Universidade Federal de Minas Gerais

Instituto de Ciências Biológicas

Departamento de Biologia Geral

Programa Interunidades de Pós-Graduação em Bioinformática

Rodrigo Profeta Silveira Santos

Análise genômica comparativa de uma linhagem de *Klebsiella pneumoniae* multirresistente recentemente isolada de um paciente no Brasil

Belo Horizonte

2018

Rodrigo Profeta Silveira Santos

Análise genômica comparativa de uma linhagem de Klebsiella pneumoniae multirresistente recentemente isolada de um paciente no Brasil

Dissertação apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do título de Mestre em Bioinformática.

Orientador: Prof. Dr. Vasco Azevedo

Coorientador: Prof. Dr. Thiago Luiz de Paula Castro

Belo Horizonte

2018

A verdadeira sabedoria é aquela que encontramos nas coisas simples da vida.

Autor Desconhecido.

AGRADECIMENTOS

Diversos foram os aprendizados obtidos durante esta jornada. O meu crescimento pessoal e acadêmico durante o período de realização deste trabalho foi imenso. Na certeza de que nada foi conquistado sozinho, agradeço a todos os envolvidos, direta ou indiretamente nessa conquista. Assim, gostaria de agradecer aos meus primeiros orientadores, Prof.ª Patricia Belini Nishiyama e Prof. Bruno Lopes Bastos, por terem me iniciado no mundo científico.

Ao Prof. Vasco Azevedo, pela orientação, apoio e por sempre ter acreditado no meu potencial.

Ao Prof. Thiago Castro, pela coorientação, ensinamentos e suporte durante a realização deste trabalho.

Ao Dr. Pablo Ramos, pela ajuda na construção e desenvolvimento deste trabalho.

Ao Dr. Sandeep Tiwari, pelo incentivo, inspiração, contribuições e apoio.

À minha família e amigos, que me apoiaram e mesmo na minha ausência me incentivaram a buscar o aperfeiçoamento desejado.

Aos integrantes do Laboratório de Genética Celular e Molecular que, como segunda família, me acolheu e incentivou o trabalho, contribuindo para realização do mesmo.

Por fim, reconheço o apoio financeiro das agências de fomento Capes e CNPq.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

BLAST	*Basic Local Alignment Search Tool.*
BLASTn	*Nucleotide-nucleotide* BLAST.
BLASTp	*Protein-protein BLAST*
De novo	Do início.
DNA	Ácido Desoxirribonucléico.
frameshift	Mudança do quadro de leitura em uma sequência codificante.
Indels	Inserção e deleção de nucleotídeos.
k-mer	Sequência de caracteres de tamanho k que se repete mais de uma vez em uma sequência.
MALDI-TOF	*Matrix Assisted Laser Desorption Ionization – Time of Fly.*
MDR	*Multidrug-resistant pathogen.*
NCBI	*National Center for Biotechnology Information.*
PATRIC	Pathosystems Resource Integration Center.
PEPR	*Phylogenomic Estimation with Progressive Refinement.*
QUAST	*Quality assessment tool.*
RAST	*Rapid Annotation using Subsystem Technology.*
Reads	Leituras provenientes do sequenciamento.

RESUMO

Klebsiella pneumoniae é uma bactéria Gram-negativa, sem motilidade, em forma de bastonete e encapsulada, que habita vários nichos ecológicos, variando do solo à água. É também um dos patógenos humanos mais importantes, emergindo como agente de graves infecções. Uma análise abrangente do genoma desse microrganismo tem o potencial de fornecer uma melhor compreensão das bases moleculares de sua virulência e patogênese, contribuindo para novas formas de controle da doença. Neste estudo, sequenciamos o genoma de uma linhagem de *K. pneumoniae* multirresistente, denominada B31, que foi isolada do lavado broncoalveolar de uma paciente admitida em uma unidade de terapia intensiva, no Brasil. As sequências de DNA obtidas da linhagem B31, a partir do sequenciamento Illumina, foram montadas usando uma abordagem *ab initio*, resultando em um cromossomo de 5,27 Mb. Um multireplicon (IncFIIk / IncFIBk), provavelmente pertencente a um plasmídeo conjugativo, e um replicon IncI1, que pertence a um segundo plasmídeo, foram encontrados usando a ferramenta PlasmidFinder. Comparações genômicas foram conduzidas entre a linhagem B31, outro isolado clínico do Brasil (Kp13), e outros 171 isolados clínicos cujas sequências completas foram recuperadas do banco de dados do NCBI. B31 foi encontrada como sendo pertencente ao *Sequence type* 15 (ST15), um dos 52 STs encontrados dentre as 173 linhagens analisadas. A busca por genes de resistência revelou que todas as linhagens selecionadas contêm pelo menos um gene codificante de beta-lactamase de espectro estreito (bla). Análises comparativas adicionais foram conduzidas com 44 linhagens representativas de *K. pneumoniae*, escolhidas com base em resultados filogenômicos e caracterização por MLST. A identificação dos genomas central e acessório foi realizada para caracterizar a conexão entre todas as linhagens analisadas. Este estudo fornece informações genômicas que podem ser úteis para auxiliar na identificação de novas drogas e no desenvolvimento de vacinas contra esse patógeno com implicações na saúde pública.

Palavras-Chave: Average Nucleotide Identity. Genômica Comparativa. Patógenos Multi-resistentes, Multilocus Sequence Type.

ABSTRACT

Klebsiella pneumoniae is a Gram-negative, non-motile, rod-shaped, and encapsulated bacterium that dwells in various ecological niches, ranging from soil to water. It is also one of the most important human pathogens, emerging as an agent of severe community infections. A comprehensive analysis of the genome of this microorganism could provide a better understanding of the molecular basis of its virulence and pathogenesis, contributing to new forms of disease control. In this study, we sequenced the genome of a multidrug resistant K. *pneumoniae* strain, herein named B31, that was isolated from the bronchoalveolar lavage of a patient admitted to an intensive care unit, in Brazil. The DNA sequences obtained from strain B31, with the Illumina sequencing technology, were assembled to form a 5.27-Mb sized genome, using an ab initio approach. A plasmid multireplicon (IncFIIk/ IncFIBk), likely to belong to a conjugative plasmid, and a replicon IncI1, which belongs to a second plasmid, were found using the PlasmidFinder tool. Genomic comparisons were conducted among strain B31, another clinical isolate from Brazil (KP13), and other 171 clinical isolates whose complete genome sequences could be retrieved from the NCBI database. B31 was found to belong to the sequence type 15 (ST15), one out of the 52 sequence types (STs) characterizing the 173 strains analyzed. An in-silico screening of resistance genes revealed that all selected strains harbor at least one narrow-spectrum beta-lactamase (bla) gene. Further comparative analyses were conducted with, 44 representative K. *pneumoniae* strains, chosen based on phylogenomics and MLST results. The identification of the core and accessory genes was conducted to disclose the connection among all analyzed strains. This study provides genomic information that might be useful to aid in the design of new drugs and vaccines against a pathogen with public health implications.

Keywords: Average Nucleotide Identity, Comparative genomics, Multidrug Resistant pathogen, Multilocus Sequence Type.

ESTRUTURA DA DISSERTAÇÃO

Essa dissertação está organizada em três seções: (I) Introdução, Justificativa e Objetivos; (II) Manuscrito completo em língua inglesa formatado de acordo com as normas do periódico científico internacional *Scientific Reports*; (III) Considerações finais e Perspectivas.

1. Introdução Geral

Klebsiella é um gênero que compreende bactérias Gram-negativas, geralmente envoltas por uma cápsula, pertencentes à família Enterobacteriaceae (DRANCOURT, 2015). A presença dessa cápsula polissacarídica é uma característica que distingue os membros desse gênero da maioria das outras bactérias da família. Além da cápsula, fímbrias foram descritas como estruturas de adesão presentes em bactérias desse gênero (DUGUID, 1959). As colônias são grandes e mucóides de consistência viscosa, enquanto as células são caracterizadas pela forma de bastonetes, sendo geralmente imóveis com ausência de flagelos (DRANCOURT, 2015). A cápsula polissacarídica da *Klebsiella* também dificulta a fagocitose e permite que a bactéria possa aderir e colonizar o trato respiratório de humanos e animais (WYRES et al., 2016).

Espécies do gênero são encontradas em diversos lugares do meio ambiente, tais como riachos e lagos, esgoto, solo e em plantas, além da superfície mucosa de mamíferos, como humanos e outros animais, sendo neles consideradas membros transitórios da microbiota (MARTIN; BACHMAN, 2018; PODSCHUN et al., 2001; ULLMANN, 1998). Em humanos, *K. pneumoniae* está presente como comensal na nasofaringe e no trato intestinal (ULLMANN, 1998). Apesar da existência de um caso de bactéria não patogênica descrito (LIMA *et al*., 2018), as bactérias desse gênero são nosocomiais, infectando principalmente indivíduos imunocomprometidos que são hospitalizados e sofrem de outras doenças (SIU *et al*., 2012). Ainda que essencialmente caracterizadas como patógenos oportunistas nosocomiais, *K. pneumoniae* pode causar pneumonia bacteriana adquirida na comunidade (LEE *et al*., 2018). Em pacientes de UTI, equipamentos tais como sonda e cateter venoso central, representam portas de entrada para infecções hospitalares, facilitando a infecção por bactérias multirresistentes. O principal reservatório da espécie é o trato gastrointestinal dos pacientes internados no hospital (DRANCOURT, 2015).

Diversos casos de *K. pneumoniae* resistentes a antibióticos foram reportados (CASPAR *et al*., 2017; HOLT *et al*., 2015; LEE *et al*., 2011; RAMOS *et al*., 2014), fato que demonstra a importância do uso adequado de antibióticos. *K. pneumoniae* produtoras de carbapenemase (KPC) e a *New Delhi* metallo-beta-lactamase (NDM) são cada vez mais frequentes no ambiente hospitalar (CALFEE, 2017) e requerem cuidados especiais no que diz respeito ao uso de antibióticos (TRUONG; YAMAKI, 2018). Estas têm se tornado um problema sério, uma vez que o desenvolvimento de novas drogas antimicrobianas é demorado

e a adaptação das bactérias ao uso dos antibióticos existentes é crescente.

Neste contexto, a variabilidade genética é um fator primordial para o sucesso adaptativo observado em microrganismos dos mais diversos nichos (HACKER *et al*., 2001). Em associação à seleção natural, essa variabilidade contribui para que determinada característica persista ou não no decorrer das gerações, determinando se um organismo pode competir pelos recursos necessários para sua a reprodução (BENNETT, 2004). Assim, a dinâmica do genoma de um determinado microrganismo depende também da adaptação a hospedeiros. Observa-se diminuição do genoma em bactérias de vida livre quando comparadas com bactérias intracelulares, por exemplo (TOFT; ANDERSSON, 2010).

Várias forças, químicas e genéticas, podem alterar o conteúdo gênico do DNA e aumentar as chances de sobrevivência, permitindo o crescimento em um ambiente anteriormente hostil. Diversos são os mecanismos associados à essas mudanças, a citar-se (1) mutações de ponto, (2) recombinação e (3) transferência horizontal de material genético entre diferentes espécies bacterianas (HACKER *et al*., 2001). Este último, representa um dos cernes da diversidade genética bacteriana, e levou a mudanças dramáticas na composição dos genomas microbianos em períodos de tempo relativamente curtos (OCHMAN; LAWRENCE; GROISMAN, 2000). Dentre essas mudanças, a propagação horizontal de genes de resistência tem elevado a taxa de sobrevivência de microrganismos mesmo com o uso destes que estão entre os medicamentos mais bem sucedidos, os antibióticos (MARTINEZ, 2014). Essa transferência ocorre principalmente por plasmídeos bacterianos (SAN MILLAN, 2018) e é causa de graves problemas de saúde pública (LONG *et al*., 2017; NICOLÁS *et al*., 2018; RAMOS *et al*., 2014). *K. pneumoniae* desempenha um papel fundamental na disseminação desses genes de resistência, a partir de micróbios ambientais para patógenos clinicamente importantes (WYRES; HOLT, 2018).Fica evidente, portanto, a importância da identificação e caracterização dos fatores genéticos relacionados à patogênese de *Klebsiella pneumoniae.* Sendo assim, análises *in silico* podem ser aplicadas para a identificação detes determinates a partir de ferramentas de bioinformática. Dentre elas destacam-se os métodos de tipagem, a filogenia, e a genômica comparativa.

Métodos de tipagem molecular fornecem meios para determinar a origem de infecções e surtos, rastrear a transmissão cruzada de patógenos associados à assistência médica, reconhecer linhagens virulentas e avaliar a eficácia das medidas de controle (RANJBAR et al., 2014). Diversos são os métodos empregados com esses propósitos. Contudo, dois métodos

se destacam com relação à tipagem em *K. pneumoniae.* O primeiro é denominado *multilocus sequence typing* (MLST). MLST é um método empregado para qualificar as relações genéticas entre isolados bacterianos, classificando linhagens e identificando grupos clonais (FEIL et al., 2004). Esta tipagem é feita com base na sequência nucleotídica de sete genes conservados que se encontram distantes no cromossomo desta bactéria: *gapA* (gliceraldeído 3-fosfato desidrogenase), *infB* (fator de iniciação da tradução 2), *mdh* (malato desidrogenase), *pgi* (fosfoglicose isomerase), *phoE* (fosforina E), *rpoB* (subunidade beta da RNA polimerase) e *tonB* (transdutor de energia periplasmico) (DIANCOURT et al., 2005). O segundo método se refere à caracterização da cápsula polissacarídica. A capsula polissacarídica é a camada mais externa da célula de *K. pneumoniae*, protegendo a bactéria da dessecação, da predação de fagos e da atuação do sistema imune do hospedeiro. Ferramentas para identificação de *loci* de síntese de cápsula a partir de dados do genoma completo já foram descritas. (WYRES et al., 2016).

A filogenia é o estudo da relação evolutiva entre organismos, muitas vezes espécies, indivíduos ou mesmo genes. Quando essa relação é feita em escala genômica é então chamada de filogenômica (CHAN; RAGAN, 2013). A representação evolucionária dessas relações é geralmente descrita por um diagrama em formato de árvore, em que os ramos representam o tempo de divergência e seus encontros (nós) representam ancentrais comuns a todas a unidades taxonômicas que são representadas pelas folhas (nós terminais) (CHOUDHURI, 2014). Os métodos de reconstrução de filogenia podem ser baseados em distância ou em caracteres (YANG; RANNALA, 2012). Nos métodos baseados em distância, a distância entre cada par de sequências é calculada, e uma matriz de distância é gerada para a reconstrução das árvores. Como exemplo temos o método *neighbour joining* (BRUNO; SOCCI; HALPERN, 1987; N SAITOU, 1987). Os métodos baseados em caracteres comparam um caractere por vez a partir de um alinhamento. Estes incluem os métodos de parcimônia, máxima verossimilhança e inferência bayesiana (YANG; RANNALA, 2012). A filogenética, quando usada em escala genômica, também pode ser usada para a genômica comparativa, que é o estudo das interrelações de diferentes genomas. A genômica comparativa ajuda a identificar regiões de similaridade e diferenças entre os genomas analisados (CHOUDHURI, 2014).

