nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-019-0492-8

# Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread

Sophia David[1], Sandra Reuter[2], Simon R. Harris[3], Corinna Glasner[4], Theresa Feltwell[3], Silvia Argimon[1], Khalil Abudahab[1], Richard Goater[1], Tommaso Giani[5], Giulia Errico[6], Marianne Aspbury[7], Sara Sjunnebo[8], the EuSCAPE Working Group[9], the ESGEM Study Group[10], Edward J. Feil[11], Gian Maria Rossolini[5,12], David M. Aanensen[1,13,14]* and Hajo Grundmann[2,4,14]*

**Public health interventions to control the current epidemic of carbapenem-resistant *Klebsiella pneumoniae* rely on a comprehensive understanding of its emergence and spread over a wide range of geographical scales. We analysed the genome sequences and epidemiological data of >1,700 *K. pneumoniae* samples isolated from patients in 244 hospitals in 32 countries during the European Survey of Carbapenemase-Producing Enterobacteriaceae. We demonstrate that carbapenemase acquisition is the main cause of carbapenem resistance and that it occurred across diverse phylogenetic backgrounds. However, 477 of 682 (69.9%) carbapenemase-positive isolates are concentrated in four clonal lineages, sequence types 11, 15, 101, 258/512 and their derivatives. Combined analysis of the genetic and geographic distances between isolates with different β-lactam resistance determinants suggests that the propensity of *K. pneumoniae* to spread in hospital environments correlates with the degree of resistance and that carbapenemase-positive isolates have the highest transmissibility. Indeed, we found that over half of the hospitals that contributed carbapenemase-positive isolates probably experienced within-hospital transmission, and interhospital spread is far more frequent within, rather than between, countries. Finally, we propose a value of 21 for the number of single nucleotide polymorphisms that optimizes the discrimination of hospital clusters and detail the international spread of the successful epidemic lineage, ST258/512.**

The bacterium *Klebsiella pneumoniae*, a major cause of both hospital- and community-acquired infections, is listed by the World Health Organization as a critical priority antibiotic-resistant bacterial pathogen for which new antibiotics are urgently needed[1]. Indeed, a recent study showed that carbapenem-resistant *K. pneumoniae* represents the fastest growing antibiotic resistance threat in Europe, in terms of human morbidity and mortality[2]. It is therefore critical to identify priority areas on which to intensify public health intervention strategies.

Rapid expansion of carbapenem resistance in *K. pneumoniae* has been attributed to the acquisition of carbapenemase enzymes that hydrolyse carbapenems (a last-line class of antibiotics) and other β-lactam antibiotics to varying degrees. Carbapenemase genes are associated with mobile elements that can spread horizontally within, and between, bacterial species; this has facilitated their widespread dissemination[3,4]. In addition to being an important pathogen, *K. pneumoniae* has been identified as a crucial entry point of antibiotic resistance genes into the Enterobacteriales (Enterobacteriaceae) family[5–7].

Emergence of carbapenem resistance seems to occur in clinical[8], urban[9] and agricultural settings[10] and is a worldwide phenomenon[11–13]. However, biased and fragmented surveillance combined with a lack of standardization in the characterization of isolates has made it difficult to discern the primary reservoirs and transmission dynamics of this global epidemic. The European Survey of Carbapenemase-Producing Enterobacteriaceae (EuSCAPE) was the first study to systematically determine the incidence and epidemiology of carbapenem-non-susceptible *K. pneumoniae* at a continental scale, enrolling 455 hospitals across Europe and neighbouring countries[14]. Between November 2013 and May 2014, hospital laboratories were asked to submit their first ten consecutive carbapenem-non-susceptible clinical isolates of either *K. pneumoniae* or *E. coli*, together with one carbapenem-susceptible same-species clinical isolate (per non-susceptible isolate) to serve as a comparator. Isolates were obtained from clinical specimens submitted for diagnostic purposes and samples obtained for screening were excluded.

To elucidate the European-wide population structure and determine the epidemiology of carbapenem-non-susceptible *K. pneumoniae* with maximum resolution, we analysed the genomes of 1,717 *K. pneumoniae* isolates submitted during the survey. This

[1]Centre for Genomic Pathogen Surveillance, Wellcome Genome Campus, Cambridge, UK. [2]Institute for Infection Prevention and Hospital Epidemiology, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [3]Pathogen Genomics, Wellcome Sanger Institute, Cambridge, UK. [4]Department of Medical Microbiology and Infection Prevention, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [5]Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy. [6]Department of Infectious Diseases, Istituto Superiore di Sanità, Rome, Italy. [7]Natural Sciences, University of Bath, Bath, UK. [8]Pathogen Informatics, Wellcome Sanger Institute, Cambridge, UK. [9]A full list of authors and their affiliations appears at the end of the paper. [10]A list of members appears in the Supplementary Note. [11]Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, UK. [12]Clinical Microbiology and Virology Unit, Florence Careggi University Hospital, Florence, Italy. [13]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [14]These authors contributed equally: David M. Aanensen, Hajo Grundmann. *e-mail: david.aanensen@sanger.ac.uk; hajo.grundmann@uniklinik-freiburg.de

unprecedented sample provides an unbiased, continental and contemporaneous population snapshot of representative clinical isolates from hospitalized patients and links lineage abundance and spatial expansion to variable genomic resistance determinants.

## Results

**Population structure and carbapenem resistance determinants.** Genome sequences were obtained for 1,717 clinical isolates that originated from 244 hospitals in 32 countries (Fig. 1, Supplementary Fig. 1 and Supplementary Tables 1–4). Of the 1,717 isolates, 944 (55.0%) were submitted by hospital laboratories as carbapenem-non-susceptible and 773 (45.0%) were submitted as carbapenem-susceptible.

Core genome phylogenetic analysis revealed a partitioning into four species, supporting recent taxonomic classifications[15,16] (Fig. 2a). Of the 1,717 isolates, 1,649 (96.0%) belong to *K. pneumoniae* sensu stricto, 48 (2.8%) to *K. variicola*, 17 (1.0%) to *K. quasipneumoniae* and 3 (0.2%) to a newly described species *K. quasivariicola*[16]. Of the 944 isolates submitted as carbapenem-non-susceptible, 939 (99.5%) belong to *K. pneumoniae* sensu stricto, 3 (0.3%) to *K. variicola* and 2 (0.2%) to *K. quasipneumoniae*.

We defined phylogenetic lineages in *K. pneumoniae* sensu stricto by their multilocus sequence type (ST)[17]. Of the 254 different STs identified among isolates of this species, 94 (37.0%) contain isolates submitted as carbapenem-non-susceptible and 15 (5.9%) contain >20 isolates (Fig. 2b). Two of the most numerous STs, 258 ($n=73$) and 512 ($n=163$), which are single-locus variants, can be considered as a single lineage as there is no substantial evolutionary separation between them (Fig. 2b).

We searched the genomes for β-lactam resistance determinants that could be relevant to carbapenem resistance and divided the entire collection into five β-lactam resistome groups. Group (1) contained one or more of any carbapenemase gene published in the literature before April 2018 (see Methods), regardless of other mechanisms; group (2) contained any extended spectrum β-lactamase (*ESBL*) gene and/or an *AmpC* gene, in combination with porin defects; group (3) contained any *ESBL* gene and/or an *AmpC* gene, without any obvious porin defects; group (4) showed an absence of any *ESBL* or *AmpC* gene but the presence of porin defects; and in group (5), all of the above determinants were absent.

Group (1) comprised the 684 isolates that carried one or more carbapenemase genes. Of these, all but two belonged to *K. pneumoniae* sensu stricto, which contained $bla_{\text{KPC-like}}$ ($n=311$), $bla_{\text{OXA-48-like}}$ ($n=248$), $bla_{\text{NDM-like}}$ ($n=79$), $bla_{\text{VIM-like}}$ ($n=56$) and $bla_{\text{IMP-like}}$ ($n=3$) genes in 28, 44, 13, 13 and 1 ST, respectively (Table 1 and Supplementary Table 4). A $bla_{\text{OXA-48-like}}$ gene was also found in one *K. quasipneumoniae* genome and a $bla_{\text{KPC-like}}$ gene in one *K. variicola* genome. Eighteen *K. pneumoniae* sensu stricto isolates carried two carbapenemase genes, the most commonly observed combination being that of $bla_{\text{NDM-like}}$ and $bla_{\text{OXA-48-like}}$ genes, which was found in ten isolates. Furthermore, in silico detection of the $bla_{\text{KPC-like}}$, $bla_{\text{OXA-48-like}}$, $bla_{\text{NDM-like}}$ and $bla_{\text{VIM-like}}$ carbapenemase genes showed 98.3–99.0% concordance with PCR results obtained previously by the national expert laboratories of individual countries (Supplementary Table 5). Of the 684 group (1) isolates, 657 (96.1%) were submitted as carbapenem-non-susceptible, including 655 of 682 (96.0%) of those belonging to *K. pneumoniae* sensu stricto (Table 1). On central retesting with reference broth microdilution, we found that group (1) isolates possessed the highest phenotypic resistance to meropenem with a median minimum inhibitory concentration (MIC) value of 32 (Fig. 3a and Supplementary Fig. 2).

We found 150 isolates belonging to group (2) that lacked a carbapenemase but harboured an *ESBL* gene and/or an *AmpC* gene, in combination with porin defects (see Methods; Table 1 and Supplementary Table 4). These resistance determinants can also account for a carbapenem-non-susceptible phenotype. All belong to *K. pneumoniae* sensu stricto and 114 of 150 (76.0%) were submitted as carbapenem-non-susceptible, a lower proportion than that observed for group (1) isolates. Central retesting of these isolates also demonstrated lower phenotypic resistance to meropenem than group (1) isolates, with a median MIC value of 1 (Fig. 3a and Supplementary Fig. 2).

