



OPEN

## Genomic epidemiological analysis of *Klebsiella pneumoniae* from Portuguese hospitals reveals insights into circulating antimicrobial resistance

Anton Spadar<sup>1</sup>, Jody Phelan<sup>1</sup>, Rita Elias<sup>2</sup>, Ana Modesto<sup>2</sup>, Cátia Caneiras<sup>3</sup>, Cátia Marques<sup>4</sup>, Luís Lito<sup>5</sup>, Margarida Pinto<sup>6</sup>, Patrícia Cavaco-Silva<sup>7,8</sup>, Helena Ferreira<sup>9</sup>, Constança Pomba<sup>10</sup>, Gabriela J. Da Silva<sup>11</sup>, Maria José Saavedra<sup>12</sup>, José Melo-Cristino<sup>5,13</sup>, Aida Duarte<sup>7,14</sup>, Susana Campino<sup>1</sup>, João Perdigão<sup>2</sup> & Taane G. Clark<sup>1,15</sup>✉

*Klebsiella pneumoniae* (Kp) bacteria are an increasing threat to public health and represent one of the most concerning pathogens involved in life-threatening infections and antimicrobial resistance (AMR). To understand the epidemiology of AMR of Kp in Portugal, we analysed whole genome sequencing, susceptibility testing and other meta data on 509 isolates collected nationwide from 16 hospitals and environmental settings between years 1980 and 2019. Predominant sequence types (STs) included ST15 (n = 161, 32%), ST147 (n = 36, 7%), ST14 (n = 26, 5%) or ST13 (n = 26, 5%), while 31% of isolates belonged to STs with fewer than 10 isolates. AMR testing revealed widespread resistance to aminoglycosides, fluoroquinolones, cephalosporins and carbapenems. The most common carbapenemase gene was *bla*<sub>KPC-3</sub>. Whilst the distribution of AMR linked plasmids appears uncorrelated with ST, their frequency has changed over time. Before year 2010, the dominant plasmid group was associated with the extended spectrum beta-lactamase gene *bla*<sub>CTX-M-15r</sub>, but this group appears to have been displaced by another carrying the *bla*<sub>KPC-3</sub> gene. Co-carriage of *bla*<sub>CTX-M</sub> and *bla*<sub>KPC-3</sub> was uncommon. Our results from the largest genomics study of Kp in Portugal highlight the active transmission of strains with AMR genes and provide a baseline set of variants for future resistance monitoring and epidemiological studies.

<sup>1</sup>Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. <sup>2</sup>Research Institute for Medicines (iMed.Ulisboa), Faculdade de Farmácia, Universidade de Lisboa, Lisbon, Portugal. <sup>3</sup>Microbiology Research Laboratory of Environmental Health (EnviHealthMicro Lab), Institute of Environmental Health (ISAMB) and Institute of Preventive Medicine and Public Health (IMP&SP), Faculty of Medicine, Universidade de Lisboa, Lisbon, Portugal. <sup>4</sup>Faculdade de Medicina Veterinária, Universidade Lusófona de Humanidades E Tecnologias, Lisbon, Portugal. <sup>5</sup>Laboratório de Microbiologia, Serviço de Patologia Clínica, Centro Hospitalar Universitário Lisboa Norte, Lisbon, Portugal. <sup>6</sup>Laboratório de Microbiologia, Serviço de Patologia Clínica, Centro Hospitalar Universitário Lisboa Central, Lisbon, Portugal. <sup>7</sup>Centro de Investigação Interdisciplinar Egas Moniz, Instituto Universitário Egas Moniz, Caparica, Portugal. <sup>8</sup>Technophage, Lisboa, Portugal. <sup>9</sup>UCIBIO, Microbiology Service, Biological Sciences Department, Faculty of Pharmacy, University of Porto, Porto, Portugal. <sup>10</sup>Centre of Interdisciplinary Research in Animal Health (CIISA), Faculty of Veterinary Medicine, University of Lisbon, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal. <sup>11</sup>Faculty of Pharmacy and Center for Neurosciences and Cell Biology, University of Coimbra, Coimbra, Portugal. <sup>12</sup>Laboratory Medical Microbiology, Department of Veterinary Sciences, CITAB-Centre for the Research and Technology Agro-Environmental and Biological Sciences, University of Trás-Os-Montes and Alto Douro, Vila Real, Portugal. <sup>13</sup>Instituto de Microbiologia, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal. <sup>14</sup>Faculdade de Farmácia, Universidade de Lisboa, Lisbon, Portugal. <sup>15</sup>Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. ✉email: taane.clark@lshtm.ac.uk

*Klebsiella pneumoniae* (Kp) is a common gram-negative pathogen associated with both hospital and community acquired infections<sup>1</sup>. During the past decades the association of Kp bacteria with antibiotic resistance has been increasingly recognized along with a variety of resistance mechanisms<sup>2</sup>. Production of laterally transferable genes encoding enzymes such as aminoglycoside-modifying enzymes drives high levels resistance to aminoglycoside antibiotics<sup>3</sup>. Resistance associated mutations in quinolone target enzymes (DNA gyrase and topoisomerase IV) are chromosomally encoded by the *gyrA*, *gyrB* and *parC* quinolone resistance-determining region as well as plasmid-mediated quinolone resistance genes<sup>4,5</sup>. Extended spectrum beta-lactamases (ESBLs) render inactive virtually all beta-lactam antibiotics (e.g., penicillins, cephalosporins, but not carbapenems). Nonetheless, of most concern is the increase in prevalence of carbapenemase producing Kp isolates<sup>6</sup> which greatly limits the options for effective therapies and drives hospital outbreaks that promote further spread of antimicrobial resistance (AMR)<sup>7,8</sup>.

In Southern Europe OXA-48-like, VIM and KPC are the dominant carbapenemase families<sup>9,10</sup>. Knowledge of the carbapenemase landscape in Portugal is incomplete. However, the prevalence of carbapenem-resistant invasive Kp isolates in Portugal has increased from 3.4% in 2015 to 10.9% in 2019. During the same period the share of isolates with combined resistance to third generation cephalosporins, fluoroquinolones and aminoglycosides (26.5% in 2019) oscillated without an upward trend<sup>6</sup>. Several recent small epidemiological studies in Portugal have focused on carbapenemase producing Kp isolates<sup>11–16</sup>, and *bla*<sub>KPC-3</sub> was the most frequently identified carbapenem resistance gene. The *bla*<sub>KPC-3</sub> gene is most frequently located within the Tn4401d transposon, and IncN, IncFII, IncFIB and IncFIIA plasmid families are the main traffickers<sup>11,13,15,17</sup>. High flexibility of the Kp accessory genome is an additional concern because it acts as gateway for the introduction of new resistance genes into a broader set of gram-negative pathogens<sup>2</sup>.

Whole genome sequencing (WGS) has revolutionised the study of pathogens, not only through the characterisation of AMR associated mutations and plasmids<sup>18,19</sup>, but also through the determination of phylogenies and transmission events<sup>20,21</sup>. The Kp phylogeny normally forms clades consistent with the commonly used multi-locus sequence typing scheme (MLST), which is based on seven gene loci (*gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB* and *tonB*)<sup>22</sup>. Lipopolysaccharide (O-type) and Capsular Polysaccharide (K-type) serotype profiles can be informative for vaccine development, and > 130 capsular serotypes have been predicted from WGS data with KL2, or O1, O2, O3, and O5 serotypes accounting for most strains<sup>23,24</sup>. Alternatively, clonal groups (CGs) based on 694 core genes are sometimes characterised<sup>25</sup>. Strains from different sequence types (STs), O and K antigen types, and CGs can differ sharply in their virulence and propensity to be antibiotic resistant<sup>25</sup>. For example, ST23, ST26, ST57 and ST163 have been linked to Kp hypervirulence<sup>26</sup>. Unfortunately, the widely used Kp MLST scheme does not allow for a high-resolution phylogeny. The recombination and horizontal gene transfer within Kp also complicates a phylogenetic analysis<sup>27</sup>, even within the same STs<sup>28</sup>. Furthermore, a large part of the core genome exhibits very limited polymorphism, and mobile genetic elements may have relatively greater clinical relevance as exemplified by virulence and AMR factors<sup>27</sup>. More generally, these include integrative conjugative elements (ICE), which are a diverse group of chromosomally integrated, self-transmissible mobile genetic elements that are active in shaping the functions of bacteria and bacterial communities, including in Kp.

The current study aims to improve understanding of the genomic and AMR landscape of Kp in Portugal by analysing the largest WGS dataset to date, which consists of 509 isolates spanning a period between years 1980 and 2019. We examine the prevalence of STs, genes associated with AMR and virulence, and how the AMR profiles relate to plasmid replicon signatures of the isolates. We found that although ST15, ST14 and ST147 predominate, almost one-third of isolates came from STs considered infrequent in Portugal. We establish there are many AMR determinants, which are evolving over time, and our work provides a baseline set of variants for future monitoring and epidemiological studies in Portugal and wider Europe.