Em bactérias, a genômica comparativa é utilizada como uma ferramenta para identificação de linhagens patogênicas (PRENTICE, 2004), mecanismos de virulência (WYRES et al., 2016) ou mesmo da disseminação de genes de resistência a antibióticos

(NICOLÁS et al., 2018; WYRES; HOLT, 2018). A identificação da homologia entre genes se faz necessária para realização desses estudos (HAGGERTY et al., 2014). A partir da definição da sequência ancestral, comparações são feitas nos genomas analisados para a identificação dos genes presentes em todas as linhagens, ou mesmo aqueles presentes em alguns grupos ou mesmo em linhagens únicas. Essa identificação é então conhecida como pangenoma (CARLOS GUIMARAES et al., 2015). Assim, todas as famílias de genes encontrados em conjunto em uma espécie são designadas como seu pangenoma (MCINERNEY; MCNALLY; O'CONNELL, 2017). O pangenoma pode ser dividido em genoma central e acessório. O primeiro se refere às famílias gênicas "essenciais" que são encontradas em todos os membros. O genoma acessório, portanto, consiste de genes parcialmente compartilhados e específicos de determinada linhagem (TETTELIN et al., 2005). Pangenomas são ainda classificados em aberto ou fechado, sendo um pangenoma caracterizado como aberto quando se espera que novos genes sejam adicionados ao repertório total de determinado organismo a partir de um novo genoma adicionado à comparação. Em oposição, quando o pool gênico não se expandir após a adição de um novo genoma à análise, o mesmo é caracterizado como fechado (CARLOS GUIMARAES et al., 2015).

Devido ao surgimento de linhagens multiresistentes a antibióticos, análises do proteoma e vias metabólicas de *K. pneumoniae* estão sendo realizadas com o objetivo de identificar novos alvos (AHMAD *et al*., 2018). Análises de genômica comparativa identificam fatores de virulência e sua a relação com a emergência e persistência das linhagens (HOLT *et al*., 2015; RAMOS *et al*., 2014; STOESSER *et al*., 2014). *K. pneumoniae* é uma causa comum de infecções nos ambientes relacionados aos cuidados de saúde. Infecções causadas por ela estão rapidamente se tornando intratáveis devido ao uso indiscriminado de antibióticos, fato que gerou resistência. Com o surgimento de linhagens hipervirulentas e resistentes, os determinantes genéticos envolvidos em suas adaptações devem ser melhor compreendidos para que novos métodos de controle possam ser desenvolvidos.

2. Justificativa

Patógenos bacterianos são causa constante de infecções em humanos e representam, portanto, grande ameaça à sua saúde. Bactérias possuem amplo repertório de genes envolvidos na sobrevivência nos diversos nichos que podem ocupar (por exemplo, no meio externo ou no hospedeiro), e que podem ocasionar, direta ou indiretamente, doenças. Produtos gênicos que contribuem para o potencial infeccioso de microrganismos são chamados fatores de virulência (HILL, 2012). Além disso, os determinantes de resistência também contribuem para o sucesso adaptativo bacteriano, uma vez que, na presença de determinado antibiótico, a bactéria resistente é selecionada em detrimento de outras que são incapazes de proliferar (MARTINEZ, 2014). A abordagem comparativa permite a identificação e caracterização inicial destes determinantes a partir da sequência genômica de isolados. Nesse contexto, *Klebsiella pneumoniae* é um patógeno com importância médica que possui várias linhagens resistentes a diversos antibióticos. Devido à dificuldade de eliminação, as linhagens resistentes podem levar a uma maior severidade das doenças, representando um risco significativo para a saúde humana. Até o presente trabalho, apenas um genoma completo havia sido obtido a partir de um isolado de *K. pneumoniae* no Brasil (Kp13) (RAMOS et al., 2014).

Além de caracterizar a linhagem B31, isolada do lavado bronco pulmonar de um paciente com infecção grave, em 2017, no Brasil, este estudo procurou estabelecer comparativos com genomas de outras linhagens de *K. pneumoniae* depositados em banco de dados público, buscando uma melhor compreensão acerca da distribuição de fatores de virulência e patogenicidade. A identificação dos determinantes de virulência em isolados nacionais é de grande importância para o entendimento dos mecanismos de patogênese relacionados à infecções por *K. pneumoniae* no país. Sendo assim, é relevante a análise genômica da linhagem *Klebsiella pneumoniae* B31 e sua comparação com os demais isolados de outros lugares do mundo e depositados em bancos de dados públicos. Apesar de alguns estudos terem abordado a genômica desta espécie, este trabalho é pioneiro em realizar uma comparação extensa envolvendo 173 isolados de *K. pneumoniae.*

3. Objetivos

3.1 Objetivo Geral

Estudar, por análises de bioinformática, o genoma do isolado clínico de *Klebsiella pneumoniae* B31, comparando-o com 172 genomas de outros isolados da mesma espécie e definindo a distribuição de características genéticas que possam contribuir na patogenicidade e resistência a antibióticos.

3.2 Objetivos Específicos

- Obter e caracterizar o genoma do isolado B31 de *K. pneumoniae*;
- Validar os genomas de *K. pneumoniae* depositados em bancos de dados como pertencentes a esta espécie, para a realização de análises comparativas;
- Realizar a tipagem *in silico* dos isolados selecionados de *K. pneumoniae*;
- Comparar o genoma do isolado B31 com os genomas de outras linhagens de *K. pneumoniae*;
- Estudar o pan-genoma de *K. pneumoniae;*
- Identificar o repertório de genes de virulência e resistência a antibióticos presentes nas linhagens selecionadas de *K. pneumoniae.*

4. Manuscrito completo

Comparative genomic analysis with a multidrug resistant *Klebsiella pneumoniae* strain recently isolated in Brazil

Rodrigo Profeta, Núbia Seyffert, Sandeep Tiwari, Marcus V. C. Viana, Arun Kumar Jaiswal, Ana Carolina Caetano, Daniel Henrique Bücker, Roselane Santos, Alfonso Gala-Garcia, Siomar C. Soares, Bertram Brenig, Pablo I. Ramos, Vasco Azevedo, e Thiago L. P. Castro

O artigo completo referente às análises realizadas nesse trabalho em língua inglesa, foi formatado de acordo com as normas da revista *Scientific Reports*. Este periódico é uma revista online de acesso aberto dos editores da *Nature* e abrange todas as áreas das ciências naturais e clínicas.

Comparative genomic analysis with a multidrug resistant *Klebsiella pneumoniae* strain recently isolated in Brazil

Rodrigo Profeta[1], Núbia Seyffert[1,2], Sandeep Tiwari[1], Marcus V. C. Viana[1], Arun Kumar Jaiswal[1,3], Ana Carolina Caetano[1], Daniel Henrique Bücker[1], Roselane Santos[1], Alfonso Gala-Garcia[1], Siomar C. Soares[1,3], Bertram Brenig[4], Pablo I. Ramos[5], Vasco Azevedo[1,†], and Thiago L. P. Castro[1,6, *,†]

[1]Department of General Biology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. [2]Institute of Biology, Federal University of Bahia, Salvador, Bahia, Brazil. [3]Department of Immunology, Microbiology and Parasitology, Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro, Uberaba, MG, Brazil. [4]Institute of Veterinary Medicine, University of Göttingen, Burckhardtweg 2, Göttingen, Germany. [5]Gonçalo Moniz Institute, Oswaldo Cruz Foundation, Salvador, Bahia, Brazil. [6]Institute of Health Sciences, Federal University of Bahia, Salvador, BA, Brazil. [*]Correspondence and requests for materials should be addressed to T.L.P.C. (email: castrotlp@gmail.com). [†]These authors share senior authorship.

ABSTRACT

Klebsiella pneumoniae is a Gram-negative, non-motile, rod-shaped, and encapsulated bacterium that dwells in various ecological niches, ranging from soil to water. It is also one of the most important human pathogens, emerging as an agent of severe community infections. A comprehensive analysis of the genome of this microorganism could provide a better understanding of the molecular basis of its virulence and pathogenesis, contributing to new forms of disease control. In this study, we sequenced the genome of a multidrug resistant *K. pneumoniae* strain, herein named B31, that was isolated from the bronchoalveolar lavage of a patient admitted to an intensive care unit, in Brazil. The DNA sequences obtained from strain B31, with the Illumina sequencing technology, were assembled to form a 5.27-Mb sized chromosome, using an ab initio approach. We identified a multireplicon plasmid (IncFIIk/ IncFIBk), likely to belong to a conjugative plasmid, and a replicon IncI1, which likely belongs to a second plasmid. Genomic comparisons were conducted among strain B31, one other clinical isolate from Brazil (Kp13), and other 171 clinical isolates whose complete genome sequences could be retrieved from the NCBI database. This study provides genomic information that might be useful to aid in the design of new drugs and vaccines against a pathogen with public health implications.

Introduction

Klebsiella is a genus of ubiquitous bacteria from the family Enterobacteriaceae. They are characterized by their rod shape, arranged singly, in pairs or short chains; often surrounded by a capsule, which gives the colonies a viscid consistency[1]. *Klebsiella pneumoniae* is a Gram-negative, non-motile, rod-shaped, oxidase-negative, lactose fermenting bacterium[2]. It dwells in various ecological niches ranging from soil to water and represents one of the most important pathogens, emerging as an agent of severe community infections[3,4].

Even though *K. pneumoniae* is part of the natural human microbiota, identified in 3.8% and 9.5% of stool and nasal samples of healthy volunteers, respectively[5], the increased number of nosocomial infections reveals their medical importance. *K. pneumoniae* infection occurs in lungs, skin pharynx or gastrointestinal tract. It causes respiratory tract infections, urinary tract infections (UTIs), and bloodstream infections. Among the hospital-acquired infections *K. pneumoniae* is one of the most common causes worldwide, owing to the high prevalence of antibiotic-resistant strains[6,7]. In consequence, it has become a serious public health concern. This species has two highly related species diverged from a common ancestor, *Klebsiella variicola* and *Klebsiella quasipneumoniae*, that are sometimes erroneously identified as *K. pneumoniae*[8].

Useful strategies to control these emerging microbial threats involve analyses of multilocus sequence types in combination with characterization of resistance and virulence genes[9]. Key virulence determinants, such as the polysaccharide capsule and siderophores are important components for the infection establishment and were identified in *Klebsiella*[9,10]. The comprehensive analyses of genome sequence provide better ways of understanding the molecular and genetic basis of the virulence of this bacterium, enabling a detailed investigation of its pathogenesis and ways to control infection onsets. Therefore, in this study we performed a genomic comparison between a newly sequenced clinical isolate of *K. pneumoniae* B31 and 172 public available *K. pneumoniae* genomes, to obtain a better biological understanding of the phylogenetic relationships and genome plasticity, and to explore insights into the resistance and virulence repertoire of these strains.

Results

Identification of a new multi-drug resistant *K. pneumoniae* strain

A previously unknown bacterial isolate was obtained from the bronchoalveolar lavage of a female patient admitted to an intensive care unit in the northeast of Brazil. This strain was able to grow well on both liquid and solid Brain Heart Infusion (BHI) media, as well as on blood agar at 37° Celsius (Figure 1). *Klebsiella pneumoniae* identification was obtained using MALDI-TOF MS, with a score of high probability. Antimicrobial Susceptibility Testing (AST) was performed and revealed that the bacterial isolate is a multidrug-resistant strain (Table 1). The isolate presented resistance to 8/17 (47%) of the tested antibiotics, including resistance to several beta-lactam antibiotics (ampicillin, cefepime, ceftazidime, ceftriaxone, cefuroxime axetil, and cefuroxime), a beta-lactamase inhibitor (sulbactam), and a fluoroquinolone (ciprofloxacin). Susceptibilities (antibiotic effectiveness) to other beta-lactams (including cefoxitin, ertapenem, imipenem, meropenem), aminoglycosides (amikacin and gentamicin), and polymyxin E (colistin) were also observed. Intermediate resistance was observed for the combination of ampicillin with the β-lactamase inhibitor tazobactam and for the tetracycline tigecycline. Intermediate resistance may be interpreted as a probable resistance only if the organism is located in some sites of the host body where the drug is not easily accessible[11]. Based on these *in vitro* antimicrobial resistance findings, the present isolate can be categorize as a multidrug-resistant (MDR) pathogen[12], highlighting its public health importance.

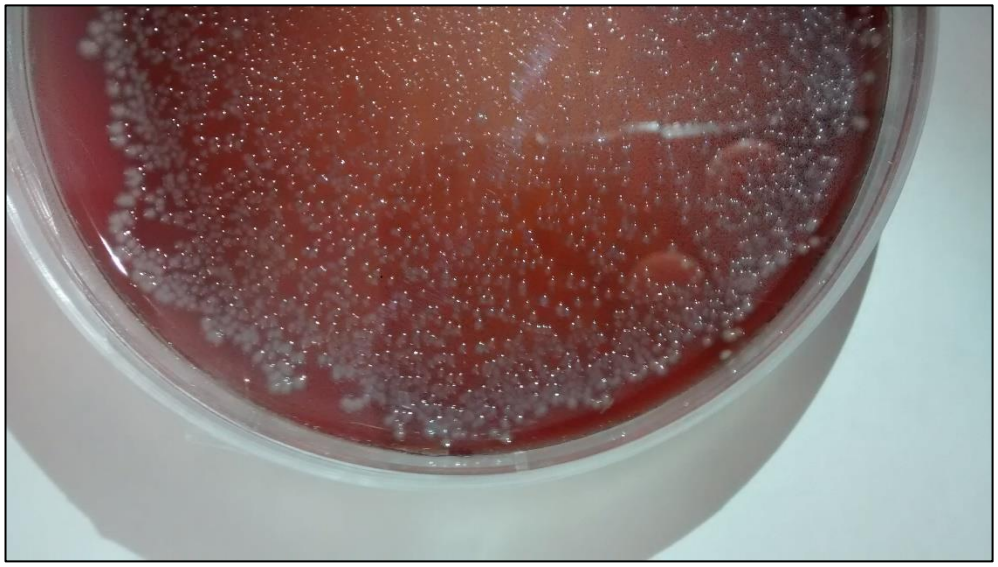


Figure 1: 10µL depletion of the bronchoalveolar lavage on blood agar plate at 37° Celsius. As noted, the plate is occupied by one-type sticky colonies. There are no colonies of another microorganism.