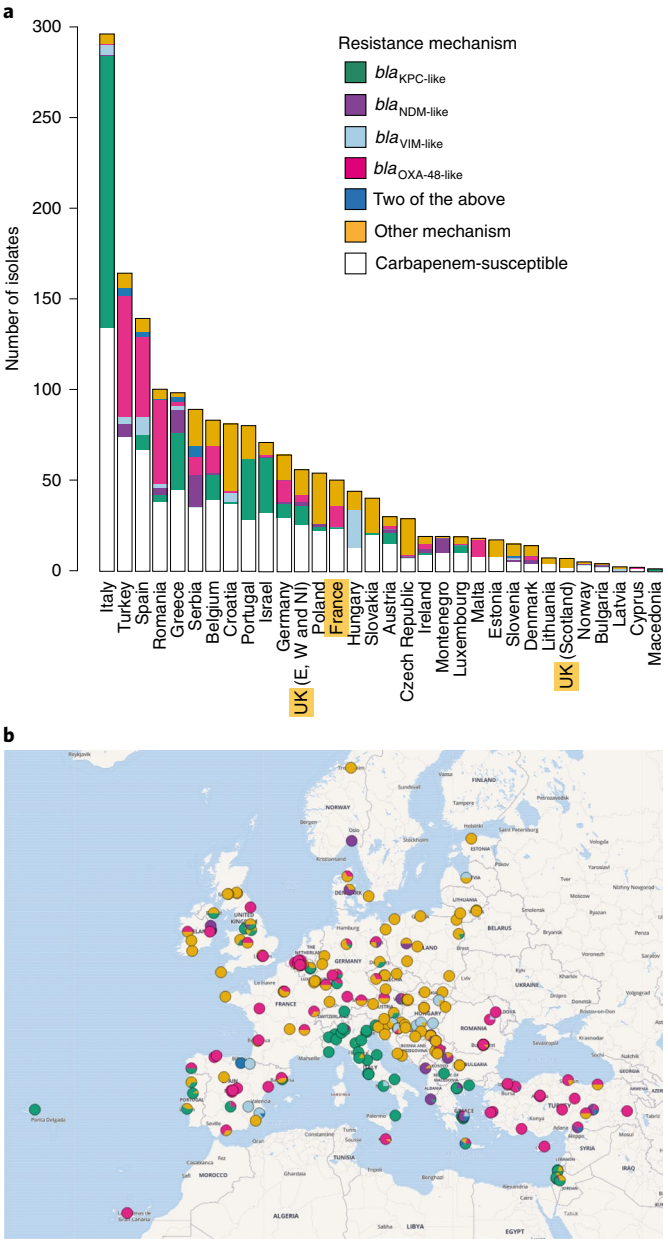


**Fig. 1 | Geographical distribution of carbapenem resistance mechanisms among isolates submitted during the EuSCAPE survey. a**, All of the isolates submitted by participating countries and analysed in this study, partitioned into: those that possess one or more of four major carbapenemase genes ($bla_{\text{KPC-like}}$, $bla_{\text{OXA-48-like}}$, $bla_{\text{NDM-like}}$ and $bla_{\text{VIM-like}}$), regardless of whether they were submitted as carbapenem-non-susceptible or -susceptible; those that lack any of the four genes and were submitted as carbapenem-non-susceptible ('Other mechanism'); and those submitted as carbapenem-susceptible. **b**, Pie charts showing the proportions of different resistance mechanisms in isolates submitted as carbapenem-non-susceptible by participating hospitals. Colour key as in **a**. E, England; W, Wales; NI, Northern Ireland.

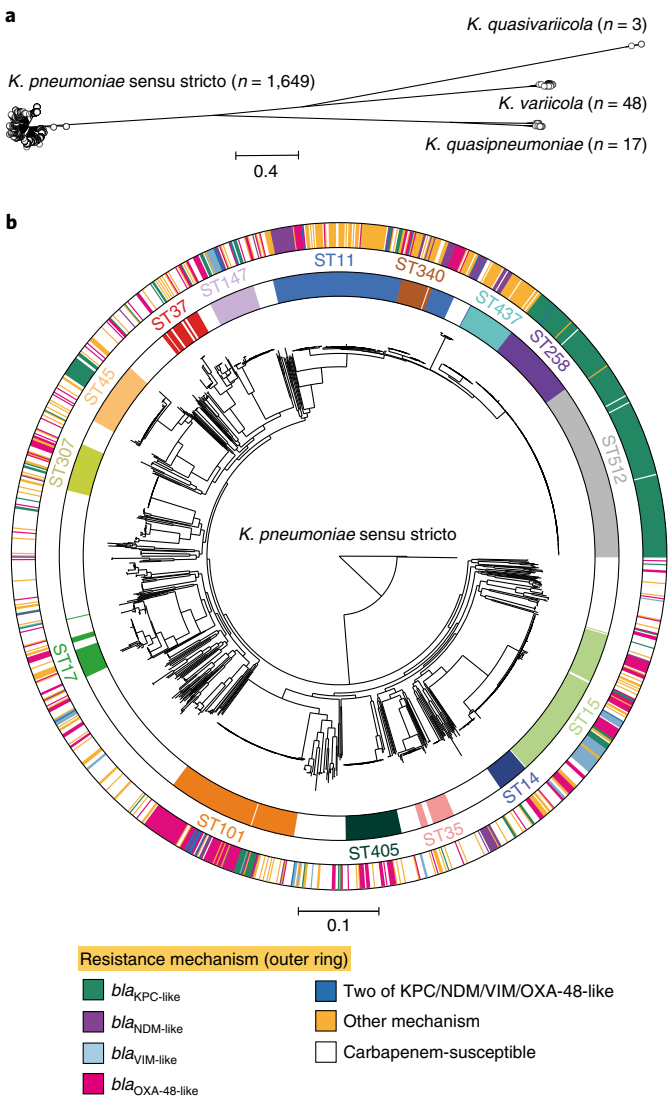


**Fig. 2 | Carbapenemase-positive isolates are concentrated in major clonal lineages of *K. pneumoniae*.** **a**, A phylogenetic tree of the 1,717 EuSCAPE isolates analysed in this study (constructed using 2,539 core genes) demonstrating the division into four species. **b**, A phylogenetic tree of 1,649 isolates belonging to *K. pneumoniae* sensu stricto; STs containing more than 20 isolates are shown in the inner ring and labelled. In the outer ring, isolates are coloured by resistance mechanism as in Fig. 1a. The scale in both trees represents the number of single nucleotide polymorphisms (SNPs) per variable site. Similar visualizations are available at: https://microreact.org/project/EuSCAPE_Kp/bc81fffe (**a**) and https://microreact.org/project/EuSCAPE_Kp/2585de34 (**b**).

A further 402, 40 and 441 isolates were assigned to groups (3), (4) and (5), respectively, including 400, 37 and 380 isolates from *K. pneumoniae* sensu stricto. In these three groups, 30.6%, 65.0% and 5.4% of isolates were submitted as carbapenem-non-susceptible, respectively (or 30.8%, 64.9% and 6.1% from *K. pneumoniae* sensu stricto only) (Table 1). However, based on conventional classification, isolates from these three groups were expected to be susceptible. Indeed, on central retesting, the three groups were associated with lower MIC values to meropenem than groups (1) and (2) (Fig. 3a and Supplementary Fig. 2), with median MIC values of ≤0.06, 0.25 and ≤0.06. Thus, discrepancies between the classifications of participating laboratories and gene contents may reflect heterogeneity in antibiotic susceptibility testing, subsequent loss of resistance elements postsubmission or incorrect classification.

**Limited genetic diversity among carbapenemase-positive isolates.** Carbapenemase-positive isolates are concentrated in the major clonal lineages of *K. pneumoniae* sensu stricto (Table 1 and Fig. 2b). In particular, 477 of the 682 (69.9%) carbapenemase-positive isolates in this species belong to four lineages comprising STs 11, 15, 101 and ST258/512, as well as their derivatives (other closely related STs that have evolved from these major STs). Among these STs (without their derivatives), the majority of isolates were carbapenemase-positive, except for ST11 (ST258/512: 97.5%; ST101: 69.9%; ST15: 55.9%; ST11: 37.9%), while the proportion of carbapenemase-positive isolates among the total *K. pneumoniae* sensu stricto sample was 39.8%. The major lineages are also characterized by low levels of core genome diversity, reflected by an average nucleotide identity (ANI) of 99.9–100% between same-ST pairs of STs 15, 101 and 258/512. This indicates a recent and common evolutionary descent. Furthermore, ST258 and ST512 harbour less diversity overall than each of the other three STs individually, supporting their combined grouping. However, ST11 (also a single-locus variant of ST258) consists of several sublineages that are distinct from ST258 and exhibits the highest diversity of the major STs with an ANI of 99.7–100% between pairs. For context, the ANI ranges from 94.1–95.4% between genomes of the four different species and from 98.9–100% between genomes of *K. pneumoniae* sensu stricto. Despite an overall limited genetic diversity, STs 11, 15, 101 and 258/512 are widely distributed across Europe, having been submitted from 22, 19, 15 and 15 countries, respectively.

Hence, the five carbapenemase genes identified among isolates in *K. pneumoniae* sensu stricto are—for the majority—confined to recently emerging and vastly successful clonal backgrounds. This can also be seen by the limited population diversity; diversity indices for carbapenemase-positive isolates were significantly lower than for the other β-lactam resistome groups (Table 1). The only exception is $bla_{\text{OXA-48-like}}$-containing isolates for which the diversity index was slightly higher and CIs overlapped with group (2) isolates (*ESBL*/*AmpC*-positive isolates with porin defects).

