AAC Journals.ASM.org

Genomic Epidemiology of *Klebsiella pneumoniae* in Italy and Novel Insights into the Origin and Global Evolution of Its Resistance to Carbapenem Antibiotics

Stefano Gaiarsa,[a,b] Francesco Comandatore,[b,c] Paolo Gaibani,[d] Marta Corbella,[a,e] Claudia Dalla Valle,[a] Sara Epis,[b] Erika Scaltriti,[f] Edoardo Carretto,[g] Claudio Farina,[h] Maria Labonia,[i] Maria Paola Landini,[d] Stefano Pongolini,[f] Vittorio Sambri,[j] Claudio Bandi,[b] Piero Marone,[a] Davide Sassera[c]

Microbiology and Virology Unit, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy[a]; Dipartimento di Scienze Veterinarie e Sanità Pubblica (DIVET), Università degli Studi di Milano, Milan, Italy[b]; Dipartimento di Biologia e Biotecnologie, Università degli Studi di Pavia, Pavia, Italy[c]; Unit of Clinical Microbiology, St. Orsola-Malpighi University Hospital, Bologna, Italy[d]; Biometric and Medical Statistics Unit, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy[e]; Sezione Diagnostica di Parma, Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna (IZSLER), Parma, Italy[f]; Clinical Microbiology Laboratory, IRCCS Arcispedale S. Maria Nuova, Reggio Emilia, Italy[g]; Microbiology Institute, AO Papa Giovanni XXIII, Bergamo, Italy[h]; Dipartimento di Diagnostica di Laboratorio e Trasfusionale, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy[i]; Unit of Clinical Microbiology, The Greater Romagna Area-Hub Laboratory, Pievesestina, Italy[j]

***Klebsiella pneumoniae* is at the forefront of antimicrobial resistance for Gram-negative pathogenic bacteria, as strains resistant to third-generation cephalosporins and carbapenems are widely reported. The worldwide diffusion of these strains is of great concern due to the high morbidity and mortality often associated with *K. pneumoniae* infections in nosocomial environments. We sequenced the genomes of 89 *K. pneumoniae* strains isolated in six Italian hospitals. Strains were selected based on antibiotypes, regardless of multilocus sequence type, to obtain a picture of the epidemiology of *K. pneumoniae* in Italy. Thirty-one strains were carbapenem-resistant *K. pneumoniae* carbapenemase producers, 29 were resistant to third-generation cephalosporins, and 29 were susceptible to the aforementioned antibiotics. The genomes were compared to all of the sequences available in the databases, obtaining a data set of 319 genomes spanning the known diversity of *K. pneumoniae* worldwide. Bioinformatic analyses of this global data set allowed us to construct a whole-species phylogeny, to detect patterns of antibiotic resistance distribution, and to date the differentiation between specific clades of interest. Finally, we detected an ~1.3-Mb recombination that characterizes all of the isolates of clonal complex 258, the most widespread carbapenem-resistant group of *K. pneumoniae*. The evolution of this complex was modeled, dating the newly detected and the previously reported recombination events. The present study contributes to the understanding of *K. pneumoniae* evolution, providing novel insights into its global genomic characteristics and drawing a dated epidemiological scenario for this pathogen in Italy.**

Multidrug resistance is currently a matter of concern worldwide. At the end of the 1970s, most *Escherichia coli* and *Klebsiella pneumoniae* strains encoded ampicillin-hydrolyzing β-lactamases, making it necessary to use third-generation cephalosporins. In the early 1980s, the first cases of resistance to these novel antibiotics were reported in *Enterobacteriaceae* (1) and were caused by genes classified as ESBL (extended-spectrum beta-lactamases). In 1985, the United States Food and Drug Administration approved the commercialization of imipenem, a molecule that showed activity against ESBL producers. This drug, and similar compounds that quickly followed (i.e., carbapenems), then were introduced into clinical practice and widely used.

In 2001, Yigit and colleagues reported a *K. pneumoniae* strain isolated in 1996 that exhibited resistance to the carbapenems imipenem and meropenem (2). The gene responsible for the resistance was identified as a group 2f, class A, carbapenem-hydrolyzing beta-lactamase, named *Klebsiella pneumoniae* carbapenemase 1 (KPC-1). Since its discovery, carbapenem resistance caused by the bla_{KPC} gene has been reported increasingly in *K. pneumoniae* isolates, initially moving through the northeastern states (3, 4) and quickly becoming the most frequently found carbapenemase in the United States (5). The spread of KPC then continued, with reports from different countries appearing ceaselessly, to the point that today this is regarded as a worldwide issue (6).

The bla_{KPC} gene is carried by a plasmid; thus, horizontal transfer between various *K. pneumoniae* strains, as well as other bacterial species, could be expected and was extensively reported (7–9). Nevertheless, most of the clinical reports to date have been caused by *K. pneumoniae* isolates belonging to clonal complex 258 (CC258) (10). This complex comprises sequence type 258 (ST258) and single-allele mutant STs based on multilocus sequence typing (MLST), such as ST11 and ST512. These epidemiological data suggest a dissemination starting from a single ances-

Received 4 September 2014 **Returned for modification** 26 September 2014
Accepted 26 October 2014

Accepted manuscript posted online 3 November 2014

Citation Gaiarsa S, Comandatore F, Gaibani P, Corbella M, Dalla Valle C, Epis S, Scaltriti E, Carretto E, Farina C, Labonia M, Landini MP, Pongolini S, Sambri V, Bandi C, Marone P, Sassera D. 2015. Genomic epidemiology of *Klebsiella pneumoniae* in Italy and novel insights into the origin and global evolution of its resistance to carbapenem antibiotics. Antimicrob Agents Chemother 59:389–396. doi:10.1128/AAC.04224-14.

Address correspondence to Davide Sassera, davide.sassera@unipv.it.

S.G. and F.C. contributed equally.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AAC.04224-14.

doi:10.1128/AAC.04224-14

tor and that CC258 presents a genomic background that is favorable both to the acquisition of plasmids bearing the bla_{KPC} gene and to the clonal spread in nosocomial environments. In 2014, Deleo and colleagues (11) presented a phylogenomic study on 85 *K. pneumoniae* isolates belonging to CC258, detecting two subclades and concluding that an ~215-kb recombination event was at the origin of the differentiation between the two. A second comparative genomic analysis, presented by Chen and colleagues (12), detected an ~1.1-Mb recombination between an ST11 recipient and an ST442 donor as the event that originated the present ST258 strain.

Since the first finding of circulation of ESBL-producing *K. pneumoniae* in Italy in 1994, a rapid and extensive dissemination of different types of ESBLs has been reported (13–15). More recently, the first Italian KPC-positive *K. pneumoniae* strain, belonging to ST258, was isolated in a hospital in Florence in 2008 from an inpatient with a complicated intra-abdominal infection (16). Since then, the diffusion of carbapenemase-producing *K. pneumoniae* in Italy has been extremely rapid and characterized mainly by isolates of CC258 (i.e., ST258 and ST512) (17–19). ST512 in particular, first reported in Israel in 2006 (20), has been spreading in southern Europe and South America (11, 19). The sporadic detection of isolates belonging to other STs (e.g., ST101 and ST147) also have characterized the epidemiology of KPC *K. pneumoniae* in Italy (19).