Table 1: Antimicrobial susceptibility testing results.

Antimicrobial(s)	MIC (µg/ml)	Interpretation
Piperacillin[b]/tazobactam[c]	32	I
Tigecycline [f]	2	I
Ampicillin[b]/sulbactam[c]	>=32	R
Ampicillin[b]	>=32	R
Cefepime[b]	>= 64	R
Ceftazidime[b]	>=64	R
Ceftriaxone[b]	>=64	R
Cefuroxime axetil[b]	>=64	R
Cefuroxime[b]	>=64	R
Ciprofloxacin[d]	>=4	R
Amikacin[a]	4	S
Cefoxitin[b]	<=4	S
Colistin[e]	<=0,5	S
Ertapenem[b]	<= 0,5	S
Gentamicin[a]	<= 1	S
Imipenem[b]	<= 0,25	S
Meropenem[b]	<= 0,25	S

[a]Aminoglycoside; [b] Beta-lactam; [c] Beta-lactamase inhibitor; [d] Fluoroquinolone; [e] Polymyxin E; [f] Tetracycline; I, intermediate; S, susceptible; R, resistant.

Genomic characterization of strain B31

Genome sequencing of strain B31 of *K. pneumoniae* revealed a single 5.27Mb circular chromosome and allowed the identification of three plasmid replicons from different compatibility groups. Two of these replicons comprised the multireplicon IncFII$_k$/Inc- FIB$_k$, which likely belongs to a conjugative plasmid. The third replicon, IncI1, is probably carried by one other plasmid. The complete assembly of plasmid sequences was not obtained, due to the presence of multiple sequence reads with repeat regions. However, the plasmid contigs were successfully used for the prediction of drug-resistance genes present in strain B31.

Misidentification of *Klebsiella* strains by MALDI-TOF MS has been recently reported, especially for *K. pneumoniae* and its closest relatives *K. quasipneumoniae* and *K. variicola*[8]. Thus,

pairwise comparisons of the Average Nucleotide Identity based on BLAST+ (ANIb) among the sequences of (1) the B31 isolate and (2) the reference type strains of these three species were performed to confirm the species. The results indicate that the present isolate fits the genomic species threshold set (95%) and shares 98.79% of ANI with the *K. pneumoniae* DSM 30104 type strain, reinforcing the previous identification of *K. pneumoniae* (Table 2). Comparisons to the type strains *K. quasipneumoniae* 01A030, *K. quasipneumoniae* subsp. *similipneumoniae* 07A044, and *K. variicola* DSM 15968 yielded ANI percentages of 93.2%, 93.5%, and 94.13%, respectively.

Table 2: Average Nucleotide Identity (ANI) results based on BLAST+ for the strain B31. The reference genomes of *K. quasipneumoniae* 01A030 (**R1**), *K. quasipneumoniae* subsp. *similipneumoniae* 07A044 (**R2**), *K. variicola* DSM 15968 **(R3)**, and *K. pneumoniae* DSM 30104 **(R4)** were used for species confirmation of the query strain B31 (**Q1**).

	R1	**R2**	**R3**	**R4**	**Q1**
R1	*	96.12 (87.12)	92.9 (87.15)	93.00 (83.58)	93.2 (84.18)
R2	96.34 (92.46)	*	93.37 (92.11)	93.38 (88.4)	93.5 (88.84)
R3	92.9 (86.63)	93.24 (86.39)	*	94.09 (83.47)	94.13 (84.89)
R4	93.07 (83.04)	93.30 (82.72)	94.15 (83.22)	*	98.79 (85.44)
Q1	93.21 (87.74)	93.49 (87.14)	94.13 (89.1)	98.67 (90.04)	*

Values above 95% were considered to be the same species (green values); values in parenthesis show aligned nucleotides (%).

The allelic profile of the sequenced strain B31 assigns the sequence type 15 (ST15). All of the seven *loci* were identified using Multilocus Sequence Typing (MLST) with 100% identity. Regarding capsule characterization, our assembly almost certainly matches a known KL type. The best matching *locus* for the isolate B31 was KL19. The coverage and identity are both high (100% and 94.91% respectively), the length discrepancy is low (+11 bp), and essentially all twenty expected genes were found with high identity. As shown in Table 3, one of the expected genes (*locus* 14) falls below the identity threshold (default 95%), but with a very close value (94.69%). It indicates that our sample has diverged a bit from the KL19 reference, possibly due to mutation and/or recombination.

Table 3: Capsular genes present in the strain B31.

Locus	Gene	Identity
1	*galF*	99.00%
2	*cpsACP*	96.19%
3	*wzi*	97.92%
4	*wza*	97.88%
5	*wzb*	100.00%
6	*wzc*	98.75%
7	*wcaJ*	98.74%
8	*wcuF*	94.86%
9	*wcmY*	96.62%
10	*wzx*	96.88%
11	*wctW*	97.17%
12	*wctX*	99.23%
13	*wzy*	98.85%
14	-	94.69%
15	*gnd*	99.57%
16	*rmlB*	99.15%
17	*rmlA*	97.92%
18	*rmlD*	97.98%
19	*rmlC*	98.92%
20	*ugd*	99.74%

Four different virulence *loci* were examined in the strains: yersiniabactin (*ybt*), colibactin (*clb*), aerobactin (*iuc*) and salmochelin (*iro*). In addition to these virulence *loci*, the presence of *rmpA* and rmpA2 genes, which cause hypermucoid phenotype by upregulating capsule production, was screened. *K. pneumoniae* B31 does not harbor any of the analyzed virulence genes.

Resistance screening showed that *K. pneuoniae* B31 does not harbor fosfomycin (*Fcyn*), glycopeptides (*Gly*), rifampin (*Rif*), carbapenemase (*Bla_Carb*), extended spectrum beta-lactamases with resistance to beta-lactamase inhibitors (*Bla_ESBL_inhR*), and broad-spectrum beta-lactamases (*Bla_broad*) genes (Table 4). The screening revealed two imprecise allele matches related to aminoglycoside (*Agly*) resistance (*AadA2*; Aph3-Ia**). As revealed by the antimicrobial susceptibility test, the strain is sensible to the aminoglycosides amikacin and gentamicin. Thus, these alleles do not confer resistance to this type of antibiotics.

Inactivation of the *mgrB* gene has been shown to promote colistin resistance in *K. pneumoniae*[13,14]. Gene coverage lower than 90% accounts for truncation/ function loss. Truncation of this gene was found in Kp B31, with 49% of coverage. Similarly, the observed resistance of the

fluoroquinolone ciprofloxacin in the AST assay was likely acquired upon mutations of the targeted type II topoisomerase (DNA gyrase) genes *gyrA* and *parC*[15]. The *mphA* gene encodes a resistance enzyme, macrolide phosphotransferases, which inactivates macrolides (MLS). Resistance to sulfonamide (Sul), tetracycline (Tet), and trimethoprim (Tmt) is conferred by the genes *sulI*, *tetA*, and *dfrA12*, respectively. In turn, the β-lactamase is encoded by the genes *OXA-1* and *CTX-M-15*.

Table 4: Resistance genes present in the strain Kp B31.

Resistance classes	Location	
	Plasmid	Chromosome
AGly	*AadA2*; Aph3-Ia**	-
Col	*MgrB-36%; PmrB-10%*	-
Fcyn	-	-
Flq	-	*ParC-80I; GyrA-83F; GyrA-87A*
Gly	-	-
MLS	*MphA*	-
Phe	*CatB4*	-
Rif	-	-
Sul	*SulI*	-
Tet	*TetA*	-
Tmt	*DfrA12*	-
Bla	*OXA-1*	*AmpH*; SHV-28**
Bla_Carb	-	-
Bla_ESBL	*CTX-M-15*	-
Bla_ESBL_inhR	-	-
Bla_broad	-	-
Bla_broad_inhR	*TEM-30**	-

Imprecise allele matches are indicated with *

The knowledge about pathogenicity islands (PI), metabolic islands (MI), resistance islands (RI) and symbiotic islands (SI), the virulence factors they encode, their mobility, and their structure is helpful in understanding the bacterial evolution and interactions. A total of 19 of all islands were classified as the four types (PI, MI, SI, and SI) (Supplementary Figure 1). All the genes found in these islands were further screened (supplementary file 1). Among them, genes that regulate length and adhesion of type 1 fimbriae, type VI secretion system (T6SS), and CRISPR-cas were found in GI2, GI4, and GI7, respectively. Prophage screening within the genome identified one incomplete prophage region with 37.7Kb. A total of 29 proteins were found to be encoded in this region, but with a low score (below 70% identity). The reduced number of prophages in the genome might be related to the presence

of genes encoding CRISPR-Cas proteins in one of the islands[16].

Comparisons among the analyzed *Klebsiella Strains*

Due to imprecise species characterization reports[8,17], species confirmation of the complete genome of 172 strains deposited as *K. pneumoniae* (Supplementary Table 1) was conducted. Misidentification of six strains was detected using the ANIb method (Table 5) and includes a strain from the American Type Culture Collection (ATCC), one of the most important culture collections of the international scenario. The best species result for the currently named *K. pneumoniae* subsp. *pneumoniae* ATCC 43816 is *K. quasipneumoniae subsp. similipneumoniae*, with an average identity of 98.8 with the reference genome *K. quasipneumoniae* subsp. *similipneumoniae* 07A044.

In order to access the evolutionary relationship between these strains, a maximum likelihood tree of the 173 strains core proteome was constructed (Figure 2). The tree was subsequently annotated with MLST, capsular type (K-type), and virulence information. Similar sequence types were grouped closely, what suggests that the ST information is an effective way of identifying the genetic relationships among bacterial isolates. Strains found not to be *K. pneumoniae* (YH43, KP5-1, KPN1481, HKUOPLC, and ATCC 43816), as demonstrated above, are better represented in Figure 3. *K. variicola* and *K. quasipneumoniae* strains formed their own clades, separated from the *K. pneumoniae* strains.

A total of 52 sequence types (STs) and 42 capsular types (KLs) were found among all 173 strains analyzed (Supplementary Table 1). One strain (*K. pneumoniae* YH43) could not be characterized using the MLST analysis. Thus, the number of strains represented in the Supplementary Table 1 is 172. Six other strains were found to belong to ST 15. This group is characterized by the presence of the allele 1 for all of the seven housekeeping genes considered. *K. pneumoniae* Goe 39795, used as the reference in the assembly process, was also found belong to ST 15. All of the ST15 strains were isolated from human in various parts of the world (Goe 39795 from Germany; PMK1 from Nepal; BR from China; and KP36 from Taiwan). No geographical information about the strains MLST-15 and 19051 could be retrieved.

Table 5: Misidentification checking using Average Nucleotide Identity (ANI), based on BLAST+ of six strains previously described as *K. pneumoniae.* The reference genomes of *K. quasipneumoniae* 01A030 (**R1**), *K. quasipneumoniae* subsp. *similipneumoniae* 07A044 (**R2**), *K. variicola* DSM 15968 (**R3)**, and *K. pneumoniae* DSM 30104 (**R4)** were used for species confirmation of the queries *Klebsiella pneumoniae* YH43 (**Q1**), *K. pneumoniae* KPN1481 (**Q2**), *K. pneumoniae* subsp. *pneumoniae* HKUOPLC (**Q3**), *K. pneumoniae* 342 (**Q4**), *K. pneumoniae* subsp. *pneumoniae* KP5-1 (**Q5**), and *K. pneumoniae* subsp. *pneumoniae* ATCC 43816 (**Q6**).

	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**R1**	**R2**	**R3**	**R4**
Q1	*	96.57 (91.31)	92.77 (83.62)	96.77 (90.67)	96.53 (88.98)	92.67 (84.49)	92.6 (85.03)	92.83 (84.22)	96.55 (91.19)	93.35 (82.52)
Q2	96.5 (82.71)	*	93.06 (77.79)	98.52 (86)	98.44 (84.75)	92.69 (77.97)	92.71 (80.8)	93 (79.33)	98.63 (84.64)	93.59 (77.1)
Q3	93.05 (90.07)	93.4 (92.16)	*	93.37 (92)	93.36 (91.21)	98.88 (94.32)	96.4 (91.28)	99.12 (94.26)	93.4 (92.08)	93.52 (88.04)
Q4	96.49 (85.48)	98.46 (88.65)	92.92 (80.33)	*	98.39 (87.04)	92.82 (79.65)	92.64 (81.61)	93 (81.58)	98.62 (87.74)	93.66 (79.45)
Q5	96.61 (87.79)	98.72 (91.62)	93.11 (83.61)	98.82 (91.08)	*	92.93 (83.17)	92.84 (84.81)	93.09 (84.34)	98.86 (91.06)	93.82 (82.82)
Q6	92.91 (88.33)	93.08 (89.41)	98.9 (91.36)	93.21 (88.59)	93.15 (88.25)	*	96.25 (88.38)	98.87 (90.92)	93.24 (89.55)	93.39 (85.19)
R1	92.66 (85.21)	92.86 (88.66)	96.09 (85.29)	92.93 (86.88)	92.78 (86.64)	95.98 (85.02)	*	96.12 (87.12)	92.9 (87.15)	93 (83.58)
R2	93.02 (89.9)	93.29 (92.59)	99.01 (93.57)	93.36 (92.42)	93.26 (91.39)	98.8 (93.01)	96.34 (92.46)	*	93.37 (92.11)	93.38 (88.4)
R3	96.65 (90.72)	98.86 (92.86)	93.17 (85.44)	98.94 (93.03)	98.81 (92.09)	93.06 (85.55)	92.9 (86.63)	93.24 (86.39)	*	94.09 (83.47)
R4	93.55 (81.73)	93.84 (84.14)	93.41 (81.39)	93.98 (84.04)	93.91 (83.75)	93.31 (81.29)	93.07 (83.04)	93.3 (82.72)	94.15 (83.22)	*

Strains with identity values above 95% are considered to belong to the same species (green values); values in parenthesis show aligned nucleotides.