**Geographic spread.** To explore the differential ecological success of *K. pneumoniae* sensu stricto with different β-lactam resistance determinants, we measured the genetic relatedness of isolates belonging to the five previously defined β-lactam resistome groups with respect to three geographic contexts: from the same hospital; from the same country but different hospitals; and from different countries. We found that for 359 of the 682 (52.6%) carbapenemase-positive (group (1)) isolates, the genetically nearest neighbour (gNN) in the collection originated from patients treated in the same hospital (Fig. 3b). When stratifying by carbapenemase, we observed that the gNN was from the same hospital for 159 of 311 (51.1%) isolates carrying $bla_{\text{KPC-like}}$ genes, 125 of 248 (50.4%) carrying $bla_{\text{OXA-48-like}}$ genes, 43 of 79 (54.4%) carrying $bla_{\text{NDM-like}}$ genes and 42 of 56 (75.0%) carrying $bla_{\text{VIM-like}}$ genes. Yet isolates from other β-lactam resistome groups ((2)–(5)) had a lower proportion (34.7%, 28.3%, 29.7%, 9.2%) of gNNs obtained from same-hospital patients, with a significant downward trend that coincided with a decreasing ability to express carbapenem resistance ($P < 0.01$, $\chi^2$ test for trend; Fig. 3b). The ANI among gNNs was very high (>99.95–100%), irrespective of the β-lactam resistome group. We tested the observed proportions against the null hypothesis of a random geographic distribution by the permutation of the hospital codes for all isolates and obtained the expected ranges by 100 repeat analyses. We could thus infer that, in the absence of geographic structure, the proportions of gNNs from the same hospital were expected to be 67.4, 56.9, 41.6, 58.2 and 14.4 times lower for all the resistome groups ((1)–(5)) considered. Assuming that gNNs with a high ANI share a recent common

**Table 1 | Characteristics of all submitted *K. pneumoniae* sensu stricto isolates with different β-lactam resistance determinants**

| Isolate subset | Number of submitted isolates (% of total) | Number of isolates submitted as carbapenem-non-susceptible (%) | Number of STs (and STs with >10% isolates)[a] | Simpson's diversity of STs (and 95% CI)[a] | Number of countries (and countries with >10% isolates) | Carbapenemase gene variants (and number of isolates) |
|---|---|---|---|---|---|---|
| $bla_{KPC\text{-like}}$ | 311 (18.9) | 304 (97.7) | 28 (ST258, *n*=69; ST512, *n*=157) | 0.68 (0.63–0.73) | 17 (Italy, *n*=150; Portugal, *n*=34; Greece, *n*=34) | KPC-2 (*n*=78), KPC-3 (*n*=232), KPC-12 (*n*=1) |
| $bla_{OXA\text{-}48\text{-like}}$ | 248 (15.0) | 237 (95.6) | 44 (ST15, *n*=38; ST101, *n*=67) | 0.88 (0.85–0.91) | 20 (Romania, *n*=47; Spain, *n*=46; Turkey, *n*=70) | OXA-48 (*n*=240), OXA-204 (*n*=1), OXA-162 (*n*=2), OXA-181 (*n*=1), OXA-232 (*n*=4) |
| $bla_{NDM\text{-like}}$ | 79 (4.8) | 76 (96.2) | 13 (ST11, *n*=25; ST101, *n*=12; ST274, *n*=10; ST395, *n*=8) | 0.84 (0.78–0.89) | 19 (Serbia, *n*=24; Greece, *n*=13; Turkey, *n*=10; Montenegro, *n*=8) | NDM-1 (*n*=79) |
| $bla_{VIM\text{-like}}$ | 56 (3.4) | 53 (94.6) | 13 (ST15, *n*=27; ST147, *n*=7) | 0.73 (0.61–0.85) | 9 (Hungary, *n*=21; Spain, *n*=12) | VIM-1 (*n*=33), VIM-4 (*n*=23) |
| $bla_{IMP\text{-like}}$ | 3 (0.2) | 3 (100) | 1 (ST15, *n*=3) | 0 (0–0) | 1 (Turkey, *n*=3) | IMP-1 (*n*=3) |
| Any carbapenemase gene (group(1))[b] | 682 (41.4) | 655 (96.0) | 69 (ST15, *n*=78; ST101, *n*=84; ST258, *n*=69; ST512, *n*=157) | 0.89 (0.88–0.91) | 29 (Italy, *n*=157; Turkey, *n*=85) | See above |
| *ESBL* and/or *AmpC* gene + porin defects, but no carbapenemase (group (2)) | 150 (9.1) | 114 (76.0) | 35 (ST11, *n*=30; ST15, *n*=18) | 0.92 (0.90–0.95) | 25 (Slovakia, *n*=19) | NA |
| *ESBL* and/or *AmpC* gene, but no carbapenemase/porin defects (group (3)) | 400 (24.3) | 123 (30.8) | 90 (ST11, *n*=52; ST15, *n*=44) | 0.95 (0.94–0.96) | 27 (Italy, *n*=43; Serbia, *n*=43) | NA |
| Porin defects, but no carbapenemase/*ESBL*/*AmpC* (group (4)) | 37 (2.2) | 24 (64.9) | 21 (ST17, *n*=5; ST437, *n*=4; ST512, *n*=4) | 0.96 (0.93–0.98) | 18 (Croatia, *n*=4; Italy, *n*=4; Portugal, *n*=4; United Kingdom, *n*=5) | NA |
| No carbapenemase/*ESBL*/*AmpC*/ porin defects (group (5)) | 380 (23.0) | 23 (6.1) | 161 | 0.98 (0.98–0.99) | 27 (Italy, *n*=71; Spain, *n*=41; Turkey, *n*=49) | NA |

[a]Ambiguous STs (those with one or more uncertain alleles) were excluded for these calculations. [b]These include three isolates that were found to carry $bla_{OXA\text{-}23}$ or $bla_{OXA\text{-}58}$. However, as these genes are usually found in only *Acinetobacter baumannii* and were found at a low level in our samples, we suspect that they are contaminants. NA, not applicable; CI, confidence interval.

ancestor, our findings suggest that carbapenemase-positive strains have the highest transmissibility and epidemic potential and that the propensity of *K. pneumoniae* sensu stricto to spread in hospitals correlates with the degree of resistance. Isolates with an *ESBL*/*AmpC* gene and/or porin defects occupy an intermediate position between carbapenemase-positive isolates and those for which none of these β-lactam resistance determinants were detected. We also observed a similar but less pronounced distribution for gNNs originating from different hospitals as long as they were from the same country. Notably, the lower bound of the CI of gNNs in the group of isolates with no β-lactam resistance determinants (group (5)) approaches that of the random distribution in this geographic context (Fig. 3b).

**Thresholds for discriminating likely transmission events.** Of the 171 hospitals that contributed carbapenemase-positive isolates, 96 (56.1%) had at least two isolates with a nearest neighbour relationship as described above, suggestive of within-hospital transmission. Given the high frequency of the nosocomial spread of carbapenemase-positive *K. pneumoniae* sensu stricto, the ability to identify likely transmission events within hospitals directly from SNP distances would present a significant advance with respect to both immediate infection control priorities and retrospective outbreak investigations. Due to widespread sampling that allows comparisons across increasing geographic scales, the EuSCAPE collection provides an opportunity to investigate levels of diversity in relation to potential transmission.

We focused on the largest epidemic clone in the EuSCAPE collection, ST258/512, and first analysed how pairwise SNP differences vary for the same geographic contexts as used previously. For each of these three contexts, we calculated the pairwise minimum SNP differences between isolates (Fig. 4a–c). The minimum SNP differences showed a central tendency and a shift of the mode to the right from 0 in the same-hospital context, to 27 in the different-hospital context and 45 in the different-country context. These results lend

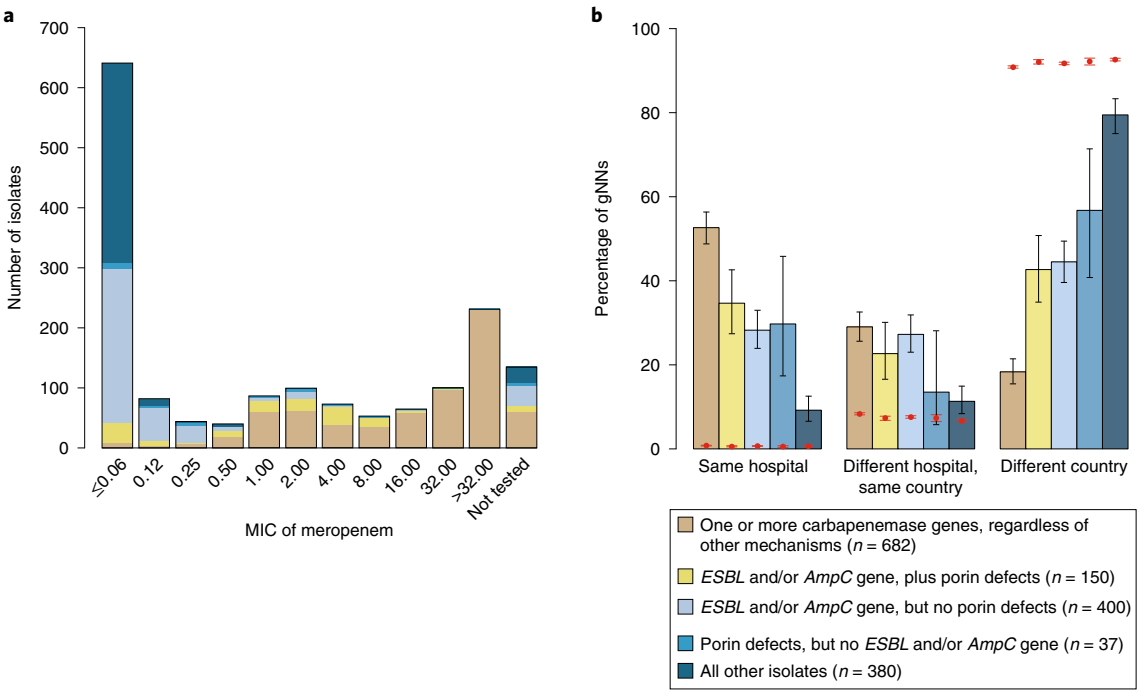


**Fig. 3 | Carbapenemase-positive isolates show strong geographical clustering. a**, Bar plots showing the distribution of MIC values among 1,518 retested *K. pneumoniae* sensu stricto isolates with different β-lactam resistance determinants. **b**, Bar plots showing the proportions of gNNs for *K. pneumoniae* sensu stricto isolates with different β-lactam resistance determinants that are from different geographic contexts (the same hospital, a different hospital in the same country, or a hospital in a different country). The black error bars represent the 95% CIs of these proportions. The proportions of gNNs in the absence of geographic structure were calculated from 100 repeat analyses in which the hospital codes were randomly permutated; the means and 95% CIs of these results are shown in red.

support to the notion that, among carbapenemase-positive isolates, geographic and evolutionary distance correspond and also provide an indication that pairwise SNP differences between isolates from a given hospital provide meaningful information about transmissions within an institution. Assuming that isolates submitted from different hospitals/countries are less likely than those from the same hospital to be closely linked in a transmission chain, we would expect an optimum SNP threshold for defining an institutional transmission to be lower than the number of SNPs typically seen between isolates from different hospitals/countries.