The aim of this study was to evaluate the geographic and phylogenetic distribution of *K. pneumoniae* isolates of different antibiotypes, both at a national and a global scale. Thus, we sequenced and analyzed the genomes from 89 *K. pneumoniae* strains, collected in six Italian hospitals from 2006 to 2013, without any *a priori* knowledge of the sequence type. We compared this national collection to all of the *K. pneumoniae* genomes available from worldwide isolations to obtain insights into both the Italian epidemiology and the global structure of the species.

MATERIALS AND METHODS

Strain sampling. Eighty-nine nonduplicate *K. pneumoniae* strains, collected from six different Italian hospitals, were included in this study without prior knowledge of the sequence type. Thirty-one were KPC producers, as demonstrated using phenotypical tests (positivity with disk diffusion synergy testing using a meropenem disk alone and in combination with aminophenylboronic acid) (21) and/or genotypical analysis (in-house methods based on reference 22); 29 were ESBL producers, as demonstrated using the procedure recommended by the CLSI (23), while 29 were susceptible to third-generation cephalosporins and carbapenems. Throughout this work, we refer to this last group of isolates as susceptible. Antimicrobial susceptibility testing was performed using a Vitek2 automated system (bioMérieux), and MICs were interpreted by following the European Committee on Antimicrobial Susceptibility Testing guidelines (24). The list of isolates, year, location of isolation, sequence type, and presence of selected antibiotic resistance genes are reported in Table S1 in the supplemental material.

Genome sequences. DNA was extracted using a QIAamp DNA mini-kit (Qiagen) by following the manufacturer's instructions. Whole genomic DNA was sequenced using an Illumina Miseq platform with a 2 by 250 paired-end run after Nextera XT paired-end library preparation. On 24 March 2014, sequences of draft and complete genomes of *K. pneumoniae* were retrieved from the NCBI ftp site, while sequencing reads of the isolates sequenced by Deleo and coworkers (11) were retrieved from the sequence read archive (SRA) database (accession no. SRP036874).

Genome assembly and retrieval. Sequencing reads from the isolates obtained in this study were assembled using MIRA 4.0 software (25) with accurate *de novo* settings. Assembled genomes are now publicly available under Bioproject (EMBL project B6543). Reads retrieved from the SRA database were checked and filtered for sequencing quality using an in-house script and then assembled using Velvet (26) with a K-mer length of 35 and automatic detection of average expected coverage and low coverage threshold.

Resistance profile and MLST determination. The MLST profile was obtained *in silico* by searching the characterizing gene variants on each genome, using an in-house Python script. The antibiotic resistance profile was determined using a BLAST search on a gene database comprising all of the most common resistance genes associated with resistance to beta-lactams, including ESBL- and KPC-producing phenotypes.

Core SNP detection and phylogeny. Single-nucleotide polymorphisms (SNPs) were detected using an in-house pipeline based on Mauve software (27), using the NJST258_1 complete genome as a reference. Each genome was individually aligned to the reference, and alignments were merged with in-house scripts. Core SNPs were defined as single-nucleotide mutations flanked by identical bases present in all of the analyzed genomes. The core SNP alignment was used to perform a phylogenetic analysis using the software RAxML (28) with a generalized time-reversible (GTR) model and 100 bootstraps. The same phylogenetic approach was used to perform the analysis on three core SNP sub-data sets (i.e., nonrecombined regions and two distinct putatively recombined regions).

Recombination. We divided the genome alignment in 5,264 windows of 1,000 nucleotides (nt) each and calculated core SNP frequency in each window for each genome, generating a matrix. The software R then was used to generate a heatmap of SNP frequency. The newly characterized strain 46AVR was used as a reference for plotting SNPs, being a member of the sister group to CC258. In parallel, we created a sub-data set of 174 CC258 genomes and 13 closely related *K. pneumoniae* genomes, removing genomes of isolates distant from the CC258 clade ($n = 103$) and the genomes within CC258 that exhibited extremely limited variability ($n = 29$), such as all but one of those obtained from single outbreaks. The choice of using a relatively large number of non-CC258 genomes ($n = 13$) was made in order to allow the detection of recombination events common to the whole clonal complex. We used this sub-data set of core SNPs in 187 genomes to perform a recombination detection analysis using the software BRATnextgen (29) with 100-iteration analysis, using 100 replicates for statistical significance.

Analysis of the recombined region. A database was created collecting protein sequences of factors previously reported to be involved in virulence and antibiotic resistance. We collected sequences from the Comprehensive Antibiotic Resistance Database (CARD) (30) and from the Antibiotic Resistance Genes Database (ARDB) (31), from proteins involved in the biosynthesis of lipopolysaccharides (LPS) and polymyxin resistance, and from the most common virulence factors and siderophores found in Gram-negative bacteria (obtained from the NCBI site). Finally, we added to our manually designed database all *K. pneumoniae* proteins described as potential virulence or resistance factors in the work by Lery and colleagues (32). Gene sequences present in the novel putative recombined region were extracted from the genome of strain NJST258_1 using an in-house Python script. Correspondence between proteins in our database and genes in the recombined region was tested using a TBLASTN search, selecting genes covering at least 75% of the database sequence with a minimum of 75% identity. Results then were manually checked (see Table S2 in the supplemental material for a complete list).

Molecular clock. We created a sub-data set of 174 CC258 genomes and 3 closely related *K. pneumoniae* genomes (used as outgroups), removing genomes of isolates distant from the CC258 clade ($n = 113$) and the genomes within CC258 that exhibited extremely limited variability ($n = 29$), such as all but one of those obtained from single outbreaks. We used the software BEAST (33) on the core SNP alignment of the 177-genome sub-data set after removing SNPs located in the potentially recombined regions. BEAST parameters used were the following: uncorrelated log-normal relaxed clock with the GTR model, with no correction for site rate