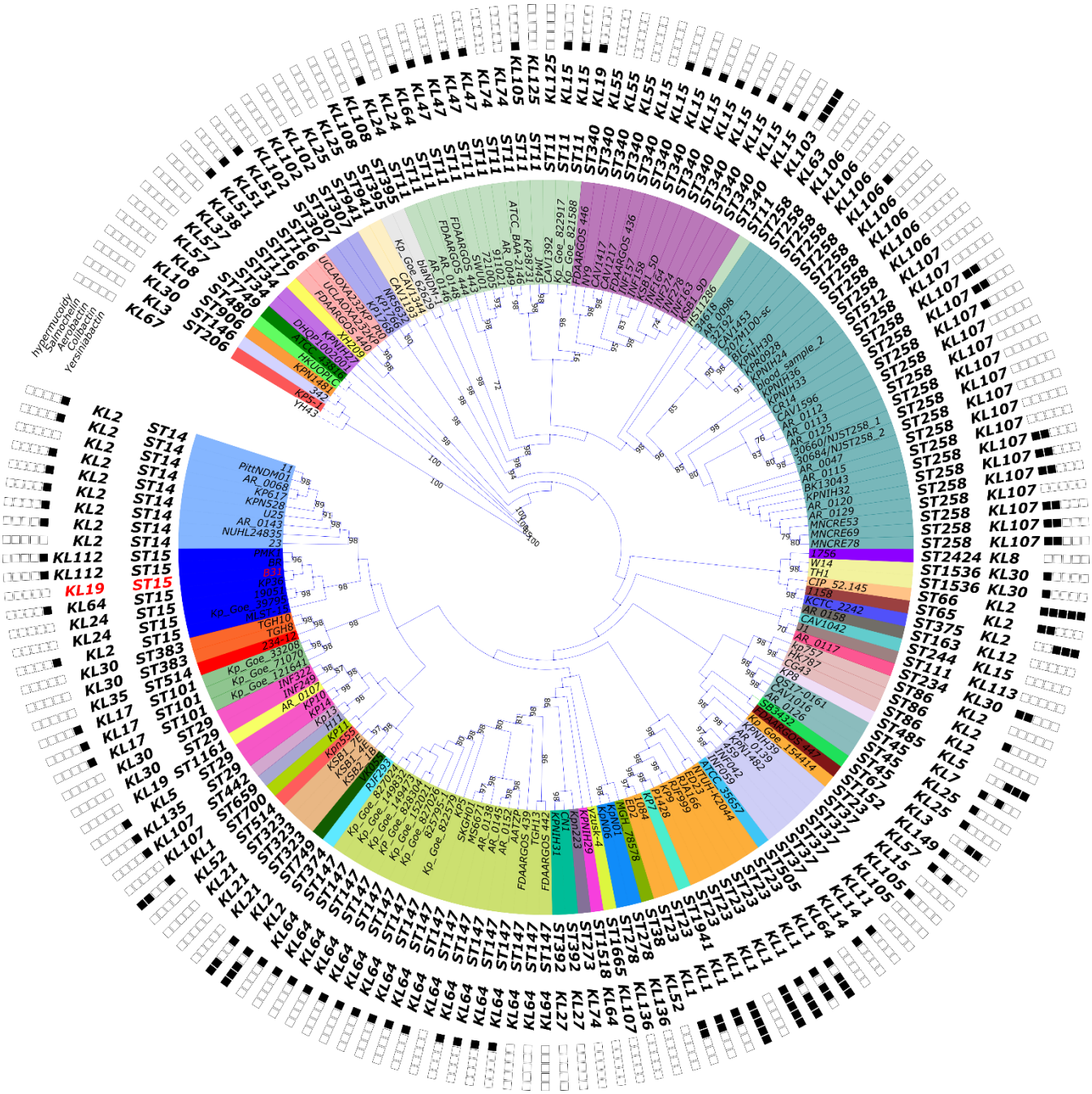


Figure 2: Maximum Likelihood tree of 173 *K. pneumoniae* strains. Colors represent sequence type (ST) and filled black squares represent presence of the virulence factor. Capsule type (KL) is also shown. The strain sequenced in this work, B31, and its MLST and capsular typing information is written in red.

ST14 strains were grouped close to the ST15 group (Figure 4). Their MLST alleles differed in only in the *infB* gene (Supplementary Table 2), forming the clonal complex (CC) 15.

There are 77 serologically defined K-type references to date (KL1 to KL77). Our results indicate that from all of the 173 strains analyzed, 125 have a capsular type *locus* associated with one of these 77 references. Their K *locus* type varies from KL1 to KL74. Likewise, KL101 and above are defined from DNA sequence data based on gene content. The other 48 strains were predicted to be in this category. Supplementary Table 1 shows the capsular characterization for all the strains analyzed.

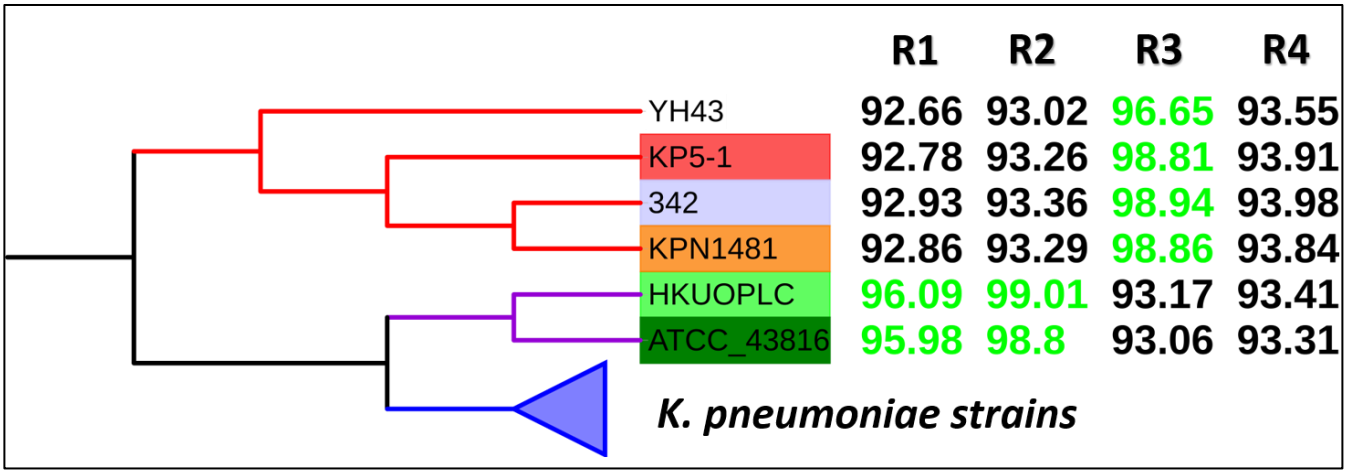


Figure 3: Cladogram of the misidentified strains of *Klebsiella pneumoniae.* The tree shows the phylogenomic relation between the strains YH43, KP5-1, 342, KPN1481, HKUOPLC, ATCC 43816, misidentified as *K. pneumoniae*, and the *K. pneumoniae* strains analyzed. Reference type strains *K. quasipneumoniae* 01A030 (**R1**), *K. quasipneumoniae* subsp. *similipneumoniae* 07A044 (**R2**), *K. variicola* DSM 15968 (**R3)**, and *K. pneumoniae* DSM 30104 (**R4)** were used to confirm species identification, by ANIb. *K. variicola* (red branches) and *K. quasipneumoniae* (purple branches) strains formed their own clades with identity above 95% (green values).

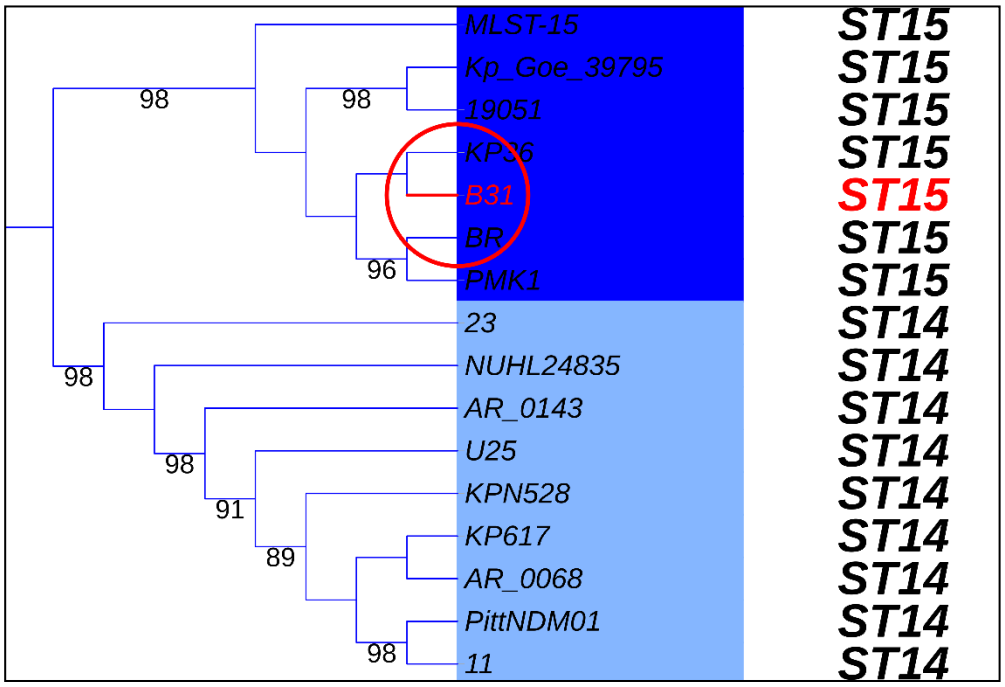


Figure 4: Clonal complex of ST15 and its single-*locus* variant ST14. Kp B31 is represented in red.

Virulence and resistance screening were performed for all *K. pneumoniae* strains. Four strains belonging to ST15, not including B31, have a high affinity iron chelating gene (*ybt*) in their genome (Figure 2). Only 3 strains harbor all of the five analyzed genes (strains P1428, RJF999, and CIP_52.145). The former two belong to ST23 and KL1, while the latter belongs to ST66 and KL2. Only few other strains also have the hypermucoid phenotype, as shown in Figure 2. The classes of resistance genes were also plotted in the phylogenomic tree (Figure 5). The screening of resistance genes showed that all strains analyzed harbor at least one narrow spectrum beta-lactamase (*Bla*) gene (red filled square), and none of them have extended spectrum beta-lactamases with resistance to beta-lactamase inhibitors (green unfilled square). All of the other resistance classes presented some strain harboring at least one gene. However, some strains do not harbor any resistance genes in addition to beta-lactamases genes.

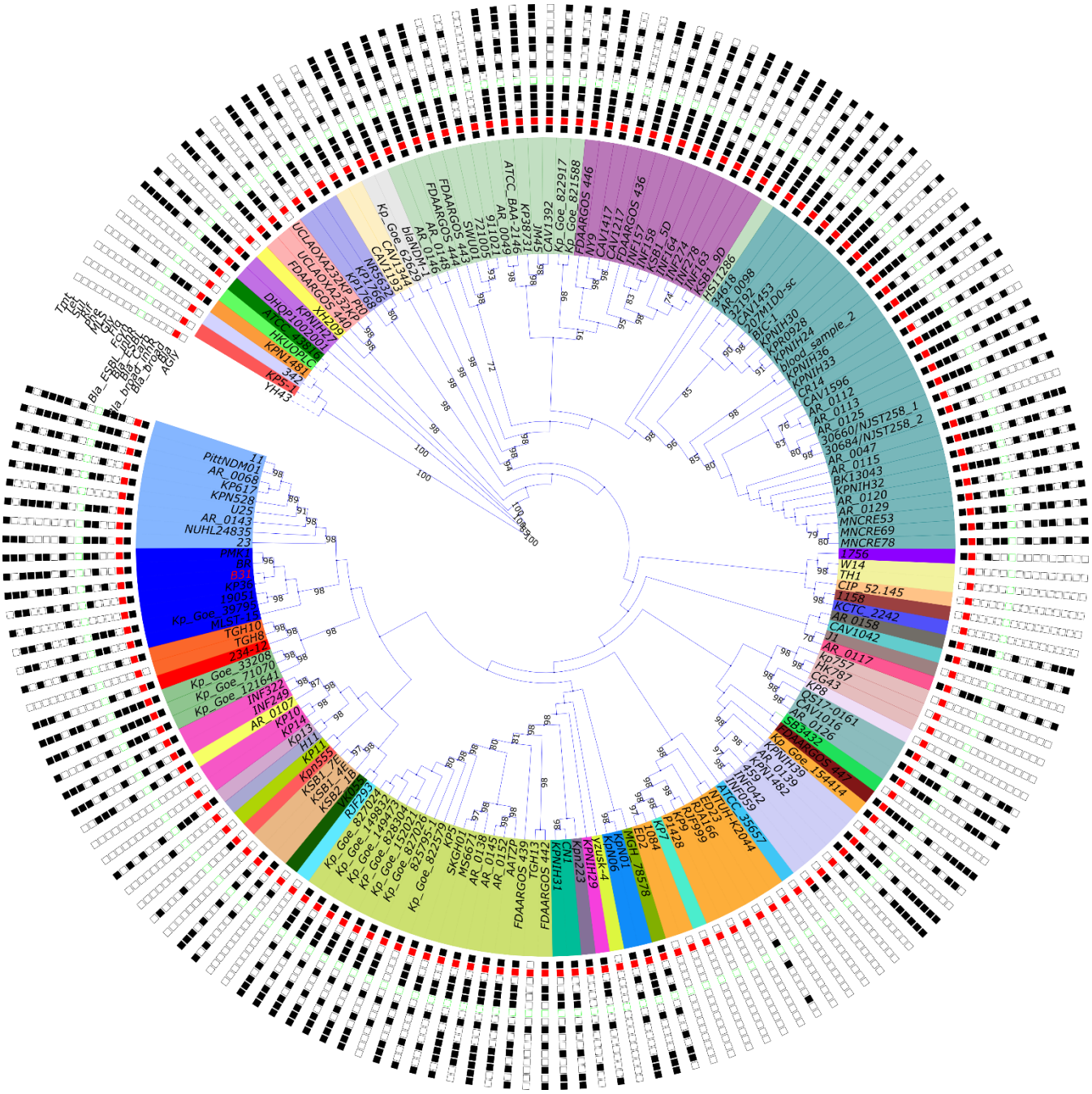


Figure 5: Phylogenomic tree including information on the presence of resistance classes of genes. Colors represent sequence type (ST) and filled black squares represent gene presence of the resistance class. The resistance classes are AGly (aminoglycosides), Bla (beta-lactamases), Bla_broad (broad spectrum beta-lactamases), Bla_broad_inhR (broad spectrum beta-lactamases with resistance to beta-lactamase inhibitors), Bla_Carb (carbapenemase), Bla_ESBL (extended spectrum beta-lactamases), Bla_ESBL_inhR (extended spectrum beta-lactamases with resistance to beta-lactamase inhibitors), Fcyn (fosfomycin), Flq (fluoroquinolones), Gly (glycopeptides), MLS (macrolides), Rif (rifampin), Sul (sulfonamides), Tet (tetracyclines), and Tmt (trimethoprim).