We next used a statistical method[18] to distinguish between intra and interhospital transmission, through the comparison of SNP profiles of a given query isolate against a reference database. We used all ST258/512 genomes from the EuSCAPE collection as the database, representing isolates from 59 hospitals, and removed one isolate at a time to be used as the query. Using a naive Bayesian classifier, the method determines the likelihood of any given query isolate originating from each of the hospitals represented in the reference database based on the presence of SNPs called against a standard reference genome. Intrahospital transmission events are assigned in those cases where the most likely source hospital of a query isolate coincides with the hospital from which the isolate was actually recovered. In contrast, if a given query isolate is found to be most similar to those from a hospital that is different to the one from which it was sampled, then it is assigned as corresponding to an interhospital transmission, although it is not possible to assign directionality.

Using these predictions as a reference, we were able to determine an optimal value of 21 SNPs to discriminate hospital clusters, which minimizes the number of false positives (pairs of isolates erroneously assigned to intrahospital transmission) and false negatives (cases of missed intrahospital transmission) (Fig. 4d–f). At the same time, this method provides a measure of uncertainty at different thresholds, which equates to a false-positive rate of 14.6% and a false-negative rate of 11.7% at 21 SNPs.

**International spread of the epidemic ST258/512 clone.** Finally, to determine the ancestral relationships between the dispersed populations of the ST258/512 clone, we analysed 236 ST258/512 genomes from the EuSCAPE collection together with 415 publicly available draft genomes of STs 258, 512 and another derivative, ST868 (Supplementary Table 6), which were isolated in 20 countries across Europe, the Middle East and North and South America. After mapping sequence reads to a ST258 reference genome (accession no. CP006923), we identified 68 recombined regions larger than 1 kb. These included a 57.1 kb region encompassing the capsular locus that introduced 563 SNPs, thereby accounting for the capsular-type (K-type) switch from K107 to K106, which coincides with the previously described evolutionary transition from clade I to clade II (ref. [19]; Supplementary Fig. 3). Another two recombination events over the capsular region that affected single isolates could also be traced to switches in K-type.

We then removed the recombined regions and used the resulting alignment to reconstruct a maximum likelihood tree of all 651 isolates (Fig. 5). ST258 isolates from the United States are basal in the phylogenetic tree, supporting the conventional view that this lineage emerged in the United States. All but three isolates ($n = 31$) from Greece, sampled between 2007 and 2014, fall into a single clade (bootstrap support, 100%) suggesting one major introduction into Greece from the United States and subsequent nationwide spread. Many isolates from other European countries (the United Kingdom and Germany, for example) also cluster among these Greek isolates, representing the spread of this lineage out of Greece, probably via human travel. All 268 isolates from Israel, sampled between 2007 and 2014, also cluster together in a single clade (bootstrap support, 97%)

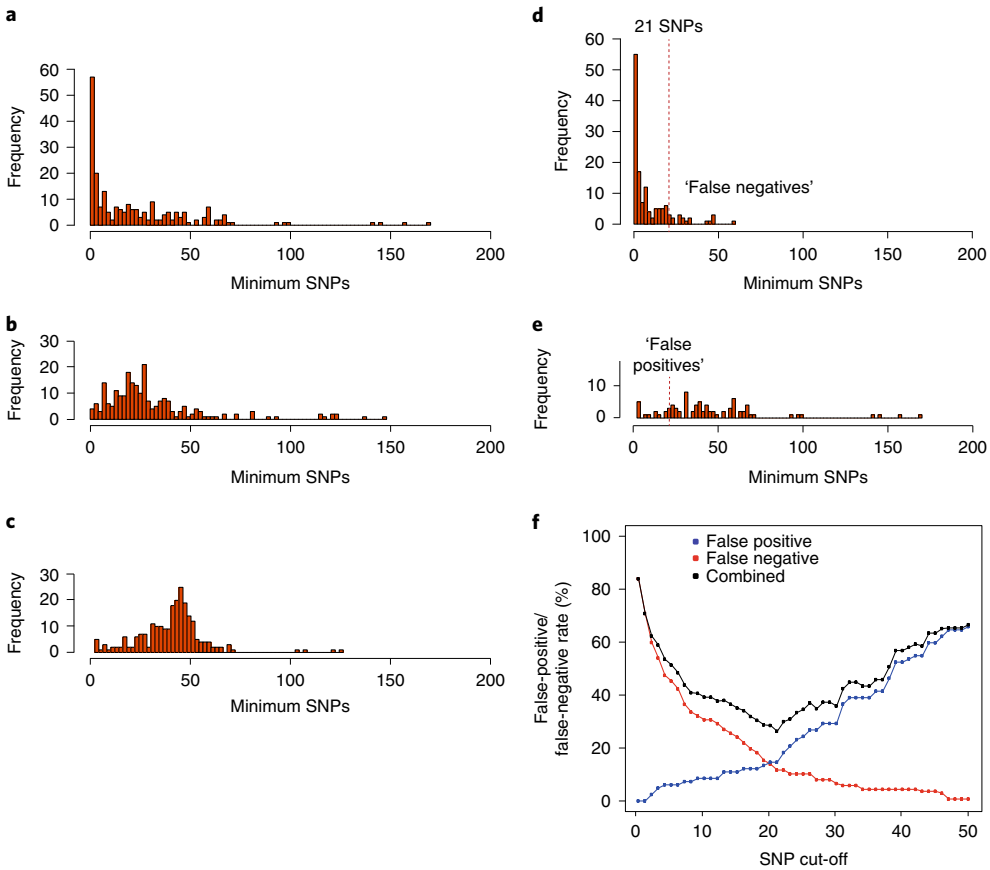


**Fig. 4 | Determination of a SNP cut-off to aid outbreak investigations of ST258/512. a–c**, The number of core chromosomal SNPs between each of the 236 ST258/512 isolates submitted during EuSCAPE and the most closely related isolate submitted by the same hospital (**a**), a different hospital in the same country (**b**) and a hospital in a different country (**c**). **d**,**e**, The number of SNPs between each ST258/512 isolate and the most closely related isolate submitted by the same hospital, when isolates are predicted to have been involved in intrahospital (**d**) or interhospital (**e**) transmission. The dotted lines are drawn at the number of SNPs at which the combined false-positive and false-negative rates for correctly identifying intrahospital transmission, calculated using the predictions as the reference, are at a minimum. **f**, False-positive and false-negative rates when different SNP cut-offs are used.

suggesting one successful introduction followed by within-country dissemination. The most basal ST512 isolates are also from Israel, implying emergence of ST512 in Israel. The tree structure demonstrates a single successful introduction of ST512 from Israel to Italy (bootstrap support, 97%), although separate introductions of ST258 to Italy are also evident. The potential of ST258/512 to spread out of endemic countries and spawn outbreaks elsewhere is also exemplified by an ST258 cluster in Spain (four patients in one hospital) and ST512 clusters in Belgium (seven patients in two nearby hospitals) and Austria (three patients in one hospital). Phylogenetic inference suggests that all three clusters probably originated from Italy.

## Discussion

The emergence of carbapenem resistance is a major setback for the effective treatment of Gram-negative bacterial infections[20]. In *K. pneumoniae*, carbapenemases are the main contributing factor to extensive drug resistance and their recent acquisition and dissemination probably forebodes pan-drug resistance in the near future[21,22]. Across Europe, resistance rates differ between countries, which may be explained in part by different levels of antibiotic use. However, critical knowledge gaps remain regarding how the characteristics of the pathogen population also contribute to the epidemic. Numerous studies have highlighted the role of animals and environmental reservoirs as important evolutionary drivers in the mobilization of resistance genes both within and between species and argued that increasing selective pressure in animals and the environment warrants a 'one-health' surveillance approach[23,24]. However, this study reasserts the central role of intra and interhospital transmission for the most burdensome carbapenemase-positive clones that are already well-established in the human population. Indeed, strong links between hospital care and the acquisition of carbapenem-non-susceptible bacteria have been noted previously[14,25–28] and the impact of community[29] or animal and environmental reservoirs[10] on clinical cases remains circumstantial. In this context, we note that a large faecal resistome analysis from slaughter pigs and broilers failed to identify carbapenemase genes in 181 pig and 178 poultry farms from nine European countries[30].