## Results

**In silico ST diversity and population structure of *K. pneumoniae* in Portugal.** This study includes a total of 509 Kp isolates. Of these, 459 are clinical isolates from hospitals in the southern (n = 378), central (n = 20) and northern (n = 61) regions of Portugal (Table 1). These were compared to isolates collected from veterinary clinics (n = 41) and environmental wastewater (n = 9) from the southern region, thereby broadening insights into the incidence of Kp in Portugal. The isolates were collected between years 1980 and 2019, but the majority (n = 455, 89%) were collected between 2000 and 2019 (Table 1). Of the O antigen types, O1 was dominant with O1v1 (44%) and O1v2 (20%) followed by O2v2 (12%) and O2v1 (10%) serotypes (Table 1). Seventy-seven different STs were inferred, with the most frequent being ST15 (n = 161), ST147 (n = 36), ST13 (n = 26) and ST14 (n = 26) (Table 1). Globally significant clonal group 258 had little presence, with 4% of isolates belonging to ST11 and none to ST258 or ST512. A high proportion of isolates (31%) were from low frequency STs (each < 2.0%). Out of 2,926 possible unique ST pairs, only 0.7% differed by a single allele, whereas 22.7%, 32.8% and 25.7% differed by 4, 5 and 6 alleles, respectively (Fig. S2). O serotypes were linked strongly with STs (Fig. S3). Only 10 STs had isolates with different O serotypes. ST15 (n = 161) had 149, 9 and 3 isolates with O1v1, O1v2 and O1/O2v1 serotypes, respectively. ST13 (n = 26) had 25 and 1 isolates with O1v2 and O1v1 serotypes, respectively. ST14 (n = 26) had 24 and 2 isolates with O1v1 and O1/O2v1 serotypes, respectively. ST348 (n = 22) had 19 and 3 isolates with O1v1 and O1/O2v1 serotypes, respectively. ST11 (n = 19) had 13, 5 and 1 isolates with O2v2, O3b and O4 serotypes, respectively. There were a further five STs (34 isolates) containing isolates with different O serotypes. The predominant capsular K serotypes were KL112 (14%) and KL24 (19%), linked to the common ST15. When considering only the unique ST and K group combinations (n = 105), the most frequent K serotypes were KL30, KL10, and KL24 observed in 7, 7 and 5 STs, respectively.

**Phylogenetic analysis.** The overall phylogenetic tree (n = 509) is largely congruent with ST type, but the propensity of Kp to recombine, means that the branch support values decline rapidly as one moves from leaves

Characteristic	N	%
<b>Region</b>		
Centre	20	3.9
South	428	84.1
North	61	12.0
<b>Collection dates</b>		
1980–1982	46	9.0
1990–1999	13	2.6
2000–2009	169	33.2
2010–2019	281	55.2
<b>Sequence type (ST)</b>		
ST15	161	31.6
ST147	36	7.1
ST13	26	5.1
ST14	26	5.1
ST348	22	4.3
ST307	20	3.9
ST11	19	3.7
ST231	16	3.1
ST70	16	3.1
ST45	12	2.4
Other**	155	30.5
<b>Inferred O serotypes</b>		
O1	329	64.6
O2	112	22.0
O3b	17	3.3
O4	13	2.6
O5	13	2.6
OL101	12	2.4
Unknown	10	2.0
O3/O3a	3	0.6
<b>Inferred K serotypes</b>		
KL24	96	16.7
KL112	71	14.8
KL64	35	7.1
KL62	32	6.4
KL3	26	5.6
Other	249	49.4
<b>Carbapenemases and carbapenem-hydrolyzing beta-lactamases</b>		
None	395	77.6
KPC-3	101	19.8
OXA-181	6	1.2
GES-5; KPC-3	5	1.0
OXA-48	1	0.2
GES-5	2	0.4
NDM-1	2	0.4
<b>Aminoglycoside resistance genotypes</b>		
AAC(3)-II; APH(3')-I; APH(6)-I;	154	30.3
AAC(6')-I; ANT(3'')-I; APH(3')-I; APH(6)-I;	64	12.6
APH(3')-I; APH(6)-I;	33	6.5
AAC(3)-II; ANT(3'')-I; APH(3')-I; APH(6)-I;	30	5.9
ANT(3'')-I;	26	5.1
AAC(3)-II; AAC(6')-I; ANT(3'')-I; APH(3')-I; APH(6)-I;	24	4.7
AAC(3)-II;	19	3.7
AAC(6')-I; ANT(3'')-I;	18	3.5
AAC(3)-II; AAC(6')-I; ANT(3'')-I;	10	2.0
ANT(3'')-I; APH(3')-I; APH(6)-I;	9	1.8
Continued		

Characteristic	N	%
AAC(3)-II; ANT(3 <sup>II</sup> )-I;	8	1.6
AAC(3)-II; ANT(3 <sup>II</sup> )-I; APH(3 <sup>I</sup> )-I;	8	1.6
AAC(3)-II; APH(3 <sup>I</sup> )-I;	8	1.6

**Table 1.** Baseline characteristics for the 509 *K. pneumoniae* study isolates. \*\*Includes ST1138 with 7 isolates.

to root (Fig. S3). Individual ST trees (Fig. 1A–D) identified several well supported geographic clades, including two previously reported<sup>8,14</sup> potential outbreaks in Lisbon and Vila Real hospitals involving representatives of ST147 and ST15. For the most abundant sequence type ST15 (n = 161), the tree shows multiple clades differentiated by K serotypes (Fig. 1A). Nearly all isolates carried *bla*<sub>CTX-M-15</sub>, except isolates from years 1980 to 1982 and 5 of 14 isolates from years 2007 to 2018 with *bla*<sub>KPC-3</sub>. Due to the low number of isolates spanning years 1990 to 1999, our dataset may not fully reflect the beta-lactamase diversity during that period. It has been suggested that *bla*<sub>TEM-10</sub> was the dominant ESBL gene at that time, and most of our isolates (11/16) carried it. The smallest clade (n = 9) was from north Portugal and was distinguished by KL19 and O1v2 serotypes with five isolates carrying *bla*<sub>KPC-3</sub>. The other *bla*<sub>KPC-3</sub> carrying ST15 isolates (n = 9) were phylogenetically very distant and part of a larger clade (n = 67) with K112 and O1v1 serotypes. Interestingly, these nine isolates carried both *bla*<sub>KPC-3</sub> and *bla*<sub>CTX-M-15</sub>, in contrast to all other relevant isolates, which carry either *bla*<sub>KPC-3</sub> or *bla*<sub>CTX-M-15</sub> (Fig. 2). Four of the nine wastewater samples belonged to ST15, and were dispersed amongst clinical samples in the ST15 phylogenetic tree (Fig. 1A).

Most ST14 isolates (n = 20/26) fell within two dominant clades (Fig. 1B); the first clade (n = 11) was sourced from northern Portugal in year 2018 and contains the K16 locus, and the second (n = 9) was distinguished by the K2 serotype and spanned years 1980 to 2010. Only the second clade had *bla*<sub>KPC-3</sub> carrying isolates (n = 6). Unlike for other STs, the genes used in phylogenetic reconstruction of ST14 were relatively concentrated within a 3.60 Mbp to 3.75 Mbp chromosomal region based on the NC\_016845.1 Kp assembly (Fig. S1).