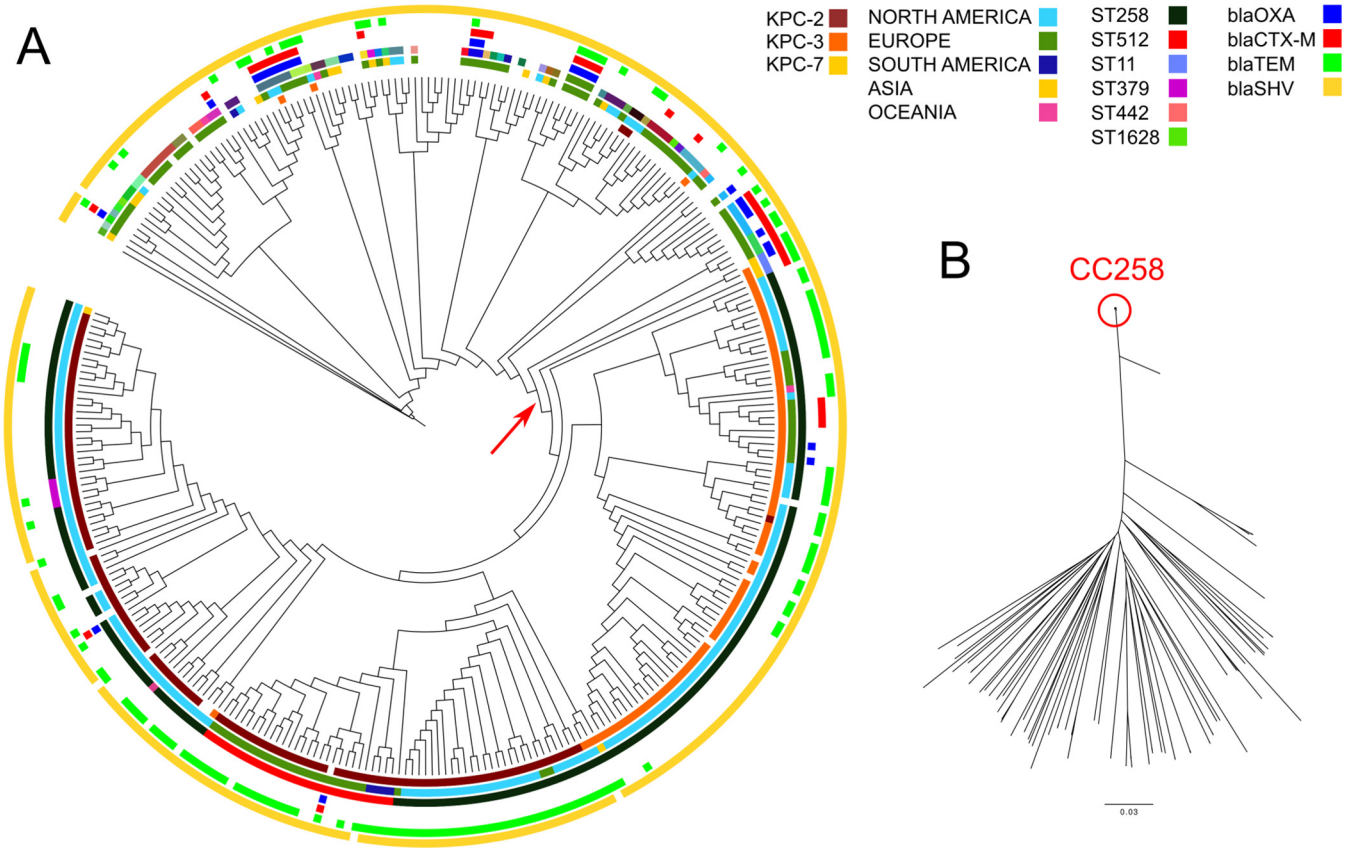

FIG 1 Maximum likelihood phylogeny of *Klebsiella pneumoniae*, based on 319 genomes. The phylogeny was reconstructed starting from an alignment of 94,812 core SNPs, using the software RAxML with a generalized time-reversible (GTR) model and 100 bootstraps, which are not shown for the sake of figure clarity. (A) Circular representation of the phylogeny, obtained using iTOL (itol.embl.de), ignoring branch length. Color circles indicate, from the innermost to the outermost, presence/absence of KPC variants, geographic location in terms of continents, ST based on multilocus sequence typing, and presence in the genome of genes from four beta-lactamase families. The red arrow indicates the origin of the clonal complex 258 clade. (B) Unrooted representation of the phylogeny showing the branch lengths, highlighting the genetic uniformity of clonal complex 258.

heterogeneity according to analyses performed in similar scenarios (34). The analysis was run for 1,000,000,000 steps, and at every 10,000 steps samples were taken. We discarded 250,000,000 steps as burn-in. The program TRACER (http://beast.bio.ed.ac.uk/tracer/) was used to evaluate the convergence of the analysis.

RESULTS

Sampling and genome sequencing. Eighty-nine *K. pneumoniae* strains were collected in six Italian hospitals, chosen based on antibiotypes regardless of sequence type, which was determined only afterwards. The data set was composed of 31 KPC producers, 29 ESBL producers, and 29 strains susceptible to carbapenems and third-generation cephalosporins, here referred to as susceptible. The genome of each of the 89 isolates was sequenced and assembled (average genome size, 5,551,959 nt; average N50, 154,414 nt; average coverage, 76.46×). All of the available *K. pneumoniae* genome sequences and reads then were retrieved from the databases ($n = 230$) to create a global data set of 319 *K. pneumoniae* genomes. All genomes in the data set were screened for genes responsible for KPC and beta-lactam resistance phenotypes, as well as for all MLST genes. A total of 55 different MLST profiles were detected, eight of which were novel; thus, they were submitted to the curators of the *K. pneumoniae* MLST database (35). Each of the eight new profiles was represented by a single newly sequenced Italian isolate (7 susceptible, 1 ESBL producer). Two of these isolates also presented a single novel allele, one for the gene *rpoB* and one for the gene *infB*. See Table S1 in the supplemental material for a list of all of the isolates sequenced in this study and their main characteristics.

Global SNP phylogeny. We used a maximum likelihood phylogenomic approach based on core SNPs to elucidate the relationships within the global genome data set comprising the newly sequenced isolates and the *K. pneumoniae* genome sequences available in the database. The presence of antibiotic resistance genes was mapped on the resulting phylogenetic tree, obtained from an alignment of 94,812 core SNPs (Fig. 1). This revealed that 97% of all KPC *K. pneumoniae* strains sequenced to date, regardless of the location of isolation, belong to a well-supported clade, corresponding to the complex CC258. On the other hand, the phylogenomic analysis showed that the isolates encoding common beta-lactam resistance genes (bla_{SHV} family, bla_{TEM} family, bla_{OXA} family, and bla_{CTX-M} family) are widespread along the tree and belong to various STs (both inside and outside CC258), with no sign of clustering. In fact, the 141 isolates encoding bla_{TEM} belong to 24 different STs, the 26 isolates encoding bla_{OXA} belong to 11 different STs, and the 37 isolates encoding bla_{CTX-M} belong to 16 different STs.