Even though the phylogenomic tree generated using Gegenees[18] revealed a clustering pattern in accordance with ST, two main clusters with strains of different STs, likely to be clonal complexes, were found (Figure 6). The first cluster comprises the strains 7 to 15 (34618, BIC-1, KPNIH36, 911021, AR 0146, FDAARGOS 446, INF157, KSB1 9D, and HS11286), with similarity percentages ranging from 93 to 100 %. This cluster includes strains of ST11, ST340, and ST258. The second cluster is formed by strains of ST14 and ST15 (11, KPN528, 19051, BR, 23, B31), with lower similarity percentages (84 - 100 %). Clustering of strains by country of isolation was not apparent. Within the first cluster, isolates from five different counties were observed. Three other strains of the same cluster did not have their isolation country reported.

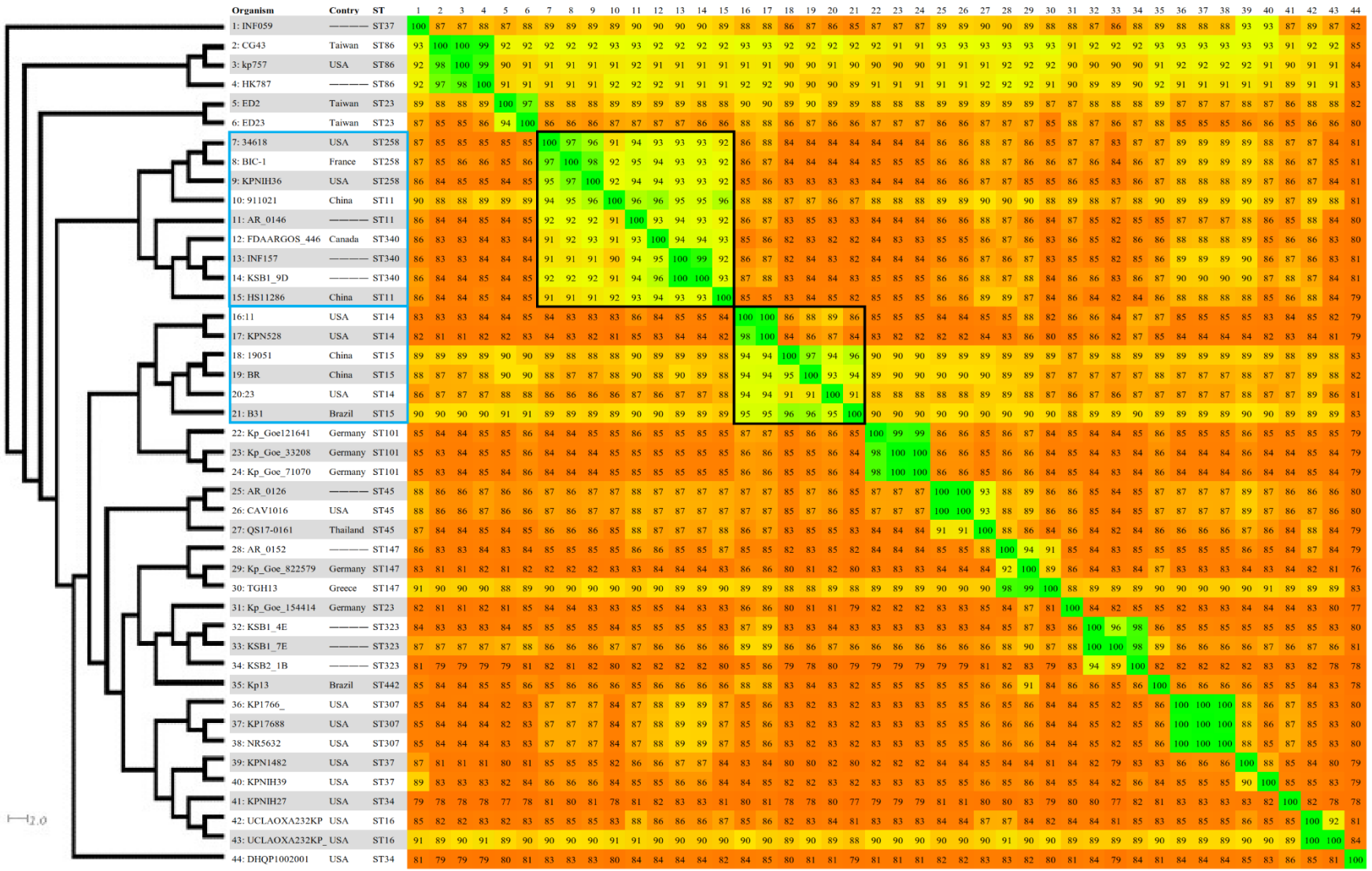


Figure 6: Phylogenomic tree and heatmap analysis of 44 Strains of *Klebsiella pneumoniae*. The figure represents the comparison between the variable content of all strains. Percentages were plotted with a color spectrum ranging from red (low similarity) to green (high similarity). The names of the strains on the left side of the figure are organized in the same order in the top part of the figure (heatmap - horizontally). STs likely to be clonal complexes are represented within blue (dendrogram - vertically) and black (heatmap - horizontally) boxes.

Analysis of *K. pneumoniae* strains indicates an open pangenome

44 representative *K. pneumoniae* strains were chosen based on phylogenomics and MLST results. The resulting pangenome of all 44 selected strains contains a total of 9,712 genes. From these, 3,588 represent the indispensable gene families that are found in all members evaluated, referred to as the core genome. Species that exhibit many gene content differences among their members are described as having extensive (or open) pangenomes, while others have closed pangenomes[19]. To evaluate whether the inclusion of more *K. pneumoniae* genomes would provide a great number of additional genes, characterizing an open pan-genome[20–22], analyses were conducted using the Heaps' Law, formulated as $n = k * N^{-\alpha}$. In a genetic context, n is the expected number of genes for a given number of genomes, N determines the number of genomes, and the k and α (α =1-γ) values are free parameters that are determined empirically. According to the Heap's Law, when α > 1 (γ < 0), the pan-genome is considered to be closed, and there will be no significant increase in the number of genes with the addition of a new genome. On the other hand, when α < 1 (0 < γ < 1), the pan-genome is open and there will be a significant increase in the number of genes for every added genome. Our results inferred that the pangenome of *Klebsiella pneumoniae* is increasing with an α of 0.843 (α = 1 – 0.157), indicating that it has an open pan-genome (Figure 7A).

The resulting core genome has 3,588 genes. 1,443 singleton genes were found (Figure 7B, and C). According to the least-squares fit of the exponential regression decay, the tg(θ) represents the point where the curve stabilizes, which may be translated to the number of genes in the core genome after stabilization and the number of singletons that will be added to the pan-genome for each newly sequenced genome. According to the tg(θ) value, only 10 genes will be added. The classification of core genome on the functional categories of the COG is represented at the Figure 8. Most of the core genes are composed of the "Metabolism" category (37.77%). A substantial number of genes were poorly characterized (32.72%). 89 genes (2.39%) were not present in any COG categories.

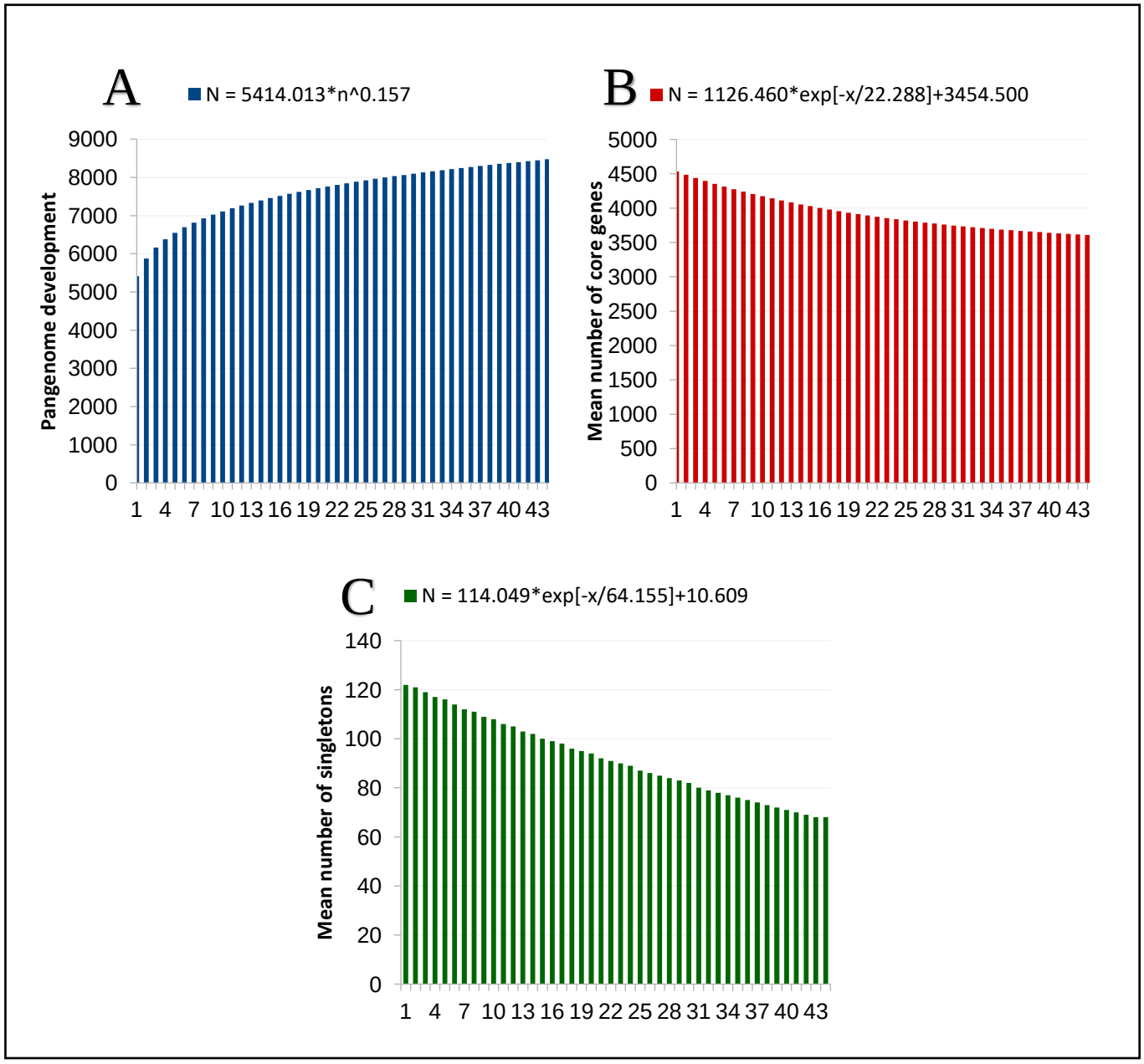


Figure 7: Pan-genome, core genome and singletons of *Klebsiella pneumoniae*. A, B, and C, respectively, showing the pangenome, core genome and singletons development using of all 44 strains of *K. pneumoniae*.

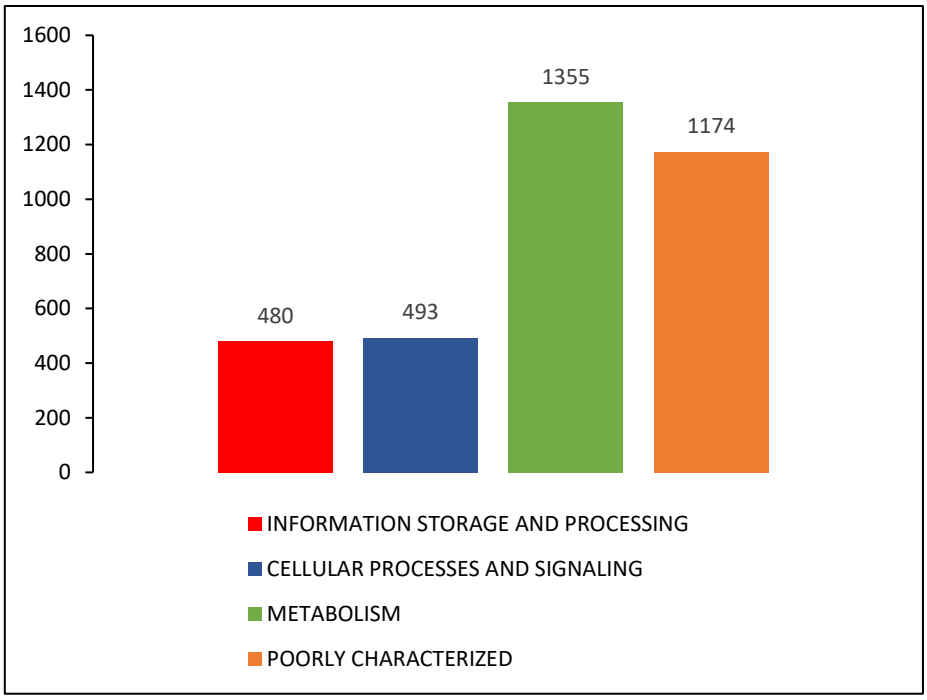


Figure 8: Core genes of the *Klebsiella pneumoniae* strains classified by COG functional category.

Discussion

Emergence of antimicrobial resistant bacteria has become a critical public health concern, as there are fewer effective antimicrobial agents available for treatment of infections caused by these bacteria[23]. In this work, we report the characterization of a multidrug resistant *K. pneumoniae* ST15 strain (named B31), which was isolated from a patient admitted to an intensive care unit in the northeast of Brazil. Resistance determinants detected in the screening explain nonsusceptibility to 10 of the 17 antimicrobials tested (Table 1). KpB31 has acquired many mechanisms of antibiotic resistance, in which their determinants are manly plasmid encoded (Table 6). A mechanism of colistin resistance is generated by inactivation of the *mgrB* and *pmrB* genes[14]. Truncation of these genes were found in kpB331, with 36% and 10% of coverage respectively, suggesting colistin resistance. The *mphA* gene encodes a macrolide phosphotransferase which inactivates macrolides (MLS)[24]. The catB4 gene encodes a group B chloramphenicol acetyltransferase, which inactivates chloramphenicol[25]. Sulfonamide (Sul), tetracycline (Tet), and trimethoprim (Tmt) resistances are generated by the genes *sulI*, *tetA*, and *dfrA12*, respectively. β-lactamases genes were found to encode *OXA-1* and *CTX-M-15*.