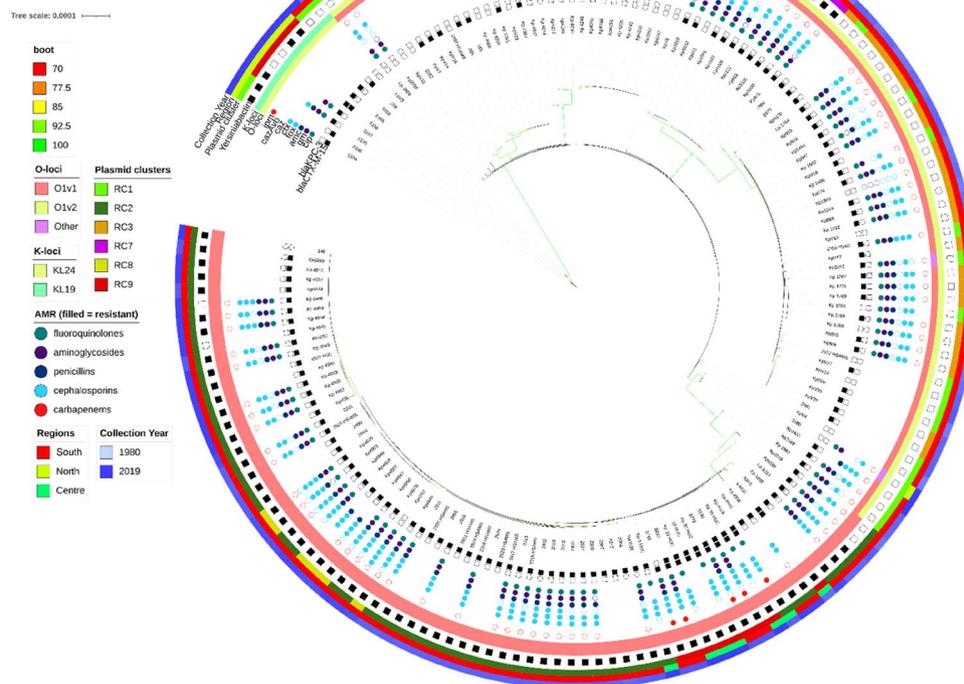
All ST147 (n = 36) isolates had KL64 and O2v1 serotypes and belonged to two main clades. One clade consisted of northern region isolates (n = 15) from year 2018, all of which carried *bla*<sub>KPC-3</sub>, but only one had *bla*<sub>CTX-M-15</sub> (Fig. 1C). All isolates in this clade carried the siderophore yersiniabactin (*ybt 16*; *ICEKp12*) locus related to virulence. A second clade (n = 20) was sourced from the southern region and spanned years 1995 to 2016. Only 11 isolates in this clade carried *bla*<sub>KPC-3</sub> and none had the yersiniabactin locus. Similarly, the ST13 tree had two clades characterised by the KL3 serotype (Fig. 1D). One clade (n = 10) was sourced from multiple hospitals in year 2019, and all isolates carried *bla*<sub>KPC-3</sub> and yersiniabactin loci. Another clade (n = 13) spanned years 1982 to 2013 and came from the southern region. Only two isolates in this clade carried *bla*<sub>CTX-M-15</sub>, and none had other ESBL genes, *bla*<sub>KPC-3</sub>, or the yersiniabactin locus. Unlike the remaining ST13 isolates, these two phylogenetically distant isolates had inferred K57 and K30 serotypes.

Our dataset included four isolates of type ST1138, an ST which was previously described at two hospitals in Portugal<sup>13</sup>. We also had previously undescribed ST6003 (n = 3), which differs from ST1138 by an *rpoB* allele. All seven isolates were collected in 2012 from the same hospital in Lisbon. Remarkably, one of the ST1138 isolates has no AMR genes or mutations apart from *bla*<sub>SHV-36</sub>. This suggests its ancestors were newly introduced into the hospital system. This original strain has split into at least two lineages, both carrying *bla*<sub>KPC-3</sub>, but it is unclear if this acquisition was independent in each lineage. There are at least two inferred lineages because the isolates from overlapping dates have one of two types of *rpoB* alleles (denoted as alleles 1 or 46). We did not identify any post-2012 ST1138 isolates in our data, but it seems to have been present in Portuguese isolates collected in year 2011<sup>13</sup>. Apart from ST6003, our dataset had two further one allele variants of ST1138: ST514 (n = 2) collected from hospital patients in the 1980s and ST323 (n = 1) collected from wastewater in 2018. Both ST514 and ST323 came from the same geographical area as sequence types ST1138 and ST6003.

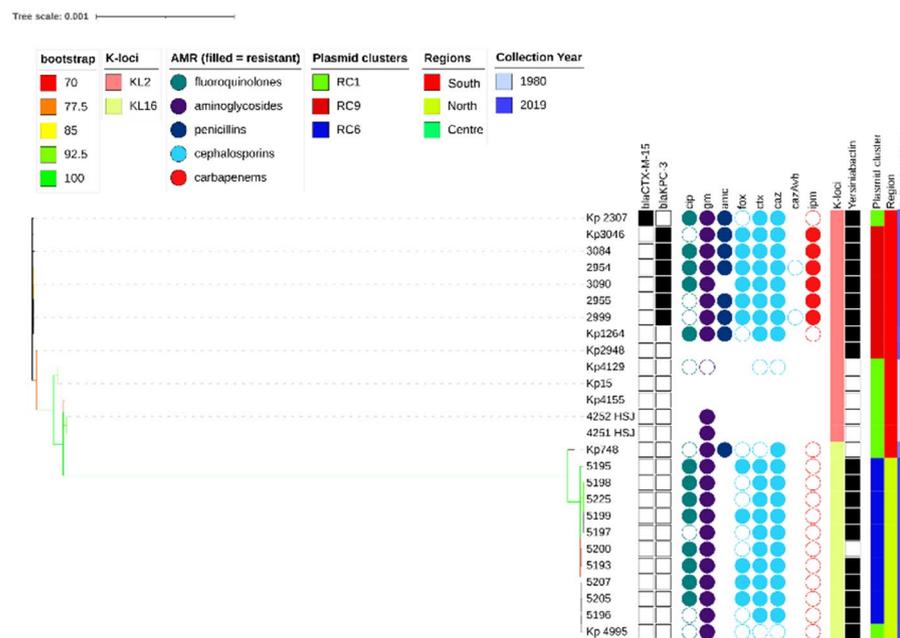
**Plasmid clustering.** Of the 48 replicon families identified by PlasmidFinder software, six occurred in at least 10% of isolates (IncFIB 98%, IncFII 96%, IncFIA 55%, IncR 42%, IncHI1B 11%, IncN 10%). An average of 4 replicons (range: 0–9) were identified per isolate with only three isolates having none. To understand the pattern of plasmid replicons across isolates we applied UMAP dimensional reduction methods<sup>29</sup> to presence-absence matrices of replicon family (n = 48) and unique replicon sequences (n = 240). The results based on families and exact replicon sequences were consistent; so here we present the higher resolution analysis using replicon sequences (Data S1). We observed nine clear and robust clusters (replicon clusters, RCs) (Fig. 3A). RC prevalence varied over time (Fig. 3B) with RC9, the cluster with 94% *bla*<sub>KPC-3</sub> carriage, becoming dominant between years 2010 and 2019. By overlaying RC type onto the overall and ST specific phylogenetic trees, mosaic distributions were revealed, further supporting the movement of plasmids among Kp lineages (Fig. 1A–D).

RC1 contained 57% of all isolates (n = 288), but very few (n = 9) isolates had *bla*<sub>KPC-3</sub>; although 59% carried ESBL encoding genes, which is consistent with early isolates dominating this cluster. The frequency of isolates in this cluster has declined substantially after year 2001. RC1 has a large diversity of plasmid replicons including IncFIB(K) (n = 214), IncFII(K) (n = 174), IncR (n = 108), and Col(pHAD28) (n = 108) with each replicon family consisting of multiple distinct phylogenetic clades making RC1 interpretation complex. In contrast, the RC9 (n = 87) cluster has 25 different STs, including the dominant ST15, ST147 and ST14 types. Almost all Kp in RC9 cluster (94%, 82/87) carry *bla*<sub>KPC-3</sub>. This gene is absent in 5 isolates: ST14 from 2007 and 2010; ST416 from 2011; ST1138 from 2012; and ST359 from 2018. The key signature of RC9 is the presence of variants of FIA(pBK30683)

(A) ST15



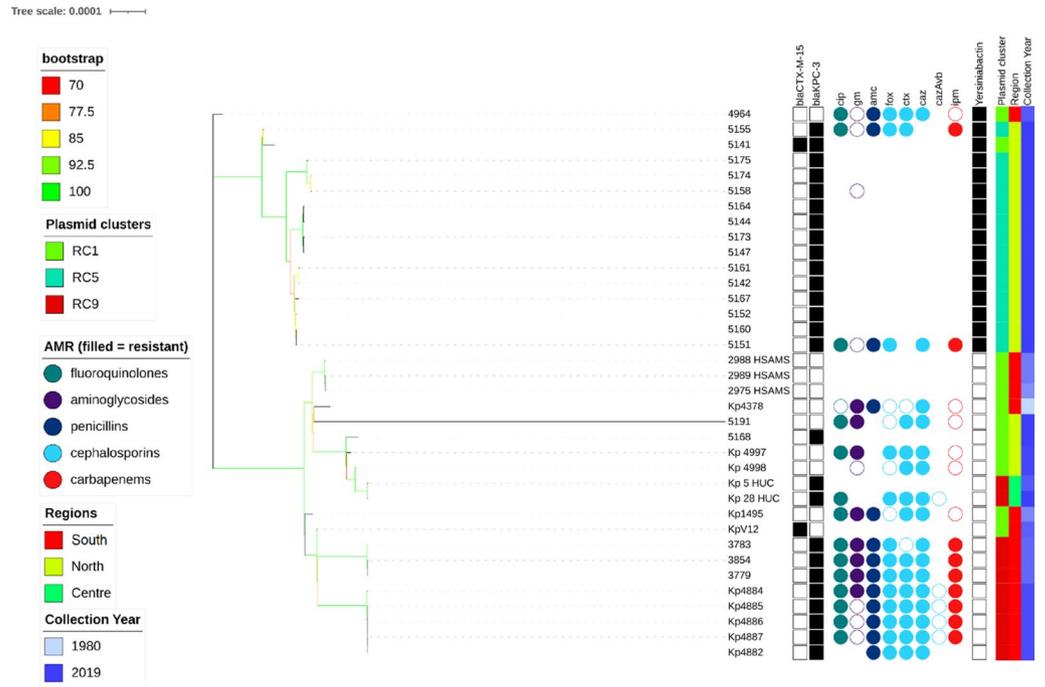
(B) ST14



**Figure 1.** Phylogenetic trees of the most common sequence types (STs), their antimicrobial resistance (AMR) phenotype, and carbapenemase and ESBL genotypic profiles. K-Loci refer to inferred K serotypes. Branch colours represent bootstrap support values.

and FII(pBK30683) replicons (Data S1). For isolates with *bla*<sub>KPC-3</sub> in RC9, more than a third (31/87) were diverse STs collected between years 2010 and 2019, but possessing neither *gyrA* mutations, nor an *aac(6′)-Ib-cr* gene.