Here we infer the epidemiology of carbapenemase-positive *K. pneumoniae* in Europe using the genome data for 1,717 carbapenem-non-susceptible and -susceptible isolates obtained from consecutive clinical samples in 244 hospitals in 32 countries over a 6 month prevalence survey (EuSCAPE)[14]. Most (99.5%) of the 944 isolates submitted as carbapenem-non-susceptible belonged to the species *K. pneumoniae* sensu stricto. Furthermore, the majority (69.6%) of isolates submitted as carbapenem-non-susceptible possessed one or more carbapenemase genes, which included the five most frequently reported carbapenemase genes worldwide: $bla_{\text{KPC-like}}$, $bla_{\text{OXA-48-like}}$, $bla_{\text{NDM-like}}$, $bla_{\text{VIM-like}}$ and $bla_{\text{IMP-like}}$. Carbapenemase-positive isolates were concentrated among four major clonal lineages of *K. pneumoniae* sensu stricto, comprising STs 11, 15, 101 and 258/512. All have recently emerged, gained abundance and, for the most part, expanded in southern or eastern European countries.

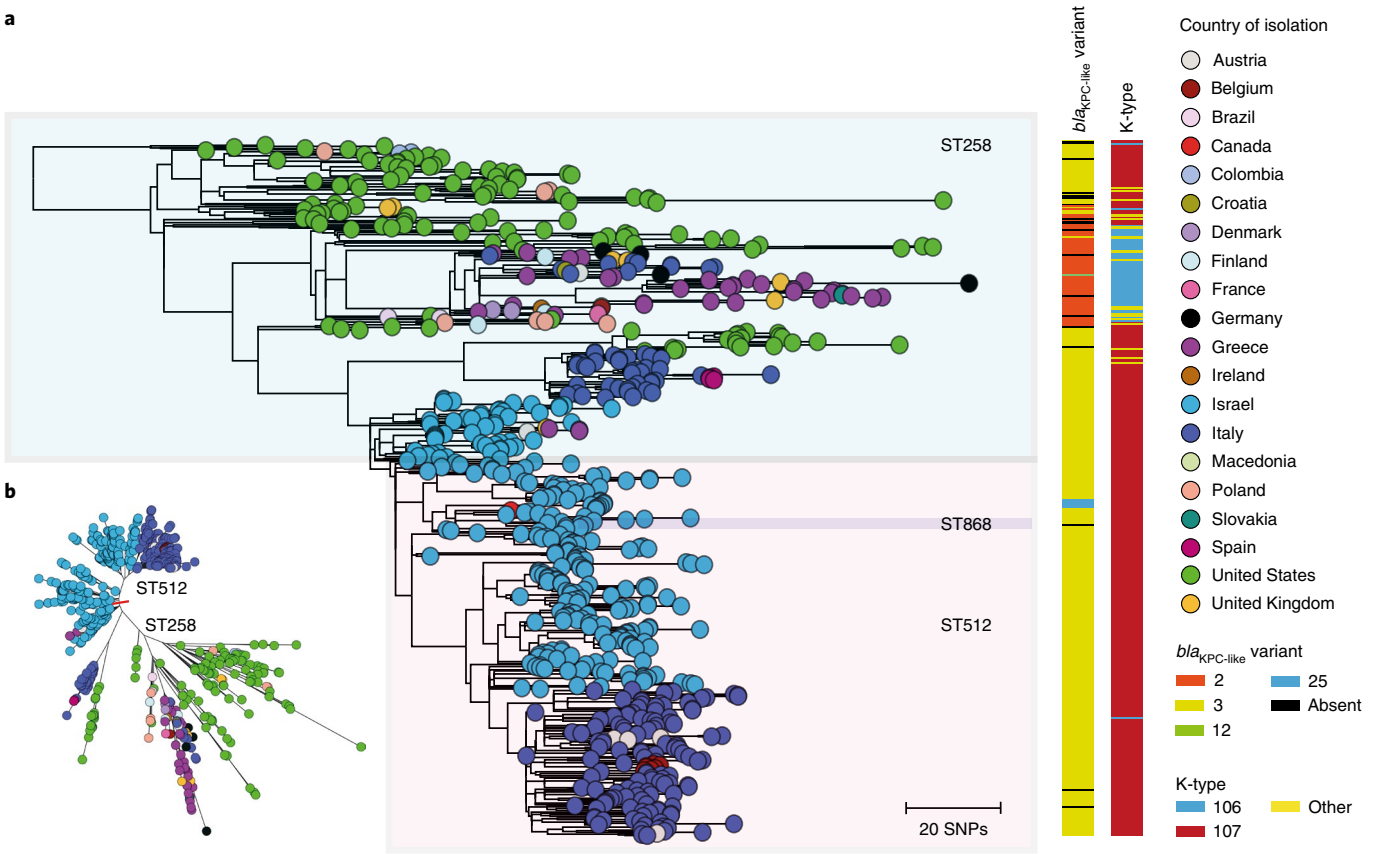


**Fig. 5 | International spread of the epidemic ST258/512 clone. a**, A phylogenetic tree of 651 isolates of ST258 and 512 (and a derivative of ST512, ST868), comprising 236 isolates submitted during the EuSCAPE survey and 415 isolates with publicly available sequence data. The tree was rooted using an ST11 isolate that was later removed. The colours of the isolate tips represent the country of isolation and metadata columns show the $bla_{KPC-like}$ variant and K-type. **b**, An unrooted version of the tree shown in **a**. Similar visualizations are available at: https://microreact.org/project/EuSCAPE_ST258/bbafcc1c (**a**) and https://microreact.org/project/EuSCAPE_ST258/dd960284 (**b**).

Hence, these lineages bear the hallmarks of so-called 'high-risk' clones, which commonly share a recent ancestor, epidemic success and a defined geographic distribution[31]. They are also often associated with outbreaks[32–35] and probably possess particular characteristics that increase their tenacity, transmissibility and population size, providing a greater opportunity for the acquisition of antibiotic resistance genes[36].

We demonstrated a strong relationship between genetic and geographic distance among carbapenemase-positive isolates, which cluster at the country level and, even more strikingly, at the hospital level. More than half of the carbapenemase-positive isolates had their gNN sampled from the same hospital and likewise the majority of hospitals that contributed carbapenemase-positive *K. pneumoniae* had at least one pair of gNNs isolated from patients treated within a 6 month period at the same hospital. This suggests frequent within-hospital transmission. At the same time, the proportion of gNNs sampled from the same hospital or same country incrementally declined with a decreasing spectrum of β-lactam resistance determinants. This suggests that, in hospitals, ecological constraints exerted by antibiotic exposure obstruct the spread of isolates with a lower capacity for expressing resistance.

In contrast to our finding that carbapenemase-positive *K. pneumoniae* is transmitted predominantly via nosocomial routes, a study that also used genomic data found that carbapenemase-producing Enterobacteriaceae (including those of the *K. pneumoniae* species complex), collected both retrospectively and prospectively in four US hospitals, could mostly not be linked to transmissions[37]. We cannot explain this difference but suggest that the lack of a consistent and purpose-designed sampling frame, ensuring that isolates from consecutive clinical samples were collected, may have obscured transmission events in these US hospitals. However, this study did find geographical structuring of ST258 isolates, demonstrating local spread of this particular lineage.

We also noted that the distributions of core genome diversities differentiated same-hospital versus different-hospital and different-country isolates. This epidemiological signature provided the means to determine an optimal cut-off value for SNP differences that could be used to aid epidemiological investigations of putative institutional outbreaks of ST258/512. In addition to this, we could also provide predictions of the sensitivity and specificity of using a cut-off at different SNP thresholds (Fig. 4f). However, one major limitation of our approach is that we had no detailed epidemiological information to help confirm whether pairs of patient isolates were linked at the ward level and by overlapping treatment intervals. Nonetheless, our analysis of a large number of contemporaneous genomes sampled over different geographical scales should improve the external validity over case studies that report epidemiological cut-off values by discriminating single hospital outbreaks from a convenience sample of unrelated isolates. Furthermore, limiting epidemiologic conclusions to the number of core genome differences will also fail to detect possible plasmid transmission events. However, we argue that plasmid spread between ST258/512 isolates probably plays a relatively minor role compared with clonal spread, considering the strong relationship between the core genome diversity and geographic distance observed.

It has been previously shown that ST258 emerged from an ST11 ancestor via a large recombination event with an ST442-like strain[19]. It has been divided into two sublineages, clade I and clade II, that are differentiated by a region of recombination affecting the capsular loci[19]. ST258 is predicted to have emerged in the 1990s in the United States where it remains endemic[38]. ST258 and its derivative, ST512, are now also endemic in Israel and in some southern European countries, most notably Italy and Greece, and have caused outbreaks elsewhere[39,40]. However, the ancestral relationships between the dispersed populations of this clone were previously unclear. Here, we observed strong geographical structuring by analysing 236 ST258/512 genomes together with 415 publicly available genomes from diverse international origins. The tree structure (Fig. 5) shows a clear country-specific partitioning with US isolates at the tips of long basal branches and monophyletic short-branched expansions in different countries, suggesting that ST258/512 originated in the United States and successively spread to Mediterranean countries. Most of the high-incidence countries that reported a rapid rise in carbapenem resistance over the past 15 yr in Europe witnessed epidemic expansion of ST258/512, as illustrated in Fig. 5 (ref. [8,11,41]).

In summary, we have shown that the epidemic of carbapenem-non-susceptible *K. pneumoniae* in Europe is driven by the expansion of a small number of clonal carbapenemase-positive lineages that propagate along nosocomial transmission routes and that antibiotic use serves as a major effect modifier. Introductions of progenitor carbapenem resistance genes are rare events in the natural history of the current epidemic and, although there is no denial that gene flow between the one-health compartments exists[42], this phenomenon has had no substantial impact on the course of the epidemiology of the major carbapenemase-positive clones of *K. pneumoniae* in Europe, in the period up to 2014. Therefore, public health efforts should focus on genomic pathogen surveillance, identifying introductions of high-risk clones and their expansion early in the course of an incipient epidemic, and reinforcing the resilience of national hospital referral networks with a no-tolerance infection control policy. All data from this publication are publicly available (see Data availability) and can be used as contextual information for public health decision-making involving investigations of outbreaks and/or isolate collections. It may also serve as a benchmark of the status of carbapenem-non-susceptible *K. pneumoniae* in Europe in 2013–14, to which future survey data can be compared, facilitating critical surveillance efforts.