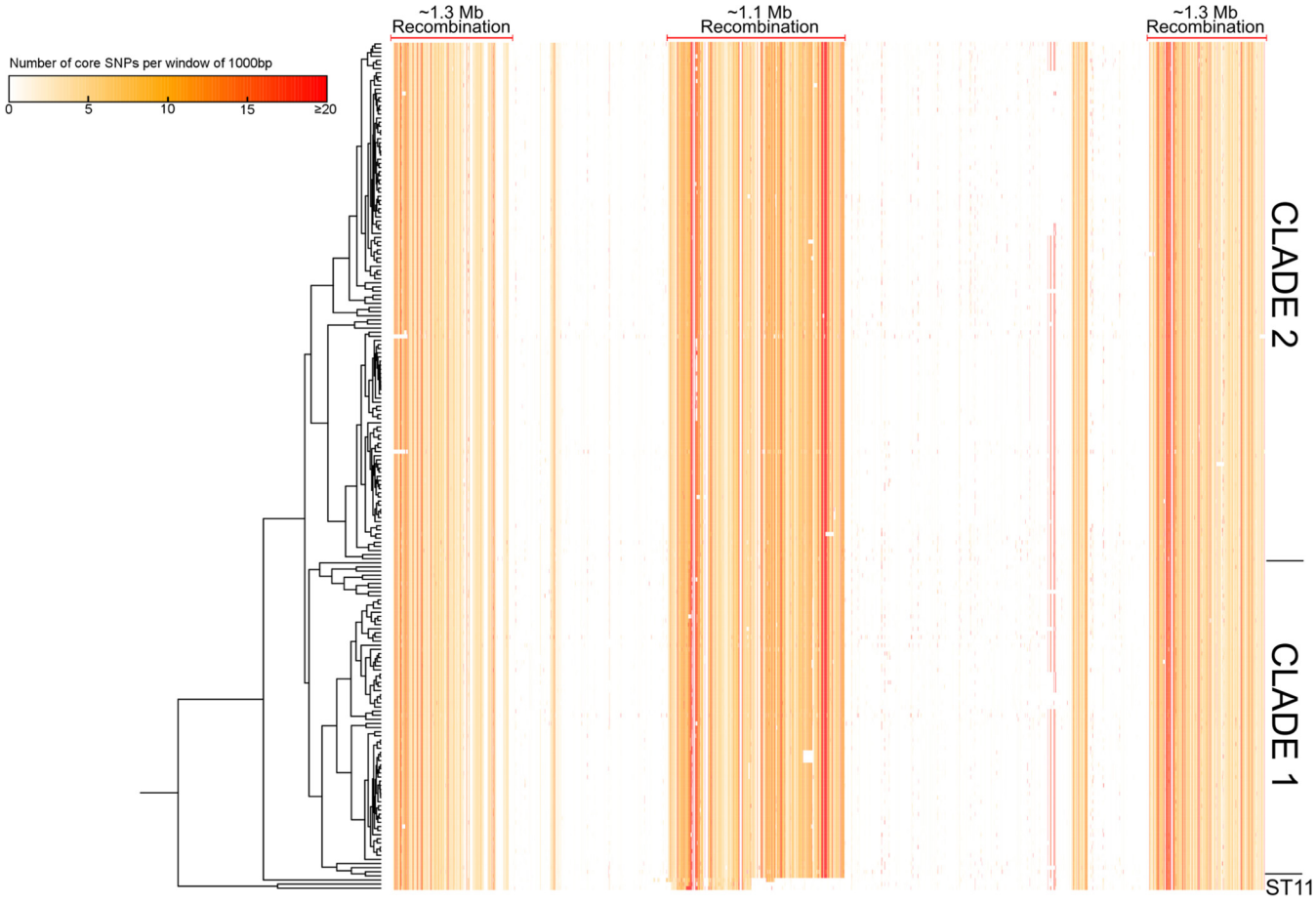

FIG 2 Uneven clustering of core SNPs in the clonal complex 258 clade. The phylogenetic reconstruction of the 206 representatives of the clonal complex 258 clade is shown on the left, while the core SNP frequency is shown on the right in shades of red, representing the number of core SNPs per 1,000-bp window for each genome. Detected recombinations are indicated at the top of the figure, and main clades of the clonal complex are indicated on the right side of the figure.

Phylogeny excluding potentially recombined regions. In a recent work by Castillo-Ramirez and coworkers (34), high-density SNPs clusters with a low ratio of nonsynonymous to synonymous evolutionary changes (*dN*/*dS*) in closely related bacterial genomes were suggested to be indicators of recombination events. Thus, we evaluated the distribution of SNPs on the genome data set, detecting a highly uneven distribution in the genomes of CC258 isolates, as most core SNPs clustered into two main regions. The first region is located between positions 1,675,550 and 2,740,033, while the second comprises the origin of replication and spans from 4,554,906 to 629,621 in strain NJST258_1 (Fig. 2) (for the distribution of core SNPs on the whole data set of 319 genomes, see Fig. S1 in the supplemental material). To further analyze the possible presence of recombination events in CC258, we used the software BRATnextgen (29), specifically intended for this purpose, on a reduced data set of 187 genomes of CC258 and closely related strains. This analysis (see Fig. S2) confirmed the presence of the two main recombination events, additionally indicating in what position of the phylogeny they could have occurred. The first event was placed between the entire CC258 clade and the non-KPC external isolates of different STs, while the second was between the outermost strains of ST11 and the inner CC258 clade. Details on these recombined regions are presented in the following paragraph.

We removed the two putative recombined regions from the core SNPs data set of 319 *K. pneumoniae* genomes and performed a phylogenetic analysis on the remaining 55,368 core SNPs. The resulting tree (see Fig. S3 in the supplemental material) is largely consistent with the one generated from the initial data set, confirming the widespread distribution of susceptible and ESBL isolates and the presence of the highly supported KPC CC258 clade. Indeed, both the analysis on all core SNPs and the one performed by removing recombining sites agree in clustering 97% of all KPC *K. pneumoniae* isolates sequenced in a well-supported clade (Fig. 1; also see Fig. S3). This monophyletic clade comprises 203 strains from Asia, Europe, Oceania, and North and South America, with isolation dates ranging from 2002 to 2013; 193 of these (95%) present the bla_{KPC} gene. Most isolates of this clade belong to ST258 ($n = 167$), but 4 other sequence types are present (i.e., ST11, SST379, ST418, and ST512), all single-nucleotide variants of ST258; thus, they belong to CC258. The second most common sequence type in the CC258 clade is 512, represented by 28 isolates that form a single monophyletic subgroup, located within the ST258 diversity. Interestingly, 24 of these 28 have been isolated in Italy, mostly in this study ($n = 19$) but also in previous works (18, 36). Within the CC258 clade, two main highly supported distinct subclades are detectable, comprising the vast majority of the genomes. Three additional CC258 genomes are located in the tree as

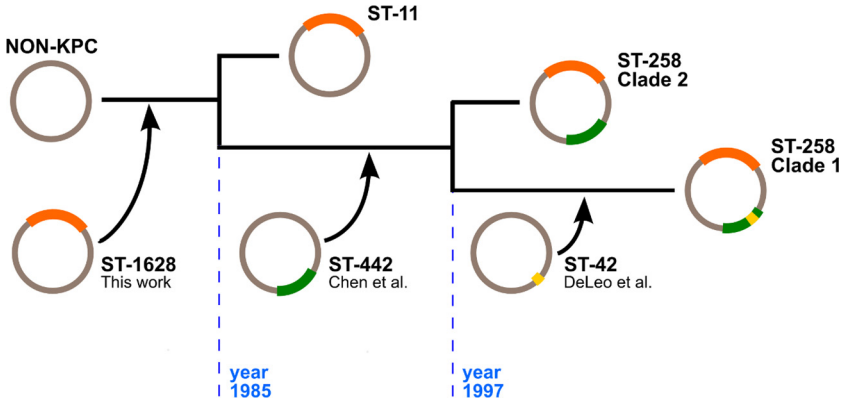

FIG 3 Hypothesis of recombinations occurring in the clonal complex 258 clade. Schematic representation based on the results of the analyses presented. Main nodes of interest are shown, highlighting the hypothesized pattern of three recombination events leading to the current state of clonal complex 258. Dates are inferred based on the molecular clock analysis depicted in Fig. 4.

sister groups of the two main clades, and all are representatives of ST11, again a single-nucleotide variation of ST258. The existence of the two main CC258 subclades was reported previously, and a single recombination event was proposed to be the cause of the differentiation between the two (11), while a subsequent work suggested multiple recombination events (37).