(C) ST147



(D) ST13

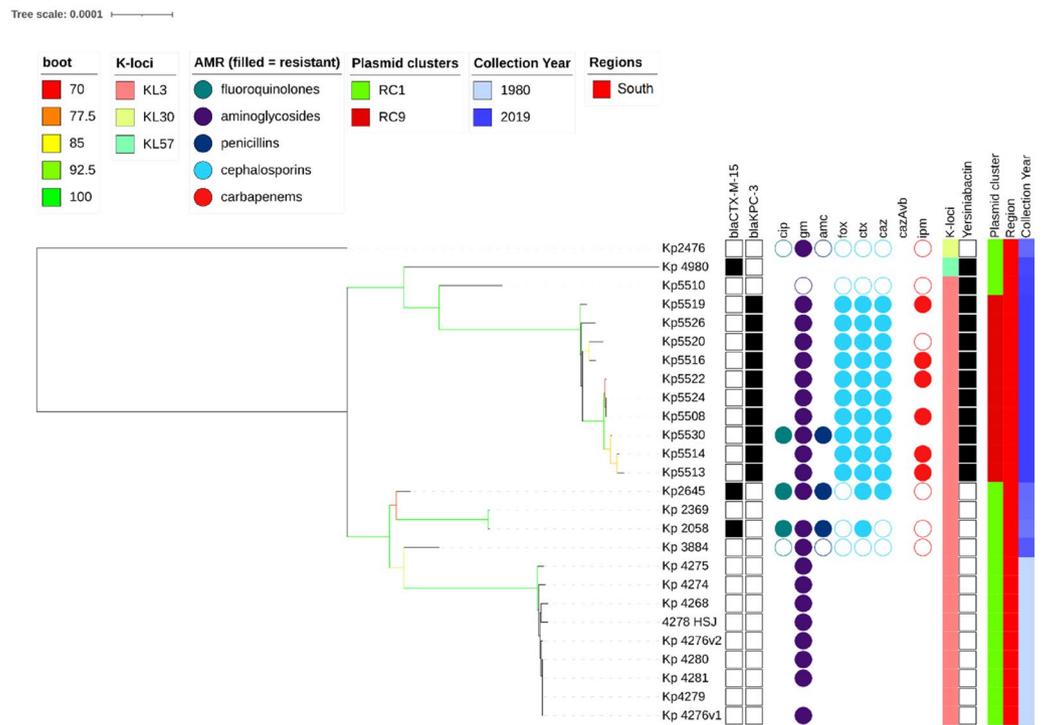
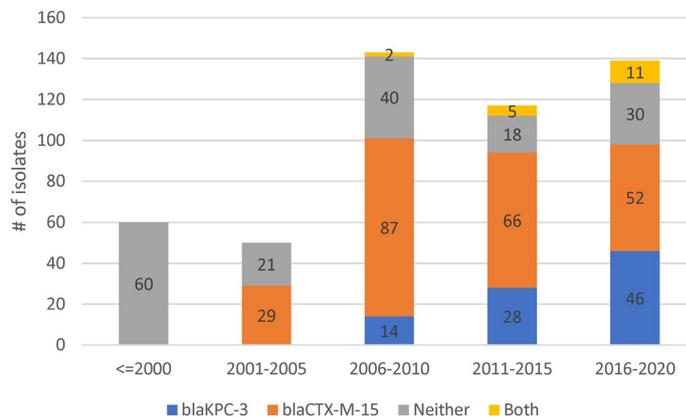


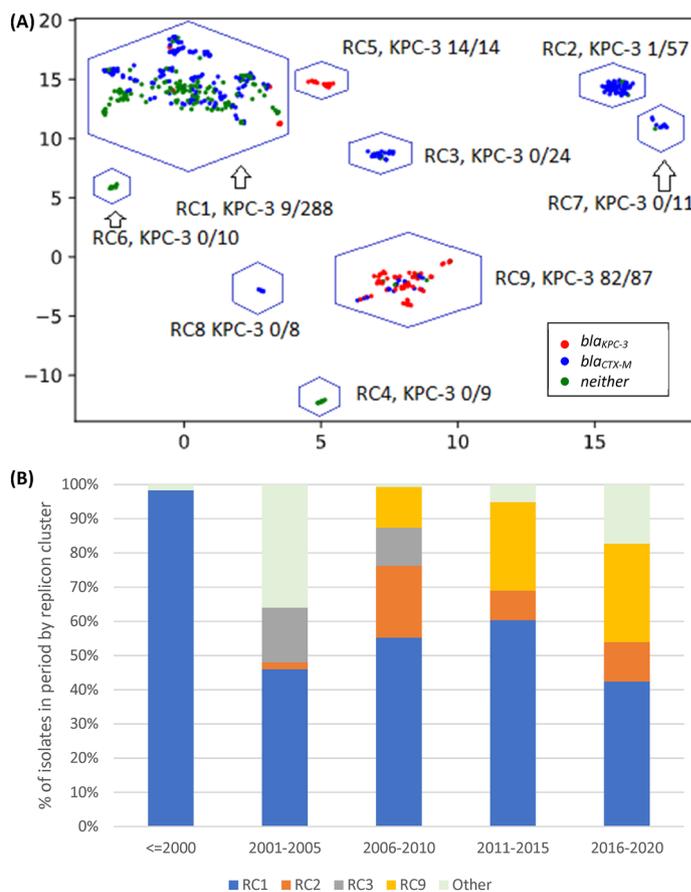
Figure 1. (continued)

These isolates were susceptible to fluoroquinolones with average inhibition zone diameters of 20 mm among 27 tested isolates.

The RC5 cluster (n = 14) consisted of ST147 carrying *bla<sub>KPC-3</sub>* and was congruent with the clade sourced from northern Portugal in year 2018 (n = 15). Of these northern Portuguese isolates, 14 belonged to RC5 and one belonged to RC1 (Fig. 1C). The RC5 cluster is identifiable by two variants of IncN and IncFIB(pKPHS1) replicons (Data S1). IncN is uncommon in our dataset (51/2222 replicons) and its strong link with *bla<sub>KPC-3</sub>* suggests, in line with earlier reports<sup>15,30</sup>, that *bla<sub>KPC-3</sub>* may be mobilized by IncN plasmids. In contrast, other clusters had



**Figure 2.** The most common ESBL (*bla<sub>CTX-M-15</sub>*) and carbapenemase (*bla<sub>KPC-3</sub>*) genes across the 509 Kp isolates by year group.



**Figure 3.** (A) Clustering of isolates by their plasmid replicon (replicon clusters, RC) and antimicrobial resistance (AMR) genotypic profiles, revealing differentiation by carriage of *bla<sub>KPC-3</sub>* and *bla<sub>CTX-M-15</sub>* genes. X and Y axis are dimensions on which full data is projected, they are useless; (B) Abundance of isolates from different plasmid clusters.

homogenous ST types: RC2, RC3, RC7 and RC8 consist of 57, 24, 11, and 8 ST15 isolates; RC6 consists of ten ST14 isolates; and RC4 of nine ST12 isolates. RC3 is interesting because despite having isolates from years 2003 to 2014, only one (1/24) had a *bla<sub>SHV</sub>* type beta-lactamase (*bla<sub>SHV-11</sub>*). These genes are very common (87.4%) and are normally chromosomal, which suggests either the loss of *bla<sub>SHV</sub>* or simultaneous circulation of several strains.

All RC2 isolates carried IncFIB(K) and IncFII(pKP91) replicons, where the associated variants were different to those from the RC1 replicons. The RC2 IncFIB(K) variant was also present in RC8 and RC7 isolates. RC7

isolates also carried a distinct variant of a Col440I replicon, and a variant of IncR is shared with RC1, RC2, RC3 and RC8. RC8 carried a distinct version of IncFII(K) and IncFIB(pQil) replicons. Finally, in addition to shared IncR and ColpVC variants, RC3 also carried IncHI1A and IncHI1B(R27) replicons, which were absent in other clusters (Data S1).