## Methods

**Culture, DNA extraction and Illumina sequencing.** Of the 2,301 samples submitted as *K. pneumoniae* during the EuSCAPE survey[14], 2,162 (94.0%) were prepared for whole-genome sequencing. The remaining 139 of 2,301 (6.0%) were discarded due to a lack of growth, contamination of the culture plates, or having multiple phenotypes suggestive of contamination with different species.

Of the 2,162 samples, 470 were plated on MacConkey agar (Sigma Aldrich) overnight at 37 °C and a single colony representing each sample was picked into PBS solution. For the remaining 1,692 of 2,162 samples, a storage bead was picked into 1.5 ml of low salt LB broth in a well of a 96-well S-block (Qiagen) and incubated at 37 °C overnight with shaking. The bacteria were pelleted and resuspended in 1× PBS solution. Lysozyme (Sigma Aldrich), RNase A (Invitrogen) and Proteinase K (Qiagen) were added to all of the samples, which were incubated for 1 h at 37 °C. DNA was extracted using the QIAamp 96-well kit on the QIAcube HT system (Qiagen). Isolates were sequenced using the Illumina HiSeq platform with 125 base pair (bp) paired-end reads.

**Quality control (QC) analysis of sequence data.** Trimmomatic v0.33 (ref. [43]) was used to trim the Illumina sequence reads. SPAdes v3.9.0 (ref. [44]) was used to generate de novo assemblies from the trimmed sequence reads using *k*-mer sizes of 41, 49, 57, 65, 77, 85 and 93 and with the –cov-cutoff flag set to 'auto'. QUAST v4.6.0 (ref. [45]) was used to generate assembly statistics. Isolates were discarded from future analysis if the size of the de novo assembly was outside of 5–7 Mb, the total number of contigs that were over 1 kb was greater than 1,000 and if <90% of the assembly comprised contigs greater than 1 kb.

Mash v2.0 (ref. [46]) was used to determine the similarity of each isolate to the genomes in the RefSeq bacterial database (https://www.ncbi.nlm.nih.gov/refseq/about/prokaryotes/), using the raw sequence reads, the full de novo assembly and each individual contig from the de novo assembly. Isolates were discarded from further analysis if the top species matches using either the raw reads or the full assembly were not *Klebsiella* spp. and if the combined percentage length of the individual contigs with a top match to *Klebsiella* spp. was <70.

For QC purposes only, all raw sequence reads were mapped to the NTUH K2044 reference genome using Burrows–Wheeler Aligner v0.7.12 (ref. [47]). The resulting BAM files were used to calculate the number of heterozygous SNP positions per sample using a Perl script (available at https://github.com/sanger-pathogens/vr-codebase/blob/master/scripts/het_snp_calculator.pl; compatible with Perl v5 onwards) and the mean depth of coverage. Isolates were discarded from future analysis if >1,000 heterozygous SNPs were identified or if the mean depth of coverage was <20×.

The QC results for all 2,162 sequenced isolates are available in Supplementary Tables 1–3 and Supplementary Fig. 1.

**Phylogenetic analysis of all isolates.** Prokka v1.5 (ref. [48]) was used to annotate the de novo assemblies with predicted genes. The annotated assemblies were used as input for Roary v3.11.3 (ref. [49]), which determined the core and accessory genes of the 1,717 isolates that passed the QC analysis using an identity threshold of 95% (BLASTp v2.6.0). An alignment of 2,539 core genes comprising 2,436,120 bp was generated, representing genes that are present in ≥95% of isolates from each of the four species (*K. pneumoniae* sensu stricto, *K. quasipneumoniae*, *K. variicola* and *K. quasivariicola*). Due to the high diversity, it was not feasible to remove recombined regions from the alignment before phylogenetic analysis. The variable positions were extracted from the alignment and used to generate a maximum likelihood tree with RAxML v7.0.4 using a general time reversible model with a gamma correction for site rate variation[50]. 100 random bootstrap replicates were also performed to assess the support for the nodes. Of the nodes, 57.6% and 66.7% have bootstrap support of over 0.9 and 0.75, respectively. The phylogenetic tree was visualized together with associated metadata using Microreact v5.99.0 (ref. [51]). The ANI between pairs of isolates was also determined from the alignment of 2,539 core genes using Panito v0.0.1 (available at https://github.com/sanger-pathogens/panito).

**Identification of STs and K-type.** The seven-gene ST (ref. [17]) of all the isolates was determined using the multilocus sequence typing (MLST) calling function of ARIBA v.2.6.1 (ref. [52]) and the '*Klebsiella pneumoniae*' database from PubMLST (ref. [17]). The K-type of each isolate was determined from the de novo assembly using Kaptive v0.5.1 (ref. [53]).

**Identification of carbapenemase genes.** Carbapenemase genes ($bla_{\text{KPC-like}}$, $bla_{\text{NDM-like}}$, $bla_{\text{VIM-like}}$, $bla_{\text{OXA-48-like}}$, $bla_{\text{IMP-like}}$, $bla_{\text{GES-like}}$, $bla_{\text{GIM-like}}$, $bla_{\text{AIM-like}}$, $bla_{\text{BIC-like}}$, $bla_{\text{DIM-like}}$, $bla_{\text{IMI-like}}$, $bla_{\text{NMC-A-like}}$, $bla_{\text{SIM-like}}$, $bla_{\text{SPM-like}}$, $bla_{\text{LMB-like}}$, $bla_{\text{FRI-like}}$, $bla_{\text{SME-like}}$, $bla_{\text{KHM-like}}$, $bla_{\text{BKC-like}}$, $bla_{\text{OXA-23-like}}$, $bla_{\text{OXA-51-like}}$ and $bla_{\text{OXA-58-like}}$) and their variants were detected using ARIBA v2.6.1 (ref. [52]) with a custom resistance gene database available at https://figshare.com/s/94437a301288969109c2. A carbapenemase was identified as present only if the gene coverage was at least 0.2× the minimum coverage of the seven MLST genes in each genome, to reduce false-positive calls. This cut-off allowed for variation in sequencing coverage between samples and provided maximum concordance with the previously obtained PCR results[14].

**Identification of *ESBL* genes, *AmpC* genes and porin alterations.** As with the carbapenemase genes, the presence of *ESBL* and *AmpC* genes was detected using ARIBA v2.6.1. Porin genes, *ompK35* and *ompK36*, were also detected with ARIBA and checked for completeness. Because ARIBA relies on raw reads, the de novo assemblies were also cross-checked for the presence of these *omp* genes, both using BLASTn v.2.7.0 (ref. [54]) and by extraction from the annotation files generated using Prokka (ref. [48]). Genes were considered defective if the gene was found to be absent, fragmented or interrupted, or if frameshift or stop codon mutations were visible.

**Phenotypic retesting.** Antimicrobial susceptibility testing was performed by broth microdilution[55] on 1,574 of 1,717 (91.7%) isolates that were included in the genomic analyses using lyophilized custom plates (Thermofisher Scientific).

**Determination of the gNNs and permutation analysis.** The gNN to each isolate was defined as that with the highest ANI, as calculated from the alignment of 2,539 core genes. The expected proportions of gNNs sampled from different geographic contexts, given an absence of geographic structure, were calculated by performing 100 random permutations of the hospital codes. The gNNs were recalculated on the basis of each permutation and the mean and 95% CIs were calculated from the distribution of the results.

**Phylogenetic analysis of ST258/512.** Sequence reads of isolates belonging to STs 258 and 512 (and other nested STs) were mapped to the reference genome, NJST258_1 (ref. [56]). Recombined regions were identified and removed from the pseudo-genome alignment using Gubbins v1.4.10 (ref. [57]) and the resulting pairwise SNP differences were calculated. A phylogenetic tree of the ST258/512 lineage (comprising 236 genomes from EuSCAPE and 415 publicly available genomes) was constructed using only the vertically inherited SNPs with RAxML

v7.0.4 (ref. [50]) and rooted using an ST11 outgroup isolate (EuSCAPE_PL007). Of the nodes, 52.9% and 65.7% have bootstrap support over 0.9 and 0.75, respectively. The phylogenetic tree was visualized together with associated metadata using Microreact v5.99.0 (ref. [51]) and with predicted recombined regions using Phandango v1.1.0 (ref. [58]).

**Predictions of intra versus interhospital transmission in the ST258/512 lineage.** A statistical method described by Aspbury et al.[18] was used to predict the hospital origin of each ST258/512 isolate in the EuSCAPE collection from the pseudo-genome alignment (excluding recombined regions). The code is available at https://github.com/marianne-aspbury/hospital-origin. However, we deviated from the approach described by Aspbury et al. in which a set of new isolate genomes were queried one at a time in chronological order against a genome database of older isolates and subsequently added to the database. Instead, we used all ST258/512 genomes from the EuSCAPE collection ($n=236$) in the database and removed one isolate at a time to be used as the query, regardless of its isolation date. Depending on whether the isolate was predicted to most likely be from the same or a different hospital as the actual submitting hospital, we inferred that the isolate was involved in intra or interhospital transmission.