Analysis of recombined regions. As described above, the SNP clustering analysis detected high SNP concentrations in two large genomic regions (Fig. 2). The smaller of the two is highly congruent with the ~1.1-Mb recombination found by Chen and colleagues (12), which represents the major evolutive change between the members of ST11 and those in the 2 main subclades of the CC258 clade. Chen and colleagues found this region to be most similar to the corresponding region of isolate Kp13 of ST442 and suggested a recombination event, with the donor strain being a close relative of Kp13. Thus, we investigated whether a recombination event is at the origin of the second, newly detected, highly mutated genomic region, located from positions 4,554,906 to 629,621. We performed a phylogenetic analysis, including all of the 319 *K. pneumoniae* genomes examined in this work, on the core SNPs located in this region and in parallel on the core SNPs located in the ~1.1-Mb region. The phylogenetic analysis of the novel ~1.3-Mb region (see Fig. S4 in the supplemental material) confirms the recombination hypothesis, as the topology of the resulting tree clearly shows that Italian isolate 67BO, of the newly described ST1628, is the sister taxon to the entire CC258 clade, suggesting that the donor was related to this isolate. The phylogenetic tree obtained from the ~1.1-Mb recombined region (see Fig. S5) confirms the published results, clustering the donor Kp13 as a sister taxon of the CC258 clade, with the exclusion of the outermost ST11 isolates. Thus, we propose an updated scenario in which a first recombination event gave origin to the first CC258 strains (represented by ST11), a second recombination subsequently originated ST258, and a third, smaller recombination initiated the split between the two main ST258 subclades (Fig. 3).

In order to investigate the potential effect of the newly discovered recombination on the phenotype of the acceptor CC258, the presence of genes possibly related to antibiotic resistance and virulence was investigated in the corresponding region of the genome of strain NJST258_1, using a specifically designed database (see Materials and Methods). Interestingly, 51 genes were detected in the region (see Table S2 in the supplemental material), grouped in three main categories: LPS modification (such as the *waa* operon), bacterial efflux transporters (i.e., efflux pumps and permeases), and regulators (e.g., *ompR-envZ* operon) (see Discussion for an analysis of the detected genes).

Molecular clock. In order to date the origin of the CC258 clade and its subclades, we performed a molecular clock analysis using the software BEAST (33). We produced a reduced data set of 3,615 core SNPs present in a selected subset of taxa (174 CC258 and three closely related non-KPC *K. pneumoniae* genomes used as outgroups), derived from the previously filtered data set, in which the potentially recombined regions of the genome were excluded (Fig. 4). Compared with the dates indicated in published reports, our estimations appear to be fairly accurate. For example, the molecular clock analysis dates the appearance of ST512 to 2007, close to the first report in Israel, i.e., 2006 (20). Additionally, the molecular clock analysis dates the radiation of American and European ST258 isolates to 1997, a time point coherent with the first report of KPC-bearing *K. pneumoniae*, i.e., 1996 (2). Thus, our calibration of the evolutionary rate, superimposed on the phylogenetic tree (Fig. 4), could be used to infer unavailable dates on the global pandemic of CC258 *K. pneumoniae*. See Discussion for further discussion of the estimated dates.

Italian strains. The structure of the phylogenomic tree allows us to depict the scenario of the epidemiology of *K. pneumoniae* in Italy (Fig. 1 and 4). While susceptible and ESBL Italian strains are homogeneously distributed on the tree and belong to a number of different STs (24 and 15, respectively), all of the KPC strains sequenced in Italy belong to CC258, indicating a strong epidemiological prevalence of this clonal complex in the Italian hospitals. Within CC258, Italian isolates are well clustered in four monophyla, three composed mostly of isolates sequenced in this study and one encompassing two isolates from a previous study (38). Of the four Italian CC258 monophyla, the one including the most isolates is composed solely of ST512 ($n = 24$), confirming the multiple reports that indicate this ST as being of great epidemiological importance, at least in this country. Our phylogenetic analysis clearly indicates that this ST512 monophylum is found within the diversity of ST258.

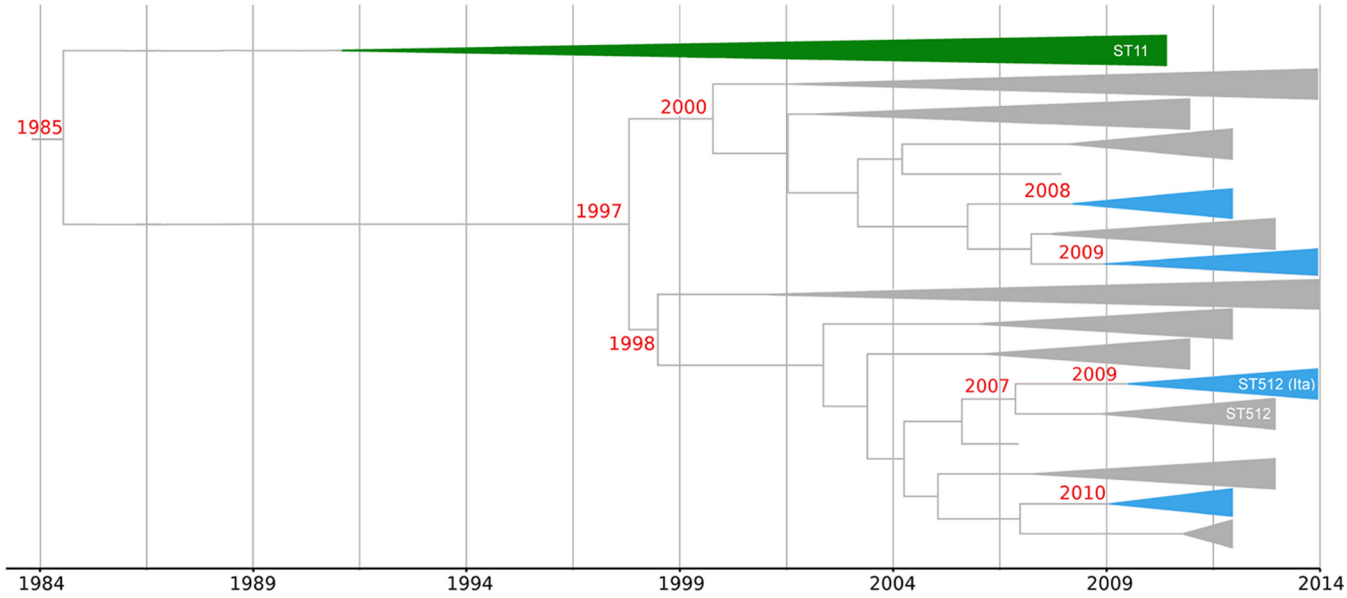

FIG 4 Estimation of divergence times in clonal complex 258. A schematic version of the time-scaled phylogeny was obtained using BEAST software with an uncorrelated log-normal relaxed clock and GTR model with no correction for site rate heterogeneity. The analysis was run for 1,000,000,000 steps, with sampling every 10,000 steps and 25% burn-in. The Italian monophyla are highlighted in blue, while the sequence type 11 (ST11) Asian clade is highlighted in green. All of the phyla with no indication of ST are comprised mainly of isolates of ST258. The dates indicated in the figure, for selected branches and nodes, were inferred from the analysis described above; for a comparison with the dates of isolation of strains, see Discussion.