**Antimicrobial resistance genotypes and phenotypes.** In vitro AMR profiles and genotypes were analysed for the 509 Kp isolates, with some gaps depending on the decade of phenotypic assessment (Table S1). Most isolates underwent antibiotic susceptibility testing for aminoglycosides (n = 356, 68.1%), cephalosporins (n = 366, 71.2%), fluoroquinolones (n = 344, 66.9%), carbapenems (n = 296, 57.6%), and penicillins (n = 311, 60.5%) (Table S1). However, 121 isolates had no AMR susceptibility data. Since AMR testing was performed over multiple years, the concentration of active compounds in disks used might vary between isolates; all breakpoints were determined based on EUCAST v11.0 (2021)<sup>31</sup>.

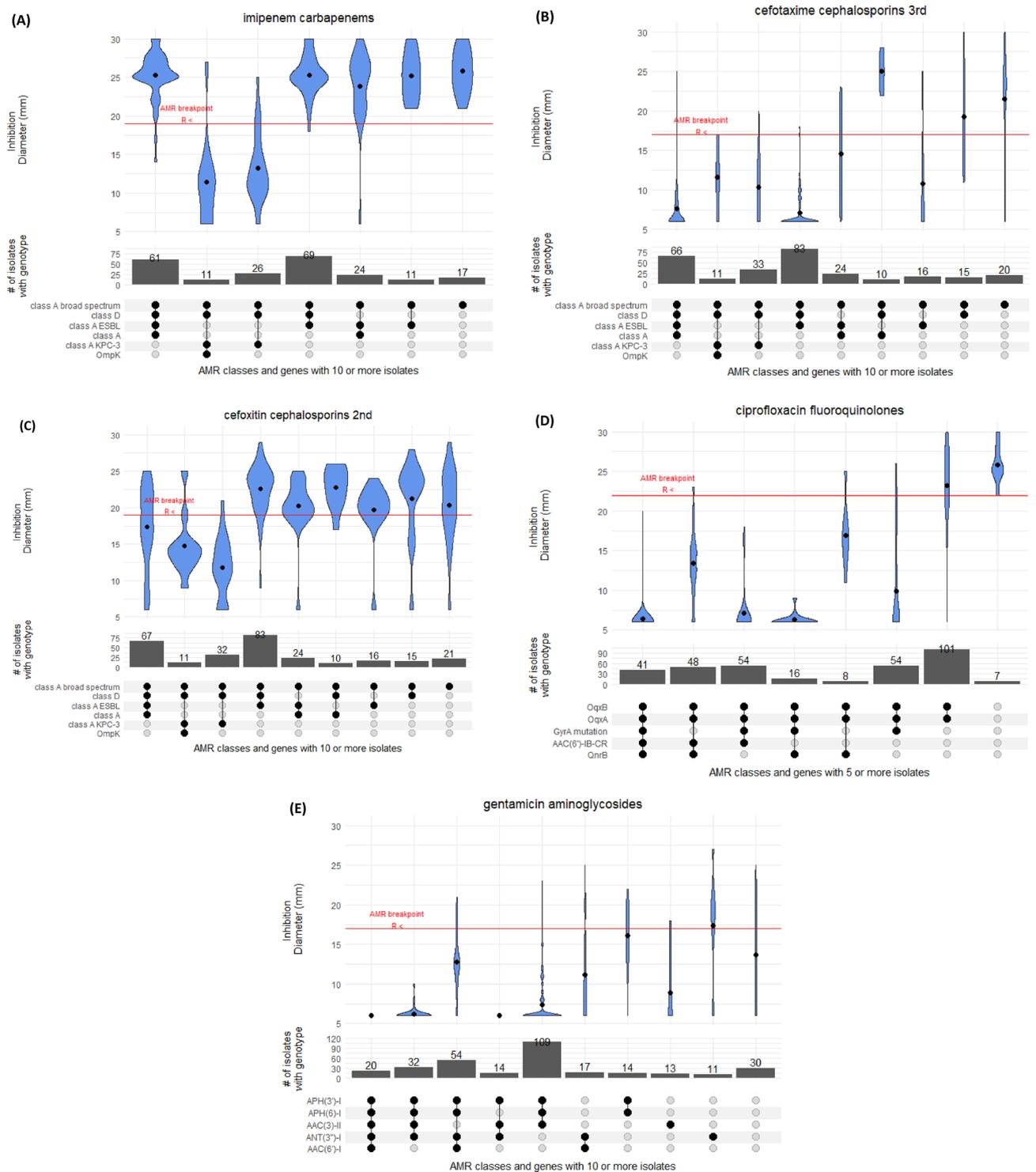
**Beta-lactams.** The majority (75%) of 304 tested isolates showed susceptibility to imipenem; an antibiotic used widely in hospital clinical practice in Portugal (Fig. 4). The resistance driver was likely *bla*<sub>KPC-3</sub> (n = 106), which was carried on a Tn4401d transposon in nearly all cases (n = 101/106). In four isolates [ST15 (Kp5149), ST147 (Kp5147), ST34 (Kp5148) and ST461 (Kp5162)], the *bla*<sub>KPC-3</sub> gene has undergone an identical inversion within the Tn4401d structure. Only one of these four isolates (ST147) was tested for imipenem resistance and was determined to be resistant (inhibition zone diameter of 6 mm). This observation suggests the inversion did not significantly impair the resistance conferred by the presence of *bla*<sub>KPC-3</sub>. Additionally, our dataset contained three isolates without carbapenemase genes but resistant to imipenem. These three isolates had no clear commonality between them, nor clear distinction from susceptible isolates with a similar genotype.

Resistance to cephalosporins was widespread with 68% of 335 tested isolates showing resistance to cefotaxime, a third-generation cephalosporin, driven likely by the presence of class A ESBL *bla*<sub>CTX-M-15</sub> (n = 213, 41%) (Fig. 4). Ceftazidime had an even higher resistance prevalence of 91%. This drug was used widely in the 1990's for *Pseudomonas* outbreaks, but is currently rarely used in monotherapies. The combination of ceftazidime and carbapenemase inhibitor avibactam (cazAvb) was rarely resistant (1/46 tests). The resistant isolate (ST15) carried *bla*<sub>OXA-1</sub>, *bla*<sub>TEM-1</sub>, *bla*<sub>SHV-28</sub> and had truncated porin gene *ompK36*. Of the imipenem resistant isolates, twelve were tested for cazAvb and all were susceptible.

Finally, the second-generation cephalosporin cefoxitin presents an interesting case because its resistance profile was frequently inconsistent with its genomic or genotypic profile (Fig. 4). Of 366 tested isolates, 8 carried AmpC beta-lactamase *bla*<sub>DHA-1</sub>, which is associated with strong inhibition of cefoxitin. Consistent with this, these 8 isolates had a cefoxitin inhibition zone diameter of 6 mm. However, the inhibition zone diameter of the tested *bla*<sub>KPC-3</sub> carrying isolates (n = 74) was on average 13 mm, which is below the 19 mm breakpoint, but well above the total inhibition diameter of 6 mm. In isolates with class A (non-broad-spectrum, broad-spectrum, ESBL) and D beta-lactamases, the susceptibility to cefoxitin varied vastly with isolates containing all four categories having almost uniform distribution of inhibition zone diameters between 25 and 6 mm. We were able to trace some of this inconsistency to 83 isolates with *bla*<sub>SHV-28</sub>. The mean diameter for isolates with class A broad spectrum, class A ESBL and class D genes is 22 mm. If this genotype also included *bla*<sub>SHV-28</sub> (class A), the mean inhibition zone diameter reduced to 16 mm, but this effect is absent for other *bla*<sub>SHV</sub> variants.

**Fluoroquinolones.** Of the 307 isolates tested for ciprofloxacin susceptibility, 288 (94%) were resistant (Fig. S4). Five isolates did not have any common fluoroquinolone resistance determinants and were all susceptible with between 22 and 30 mm inhibition zone diameters. By far the strongest determinant of resistance was the presence of mutations in type II topoisomerase *gyrA*. In silico screening with Kleborate software identified simultaneous mutations GyrA83F and GyrA87A (n = 150), and single mutations GyrA83I (n = 91) and GyrA83Y (n = 15) as the most common ones. Nearly all isolates with *gyrA* mutations had inhibition zone diameters below 10 mm, whereas the resistance breakpoint is 22 mm. In 54 tested isolates that had both the *gyrA* mutation and *qnrB* gene, nearly all had inhibition zone diameters of 6 mm, demonstrating clear compounding of resistance. The *oqxAB* genotype gives a small decrease in susceptibility, but most of such isolates were still susceptible. Finally, *qnrB* and *qnrS* each substantially decreased inhibition zone diameter. In our dataset these genes were present only in combination with *oqxAB*, leading to reduced inhibition zone diameters of ~50% compared to *oqxAB* alone. As we did not have isolates with both *qnr* genes, we could not confirm if their effect is cumulative. Further, 36 and 24 isolates were tested with levofloxacin and norfloxacin, respectively, with 42% of each set showing susceptibility. In contrast to ciprofloxacin, levofloxacin was only moderately affected by *gyrA* mutations with an average inhibition zone diameter reduced to 14 mm (Fig. S4).