**Statistical tests.** A $\chi^2$ test for trend was used to test for a significant downward trend in the proportion of gNNs obtained from the same hospital across the β-lactame resistance groups ((1)−(5)). The test was implemented in R v3.1.2 (https://www.r-project.org/).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All raw and assembled Illumina sequence data are available from the European Nucleotide Archive under the study accession no. PRJEB10018/ERP011196. Individual accession numbers for raw sequence data and de novo assemblies are also available in Supplementary Table 4. Phylogenetic analysis of the 1,717 *K. pneumoniae* isolates (EuSCAPE only) and the ST258/512 isolates (EuSCAPE and public data) together with all metadata and links to raw sequence data are available at the project URLs https://microreact.org/project/EuSCAPE_Kp and https://microreact.org/project/EuSCAPE_ST258 within Microreact (ref. [51]). Phylogenetic analyses and the metadata of the isolates are also available separately for each of the contributing countries in Microreact (see Supplementary Table 7 for project URLs).

Received: 16 October 2018; Accepted: 20 May 2019;
Published online: 29 July 2019

## References

1. *Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics* (World Health Organization, 2017).
2. Cassini, A. et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect. Dis.* **19**, 56–66 (2019).
3. Mathers, A. J. et al. Molecular dissection of an outbreak of carbapenem-resistant Enterobacteriaceae reveals intergenus KPC carbapenemase transmission through a promiscuous plasmid. *mBio* **2**, e00204-11 (2011).
4. 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.* **72**, 3025–3034 (2017).
5. Yigit, H. et al. Novel carbapenem-hydrolyzing β-lactamase, KPC-1, from a carbapenem-resistant strain of *Klebsiella pneumoniae*. *Antimicrob. Agents Chemother.* **45**, 1151–1161 (2001).
6. Poirel, L., Heritier, C., Tolun, V. & Nordmann, P. Emergence of oxacillinase-mediated resistance to imipenem in *Klebsiella pneumoniae*. *Antimicrob. Agents Chemother.* **48**, 15–22 (2004).
7. Yong, D. et al. Characterization of a new metallo-β-lactamase gene, $bla_{NDM-1}$, and a novel erythromycin esterase gene carried on a unique genetic structure in *Klebsiella pneumoniae* sequence type 14 from India. *Antimicrob. Agents Chemother.* **53**, 5046–5054 (2009).
8. Munoz-Price, L. S. et al. Clinical epidemiology of the global expansion of *Klebsiella pneumoniae* carbapenemases. *Lancet Infect. Dis.* **13**, 785–796 (2013).
9. Walsh, T. R., Weeks, J., Livermore, D. M. & Toleman, M. A. Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study. *Lancet Infect. Dis.* **11**, 355–362 (2011).
10. Wang, Y. et al. Comprehensive resistome analysis reveals the prevalence of NDM and MCR-1 in Chinese poultry production. *Nat. Microbiol.* **2**, 16260 (2017).
11. Albiger, B. et al. Carbapenemase-producing Enterobacteriaceae in Europe: assessment by national experts from 38 countries, May 2015. *Eurosurveillance* **20**, 17–34 (2015).
12. Kazi, M. et al. Molecular characterization of carbapenem-resistant Enterobacteriaceae at a tertiary care laboratory in Mumbai. *Eur. J. Clin. Microbiol. Infect. Dis.* **34**, 467–472 (2015).
13. Singh-Moodley, A. & Perovic, O. Antimicrobial susceptibility testing in predicting the presence of carbapenemase genes in Enterobacteriaceae in South Africa. *BMC Infect. Dis.* **16**, 536 (2016).
14. Grundmann, H. et al. Occurrence of carbapenemase-producing *Klebsiella pneumoniae* and *Escherichia coli* in the European Survey of Carbapenemase-Producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. *Lancet Infect. Dis.* **17**, 153–163 (2017).
15. 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. USA* **112**, E3574–E3581 (2015).
16. Long, S. W. et al. Whole-genome sequencing of a human clinical isolate of the novel species *Klebsiella quasivariicola* sp. nov. *Genome Announc.* **5**, e01057-17 (2017).
17. Diancourt, L., Passet, V., Verhoef, J., Grimont, P. A. & Brisse, S. Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *J. Clin. Microbiol.* **43**, 4178–4182 (2005).
18. Aspbury, M. et al. Rapid statistical methods for inferring intra- and inter-hospital transmission of nosocomial pathogens from whole genome sequence data. Preprint at https://www.biorxiv.org/content/10.1101/442319v1 (2018).
19. Chen, L., Mathema, B., Pitout, J. D. D., DeLeo, F. R. & Kreiswirth, B. N. Epidemic *Klebsiella pneumoniae* ST258 is a hybrid strain. *mBio* **5**, e01355-14 (2014).
20. Papp-Wallace, K. M., Endimiani, A., Taracila, M. A. & Bonomo, R. A. Carbapenems: past, present, and future. *Antimicrob. Agents Chemother.* **55**, 4943–4960 (2011).
21. de Man, T. J. B. et al. Genomic analysis of a pan-resistant isolate of *Klebsiella pneumoniae*, United States 2016. *mBio* **9**, e00440-18 (2018).
22. Meletis, G. Carbapenem resistance: overview of the problem and future perspectives. *Ther. Adv. Infect. Dis.* **3**, 15–21 (2016).
23. Bush, K. et al. Tackling antibiotic resistance. *Nat. Rev. Microbiol.* **9**, 894–896 (2011).
24. Murray, A. K. et al. Novel insights into selection for antibiotic resistance in complex microbial communities. *mBio* **9**, e00969-18 (2018).
25. Schwaber, M. J. et al. Containment of a country-wide outbreak of carbapenem-resistant *Klebsiella pneumoniae* in Israeli hospitals via a nationally implemented intervention. *Clin. Infect. Dis.* **52**, 848–855 (2011).
26. Falagas, M. E. et al. Risk factors of carbapenem-resistant *Klebsiella pneumoniae* infections: a matched case-control study. *J. Antimicrob. Chemother.* **60**, 1124–1130 (2007).
27. Mohan, B. et al. Fecal carriage of carbapenem-resistant Enterobacteriaceae and risk factor analysis in hospitalised patients: a single centre study from India. *Indian J. Med. Microbiol.* **35**, 555–562 (2017).
28. Bartolini, A. et al. Prevalence, molecular epidemiology and intra-hospital acquisition of *Klebsiella pneumoniae* strains producing carbapenemases in an Italian teaching hospital from January 2015 to September 2016. *Int. J. Infect. Dis.* **59**, 103–109 (2017).
29. Yusuf, E. et al. OXA-48 producing *Klebsiella pneumoniae* in a household contact of a previously infected patient: person-to-person transmission or coincidental community acquisition? *Microb. Drug Resist.* **22**, 134–136 (2016).
30. Munk, P. et al. Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nat. Microbiol.* **3**, 898–908 (2018).
31. Aanensen, D. M. et al. Whole-genome sequencing for routine pathogen surveillance in public health: a population snapshot of invasive *Staphylococcus aureus* in Europe. *mBio* **7**, e00444-16 (2016).
32. Pitart, C. et al. First outbreak of a plasmid-mediated carbapenem-hydrolyzing OXA-48 β-lactamase in *Klebsiella pneumoniae* in Spain. *Antimicrob. Agents Chemother.* **55**, 4398–4401 (2011).
33. Snitkin, E. S. et al. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci. Transl. Med.* **4**, 148rr116 (2012).
34. Zhou, K. et al. Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum β-lactamase producing ST15 *Klebsiella pneumoniae*. *Sci. Rep.* **6**, 20840 (2016).
35. Gu, D. et al. A fatal outbreak of ST11 carbapenem-resistant hypervirulent *Klebsiella pneumoniae* in a Chinese hospital: a molecular epidemiological study. *Lancet Infect. Dis.* **18**, 37–46 (2018).
36. Mathers, A. J., Peirano, G. & Pitout, J. D. D. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. *Clin. Microbiol. Rev.* **28**, 565–591 (2015).

37. Cerqueira, G. C. et al. Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. *Proc. Natl Acad. Sci. USA* **114**, 1135–1140 (2017).
38. Bowers, J. R. et al. Genomic analysis of the emergence and rapid global dissemination of the clonal group 258 *Klebsiella pneumoniae* pandemic. *PLoS ONE* **10**, e0133727 (2015).
39. Rodriguez-Zuluet, P. et al. First outbreak of KPC-3-producing *Klebsiella pneumoniae* (ST258) clinical isolates in a Mexican medical center. *Antimicrob. Agents Chemother.* **57**, 4086–4088 (2013).
40. Lopez-Cerero, L. et al. Characterisation of the first ongoing outbreak due to KPC-3-producing *Klebsiella pneumoniae* (ST512) in Spain. *Int. J. Antimicrob. Agents* **44**, 538–540 (2014).
41. Glasner, C. et al. Carbapenemase-producing Enterobacteriaceae in Europe: a survey among national experts from 39 countries, February 2013. *Eurosurveillance* **18**, 9–15 (2013).
42. Kock, R. et al. Carbapenem-resistant Enterobacteriaceae in wildlife, food-producing, and companion animals: a systematic review. *Clin. Microbiol. Infect.* **24**, 1241–1250 (2018).
43. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
44. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
45. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
46. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
48. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
49. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
50. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
51. Argimón, S. et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genom.* **2**, e000093 (2016).
52. Hunt, M. et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genom.* **3**, e000131 (2017).
53. Wyres, K. L. et al. Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb. Genom.* **2**, e000102 (2016).
54. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
55. *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically* 10th edn (CLSI, 2015).
56. Deleo, F. R. et al. Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. *Proc. Natl Acad. Sci. USA* **111**, 4988–4993 (2014).
57. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
58. Hadfield, J. et al. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* **34**, 292–293 (2018).


## Acknowledgements

We thank the Pathogen Informatics Group and Core Sequencing Facility at the Wellcome Sanger Institute for their contributions to the study. We also thank F. Bosma and M. Zigterman from the Medical Microbiology and Infection Prevention Department of the University Medical Center Groningen for their support in assembling the isolate collection. This work was funded by The Centre for Genomic Pathogen Surveillance, Wellcome Genome Campus, Wellcome (grant nos. 098051 and 099202) and the NIHR Global Health Research Unit on Genomic Surveillance of Antimicrobial Resistance (NIHR 16/136/111). The EuSCAPE project was funded by ECDC through a specific framework contract (ECDC/2012/055) following an open call for tender (OJ/25/04/2012-PROC/2012/036). This work was carried out on behalf of the ESCMID Study Group for Epidemiological Markers (ESGEM); the full list of members is provided in the Supplementary Information.