DISCUSSION

***Klebsiella pneumoniae* in Italy.** We sequenced the genomes of 89 *K. pneumoniae* strains isolated in Italy, among them 31 KPC producers, 29 ESBL producers, and 29 strains susceptible to beta-lactams and carbapenems. Based on our phylogenomic analysis, the 29 genomes from susceptible *K. pneumoniae* strains isolated in Italy are scattered along the tree, showing no evident sign of clusterization. The sequencing of these isolates allowed us to expand the known diversity of the *K. pneumoniae* species, detecting seven novel MLST profiles and contributing to the overall robustness of current and future phylogenetic analyses. The genomes obtained from 29 ESBL isolates also show a considerable diversity, as they are distributed on the phylogenetic tree and belong to 15 different STs, among them a newly found ST.

Regarding KPC isolates, all Italian sequenced strains are found in CC258. Since no *a priori* selection of STs was performed, this result indicates a strong prevalence of CC258 among KPC *K. pneumoniae* isolates in Italy, even though isolates from different STs have been reported previously by nongenomic studies (e.g., reference 19), and a wider genomic sampling surely would allow us to obtain genomes of KPC isolates belonging to other STs. The genomes of KPC-producing *K. pneumoniae* strains isolated in Italy cluster in four monophyletic groups. If we consider that the first reported case of KPC in Italy occurred in 2008, we can use the dates obtained from the molecular clock to conclude that these monophyletic groups represent four different entrances of KPC *K. pneumoniae* in Italy (Fig. 4). This indicates that KPC strains can move effectively among different countries and continents, and that the current Italian scenario of widespread KPC resistance has been caused by multiple overlapping outbreaks. Additional sampling from Italian CC258 isolates could either confirm these results or detect novel monophyla, possibly discovering additional entrance events.

Among the four Italian CC258 monophyla, one is composed entirely of isolates of ST512. This KPC sequence type was first reported in Israel in 2006 (20) but has been spreading since then, mostly in Italy and South America (11, 17). In accordance with these reports, the four available ST512 genomes from South American isolates cluster in our phylogeny as a sister group of the Italian ST512 clade (Fig. 1 and 4). The molecular clock analysis dates the common ancestor of all members of ST512 to 2007, in relative agreement with the first report of this ST, i.e., 2006 (20). Considering that this ST is known to be a single-nucleotide variant of ST258, these results indicate that a mutational event occurred around 2006, giving rise to this sequence type, that then spread to Israel, South America, and Italy. Genome sequencing of isolates of this ST from Israel, currently unavailable, could allow us to perform phylogenetic analyses aimed at better understanding the geographical and temporal origin of the ST512 clade.

Origin of the CC258 clade. Our phylogenomic analysis, coupled with the detection of recombination events and with the molecular clock analysis, allow us to update the hypothesis regarding the origin and evolution of CC258, the most widespread bearer of KPC resistance worldwide (Fig. 3). We postulate a first recombination event that occurred before 1985 between a donor similar to ST1628 and a receiver, an ancestor of ST11. This event, which transferred a region of ~1.3 Mb to the current ST11, gave rise to the basal lineage of CC258. Since only three genomes of ST11 currently are available, all isolated from Asian patients, the current phylogeny suggests that this first recombination event occurred on the Asian continent. However, additional genome sequences of ST11 from different geographic locations are necessary to support or falsify this hypothesis. Our molecular clock analysis also can be useful to date the two subsequent, previously reported (11, 12) recombination events. The second recombination event, confirmed by our phylogenies, gave rise to ST258, having as a recipi-

ent ST11 and a donor similar to ST442 (12). Our molecular clock analysis dates this event to between 1985 and 1997. Considering that all of the known genomic CC258 diversity from the American and European continents is included within the subclade that originated in 1997 (Fig. 4), this second event could have been pivotal in the subsequent pandemic of KPC-bearing CC258. Finally, we can date the third smaller recombination event, the one that gave origin to the differentiation between the two main CC258 subclades (11), to between 1999 and 2001. Thus, we can hypothesize that these three events have produced a genomic background apt to bear and diffuse KPC plasmids, contributing to the success of the KPC pandemic.

The proposed scenario suggests that the genomic diversity of the whole *K. pneumoniae* species constitutes a reservoir of genetic variability capable of recombination events of large portions of the genome, with subsequent generation of novel variants. In this scenario, we hypothesize that large genomic recombinations are at the basis of important phenotypic/functional changes that, together with the acquisition and diffusion of plasmids bearing antibiotic resistance genes, have led to the current global epidemic. This hypothesis is supported by the multiple detected recombination events, as well as by the limited number of SNPs identified outside the recombined regions (a total of 1,086 core SNPs in the 206 analyzed CC258 genomes), and finally by the current impossibility to phenotypically differentiate the isolates of subclade ST512 from those of ST258. An alternative hypothesis is that the main reason for the diffusion of CC258 is simply the acquisition of the resistance to carbapenemic antibiotics, and that the genomic variations, whether they are recombinations or point mutations, do not provide a specific fitness benefit but are merely an example of genetic hitchhiking.

In order to investigate the importance of the recombination event described in this work, the gene content of the ~1.3-Mb region was analyzed. Fifty-one genes in this genomic context were found to be potentially related to virulence or antibiotic resistance (see Table S2 in the supplemental material). The presence of LPS synthesis genes is worth a mention because of the multiple linkages between the outer membrane and virulence (39). Genes of the operon *waa* (also known as *rfa*) are responsible for the biogenesis of the core LPS, while genes of the family *arn* control the modifications of lipid A. Modifications in membrane composition can lead to changes in surface charge and interfere with the activity of antibiotics that act on LPS, such as polymyxins and novobiocin (40). Moreover, the presence of *mla* genes in the recombined region is worth being highlighted. These genes are presumed to maintain lipid asymmetry in the Gram-negative outer membrane, as they transport phospholipids to the inner side of the membrane. *mla* genes were reported as virulence factors in *Escherichia coli* and in other Gram-negative bacteria, as mutations in these genes can lead to a change in the permeability of the outer membrane and to a subsequent variation in virulence (41). The presence of fumarate reductase genes of the family *fmr* in the recombined region suggests a link with the variation of virulence of CC258. In fact, fumarate reductase is a virulence determinant in *Helicobacter pylori*, *Mycobacterium tuberculosis*, *Actinobacillus pleuropneumoniae*, and *Salmonella enterica*, as mutants of these genes show variations in virulence (32). Finally, the *ompR-envZ* operon, present in the recombined region, is a two-component system that acts as a transcription regulator, affecting the expression of the genes *ompF* and *ompC* (42). Mutations in the *ompR* and *envZ* genes have been shown to reduce the expression of outer membrane porins OmpF and OmpC (43). This in turn can have drastic effects on both the virulence and antibiotic resistance of mutant strains. It has been reported in particular that *OmpR* mutations can lead to reduced susceptibility to carbapenemic antibiotics in *Enterobacteriaceae* (44).