We examined the diversity of the *parC* gene which has been associated with fluoroquinolone resistance<sup>5,32</sup>, and found 244 isolates carrying both the ParC80I mutation and *gyrA* mutations. The *parC* alleles were ST specific except for three ST307 outlying isolates that differ by a single SNP from the other seventeen ST307 and fifteen ST15 isolates (Fig. S5) which suggests a lack of selective pressure on the gene. A tree constructed using *gyrA* sequences formed clear clades for the abundant ST14 and ST15 types (Fig. S5). ST14 isolates formed two major clades, one linked to decades 1980's and 2010's (n = 14) and another from the 2000's (n = 15). ST15 formed three clades, one linked to the 1980s (n = 13), and two others post-year-2000 (n = 18, n = 131). There was additional evidence of selective pressure in other STs for which we had fewer isolates (Fig. S5). We could not test the effect of *parC* mutations on resistance because all isolates with mutations in this gene also had mutations in *gyrA*. Finally, 237 isolates (46%) had an *aac(6)-Ib-cr* gene, of which 132 were tested for ciprofloxacin resistance and 128 were resistant. This gene had moderate impact on disk diameter, but this impact is sufficient for isolates to fall below resistance threshold. While the majority of *aac(6)-Ib-cr* carrying isolates also had *gyrA* mutations,



**Figure 4.** Distribution of inhibition zone diameters for different genotypes for (A) imipenem, (B) cefotaxime, (C) ceftioxin, (D) ciprofloxacin, and (E) gentamicin antimicrobials.

we found 48 isolates with an *oqxAB/qnrB/aac(6)-Ib-cr* genotype for which mean inhibition zone diameter was 13 mm compared to 17 mm for 8 isolates with an *oqxAB/qnrB* genotype.

**Tetracycline.** We tested Kp isolates for tetracycline (n = 57) and tigecycline (n = 48) resistance. While breakpoints for tetracycline are not standardised<sup>31</sup>, the results indicate very high resistance to the antibiotic (Fig. S4). The only five isolates susceptible to tetracycline were collected between years 1980 and 1982, which suggests in the remaining 52 isolates that resistance is acquired rather than intrinsic. Isolates with *tetABD* efflux pumps (n = 31/57) had marginally more resistant profiles. In contrast to tetracycline, most of tigecycline tests inhibition

zone diameters were located just below the breakpoint, with only a few isolates with < 10 mm (Fig. S4). Kp does not have an EUCAST zone diameter breakpoint, so we used the *E. coli* 18 mm breakpoint instead. Interestingly, the *tetABD* efflux pumps did not reduce the inhibition zone diameter. Isolates without these efflux pumps were marginally more (mean of 14 mm versus 17 mm), not less, resistant.

**Aminoglycosides.** Among isolates tested with gentamicin (n = 349) and amikacin (n = 29), 89% and 55% were resistant, respectively. In those isolates that had no known gentamicin resistance determinants (n = 28), the average inhibition zone diameter was 15 mm versus an established breakpoint of 18 mm (Fig. 4). Two acetyltransferases were present in our isolates ((AAC(6')-I, n = 283; AAC(3)-II, n = 291), of which AAC(3)-II was a stronger determinant of resistance, with nearly all inhibition zone diameters below 10 mm. AAC(6')-I reduced the inhibition zone diameter from 15 to 12 mm, and the effect was cumulative with AAC(3)-II. Nucleotidyltransferase ANT(3'')-I and phosphatases APH(6')-I and APH(3')-I, even combined, appear to have a marginal effect on resistance. The 16S rRNA methyltransferase gene *armA* was present in two isolates, and the gene *rmtB* was present in two additional isolates.

**Virulence determinants.** Using VFDB, Kleborate and BLAST tools, we examined the dataset for virulence genes<sup>33–35</sup>. The type 1 fimbriae locus, *fimABCDH*, was present in nearly all isolates (501/509, 98%). It was absent in isolates from ST960 (4/4), ST15 (3/161) and ST76 (1/3). Similarly, the type 3 fimbriae locus, *mrkABCDF*, was present in all isolates except for one (ST147). In contrast, iron uptake locus *kfu* was present in 245 (50%) isolates. This locus was perfectly correlated with ST; no ST had simultaneously *kfu* positive and negative isolates. Among the most frequent STs, ST15 (n = 161), ST13 (n = 26) and ST14 (n = 26) isolates had a *kfu* locus, but it was absent in ST147 (n = 36), ST348 (n = 22), and ST307 (n = 20) isolates. There were only a few uncommon virulence genes. For example, we did not find any *rmpA* and *rmpA2* regulators of hypermucoviscosity. Two isolates (Kp4248 and KpV9) had aerobactin siderophores: one with *iucA2* (ST3) and the other with *iuc3* (ST3027). Both isolates had limited known AMR determinants. ST3 had only *bla<sub>SHV-1</sub>* and ST3027 had APH(6')-Ia/d, *tet(A)* and *bla<sub>SHV-33</sub>*. The salmochelin locus *iroBCDEN* was found only in some ST48 isolates (n = 7/9), all of which carried ESBL *bla<sub>CTX-M-15</sub>*. Of these 7 isolates, 6 came from the same Lisbon hospital between years 2005 and 2009. Yersiniabactin was present in 56% of isolates, and was carried most frequently on integrative conjugative elements ICEKp3, ICEKp4 and ICEKp12 at 21%, 15% and 6% of all isolates, respectively. Additional screening against the VFDB virulence database only revealed *astA* and *cseA* genes in one isolate.

## Discussion

In this work we have analysed whole genome sequence data from 509 Kp isolates collected between years 1980 and 2019 from the hospital systems across southern, central, and northern regions of Portugal, as well as from veterinary clinics and a sewage treatment plant in the southern region. Because this is one of the largest in-country collections sequenced, it allowed us to investigate and understand the temporal and spatial genetic diversity of this important pathogen. We observed that 31% of isolates belonged to STs that are considered infrequent in Portugal. If the ST diversity of the dataset was driven by mutations within ST determining genes, we would expect that most samples differ by a single allele, but this was not the case. This observation is further supported by SNP distances between different STs. Instead, the diversity is more likely to be driven by either recombination or coexistence of many strains. The simultaneous presence of so many STs in the country is suggestive of importation and large environmental or human reservoirs of infection. The latter is consistent with high rates of colonisation observed in different countries and settings<sup>36</sup>. The non-human sourced isolates from wastewater and animal settings did not stand out in the analysis, which suggests a flow of Kp between humans and environmental reservoirs. However, the limited number of non-human sourced isolates did not allow us to determine the direction of this flow.

Our WGS sequencing analysis revealed insights into AMR genotyping and phenotyping. During the period with best isolate coverage, years 2000 to 2019, the dominant beta-lactam and carbapenem resistance determinants were *bla<sub>CTX-M-15</sub>* (41%) and *bla<sub>KPC-3</sub>* (21%). In isolates from years 1990 to 1999, a majority (11 of 16) had *bla<sub>TEM-10</sub>*, which is thought to be the dominant ESBL during that period. Our analysis of plasmids reveals a complex and mosaic distribution across isolates suggestive of active selection between them. The increase in prevalence of *bla<sub>KPC-3</sub>* has been accompanied by a decrease in the frequency of older *bla<sub>CTX-M-15</sub>*, which encodes narrower spectrum beta-lactamase. We observed very strong clustering of isolates by their detected plasmid replicons. Plasmid naming nomenclature is based on shared replication mechanisms and incompatibility<sup>37</sup>, so we expected to observe some structure. The replicons themselves revealed a rigid pattern that is clinically relevant due to carriage of *bla<sub>KPC-3</sub>* on two types of plasmids. Isolates with FIA (pBK30683) and FII (pBK30683) replicons were possible sources of *bla<sub>KPC-3</sub>*, while isolates with IncN and IncFIB (pKPHS1) were restricted to ST147 types. However, the isolates with *bla<sub>KPC-3</sub>* had an exact same replicon allele present for at least 10 years and formed a very clear cluster of isolates (denoted as replicon cluster 9).

While our AMR test results may suffer from changes in the amount of active compound in test disk assays across the years, we did find interesting and robust results. As expected, we found that *bla<sub>KPC-3</sub>* was the dominant carbapenemase gene<sup>8,12,13</sup>, but we also observed a reduction in carriage of ESBL *bla<sub>CTX-M</sub>* in those isolates that acquired *bla<sub>KPC-3</sub>*. This displacement of *bla<sub>CTX-M</sub>* indicates active selection, and co-carriage of *bla<sub>CTX-M</sub>* and *bla<sub>KPC-3</sub>* was very rare. While AMR genotype was largely consistent with phenotype, we observed that ceftiofime, a retired second-generation cephalosporin shows moderate activity in isolates lacking class C beta-lactamases. We also found that nearly half of the isolates in replicon cluster 9, which had almost universal carriage of *bla<sub>KPC-3</sub>*, were susceptible to fluoroquinolones as they lacked *gyrA* mutations or other resistance factors. More generally, such insights may offer opportunities for additional treatment of infections with *bla<sub>KPC-3</sub>* carrying Kp.