## Author contributions

H.G. and D.M.A. conceived the study. The EuSCAPE Working Group collected the bacterial isolates and epidemiological data and performed preliminary laboratory analyses. The ESGEM facilitated the training and capacity building for the collection of bacterial isolates and preliminary analyses. S.D., S.R., S.R.H., C.G., T.F., S.A., K.A., R.G., T.G., G.E., M.A., S.S., E.J.F., G.M.R., H.G. and D.M.A. performed the data analysis. S.D., S.R., E.J.F., G.M.R., H.G. and D.M.A. wrote the manuscript. All authors read and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41564-019-0492-8.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to D.M.A. or H.G.

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



## the EuSCAPE Working Group

**Andi Koraqi[15], Denada Lacej[15], Petra Apfalter[16], Rainer Hartl[16], Youri Glupczynski[17], Te-Din Huang[17], Tanya Strateva[18], Yuliya Marteva-Proevska[19], Arjana Tambic Andrasevic[20], Iva Butic[20], Despo Pieridou-Bagatzouni[21], Panagiota Maikanti-Charalampous[21], Jaroslav Hrabak[22], Helena Zemlickova[23], Anette Hammerum[24], Lotte Jakobsen[24], Marina Ivanova[25], Anastasia Pavelkovich[25], Jari Jalava[26], Monica Österblad[26], Laurent Dortet[27], Sophie Vaux[28], Martin Kaase[29], Sören G. Gatermann[30], Alkiviadis Vatopoulos[31], Kyriaki Tryfinopoulou[31], Ákos Tóth[32], Laura Jánvári[32], Teck Wee Boo[33], Elaine McGrath[33], Yehuda Carmeli[34], Amos Adler[34], Annalisa Pantosti[35], Monica Monaco[35], Lul Raka[36], Arsim Kurti[36], Arta Balode[37], Mara Saule[37], Jolanta Miciuleviciene[38], Aiste Mierauskaite[38], Monique Perrin-Weniger[39], Paul Reichert[39], Nina Nestorova[40], Sonia Debattista[40], Gordana Mijovic[41], Milena Lopicic[41], Ørjan Samuelsen[42], Bjørg Haldorsen[42], Dorota Zabicka[43], Elzbieta Literacka[44], Manuela Caniça[45], Vera Manageiro[45], Ana Kaftandzieva[46], Elena Trajkovska-Dokic[46], Maria Damian[47], Brandusa Lixandru[47], Zora Jelesic[48], Anika Trudic[49], Milan Niks[50], Eva Schreterova[51], Mateja Pirs[52], Tjasa Cerar[52], Jesús Oteo[53], Belén Aracil[53], Christian Giske[54], Karin Sjöström[55], Deniz Gür[56], Asli Cakar[56], Neil Woodford[57], Katie Hopkins[57], Camilla Wiuff[58] and Derek J. Brown[59]**

[15]University Hospital Center Mother Theresa, Tirana, Albania. [16]Elisabethinen Hospital Linz, Linz, Austria. [17]CHU Dinant-Godinne, UCL Namur, Namur, Belgium. [18]Faculty of Medicine, Medical University of Sofia, Sofia, Bulgaria. [19]Alexandrovska University Hospital, Sofia, Bulgaria. [20]University Hospital for Infectious Diseases, Zagreb, Croatia. [21]Nicosia General Hospital, Nicosia, Cyprus. [22]Faculty of Medicine in Plzen, Charles University in Prague, Plzen, Czech Republic. [23]National Institute of Public Health, Prague, Czech Republic. [24]Statens Serum Institut, Copenhagen, Denmark. [25]East Tallinn Central Hospital, Tallinn, Estonia. [26]National Institute for Health and Welfare, Turku, Finland. [27]Bicêtre Hospital, Le Kremlin-Bicêtre, France. [28]Institut de Veille Sanitaire, Saint-Maurice, France. [29]Universitätsmedizin Göttingen, Göttingen, Germany. [30]Ruhr-University Bochum, Bochum, Germany. [31]National School of Public Health, Athens, Greece. [32]National Center for Epidemiology, Budapest, Hungary. [33]Galway University Hospitals, Galway, Ireland. [34]Ministry of Health, Tel Aviv, Israel. [35]Istituto Superiore di Sanità, Rome, Italy. [36]National Institute of Public Health of Kosovo, Prishtina, Kosovo. [37]Pauls Stradins Clinical University Hospital, Riga, Latvia. [38]National Public Health Surveillance Laboratory, Vilnius, Lithuania. [39]Laboratoire National De Sante, Düdelingen, Luxembourg. [40]Mater Dei Hospital, Msida, Malta. [41]Institute of Public Health, Podgorica, Montenegro. [42]University Hospital of North Norway, Tromsø, Norway. [43]Narodowy Instytut Lekow, Warsaw, Poland. [44]National Medicines Institute, Warsaw, Poland. [45]National Institute of Health Dr. Ricardo Jorge, Lisbon, Portugal. [46]Medical Faculty, Institute of Microbiology and Parasitology, Skopje, Republic of Macedonia. [47]Cantacuzino National Research Institute, Bucharest, Romania. [48]Institute for Public Health of Vojvodina, Novi Sad, Serbia. [49]Institute for Pulmonary Diseases of Vojvodina, Sremska Kamenica, Serbia. [50]Public Health Authority of the Slovak Republic, Bratislava, Slovakia. [51]Faculty of Medicine, P. J. Šafárik University, Kosice, Slovakia. [52]Institute of Microbiology and Immunology, Ljubljana, Slovenia. [53]Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain. [54]Karolinska Institutet, Stockholm, Sweden. [55]Public Health Agency of Sweden, Stockholm, Sweden. [56]Hacettepe University, Ankara, Turkey. [57]Reference Microbiology Services, Public Health England, London, UK. [58]Sydvestjysk Hospital, Esbjerg, Denmark. [59]Scottish Microbiology Reference Laboratories, Glasgow, UK.

nature research

Corresponding author(s): Hajo Grundmann & David Aanensen

Last updated by author(s): Apr 29, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☐ | ☒ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software was used. |
| Data analysis | Trimmomatic v0.33<br>SPAdes v3.9.0<br>QUAST v4.6.0<br>MASH v2.0<br>Burrows Wheeler Aligner v0.7.12<br>https://github.com/sanger-pathogens/vr-codebase/blob/master/scripts/het_snp_calculator.pl<br>Prokka v1.5<br>Roary v3.11.3<br>RAxML v7.0.4<br>https://github.com/sanger-pathogens/panito<br>Ariba v.2.6.1<br>Kaptive (no version available)<br>BLASTn v2.7.0<br>Gubbins v1.4.10<br>Microreact (www.microreact.org)<br>Phandango v1.1.0<br>R v3.1.2<br>https://github.com/marianne-aspbury/hospital-origin |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw and assembled Illumina sequence data are available from the European Nucleotide Archive (ENA) under the study accession number, PRJEB10018/ ERP011196. Individual accession numbers for raw sequence data and de novo assemblies are also available in Supplementary Table 4. Phylogenetic analyses, metadata and links to raw sequence data are available in Microreact (see Data Availability).

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | Genomic analysis of 1717 isolates of Klebsiella pneumoniae submitted during the European Survey of Carbapenemase-Producing Enterobacteriaceae (EuSCAPE) (Grundmann et al. 2017). |
| Research sample | 1717 genomes of K. pneumoniae submitted during the EuSCAPE survey between November 2013 and May 2014 from 244 hospitals in 32 countries across Europe and neighbouring countries. |
| Sampling strategy | All isolates collected during the EuSCAPE survey were used, except where there was a lack of growth upon re-culture, contamination of the culture plates, or multiple phenotypes suggestive of contamination with other species. |
| Data collection | Isolates were collected by individual hospitals participating in the EuSCAPE survey, sent to the national expert laboratories for each country, and then sent to Hajo Grundmann to create a central collection. |
| Timing and spatial scale | Between November 2013 and May 2014, each participating hospital was asked to provide the first ten carbapenem non-susceptible isolates of either K. pneumoniae or E. coli followed by a susceptible comparator isolate. Each hospital could submit a maximum of 20 isolates. |
| Data exclusions | Samples that were submitted during EuSCAPE that were found to have a lack of growth, contamination in the culture plates, or multiple phenotypes, were not sequenced. These exclusion criteria were pre-established. Furthermore, sequenced samples that failed to pass one or more of a number of QC criteria were discarded. The criteria taken into account were genome assembly size, number of contigs, number of heterozygous SNPs, species identification and mapping coverage. The particular exclusion criteria used were established upon data analysis. |
| Reproducibility | Reproducibility is inherent to short-read sequencing provided that the sequencing coverage is high enough and the sequenced samples are not contaminated. To ensure reproducibility, we therefore applied rigorous QC filters to the sequence data as described in the Methods. Furthermore, we found that the carbapenemase content of the samples, as determined from the sequence data, was highly concordant (98.3-99.0% for individual genes) with the results obtained by PCR from the national expert laboratories. |
| Randomization | Hospitals and laboratories that submitted isolates during the EuSCAPE survey were selected for their geo-demographic representativeness within participating countries (see Grundmann et al. 2017). All submitted K. pneumoniae isolates were included in this study except for those failing QC criteria (i.e. random sampling did not apply). |
| Blinding | Blinding was not relevant because sources of bias that could be avoided by blinding did not play a role in our investigation. |

Did the study involve field work? ☐ Yes ☒ No

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |