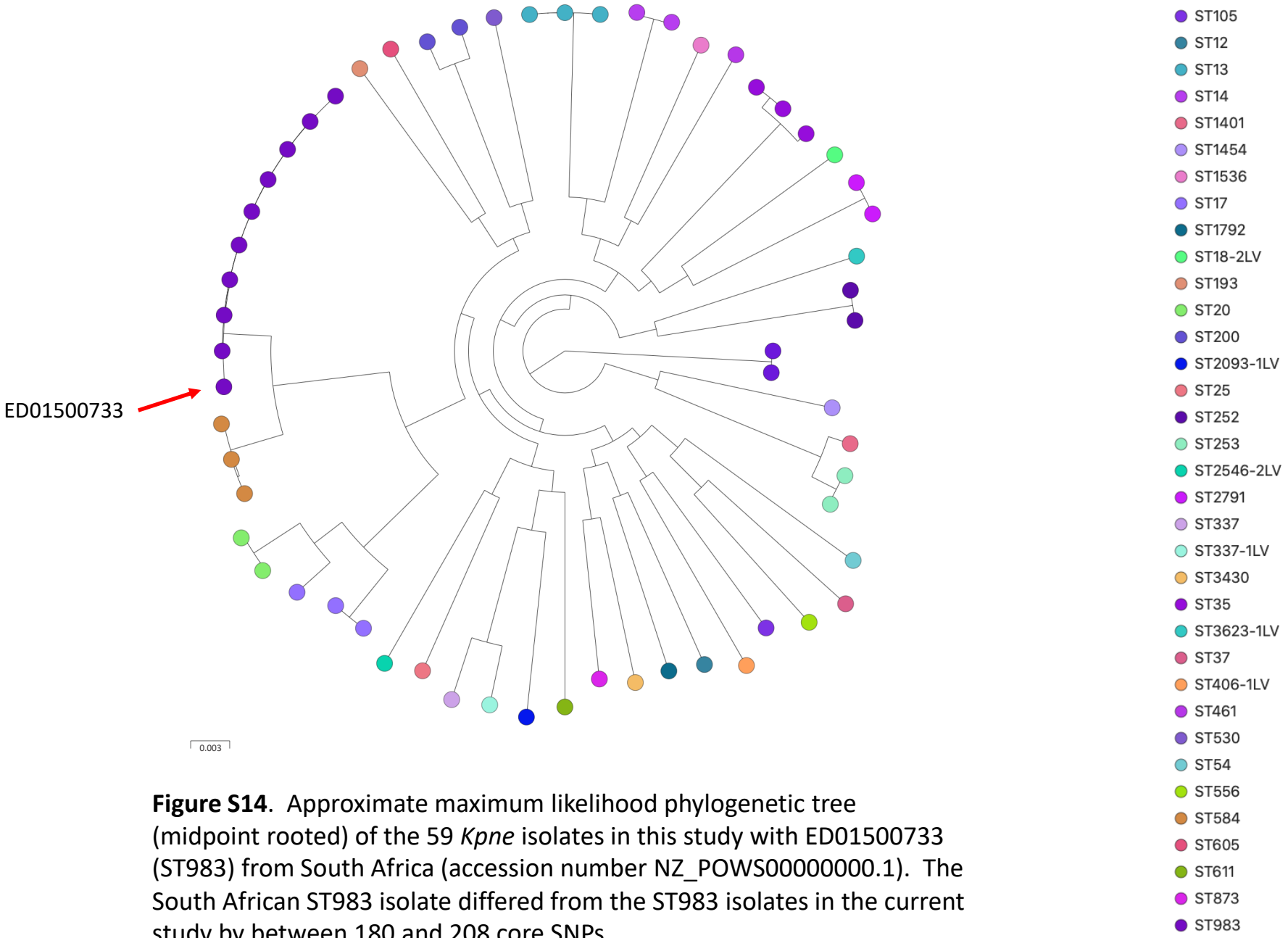


Figure S14. Approximate maximum likelihood phylogenetic tree (midpoint rooted) of the 59 *Kpne* isolates in this study with ED01500733 (ST983) from South Africa (accession number NZ_POWS00000000.1). The South African ST983 isolate differed from the ST983 isolates in the current study by between 180 and 208 core SNPs.

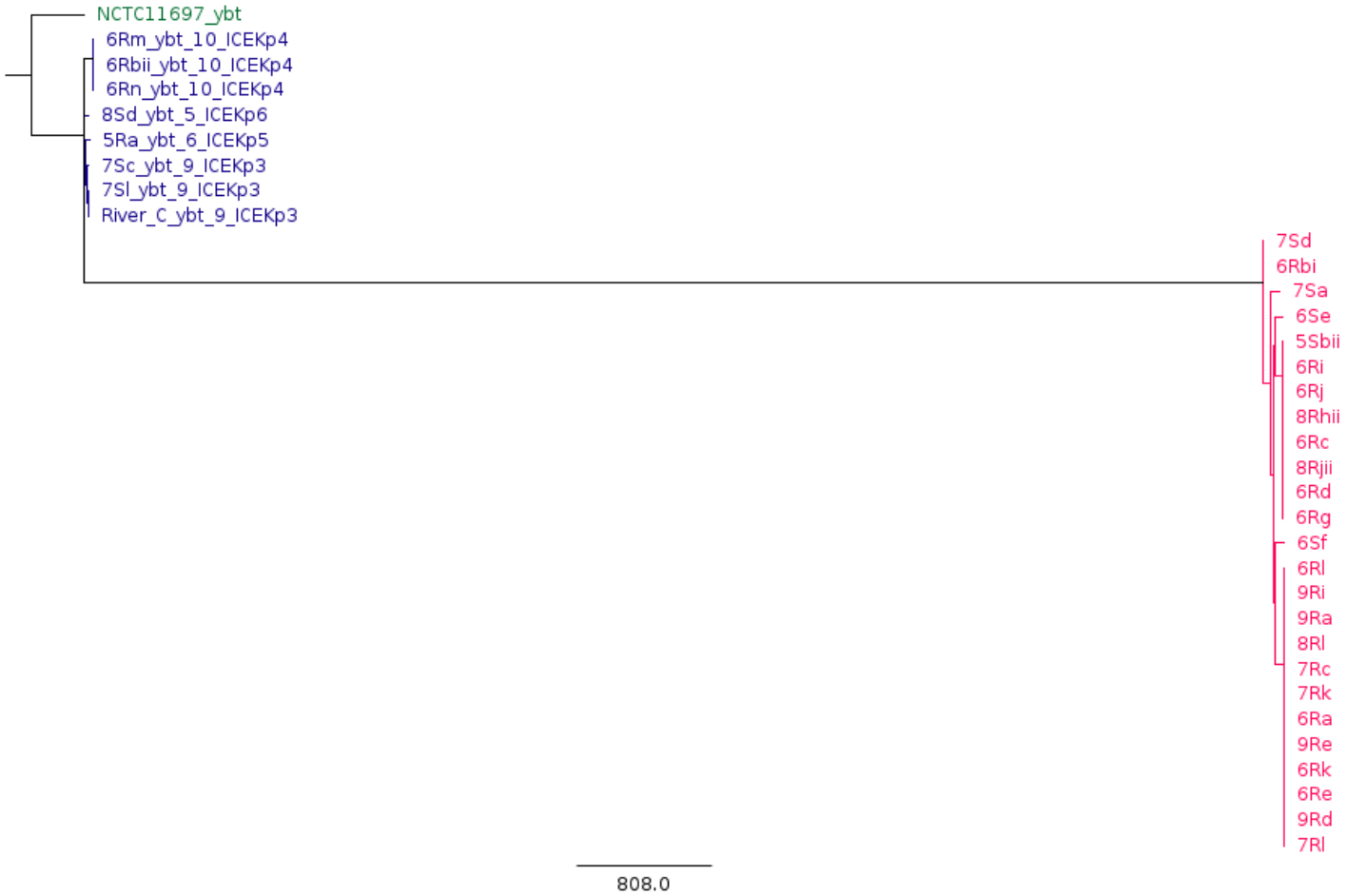


Figure S15. Phylogenetic reconstruction of the *ybt* locus identified in our samples (n=33). Blue leaves correspond to *Kpne* isolates and pink to *Korn*.