Overall, our work has provided temporal and spatial insights into Kp STs and AMR related genes and plasmids circulating in Portugal. We found a large diversity of STs, with ST15 and ST147 being the most frequent (< 40%), but almost one-third of isolates had uncommon types (< 2% frequency). Dominant beta-lactamase genes are changing over time due to changes in drug utilisation and plasmid changes. The *bla*<sub>OXA-9</sub> and *bla*<sub>TEM-1</sub> of the 1980s were displaced by *bla*<sub>CTX-M-15</sub> in 2000's which in turn were replaced by *bla*<sub>KPC-3</sub>. These insights reinforce the need for genomic sequencing and tools to assist surveillance and clinical decision making.

## Methods

**Isolate collection, library preparation and sequencing.** The isolates (n = 509) were identified between years 1980 and 2019 from 16 hospitals in Lisbon and its metropolitan area (Southern region), Coimbra (Central region), and Porto and Vila Real (Northern Portugal), except for 9 isolates from Beirolas wastewater (Lisbon) and 41 samples from veterinary clinics (Lisbon) (Table 1). The isolates were cultured as described previously<sup>8</sup>. The set of isolates represent a convenience sample accumulated over 40 years, with the collection site known for the majority (74%). Isolates with known collection site were sourced from blood (32%), urine (31%), rectal screening swabs (10%), pus (7%) and wastewater (7%). Clinical isolates obtained from hospitals were identified at local clinical microbiology laboratories and sent to the Faculty of Pharmacy (University of Lisbon; FFUL) for further phenotypic and genotypic analysis. Given the wide temporal span of the isolates, the initial identification methods that were employed at local laboratories vary across isolates, but were based on biochemical identification methods (e.g., API, Vitek). DNA was extracted from strain cultures grown overnight at 37°C on Mueller–Hinton Agar. DNA extraction was carried out using the Cetyl trimethylammonium bromide method<sup>38</sup>. Library preparation of the DNA samples was performed using a QIAseq FX DNA library kit, following the manufacturer's protocol. WGS was performed on Illumina HiSeq (paired end 150 bp) through The Applied Genomics Centre (London School of Hygiene and Tropical Medicine)<sup>39</sup>. Only those isolates which Kleborate software (v 2.1.0)<sup>40</sup> identified as Kp were used for further analysis.

**Antimicrobial susceptibility testing.** Antimicrobial susceptibility testing at FFUL was carried out using the Kirby-Bauer disk diffusion method as per the European Committee on Antimicrobial Susceptibility Testing (EUCAST) guidelines for performance and interpretation of antimicrobial susceptibility testing (v11.0, 2021)<sup>31</sup>. AMR testing was performed over multiple years, and therefore the concentration of active compounds in disks used might vary between isolates. The AMR testing was not performed for 121 isolates, and not all isolates were tested for the same antimicrobials. The Pearson correlation coefficient was used to assess correlations between disk inhibition zone diameters for different antimicrobials.

**Genome assembly, annotation, and genotyping.** Raw Illumina reads were assembled using Uni-cycler software (v0.4.8)<sup>41</sup>, with assembly fragmentation and completeness assessed against 440 core genes of enterobacterales (enterobacterales\_odb9) using Busco software (v4)<sup>42</sup>. The assessed quality of assemblies was high (median N50: 284Kbp) with all but one having complete single copies of > 97% genes in the Busco reference set. Assemblies were annotated using Prokka software (v 1.14.6)<sup>43</sup>, combined with the *Klebsiella* specific reference genes set<sup>44</sup>. O and K antigen serotypes, genomic AMR, virulence (e.g., ICEKps), and sequence type (ST) profiles were analysed and inferred in silico using Kleborate<sup>40</sup> and AMRfinder (v3.8.4)<sup>45</sup> software with associated databases (accessed October 2020). We have also used Abricate software (v1.0.1)<sup>46</sup> with the virulence factor database VFDB (accessed March 2021) to find additional virulence genes. Plasmid detection and classification was performed using Plasmidfinder software (v2.1.1)<sup>43</sup>.

**Phylogenetic analysis.** The recombination and horizontal gene transfer within Kp can complicate phylogenetic analysis, and a widely used Kp MLST scheme<sup>22</sup> does not allow for a high-resolution phylogeny. We have observed that regions linked to transposons and other genes relating to mobile genetic elements tended to produce high number of SNPs, which would bias phylogenetic reconstruction. For this reason, instead of performing phylogenetic reconstruction using SNPs called against a reference genome, we used a reference-free method. We focused on a subset of core genes defined by two conditions. First, in all isolates, the gene coding sequence length is within 1% of median gene length across isolates. Second, the genes have pairwise identity of > 99%. Based on these criteria, out of ~ 5,000 genes in Kp isolates, 1424, 2212, 2802, and 3170 genes were present in 100%, 99%, 95% and 90% of all 509 isolates. We used Shannon's entropy to identify a 100 (from the 1424) genes with the most diverse nucleotide sequences. These 100 genes were aligned using MAFFT (v7.467)<sup>48</sup> and the resulting alignments used to construct phylogenetic trees for the entire isolate set (n = 509). The trees were reconstructed using IQTREE (v2.0.3)<sup>49</sup> with 1000 bootstrap replicates for each tree. The scripts can be found at <https://github.com/AntonS-bio/entropy>. For each ST, the gene selection was performed separately. The location of each chromosomal gene on the Kp reference genome (NC\_016845.1) (Fig. S1) was generated with BRIG software (v 0.95)<sup>50</sup>. Phylogenetic trees were visualised in ITOL and are available (<https://itol.embl.de/shared/Zp28yLE9IuWB>).

**Statistical analysis.** Statistical analysis was performed using R software (v4.0.3)<sup>51</sup>. Additional analysis was performed in Python (v3.6). For detection of replicon clusters, we created a presence/absence matrix with one row per isolate and one column per each unique replicon sequence. We performed dimensional reduction on this matrix using the Uniform Manifold Approximation and Projection (UMAP) algorithm<sup>29</sup> implemented in the R uwot package. We used both hamming and jaccard distance measures and a broad range of parameters to establish the robustness of our results. The cluster detection was performed using the DBSCAN algorithm implemented in R<sup>52</sup>. Analysis scripts are available on <https://github.com/AntonS-bio>.

**Consent for publication.** All authors have consented to the publication of this manuscript.

## Data availability

All sequencing data is available from European Nucleotide Archive project PRJEB47288. All AMR testing data is available in PATRIC database. Analysis scripts are available at <https://github.com/AntonS-bio>. High resolution phylogenetic trees are available at <https://itol.embl.de/shared/Zp28yLE9IuWB>.

Received: 1 November 2021; Accepted: 3 August 2022

Published online: 13 August 2022

## References

- Holt, K. E. *et al.* Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *K. Pneumoniae*, an urgent threat to public health. *Proc. Natl. Acad. Sci. USA*. **112**(27), E3574–E3581 (2015).
- Wyres, K. L. & Holt, K. E. *K. Pneumoniae* as a key trafficker of drug resistance genes from environmental to clinically important bacteria. *Curr. Opin. Microbiol.* **45**, 131–139 (2018).
- Ramirez, M. S. & Tolmasky, M. E. Aminoglycoside modifying enzymes. *Drug Res. Updates*. **13**(6), 151–171 (2010).
- Redgrave, L.S., Sutton, S.B., Webber, M.A., & Piddock, L.J.V. *Fluoroquinolone Resistance: Mechanisms, Impact on Bacteria, and Role in Evolutionary Success*, Vol. 22, Trends in Microbiology 438–45 (Elsevier Ltd, 2014). <https://doi.org/10.1016/j.tim.2014.04.007>.
- Hooper, D. C. & Jacoby, G. A. Topoisomerase inhibitors: Fluoroquinolone mechanisms of action and resistance. *Cold Spring Harbor Perspect. Med.* **6**(9), a025320 (2016).
- European Centre for Disease Prevention and Control. Antimicrobial resistance in the EU/EEA (EARS-Net)- Annual Epidemiological Report for 2019 [Internet]. 2020. <https://www.ecdc.europa.eu/en/publications-data/surveillance-antimicrobial-resistance-europe-2019>.
- Ferrari, C. *et al.* Multiple *K. Pneumoniae* KPC clones contribute to an extended hospital outbreak. *Front. Microbiol.* **10**, 556 (2019).
- Perdigão, J. *et al.* Whole-genome sequencing resolves a polyclonal outbreak by extended-spectrum beta-lactam and carbapenem-resistant *K. Pneumoniae* in a Portuguese tertiary-care hospital. *Microbiol. Genom.* **6**, 8896 (2020).
- Logan, L. K. & Weinstein, R. A. The epidemiology of carbapenem-resistant enterobacteriaceae: The impact and evolution of a global menace. *J. Infect. Dis.* **215**(Suppl 1), S28–36 (2017).
- Munoz-Price, L.S., Poirel, L., Bonomo, R.A., Schwaber, M.J., Daikos, G.L., Cormican, M. *et al.* *Clinical epidemiology of the global expansion of K. Pneumoniae carbapenemases*, Vol. 13, The Lancet Infectious Diseases 785–96 (NIH Public Access, 2013).
- Aires-De-Sousa, M. *et al.* Epidemiology of carbapenemase-producing *K. Pneumoniae* in a hospital, Portugal. *Emerg. Infect. Dis.* **25**(9), 1632–1638 (2019).
- Guerra, A. M. *et al.* Multiplicity of carbapenemase-producers three years after a kpc-3-producing *K. Pneumoniae* st147-k64 hospital outbreak. *Antibiotics* **9**(11), 1–11 (2020).
- Manageiro, V. *et al.* Predominance of KPC-3 in a survey for carbapenemase-producing Enterobacteriaceae in Portugal. *Antimicrob. Agents Chemother.* **59**(6), 3588–3592 (2015).
- Perdigão, J. *et al.* Genomic epidemiology of carbapenemase producing *K. Pneumoniae* strains at a northern portuguese hospital enables the detection of a misidentified klebsiella variicola kpc-3 producing strain. *Microorganisms*. **8**(12), 1–18 (2020).
- Rodrigues, C. *et al.* KPC-3-producing *K. Pneumoniae* in Portugal linked to previously circulating non-CG258 lineages and uncommon genetic platforms (Tn4401d-IncFIA and Tn4401d-IncN). *Front. Microbiol.* **7**, 665 (2016).
- Pires, D. *et al.* Evolving epidemiology of carbapenemase-producing Enterobacteriaceae in Portugal: 2012 retrospective cohort at a tertiary hospital in Lisbon. *J. Hosp. Infect.* **92**(1), 82–85 (2016).
- Caneiras, C., Lito, L., Melo-Cristino, J. & Duarte, A. Community-and hospital-acquired *K. Pneumoniae* urinary tract infections in Portugal: Virulence and antibiotic resistance. *Microorganisms*. **7**(5), 83356 (2019).
- Phelan, J. *et al.* The variability and reproducibility of whole genome sequencing technology for detecting resistance to anti-tuberculous drugs. *Genome Med.* **8**(1), 2258 (2016).
- Phelan, J. E. *et al.* Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med.* **11**(1), 882 (2019).
- Napier, G. *et al.* Robust barcoding and identification of Mycobacterium tuberculosis lineages for epidemiological and clinical studies. *Genome Med.* **12**(1), 114 (2020).
- Guerra-Assunção, J. A. *et al.* Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *eLife*. **2**(4), 558 (2015).
- Diancourt, L., Passet, V., Verhoef, J., Grimont, P. A. D. & Brisse, S. Multilocus sequence typing of *K. Pneumoniae* nosocomial isolates. *J. Clin. Microbiol.* **43**(8), 4178–4182 (2005).
- Wyres, K. L. *et al.* Identification of Klebsiella capsule synthesis loci from whole genome data. *Microbiol Genom.* **2**(12), e000102 (2016).
- Follador, R. *et al.* The diversity of *K. Pneumoniae* surface polysaccharides. *Microbiol Genom.* **2**(8), e000073 (2016).
- Bialek-Davenet, S. *et al.* Genomic definition of hypervirulent and multidrug-resistant *K. Pneumoniae* clonal groups. *Emerg. Infect. Dis.* **20**(11), 1812–1820 (2014).
- Brisse, S. *et al.* Virulent clones of *K. Pneumoniae*: Identification and evolutionary scenario based on genomic and phenotypic characterization. *PLoS ONE* **4**(3), 2286 (2009).
- Wyres, K. L. *et al.* Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *K. Pneumoniae*. *PLoS Genet.* **15**(4), e1008114 (2019).
- Comandatore, F. *et al.* Gene composition as a potential barrier to large recombinations in the bacterial pathogen *K. Pneumoniae*. *Genome Biol. Evol.* **11**(11), 3240–51 (2019).
- McInnes, L., Healy, J., & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv. 2018; <http://arxiv.org/abs/1802.03426>
- Lopes, E. *et al.* Epidemiology of carbapenemase-producing *K. Pneumoniae* in northern Portugal: Predominance of KPC-2 and OXA-48. *J. Global Antimicrob. Res.* **22**, 349–353 (2020).
- EUCAST: Clinical breakpoints and dosing of antibiotics-v 11.0. 2021. [https://www.eucast.org/clinical\\_breakpoints/](https://www.eucast.org/clinical_breakpoints/).
- Geetha, P. V., Aishwarya, K. V. L., Mariappan, S. & Sekar, U. Fluoroquinolone resistance in clinical isolates of *K. Pneumonia e. J. Lab. Phys.* **12**(2), 121 (2020).
- Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: Hierarchical and refined dataset for big data analysis–10 years on. *Nucleic Acids Res.* **44**(D1), D694–D697 (2016).
- Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E. *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2018.
- Lam, M. M. C. *et al.* A genomic surveillance framework and genotyping tool for *K. Pneumoniae* and its related species complex. *Nat. Commun.* **12**(1), 1–16 (2021).

36. Huynh, B. T. *et al.* *K. Pneumoniae* carriage in low-income countries: Antimicrobial resistance, genomic diversity and risk factors. *Gut Microbes*. **11**(5), 1287–1299 (2020).
37. Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. *Mobile Genetic Elements Associated with Antimicrobial Resistance* Vol. 31 (American Society for Microbiology, 2018).
38. Parish, T., Stoker, N.G., van Soolingen, D., de Haas, P.E.W., & Kremer, K. Restriction Fragment Length Polymorphism Typing of Mycobacteria. In *Mycobacterium Tuberculosis Protocols* 165–203 (Humana Press, 2003).
39. Spadar, A. *et al.* Methylation analysis of *K. Pneumoniae* from Portuguese hospitals. *Sci. Rep.* **11**(1), 6491 (2021).
40. Lam, M. M. C., Wick, R. R., Wyres, K. L. & Holt, K. E. Genomic surveillance framework and global population structure for *K. pneumoniae*. *bioRxiv*. **2**, 202012 (2020).
41. Wick, R. R., Heinz, E., Holt, K. E. & Wyres, K. L. Kaptive web: User-Friendly capsule and lipopolysaccharide serotype prediction for Klebsiella genomes. *J. Clin. Microbiol.* **56**(6), 197–215 (2018).
42. Seppey, M., Manni, M., & Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness. In *Methods in Molecular Biology* 227–45 (Humana Press Inc., 2019).
43. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**(14), 2068–2069 (2014).
44. Ehrlich, R. Prokka database maker. 2019. [https://github.com/rehrlich/prokka\\_database\\_maker](https://github.com/rehrlich/prokka_database_maker).
45. Feldgarden, M. *et al.* Validating the AMRFINDER tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* **63**(11), 550 (2019).
46. Seemann, T. Abricate: Mass screening of contigs for antimicrobial and virulence genes. <https://github.com/tseemann/abricate>.
47. Carattoli, A. *et al.* In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**(7), 3895–3903 (2014).
48. Katoh, K. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**(14), 3059–3066 (2002).
49. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**(1), 268–274 (2015).
50. Alikhan, N.-F., Petty, N. K., Zakour, N. L. B. & Beatson, S. A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genom.* **12**(1), 1–10 (2011).
51. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.
52. Hahsler, M., Piekenbrock, M. & Doran, D. dbscan: Fast density-based clustering with R. *J. Stat. Softw.* **91**(1), 71158 (2019).

## Acknowledgements

SC is funded by the Medical Research Council UK (Grant no. MR/M01360X/1) and BBSRC UK (BB/R013063/1). TGC is supported by the Medical Research Council UK (Grant no. MR/K000551/1, MR/M01360X/1, MR/N010469/1, MR/R020973/1) and BBSRC (BB/R013063/1). JPe is supported by Fundação para a Ciência e Tecnologia (ref. CEECIND/00394/2017 and UID/DTP/04138/2019). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

A.D., S.C., J.Pe. and T.G.C. designed the study. A.M., R.E., J.M.-C., G.L.S., C.P., C.M., M.P., and M.J.S. collected and processed isolates, including performing A.M.R. testing. S.C. and T.G.C. generated the sequence data. A.S. analysed the data, under the supervision of J.Ph., J.Pe., and T.G.C. A.S. wrote the first draft of the manuscript, with contributions from A.D., J.Pe. and T.G.C. All authors have edited manuscript drafts and agreed on the contents of the final version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-17996-1>.

**Correspondence** and requests for materials should be addressed to T.G.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022