Further functional investigations aimed at unveiling the reasons for the success of the CC258 clade, possibly focusing on the detected recombinant regions, would greatly improve our understanding of the *K. pneumoniae* pandemic and would provide important tools in the fight against KPC-producing strains. Finally, our conclusions should lead to additional studies focused on the recombination potential of other STs of *K. pneumoniae*. If this capacity were found to be widespread, we should be aware that future recombination events could lead to the diffusion of novel epidemic clones.

ACKNOWLEDGMENTS

This work was supported by Ricerca Corrente 2013 funding from Fondazione IRCCS Policlinico S. Matteo to P.M.

We thank Simone Ambretti for providing samples and Rosa Visiello for her assistance in correcting the manuscript.

REFERENCES

1. **Knothe H, Shah P, Krcmery V, Antal M, Mitsuhashi S.** 1983. Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of *Klebsiella pneumoniae* and *Serratia marcescens*. Infection **11:**315–317. http://dx.doi.org/10.1007/BF01641355.
2. **Yigit HA, Queenan M, Anderson GJ, Domenech-Sanchez A, Biddle JW, Steward CD, Alberti S, Bush K, Tenover FC.** 2001. Novel carbapenem-hydrolyzing beta-lactamase, KPC-1, from a carbapenem-resistant strain of *Klebsiella pneumoniae*. Antimicrob Agents Chemother **45:**1151–1161. http://dx.doi.org/10.1128/AAC.45.4.1151-1161.2001.
3. **Woodford N, Tierno PM, Young K, Tysall L, Palepou MF, Ward E, Painter RE, Suber DF, Shungu D, Silver LL, Inglima K, Kornblum J, Livermore DM.** 2004. Outbreak of *Klebsiella pneumoniae* producing a new carbapenem-hydrolyzing class A beta-lactamase, KPC-3, in a New York medical center. Antimicrob Agents Chemother **48:**4793–4799. http://dx.doi.org/10.1128/AAC.48.12.4793-4799.2004.
4. **Bratu S, Landman D, Haag R, Recco R, Eramo A, Alam M, Quale J.** 2005. Rapid spread of carbapenem-resistant *Klebsiella pneumoniae* in New York City: a new threat to our antibiotic armamentarium. Arch Intern Med **165:**1430–1435. http://dx.doi.org/10.1001/archinte.165.12.1430.
5. **Nordmann P, Cuzon G, Naas T.** 2009. The real threat of *Klebsiella pneumoniae* carbapenemase-producing bacteria. Lancet Infect Dis **9:**228–236. http://dx.doi.org/10.1016/S1473-3099(09)70054-4.
6. **Cantón R, Akóva M, Carmeli Y, Giske CG, Glupczynski Y, Gniadkowski M, Livermore DM, Miriagou V, Naas T, Rossolini GM, Samuelsen Ø, Seifert H, Woodford N, Nordmann P.** 2012. Rapid evolution and spread of carbapenemases among Enterobacteriaceae in Europe. Clin Microbiol Infect **18:**413–431. http://dx.doi.org/10.1111/j.1469-0691.2012.03821.x.
7. **Richter SN, Frasson I, Bergo C, Parisi S, Cavallaro A, Palù G.** 2011. Transfer of KPC-2 carbapenemase from *Klebsiella Pneumoniae* to *Escherichia Coli* in a patient: first case in Europe. J Clin Microbiol **49:**2040–2042. http://dx.doi.org/10.1128/JCM.00133-11.
8. **Luo Y, Yang J, Ye L, Guo L, Zhao Q, Chen R, Chen Y, Han X, Zhao J, Tian S, Han L.** 2014. Characterization of KPC-2-producing *Escherichia coli*, *Citrobacter freundii*, *Enterobacter cloacae*, *Enterobacter aerogenes*, and *Klebsiella oxytoca* isolates from a Chinese hospital. Microb Drug Resist **4:**264–269. http://dx.doi.org/10.1089/mdr.2013.0150.
9. **Chen L, Chavda KD, Melano RG, Hong T, Rojtman AD, Jacobs MR, Bonomo RA, Kreiswirth BN.** 2014. A molecular survey of the dissemination of two blaKPC-harboring IncFIA plasmids in New Jersey and New York hospitals. Antimicrob Agents Chemother **58:**2289–2294. http://dx.doi.org/10.1128/AAC.02749-13.
10. **Andrade LN, Curiao T, Ferreira JC, Longo JM, Clímaco EC, Martinez R, Bellissimo-Rodrigues F, Basile-Filho A, Evaristo MA, Del Peloso PF,**

Ribeiro VB, Barth AL, Paula MC, Baquero F, Cantón R, Darini AL, Coque TM. 2011. Dissemination of blaKPC-2 by the spread of *Klebsiella pneumoniae* clonal complex 258 clones (ST258, ST11, ST437) and plasmids (IncFII, IncN, IncL/M) among *Enterobacteriaceae* species in Brazil. Antimicrob Agents Chemother **55:**3579–3583. http://dx.doi.org/10.1128/AAC.01783-10.

11. **Deleo FR, Chen L, Porcella SF, Martens CA, Kobayashi SD, Porter AR, Chavda KD, Jacobs MR, Mathema B, Olsen RJ, Bonomo RA, Musser JM, Kreiswirth BN.** 2014. Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 Klebsiella pneumoniae. Proc Natl Acad Sci U S A **111:**4988–4993. http://dx.doi.org/10.1073/pnas.1321364111.
12. **Chen L, Mathema B, Pitout JD, DeLeo FR, Kreiswirth BN.** 2014. Epidemic *Klebsiella pneumoniae* ST258 is a hybrid strain. mBio **5:**e01355-14. http://dx.doi.org/10.1128/mBio.01355-14.
13. **Pagani L, Ronza P, Giacobone E, Romero E.** 1994. Extended-spectrum beta-lactamases from *Klebsiella pneumoniae* strains isolated at an Italian hospital. Eur J Epidemiol **10:**533–540. http://dx.doi.org/10.1007/BF01719569.
14. **Perilli M, Dell'Amico E, Segatore B, de Massis MR, Bianchi C, Luzzaro F, Rossolini GM, Toniolo A, Nicoletti G, Amicosante G.** 2002. Molecular characterization of extended-spectrum beta-lactamases produced by nosocomial isolates of Enterobacteriaceae from an Italian nationwide survey. J Clin Microbiol **40:**611–614. http://dx.doi.org/10.1128/JCM.40.2.611-614.2002.
15. **D'Andrea MM, Arena F, Pallecchi L, Rossolini GM.** 2013. CTX-M-type β-lactamases: a successful story of antibiotic resistance. Int J Med Microbiol **303:**305–317. http://dx.doi.org/10.1016/j.ijmm.2013.02.008.
16. **Giani T, D'Andrea MM, Pecile P, Borgianni L, Nicoletti P, Tonelli F, Bartoloni A, Rossolini GM.** 2009. Emergence in Italy of *Klebsiella pneumoniae* sequence type 258 producing KPC-3 carbapenemase. J Clin Microbiol **47:**3793–3794. http://dx.doi.org/10.1128/JCM.01773-09.
17. **Gaibani P, Ambretti S, Berlingeri A, Gelsomino F, Bielli A, Landini MP, Sambri V.** 2011. Rapid increase of carbapenemase-producing *Klebsiella pneumoniae* strains in a large Italian hospital: surveillance period 1 March–30 September 2010. Euro Surveill **16:**19800.
18. **Comandatore F, Gaibani P, Ambretti S, Landini MP, Daffonchio D, Marone P, Sambri V, Bandi C, Sassera D.** 2013. Draft genome of Klebsiella pneumoniae sequence type 512, a multidrug-resistant strain isolated during a recent KPC outbreak in Italy. Genome Announc **1:**e00035-12. http://dx.doi.org/10.1128/genomeA.00035-12.
19. **Giani T, Pini B, Arena F, Conte V, Bracco S, Migliavacca R, Pantosti A, Pagani L, Luzzaro F, Rossolini GM.** 2013. Epidemic diffusion of KPC carbapenemase-producing *Klebsiella pneumoniae* in Italy: results of the first countrywide survey, 15 May to 30 June 2011. Euro Surveill **18:**20489.
20. **Warburg G, Hidalgo-Grass C, Partridge SR, Tolmasky ME, Temper V, Moses AE, Block C, Strahilevitz J.** 2012. A carbapenem-resistant *Klebsiella pneumoniae* epidemic clone in Jerusalem: sequence type 512 carrying a plasmid encoding aac(6′)-Ib. J Antimicrob Chemother **67:**898–901. http://dx.doi.org/10.1093/jac/dkr552.
21. **Doyle D, Peirano G, Lascols C, Lloyd T, Church DL, Pitout JD.** 2012. Laboratory detection of Enterobacteriaceae that produce carbapenemases. J Clin Microbiol **50:**3877–3880. http://dx.doi.org/10.1128/JCM.02117-12.
22. **Poirel L, Walsh TR, Cuvillier V, Nordmann P.** 2011. Multiplex PCR for detection of acquired carbapenemase genes. Diagn Microbiol Infect Dis **70:**119–123. http://dx.doi.org/10.1016/j.diagmicrobio.2010.12.002.
23. **Clinical and Laboratory Standards Institute.** 2011. Performance standards for antimicrobial susceptibility testing; 21st informational supplement. CLSI M100-S121, vol 31. Clinical and Laboratory Standards Institute, Wayne, PA.
24. **European Committee on Antimicrobial Susceptibility Testing.** 2014. Breakpoint tables for interpretation of MICs and zone diameters. Version 4.0. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/Breakpoint_table_v_4.0.pdf.
25. **Chevreux B, Wetter T, Suhai S.** 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. Proceedings of the 1999 German Conference on Bioinformatics.
26. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
27. **Darling AE, Mau B, Perna NT.** 2010. Progressivemauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5:**e11147. http://dx.doi.org/10.1371/journal.pone.0011147.
28. **Stamatakis A.** 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30:**1312–1313. http://dx.doi.org/10.1093/bioinformatics/btu033.
29. **Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, Corander J.** 2012. Detection of recombination events in bacterial genomes from large population samples. Nucleic Acids Res **40:**e6. http://dx.doi.org/10.1093/nar/gkr928.
30. **McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJ, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD.** 2013. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother **57:**3348–3357. http://dx.doi.org/10.1128/AAC.00419-13.
31. **Liu B, Pop M.** 2009. ARDB-Antibiotic Resistance Genes Database. Nucleic Acids Res **37:**D443–D447. http://dx.doi.org/10.1093/nar/gkn656.
32. **Lery LM, Frangeul L, Tomas A, Passet V, Almeida AS, Bialek-Davenet S, Barbe V, Bengoechea JA, Sansonetti P, Brisse S, Tournebize R.** 2014. Comparative analysis of *Klebsiella pneumoniae* genomes identifies a phospholipase D family protein as a novel virulence factor. BMC Biol **12:**41. http://dx.doi.org/10.1186/1741-7007-12-41.
33. **Drummond AJ, Rambaut A.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol **7:**214. http://dx.doi.org/10.1186/1471-2148-7-214.
34. **Castillo-Ramirez S, Harris SR, Holden MTG, He M, Parkhill J, Bentley SD, Feil EJ.** 2011. The impact of recombination on dN/dS within recently emerged bacterial clones. PLoS Pathog **7:**e1002129. http://dx.doi.org/10.1371/journal.ppat.1002129.
35. **Diancourt L, Passet V, Verhoef J, Grimont PAD, Brisse S.** 2005. Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. J Clin Microbiol **43:**4178–4182. http://dx.doi.org/10.1128/JCM.43.8.4178-4182.2005.
36. **Villa L, Feudi C, Fortini D, García-Fernández A, Carattoli A.** 2014. Genomics of KPC-producing *Klebsiella pneumoniae* sequence type 512 clone highlights the role of RamR and ribosomal S10 protein mutations in conferring tigecycline resistance. Antimicrob Agents Chemother **58:**1707–1712. http://dx.doi.org/10.1128/AAC.01803-13.
37. **Wright MS, Perez F, Brinkac L, Jacobs MR, Kaye K, Cober E, van Duin D, Marshall SH, Hujer AM, Rudin SD, Hujer KM, Bonomo RA, Adams MD.** 2014. Population structure of KPC-producing *Klebsiella pneumoniae* isolates from midwestern U.S. hospitals. Antimicrob Agents Chemother **58:**4961–4965. http://dx.doi.org/10.1128/AAC.00125-14.
38. **Comandatore F, Sassera D, Ambretti S, Landini MP, Daffonchio D, Marone P, Sambri V, Bandi C, Gaibani P.** 2013. Draft genome sequences of two multidrug resistant *Klebsiella pneumoniae* ST258 isolates resistant to colistin. Genome Announc **1:**e00113–12. http://dx.doi.org/10.1128/genomeA.00113-12.
39. **Heinrichs DE, Yethon JA, Whitfield C.** 1998. Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. Mol Microbiol **30:**221–232. http://dx.doi.org/10.1046/j.1365-2958.1998.01063.x.
40. **Goldberg JB.** 1999. Genetics of bacterial polysaccharides. CRC Press, London, United Kingdom.
41. **Malinverni JC, Silhavy TJ.** 2009. An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. Proc Natl Acad Sci U S A **12:**8009–8014. http://dx.doi.org/10.1073/pnas.0903229106.
42. **Buckler DR, Anand GS, Stock AM.** 2000. Response-regulator phosphorylation and activation: a two-way street? Trends Microbiol **8:**153–156. http://dx.doi.org/10.1016/S0966-842X(00)01707-8.
43. **Yuan J, Wei B, Shi M, Gao H.** 2011. Functional assessment of EnvZ/OmpR two-component system in *Shewanella oneidensis*. PLoS One **6:**e23701. http://dx.doi.org/10.1371/journal.pone.0023701.
44. **Tängdén T, Adler M, Cars O, Sandegren L, Löwdin E.** 2013. Frequent emergence of porin-deficient subpopulations with reduced carbapenem susceptibility in ESBL-producing *Escherichia coli* during exposure to ertapenem in an in vitro pharmacokinetic model. J Antimicrob Chemother **68:**1319–1326. http://dx.doi.org/10.1093/jac/dkt044.