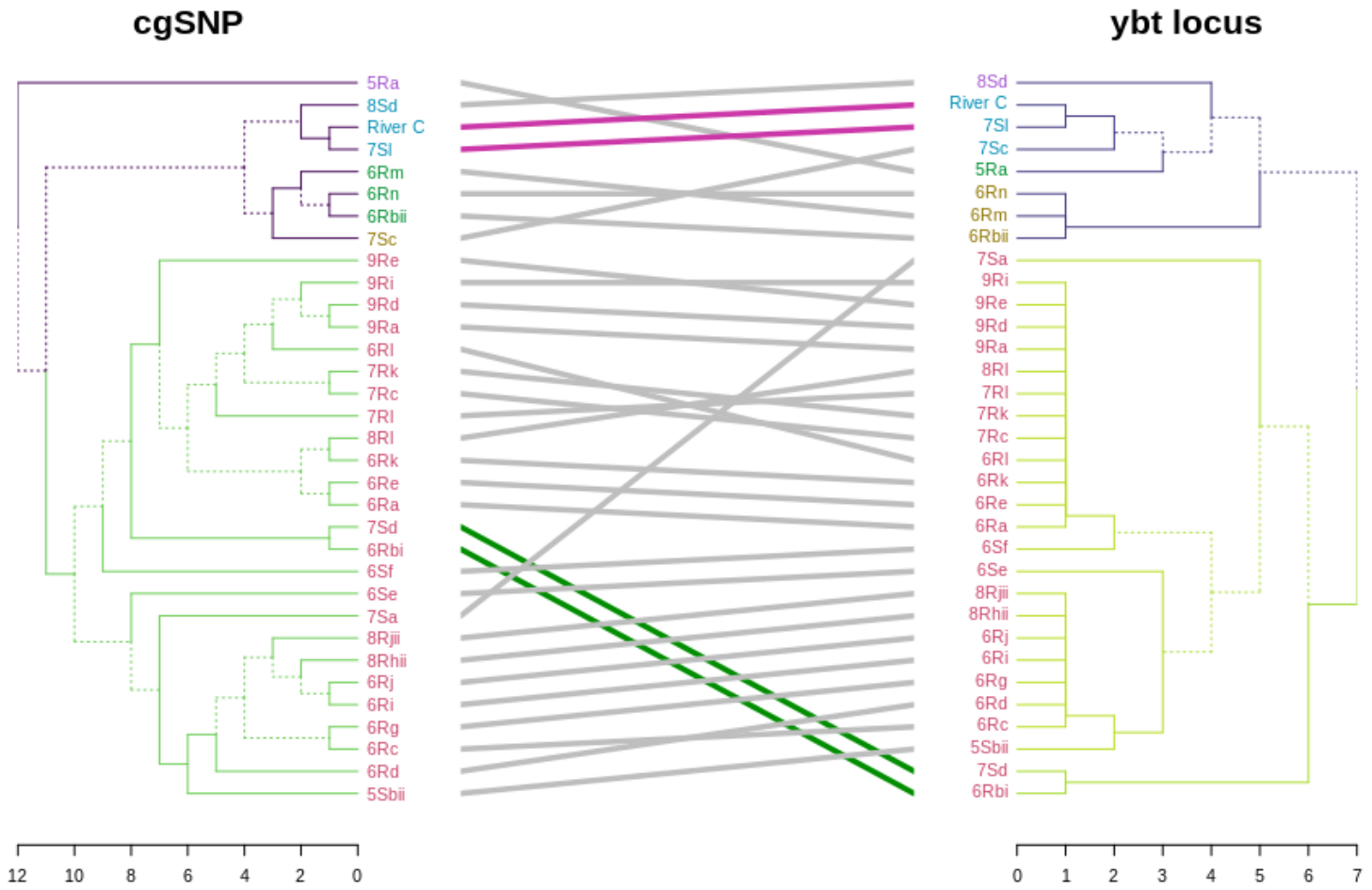


Figure S16. Tanglegram linking the phylogenetic trees constructed using SNPs in the core genome (Left) and the *ybt* locus (Right). Both trees are midpoint rooted and include 33 isolates. The purple branches include *Kpne ybt*-positive isolates, while the green branches include *Korn ybt*-positive isolates. Lines have been drawn between tips in the trees representing the same isolate, while the tree branches were sorted to minimize the number of overlapping lines required. The lines are coloured by the leaves that are common in subtrees of both trees. The lines of the leaves that are not common in subtrees of both trees were left in grey. The leaves are coloured according to the maximum number of matching clusters the trees contain (k=5).

Click here to access/download
Supplementary Material - Excel file
supplementary tables.xlsx

CONFIDENTIAL