Chromosomal-encoded resistance determinants were observed only for fluoroquinolones and Beta lactams. The observed resistance against the fluoroquinolone ciprofloxacin, in the AST assay, was acquired upon mutations of the targeted type II topoisomerase (DNA gyrase) genes *gyrA* and *parC,* as described before[15]. Beta-lactamases, such as *AmpH* and *SHV,* are universally found in *K. pneumoniae*[26] and so their genes should be detected in most genomes, as it is in KpB31. This fact confirms our finding that all strains analyzed harbor at least one narrow spectrum beta-lactamases (Bla) gene.

Albeit there are hypervirulent strains, which is not the case of Kp B31, it has been demonstrated that several other strains have the capacity to cause invasive diseases by the acquisition of several virulence factors by horizontal transference[2,27–29]. None of the most common virulence genes were found in the strain B31, presence of niche and virulence factor genes were. The type 1 fimbriae genes found mediates bacterial adherence to eukaryotic cells and was associated with urinary tract infection[30]. This gene was found inside a genomic island, which is a mark of horizontal DNA transfer. Type VI secretion system (T6SS) genes are also present in B31. This system allows bacteria to battle other cells by injection of effector molecules, such as toxins into their cytoplasm[31]. CRISPR-Cas (clustered, regularly interspaced short palindromic repeats–CRISPR-associated proteins) system is an adaptive

immune system in prokaryotes that provides protection against invading DNA, such as bacteriophages[16]. The reduced number of prophages found in the B31 genome can be related to the presence of CRISPR-Cas proteins found to be present in one of the islands.

Our results show that MLST is a useful tool to assess the relatedness among bacterial isolates as the STs grouped accordingly in the phylogenomic tree (Figure 2 and 6). Six closely related genotypes clustered together with it in the same sequence type, forming a clonal complex (Figure 4). Clonal complexes, other than CC15, have been previously characterized[32,33].

Capsule typing is another useful way to link genetic diversity with phenotypical characteristics. Polysaccharide capsules play roles in (1) enhancing resistance to phagocytosis and complement-mediated killing during infection, and (2) preventing desiccation while bacteria is not in the host[34]. Despite the ambiguous definition as a niche or a virulence factor[35,36], capsules typing is noteworthy for serotyping characterization. Therefore, capsule *locus* (KL) type has been repeatedly identified for *K. pneumoniae*. There are 77 immunologically distinct *Klebsiella* capsule types (K-types) defined by serology[37]. However, due to specialist requirement and lack of reagents in the majority of the microbiology laboratories, it is seldom applied. Therefore, in a genomic era, capsular screening has become a valuable tool. No previews reports of a *K. pneumoniae* strain harboring capsular synthesis K type *locus* (KL) KL19 was found. However, KL19 capsular serotype strains belonging to the related *K. pneumoniae* species, denominated *K. variicola* were recently isolated[8].

Species characterization was confirmed by ANI analysis, as described before[38–40]. It was essential for the discovery that some genomes have been deposited as a wrong species, in agreement with previous finding[8]. Given the similarity between *K. pneumoniae, K. quasipneumoniae* and *K. variicola* species and the absence of clinical differences between individuals infected with one or the other, our result suggests that *K. variicola* and *K. quasipneumoniae* reporting cases are probably not accurate due to misidentification with *K. pneumoniae.* Hence, a greater care for their characterization should be taken.

The phylogenomic tree produced using Gegenees reveled two main clusters representing two clonal complexes, namely CC258 and CC15. The later clonal group contains the strain B31.These two clonal complexes, CC258 (ST258, ST340, and ST11) and CC15 (ST15 and ST14) and, were reported to be the predominant *K. pneumoniae* clones resistant to third-generation cephalosporins in five African and two Vietnamese major towns[41]. *K. pneumoniae* ST15 has been spread globally and is responsible for the high prevalence of CTX-M-15 in Europe and Asia[42]. This sequence type often displays resistance to beta-lactams and fluoroquinolones and was recently associated with transferable mcr-1

colistin resistance[28]. Therefore, the appearence of one isolate belonging to this ST in Brazil, present report, shows that multidrugresistant *K. pneumoniae* is spread worldwide and infections caused by this organism should be and more carefully treated as resistance is increasing. Strain Kp13, also isolated from brazil[2], appeared in closer relation with strains from ST323 (KSB1_4E, KSB1_7E, and KSB2_1B). Even though strains B31 and kp13 were isolated in the same country, their genomes do not look more alike as they do with other strains. It shows the variable characteristics of *K. pneumonie* genome, as conjointly highlighted by the heatmap.

All the gene families found in a species is designated as its pangenome[19]. It is worth noting that the term pangenome should be used when all species members are taken into consideration, at least the sequenced ones. However, the pangenome analysis here executed aimed the genomic comparison of limited number of 44 strains to achieve a global view of their genome repertoire. Species that exhibit many gene content differences among their members are described as having extensive (or open) pangenomes, while others have closed pangenomes[19]. An open pangenome was found in our analysis, which means new genes are expected to be added to the total repertoire of this organism. This result agrees with the presence of this species in various ecological niches, which contributes to the flexible genetic content observed due to the extent contact with potential gene donors[22]. Regarding the core genome, the majority of the genes were classified as belonging to the metabolism category. This category contains genes that are usually involved in essential cellular processes, such as the energy production and conversion, along with the transport and metabolism of amino acids, nucleotides, carbohydrates, coenzymes, lipids, inorganic ions, and secondary metabolites, as well as the later biosynthesis[43]. Given the essentiality of the core genome, this functional category is expected to be highly represented among their genes. Although a substantial number of genes were poorly characterized, this result is in agreement with previous gene functional classification of *K. pneumoniae*[44].

Whole-genome sequencing of multidrug-resistant (MDR) pathogens has been an important instrument to better understand the genesis, potential mobility, and spectrum of antimicrobial resistance mechanisms. Hence, information provided by sequencing might assist in the identification of resistance genotypes and their global dissemination. These data provide additional information about the pathogenesis of a *K. pneumoniae* strain isolated from Brazil. The identification of virulence and multiple resistance determinants of this strain disclose its enormous potential to install infection and survive even after the action of many antibiotics used to treat patients. In addition, new insights into the typical characteristics of the species by comparative genomics are given. Lastly, Misidentification of *K.*

pneumoniae cases is probable to be common and should be further analyzed.

Methods

Isolate characterization, genomic DNA extraction and sequencing

Klebsiella pneumoniae isolate B31 was obtained from the bronchoalveolar lavage of an infected patient in 2017, Brazil. Following the classical methods for speciation of bacteria when patients are suspected of having a bacterial infection, visible colonies of the organism were isolated in pure culture (on blood agar plate), as shown in Figure 1. To confirm the organism identification, MALDI-TOF was adopted using the MALDI Biotyper 3.0 software version (Bruker Daltoniks, Germany). *Klebsiella pneumoniae* antimicrobial susceptibility tests were performed by Vitek (bioMérieux). DNA was extracted through mechanical lysis followed by purification with a phenol solution (phenol:chloroform:isoamyl alcohol 25:24:1, v/v, respectively) and precipitation with ethanol 70% and sodium acetate 3M. Whole genome sequencing was performed using Illumina HiSeq technology (Illumina, San Diego, CA, USA) from paired-end libraries (2x 150 bp) and fragment of approximately 450 bp using the Agilent 2100 Bioanalyzer system as a control.

Genome assembly and annotation

Reads had the quality checked using FastQC[45]. *De novo* assembly was performed using Newbler (v2.9) and SPAdes (v3.9.1). For the Spades assembly, the best k-mer length was estimated using KmerGenie (v1.7044)[46]. The quality of each assembly was observed in QUAST[47] and the best assembly was chosen. Contigs were ordered and the scaffold was generated via CONTIGuator using *K. pneumoniae* strain Kp_Goe-39795 as reference (CP018458.1). Part of the gaps were filled with others assemblies via FGAP[48]. The remaining gaps were evaluated individually and then they were filled manually by reference comparison using CLC Genomics Workbench (v6.5) to obtain a consensus sequence after mapping the reads against the assembled genome. The presence of plasmids was searched using PlasmidFinder (v1.3)[49]. To assemble the plasmids, PlasmidSpades[50] was used. The contigs generated by this approach had the plasmid similarity checked via BLASTn. The scaffolding generation and the gap-closure steps were proceeded as for the genome.

Annotation was done via RASTtk (Rapid Annotation Using Subsystem Technology)[51] annotation service in PATRIC (Pathosystems Resource Integration Center)[52]. Frameshifts were cautiously analyzed by a manual curation to confirm their authenticity. Since any sequencing platform has its own drawback, fixing a maximum number of any kind of sequencing errors is vital. Therefore,

the manual curation applied in this work was done as follows. Initially, annotation of the reference genome was transferred to the assembled genome using an in-house script to identify the pseudogenes. Artemis v16.0.0[53] was used to examine the obtained frameshifted genes, and they were then compared to other genes by BLASTn against the NR database at NCBI (National Center for Biotechnology Information) to inspect if the mutation was conserved. In addition, insertions/deletions (indels) caused by sequencing artifacts were corrected by mapping the reads against the assembled genome using CLC Genomics Workbench (v6.5) BLASTp was subsequently performed against the UniProt database[54] to check the translated sequence.

Bioinformatics Analyses for Kp B31 characterization

In silico confirmation of the species was done using the web server JSpeciesWS[55]. JSpeciesWS is an online service designed to estimate the probability if two or more genomes belong to the same species by pairwise comparison of their Average Nucleotide Identity (ANI). ANI measures the nucleotide-level genomic similarity between the coding regions of two or more (all vs. all) genomes. ANIb (based on BLAST+) was chosen for the analysis. Reference type strain genomes of *K. quasipneumoniae* subsp. quasipneumoniae 01A030, *K. quasipneumoniae* subsp. *similipneumoniae* 07A044, *K. variicola* DSM 15968, and *K. pneumoniae* subsp. *pneumoniae* DSM 30104 were chosen for comparison from the curated database GenomesDB. Genomes with sequence identity above 95% were considered to be the same species.

Profiling of the allele sequences for seven housekeeping genes: *gapA* (glyceraldehyde 3-phosphate dehydrogenase), *infB* (translation initiation factor 2), *mdh* (malate dehydrogenase), *pgi* (phosphoglucose isomerase), *phoE* (phosphorine E), *rpoB* (betasubunit of RNA polymerase), and *tonB* (periplasmic energy transducer) and prediction of sequence types (STs) were determined via MLST inside the *Klebsiella pneumoniae* BIGSdb (Bacterial Isolate Genome Sequence Database) available at the Pasteur Institute (http://bigsdb.pasteur.fr/*Klebsiella*/*Klebsiella*.html) (Supplementary Table 2). In order to assign a capsule *locus* (KL) type in the genome data, the tool Kaptive[9] inside kleborate (v2.0)[56] was used. The best matching *locus* from the Kaptive's reference database was reported.

Virulence and resistance gene screening was also done using kleborate (v2.0)[56]. Kleborate examines four different virulence *loci* in *Klebsiella*: yersiniabactin (*ybt*), colibactin (*clb*), aerobactin (*iuc*) and salmochelin (*iro*). In addition to the virulence *loci*, Kleborate screens for the *rmpA* and *rmpA2* genes which result in a hypermucoid phenotype by upregulating capsule production. Resistance genes

are screened against the ARG-Annot database of acquired resistance genes (SRST2 version)[57,58]. Table 6 shows the drug classes kleborate scans for.

Identification of Genomic Islands (pathogenic island, metabolic island, resistance island and symbiotic island) in the genome of *Klebsiella pneumoniae* strain B31 was performed with GIPSy (Genomic Island Prediction Software)[59] through the detection of regions presenting (1) deviations in genomic signature (i.e., anomalous G+C and/or codon usage deviation); (2) presence of transposase, virulence or flanking tRNA genes; and (3) absence in the non-pathogenic organism *Klebsiella oxytoca* strain AR380. Presence of prophage inside the genome was identified using PHASTER[60].

Table 6: Drug classes of resistance genes screened by Kleborate.

Drug class	Abbreviation
Aminoglycosides	Agly
beta-lactamases*	Bla
broad spectrum beta-lactamases	Bla_broad
broad spectrum beta-lactamases with resistance to beta-lactamase inhibitors	Bla_broad_inhR
Carbapenemase	Bla_Carb
extended spectrum beta-lactamases	Bla_ESBL
extended spectrum beta-lactamases with resistance to beta-lactamase inhibitors	Bla_ESBL_inhR
Fosfomycin	Fcyn
Fluoroquinolones	Flq
Glycopeptides	Gly
Macrolides	MLS
Phenicols	Phe
Rifampin	Rif
Sulfonamides	Sul
Tetracyclines	Tet
Trimethoprim	Tmt

* Classification of beta-lactamases follows Lahey's classes (https://www.lahey.org/studies/).

Comparative genomics

To perform the comparative analysis, the complete chromosome sequence of additional 172 strains of *K. pneumoniae* (Supplementary Table 1) were selected. MLST, KL- typing, and virulence and resistance screening of these selected strains were performed as for the strain B31 (see subtopic above). A phylogenomic tree based on their core proteome was generated by the software PEPR (Phylogenomic

Estimation with Progressive Refinement) (https://github.com/enordber/pepr.git). The genomes of *Klebsiella oxytoca* KCTC 1686 and *Klebsiella oxytoca* E718 were used as the outgroup. The resulting Newick tree file was visualized and annotated with MLST, KL type, and virulence and resistance gene presence using iTOL v4.2[61] (Figures 1 and 2).

Based on this phylogenomic tree, 44 representative strains were chosen according to the most prevalent sequence types, with approximately 3 strains selected from each of these STs. The genome of the only strain also isolated in Brazil (Kp13) was also included. Gegenees (version 2.1)[18] was used to configure an identity matrix between these 44 genomes. Similarity percentages of their variable content were plotted in a heatmap chart with a spectrum ranging from low similarity (red) to high similarity (green). A dendrogram for the matrix was generated with the program SplitsTree (version 4.14.5)[62], using the neighbor joining method[63].

Prediction of orthologous genes was performed with the software OrthoFinder[64] and further classified with the use of an in-house script in: core genome, containing only genes that are commonly shared by all strains; shared genome, containing genes that are shared by two or more strains, but not all; and singletons, with strain-specific genes. Briefly, OrthoFinder uses the amino acid fasta files (.faa) of all Coding DNA Sequences (CDSs) in each genome to perform an all-vs-all BLASTp analysis. The sequences were then grouped using the Markov Clustering (MCL) algorithm to determine the orthologous genes. The pan-genome, core genome and singletons developments were calculated based on the mean values of the permutations of all genomes using, as described by Soares *et al*. in 2013[21]. The final curves were then fitted using an *in-house* script to estimate the fixed parameters for Heap's Law (pan-genome analyses) and least-squares fit of the exponential regression decay (core-genome and singletons). Briefly, we compared 44 genomes of *Klebsiella pneumoniae* to calculate the Pan-genome, core genome and singleton data set. The core genome was classified according to the functional categories of cluster of orthologous groups (COG) into (1) Information storage and processing; (2) signaling and cellular processes; (3) metabolism; and, (4) poorly characterized. To perform this classification, the CDSs of core genome were BLAST aligned against the myva database of COG, using an e-value of 1e-6.

References

1. Drancourt, A. L. Klebsiella. *Bergey's Man. Syst. Archaea Bact.* 1–26 (2015). doi:10.1002/9781118960608.gbm01150.
2. Ramos, P. I. P. *et al.* Comparative analysis of the complete genome of KPC-2-producing Klebsiella pneumoniae Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. *BMC Genomics* **15,** 1–16 (2014).
3. ULLMANN, R. P. A. U. Klebsiella spp. as Nosocomial Pathogens: Epidemiology, Taxonomy, Typing Methods, and Pathogenicity Factorso Title. *R. Pod. U. ULLMANN Dep. Med. Microbiol. Virol. Univ. Kiel, Kiel, Ger.* **11,** 589–603 (1998).
4. Podschun, R., Pietsch, S., Ho, C. & Ullmann, U. Incidence of Klebsiella Species in Surface Waters and Their Expression of Virulence Factors. *Appl. Environ. Microbiol.* **67,** 3325–3327 (2001).
5. Conlan, S., Kong, H. H. & Segre, J. A. Species-Level Analysis of DNA Sequence Data from the NIH Human Microbiome Project. *PLoS One* **7,** (2012).
6. Martin, R. M. & Bachman, M. A. Colonization, Infection, and the Accessory Genome of Klebsiella pneumoniae. *Front. Cell. Infect. Microbiol.* **8,** 1–15 (2018).
7. Dorman, M. J. & Short, F. L. Genome watch: Klebsiella pneumoniae: When a colonizer turns bad. *Nat. Rev. Microbiol.* **15,** 1 (2017).
8. Long, S. W. *et al.* Whole-Genome Sequencing of Human Clinical Klebsiella pneumoniae Isolates Reveals Misidentification and Misunderstandings of Klebsiella pneumoniae, Klebsiella variicola, and Klebsiella quasipneumoniae. **2,** 1–15 (2017).
9. Wyres, K. L. *et al.* Identification of Klebsiella capsule synthesis loci from whole genome data. *Microb. Genomics* **2,** 1–15 (2016).
10. Lam, M. M. C. *et al.* Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICEKp in Klebsiella pneumoniae populations. *bioRxiv* 98178 (2018). doi:10.1101/098178
11. Rodloff, A., Bauer, T., Ewig, S., Kujath, P. & Müller, E. Susceptible, intermediate, and resistant - the intensity of antibiotic action. *Dtsch. Ärzteblatt Int.* **105,** 657–62 (2008).
12. Magiorakos, a *et al.* Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Microbiology* **18,** 268–281 (2011).

13. Cannatelli, A. *et al.* MgrB inactivation is a common mechanism of colistin resistance in KPC-producing klebsiella pneumoniae of clinical origin. *Antimicrob. Agents Chemother.* **58,** 5696–5703 (2014).
14. Poirel, L. *et al.* The mgrB gene as a key target for acquired resistance to colistin in Klebsiella pneumoniae. *J. Antimicrob. Chemother.* **70,** 75–80 (2015).
15. Bagel, S., Hüllen, V., Wiedemann, B. & Heisig, P. Impact of gyrA and parC mutations on quinolone resistance, doubling time, and supercoiling degree of Escherichia coli. *Antimicrob. Agents Chemother.* **43,** 868–875 (1999).
16. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482,** 331–338 (2012).
17. Nicolás, M. F. *et al.* Comparative genomic analysis of a clinical isolate of Klebsiella quasipneumoniae subsp. similipneumoniae, a KP -2 and OKP-B-6 beta-lactamases producer harboring two drug-resistance plasmids from Southeast Brazil. *Front. Microbiol.* **9,** (2018).
18. Ågren, J., Sundström, A., Håfström, T. & Segerman, B. Gegenees: Fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One* **7,** (2012).
19. McInerney, J. O., McNally, A. & O'Connell, M. J. Why prokaryotes have pangenomes. *Nat. Microbiol.* **2,** 1–5 (2017).
20. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial 'pan-genome'. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 13950 LP-13955 (2005).
21. Soares, S. C. *et al.* The Pan-Genome of the Animal Pathogen Corynebacterium pseudotuberculosis Reveals Differences in Genome Plasticity between the Biovar ovis and equi Strains. *PLoS One* **8,** (2013).
22. Carlos Guimaraes, L. *et al.* Inside the Pan-genome - Methods and Software Overview. *Curr. Genomics* **16,** 245–252 (2015).
23. Anne-Sophie, B. & Benoit, G. Multidrug resistant (or antimicrobial-resistant) pathogens - Alternatives to new antibiotics? *Swiss Med. Wkly.* **147,** (2017).
24. Shakya, T. & Wright, G. D. Nucleotide selectivity of antibiotic kinases. *Antimicrob. Agents Chemother.* **54,** 1909–1913 (2010).
25. Bunny, K. L., Hall, R. M. & Stokes, H. W. New mobile gene cassettes containing an

aminoglycoside resistance gene, aacA7, and a chloramphenicol resistance gene, catB3, in an integron in pBWH301. *Antimicrob. Agents Chemother.* **39,** 686–693 (1995).

26. Stürenburg, E. & Mack, D. Extended-spectrum β-lactamases: Implications for the clinical microbiology laboratory, therapy, and infection control. *J. Infect.* **47,** 273–295 (2003).
27. 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.* **112,** E3574–E3581 (2015).
28. Caspar, Y. *et al.* Mcr-1 colistin resistance in ESBL-producing Klebsiella pneumoniae, France. *Emerg. Infect. Dis.* **23,** 874–876 (2017).
29. Wyres, K. L. & Holt, K. E. Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria. *Curr. Opin. Microbiol.* **45,** 131–139 (2018).
30. Struve, C., Bojer, M. & Krogfelt, K. A. Characterization of Klebsiella pneumoniae type 1 fimbriae by detection of phase variation during colonization and infection and impact on virulence. *Infect. Immun.* **76,** 4055–4065 (2008).
31. Gallique, M., Bouteiller, M. & Merieau, A. The type VI secretion system: A dynamic system for bacterial communication? *Front. Microbiol.* **8,** 1–10 (2017).
32. DeLeo, F. R. *et al.* Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. *Proc. Natl. Acad. Sci.* **111,** 4988–4993 (2014).
33. Gomez-Simmonds, A. & Uhlemann, A.-C. Clinical Implications of Genomic Adaptation and Evolution of Carbapenem-Resistant Klebsiella pneumoniae. *J. Infect. Dis.* **215,** S18–S27 (2017).
34. Moxon, E. R. & Kroll, J. S. The Role of Bacterial Polysaccharide Capsules as Virulence Factors. in *Bacterial Capsules* (eds. Jann, K. & Jann, B.) 65–85 (Springer Berlin Heidelberg, 1990).
35. Hill, C. Virulence or niche factors: What's in a name? *J. Bacteriol.* **194,** 5725–5727 (2012).
36. Tauch, A. & Burkovski, A. Molecular armory or niche factors: Virulence determinants of corynebacterium species. *FEMS Microbiol. Lett.* **362,** 1–6 (2015).
37. ØRSKOV, I. D. A. & FIFE-ASBURY, M. A. New Klebsiella Capsular Antigen, K82, and the Deletion of Five of Those Previously Assigned. *Int. J. Syst. Bacteriol.* **27,** 386–387

(1977).

38. Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. B Biol. Sci.* **361,** 1929–1940 (2006).
39. Goris, J. *et al.* DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57,** 81–91 (2007).
40. Kim, M., Oh, H. S., Park, S. C. & Chun, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **64,** 346–351 (2014).
41. Breurec, S. *et al.* Klebsiella pneumoniae resistant to third-generation cephalosporins in five African and two Vietnamese major towns: Multiclonal population structure with two major international clonal groups, CG15 and CG258. *Clin. Microbiol. Infect.* **19,** 349–355 (2013).
42. Lee, M. Y. *et al.* High prevalence of CTX-M-15-producing Klebsiella pneumoniae isolates in Asian countries: Diverse clones and clonal dissemination. *Int. J. Antimicrob. Agents* **38,** 160–163 (2011).
43. Tatusov, R. L. A Genomic Perspective on Protein Families. *Science (80-.).* **278,** 631–637 (1997).
44. Guilhen, C. *et al.* Transcriptional profiling of Klebsiella pneumoniae defines signatures for planktonic, sessile and biofilm-dispersed cells. *BMC Genomics* **17,** (2016).
45. Andrews, S. FastQC: A quality control tool for high throughput sequence data. (2010).
46. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30,** 31–37 (2014).
47. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075 (2013).
48. Piro, V. C. *et al.* FGAP: An automated gap closing tool. *BMC Res. Notes* **7,** 1–5 (2014).
49. Carattoli, A. *et al.* In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58,** 3895–3903 (2014).
50. Antipov, D. *et al.* PlasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics* **32,** 3380–3387 (2016).
51. Brettin, T. *et al.* RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes.

Sci. Rep. **5,** (2015).

52. Wattam, A. R. *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42,** 581–591 (2014).
53. Carver, T. *et al.* Artemis and ACT: Viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24,** 2672–2676 (2008).
54. Apweiler, R. *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32,** D115-9 (2004).
55. Richter, M., Rosselló-Móra, R., Oliver Glöckner, F. & Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **32,** 929–931 (2015).
56. Lam, M. M. C. *et al.* Frequent emergence of pathogenic lineages of Klebsiella pneumoniae via mobilisation of yersiniabactin and colibactin. *bioRxiv* (2017). doi:http://dx.doi.org/10.1101/098178
57. Gupta, S. K. *et al.* ARG-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **58,** 212–220 (2014).
58. Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **6,** 1–16 (2014).
59. Soares, S. C. *et al.* GIPSy: Genomic island prediction software. *J. Biotechnol.* **232,** 2–11 (2016).
60. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44,** W16–W21 (2016).
61. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44,** W242–W245 (2016).
62. Kloepper, T. H. & Huson, D. H. Drawing explicit phylogenetic networks and their integration into SplitsTree. *BMC Evol. Biol.* **8,** 22 (2008).
63. Huson, D. H. & Bryant, D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* **23,** 254–267 (2006).
64. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16,** 1–14 (2015).

5. Considerações finais e perspectivas

A montagem do genoma do isolado B31 e consequentes análises comparativas realizadas neste trabalho evidenciaram a grande plasticidade existente entre os genomas de *K. pneumoniae*, o que provavelmente está relacionado à capacidade de adaptação aos diferentes ambientes em que a bactéria se encontra. A linhagem B31, isolada do lavado broncoalveolar de uma paciente infectada no Brasil, foi confimada como sendo *K. pneumoniae* e apresenta um repertório de genes de resistência que a caracteriza como um patógeno multirressistente. Apesar disso, não foram encontrados genes de virulência que a caracterizem como hipervirulenta, salientando o aspecto oportunista da espécie.

O melhor *locus* correspondente para tipagem capsular para o isolado B31 foi o KL19, que corresponde a um dos 77 tipos sorologicamente testados. Além disso, os resultados obtidos mostram que B31 pertence a um ST ainda não descrito no Brasil (ST15), mas que já foi descrito em vários outros locais do mundo. Assim, análises futuras que possam determinar a dinâmica de circulação desta bactéria no Brasil serão de grande importância.

Devido ao fato do pangenoma obtido estar aberto, novos genomas deve ser acrescido às análises para confirmar tal achado. Assim, além disso, estudar a rede metabólica do isolado B31 para identificar reações importantes e potenciais alvos de controle deste patógeno, bem como cruzar estes dados com dados de análises da expressão diferencial deste e de outros isolados para o melhor entendimento dos mecanismos associados à sua patogênese são perspectivas deste trabalho.

6. Referências bibliográficas

AHMAD, S. et al. Subtractive Genomics, Molecular Docking and Molecular Dynamics Simulation Revealed LpxC as a Potential Drug Target Against Multi-Drug Resistant Klebsiella pneumoniae. **Interdisciplinary Sciences: Computational Life Sciences**, v. 0, n. 0, p. 0, 2018.

BENNETT, P. M. Genome Plasticity. In: WOODFORD, N.; JOHNSON, A. P. (Eds.). . **Genomics, Proteomics, and Clinical Bacteriology: Methods and Reviews**. Totowa, NJ: Humana Press, 2004. p. 71–113.

BRUNO, W. J.; SOCCI, N. D.; HALPERN, A. L. Weighted Neighbor Joining: A Likelihood-Based Approach to Distance-Based Phylogeny Reconstruction. n. July, p. 189–197, 1987.

CALFEE, D. P. Recent advances in the understanding and management of Klebsiella pneumoniae. **F1000Research**, v. 6, p. 1760, 2017.

CARLOS GUIMARAES, L. et al. Inside the Pan-genome - Methods and Software Overview. **Current Genomics**, v. 16, n. 4, p. 245–252, 2015.

CASPAR, Y. et al. Mcr-1 colistin resistance in ESBL-producing Klebsiella pneumoniae, France. **Emerging Infectious Diseases**, v. 23, n. 5, p. 874–876, 2017.

CHAN, C. X.; RAGAN, M. A. Next-generation phylogenomics. p. 1–6, 2013.

CHOUDHURI, S. Phylogenetic Analysis. In: **Bioinformatics for Beginners: genes, genomes, molecular evolution, databases and analytical tool**. [s.l.] Elsevier Inc., 2014. p. 209–218.

DIANCOURT, L. et al. Multilocus Sequence Typing of Klebsiella pneumoniae Nosocomial Isolates Multilocus Sequence Typing of Klebsiella pneumoniae Nosocomial Isolates. **Journal of clinical microbiology**, v. 43, n. 8, p. 4178–4182, 2005.

DRANCOURT, A. L. Klebsiella. **Bergey's Manual of Systematics of Archaea and Bacteria**, p. 1–26, 2015.

DUGUID, J. P. Fimbriae and Adhesive Properties in Klebsiella Strains. **Journal of General Microbiology**, v. 21, n. 1, p. 271–286, 1959.

FEIL, E. J. et al. eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. **Journal of Bacteriology**, v. 186, n. 5, p. 1518–1530, 2004.

HACKER, J. et al. Ecological fitness, genomic islands and bacterial pathogenicity A

Darwinian view of the evolution of microbes. **EMBO reports**, v. 2, n. 5, p. 376–381, 2001.

HAGGERTY, L. S. et al. A pluralistic account of homology: Adapting the models to the data. **Molecular Biology and Evolution**, v. 31, n. 3, p. 501–516, 2014.

HOLT, K. E. et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae* , an urgent threat to public health. **Proceedings of the National Academy of Sciences**, v. 112, n. 27, p. E3574–E3581, 2015.

LEE, B. et al. Case report Community acquired Klebsiella pneumoniae meningitis : a case report. v. 8, n. June, p. 92–95, 2018.

LEE, M. Y. et al. High prevalence of CTX-M-15-producing Klebsiella pneumoniae isolates in Asian countries: Diverse clones and clonal dissemination. **International Journal of Antimicrobial Agents**, v. 38, n. 2, p. 160–163, 2011.

LIMA, W. C. et al. Genome sequencing and functional characterization of the non-pathogenic Klebsiella pneumoniae KpGe bacteria. **Microbes and Infection**, v. 20, n. 5, p. 293–301, 2018.

LONG, S. W. et al. Whole-Genome Sequencing of Human Clinical Klebsiella pneumoniae Isolates Reveals Misidentification and Misunderstandings of Klebsiella pneumoniae, Klebsiella variicola, and Klebsiella quasipneumoniae. v. 2, n. 4, p. 1–15, 2017.

MARTIN, R. M.; BACHMAN, M. A. Colonization, Infection, and the Accessory Genome of Klebsiella pneumoniae. **Frontiers in Cellular and Infection Microbiology**, v. 8, n. January, p. 1–15, 2018.

MARTINEZ, J. L. General principles of antibiotic resistance in bacteria. **Drug Discovery Today: Technologies**, v. 11, n. 1, p. 33–39, 2014.

MCINERNEY, J. O.; MCNALLY, A.; O'CONNELL, M. J. Why prokaryotes have pangenomes. **Nature Microbiology**, v. 2, n. March, p. 1–5, 2017.

N SAITOU, M. N. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biol ogy and Evolution**, v. 4, n. 4, p. 406–425, 1987.

NICOLÁS, M. F. et al. Comparative genomic analysis of a clinical isolate of Klebsiella quasipneumoniae subsp. similipneumoniae, a KP -2 and OKP-B-6 beta-lactamases producer harboring two drug-resistance plasmids from Southeast Brazil. **Frontiers in Microbiology**, v. 9, n. FEB, 2018.

OCHMAN, H.; LAWRENCE, J. G.; GROISMAN, E. Lateral gene transfer and the nature of bacterial. **innovation, Nature**, v. 405, p. 299–304, 2000.

PODSCHUN, R. et al. Incidence of Klebsiella Species in Surface Waters and Their Expression of Virulence Factors. **Applied and Environmental Microbiology**, v. 67, n. 7, p.

3325–3327, 2001.

PRENTICE, M. B. Bacterial comparative genomics. 2004.

RAMOS, P. I. P. et al. Comparative analysis of the complete genome of KPC-2-producing Klebsiella pneumoniae Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. **BMC Genomics**, v. 15, n. 1, p. 1–16, 2014.

RANJBAR, R. et al. Typing methods used in the molecular epidemiology of microbial pathogens: a how-to guide. **The new microbiologica**, v. 37, n. 1, p. 1–15, 2014.

SAN MILLAN, A. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. **Trends in Microbiology**, v. xx, p. 1–8, 2018.

SIU, L. K. et al. Klebsiella pneumoniae liver abscess: A new invasive syndrome. **The Lancet Infectious Diseases**, v. 12, n. 11, p. 881–885, 2012.

STOESSER, N. et al. Genome sequencing of an extended series of NDM-producing Klebsiella pneumoniae isolates from neonatal infections in a Nepali hospital characterizes the extent of community- Versus hospital- associated transmission in an endemic setting. **Antimicrobial Agents and Chemotherapy**, v. 58, n. 12, p. 7347–7357, 2014.

TETTELIN, H. et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome". **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 39, p. 13950 LP-13955, 27 set. 2005.

TOFT, C.; ANDERSSON, S. G. E. Evolutionary microbial genomics: Insights into bacterial host adaptation. **Nature Reviews Genetics**, v. 11, n. 7, p. 465–475, 2010.

TRUONG, W. R.; YAMAKI, J. The Hospital Antimicrobial Use Process: From Beginning to End. **Open Forum Infectious Diseases**, v. 5, n. 6, p. 1–5, 2018.

ULLMANN, R. P. A. U. Klebsiella spp. as Nosocomial Pathogens: Epidemiology, Taxonomy, Typing Methods, and Pathogenicity Factorso Title. **R. PODSCHUN* AND U. ULLMANN Department of Medical Microbiology and Virology, University of Kiel, Kiel, Germany**, v. 11, n. 4, p. 589–603, 1998.

WYRES, K. L. et al. Identification of Klebsiella capsule synthesis *loci* from whole genome data. **Microbial Genomics**, v. 2, n. 12, p. 1–15, 2016.

WYRES, K. L.; HOLT, K. E. Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria. **Current Opinion in Microbiology**, v. 45, p. 131–139, 2018.

YANG, Z.; RANNALA, B. Molecular phylogenetics: Principles and practice. **Nature Reviews Genetics**, v. 13, n. 5, p. 303–314, 2012.

7. Material Suplementar

Supplementary Table 1. All strains information. ST and the best match *locus* for K-type are shown.

Kp Strain	ST	K-type	Match confidence	Identity
1084	ST23	KL1	None	99.94%
ED2	ST23	KL1	Very high	99.98%
ED23	ST23	KL1	Very high	99.99%
KP11	ST700-1LV	KL1	Very high	97.07%
KP7	ST1941	KL1	Very high	99.99%
KP9	ST23	KL1	Very high	99.99%
NTUH-K2044	ST23	KL1	Perfect	100.00%
P1428	ST23	KL1	Very high	99.99%
RJA166	ST23	KL1	Perfect	100.00%
RJF999	ST23	KL1	Perfect	100.00%
1158	ST65	KL2	Very high	98.39%
HK787	ST86	KL2	Very high	99.92%
isolate 11	ST14	KL2	Very high	99.74%
MLST-15	ST15	KL2	Very high	99.53%
AR_0068	ST14	KL2	Very high	99.74%
AR_0143	ST14	KL2	Very high	99.74%
ATCC 43816 KPPR1	ST493	KL2	Very high	99.72%
CG43	ST86	KL2	None	99.88%
isolate 23	ST14	KL2	Very high	99.75%
KCTC 2242	ST375	KL2	Very high	98.40%
Kp52.145	ST66	KL2	High	98.32%
KP617	ST14	KL2	Good	99.73%
kp757	ST86	KL2	Very high	99.93%
KPN528	ST14	KL2	Very high	99.74%
NUHL24835	ST14	KL2	Good	99.74%
PittNDM01	ST14	KL2	Very high	99.74%
RJF293	ST374	KL2	High	99.71%
U25	ST14	KL2	Very high	99.74%
KP5-1	ST206-2LV	KL3	High	97.02%
SB3432	ST67	KL3	Very high	99.48%
KP10	ST29	KL5	None	89.02%
KP8	ST485	KL5	Low	99.21%
QS17-0161	ST45	KL7	Very high	98.51%
1756	ST2424-1LV	KL8	Very high	99.85%
HKUOPLC	ST480-2LV	KL8	Very high	98.68%
KPN1481	ST906	KL10	Very high	96.33%
459	ST37	KL11	None	74.88%
AR_0158	ST163	KL12	High	99.90%
INF042	ST37	KL14	Very high	99.98%
INF059	ST37	KL14	Very high	99.98%
INF163	ST340	KL15	High	99.72%
INF164	ST340	KL15	Very high	99.91%
CAV1042	ST244	KL15	Very high	99.97%
FDAARGOS_436	ST340	KL15	Very high	99.76%

Kp Strain	ST	K-type	Match confidence	Identity
INF157	ST340	KL15	Very high	99.91%
INF158	ST340	KL15	Very high	99.91%
INF274	ST340	KL15	Very high	99.90%
INF278	ST340	KL15	Very high	99.90%
Kp_Goe_821588	ST11	KL15	Very high	99.78%
Kp_Goe_822917	ST11	KL15	Very high	99.78%
KPNIH39	ST37	KL15	Good	99.78%
KSB1_5D	ST340	KL15	Very high	99.91%
KSB1_9D	ST340	KL15	Very high	99.91%
Kp_Goe_121641	ST101	KL17	Very high	99.99%
Kp_Goe_33208	ST101	KL17	Very high	99.98%
Kp_Goe_71070	ST101	KL17	Very high	99.98%
AR_0107	ST1161	KL19	High	94.92%
B31	ST15	KL19	High	94.91%
FDAARGOS_446	ST340	KL19	High	94.92%
KSB1_4E	ST323	KL21	Perfect	100.00%
KSB1_7E	ST323	KL21	Perfect	100.00%
KSB2_1B	ST323	KL21	Perfect	100.00%
19051	ST15	KL24	Good	99.98%
AR_0146	ST11	KL24	Very high	99.99%
AR_0148	ST11	KL24	Very high	99.99%
Kp_Goe_39795	ST15	KL24	High	99.99%
AR_0126	ST45	KL25	High	98.53%
CAV1016	ST45	KL25	High	98.53%
CAV1193	ST941	KL25	High	99.85%
CAV1344	ST941	KL25	High	99.85%
CN1	ST392	KL27	Very high	99.27%
KPNIH31	ST392	KL27	Very high	99.27%
342	ST146	KL30	Very high	95.40%
AR_0117	ST234	KL30	Good	96.36%
INF249	ST29	KL30	Good	96.55%
INF322	ST29	KL30	Good	96.55%
TGH10	ST383	KL30	Good	96.40%
TGH8	ST383	KL30	Very high	96.40%
TH1	ST1536	KL30	Very high	99.17%
W14	ST1536	KL30	Very high	99.17%
strain 234-12	ST514	KL35	High	99.31%
KPNIH27	ST34	KL37	Good	99.92%
XH209	ST17	KL38	Very high	99.00%
721005	ST11	KL47	High	99.30%
911021	ST11	KL47	Very high	99.32%
FDAARGOS_443	ST11	KL47	Very high	99.41%
strain SWU01	ST11	KL47	Very high	99.40%
UCLAOXA232KP_Pt0	ST16	KL51	High	99.92%
FDAARGOS_440	ST16	KL51	Very high	99.94%
UCLAOXA232KP	ST16	KL51	High	99.93%
Kp Strain	**ST**	**K-type**	**Match confidence**	**Identity**

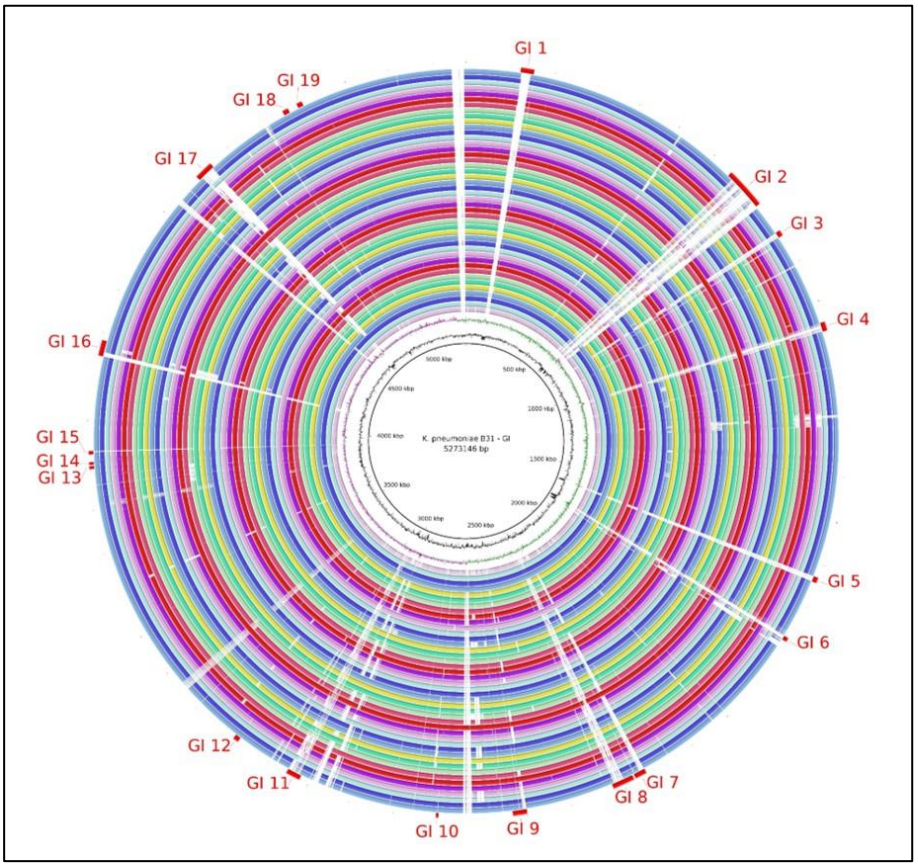
Kpn555	ST514-2LV	KL52	Very high	96.60%
MGH 78578	ST38	KL52	Perfect	100.00%
CAV1217	ST340	KL55	High	96.27%
CAV1417	ST340	KL55	Very high	96.27%
NY9	ST340	KL55	Very high	96.27%
VK055	ST749-2LV	KL57	Very high	96.27%
DHQP1002001	ST34	KL57	High	97.37%
Kp_Goe_154414	ST23	KL57	Good	98.81%
34618	ST258	KL63	Very high	98.81%
825795-1	ST147	KL64	Perfect	100.00%
AATZP	ST147	KL64	Very high	99.99%
AR_0138	ST147	KL64	Perfect	100.00%
AR_0145	ST147	KL64	Perfect	100.00%
AR_0152	ST147	KL64	Perfect	100.00%
ATCC 35657	ST505	KL64	Very high	99.90%
FDAARGOS_439	ST147	KL64	Perfect	100.00%
FDAARGOS_442	ST147	KL64	Very high	100.00%
FDAARGOS_444	ST11	KL64	Very high	99.91%
Kp_Goe_149473	ST147	KL64	Perfect	100.00%
Kp_Goe_149832	ST147	KL64	Perfect	100.00%
Kp_Goe_152021	ST147	KL64	Perfect	100.00%
Kp_Goe_822579	ST147	KL64	Perfect	100.00%
Kp_Goe_827024	ST147	KL64	Perfect	100.00%
Kp_Goe_827026	ST147	KL64	Very high	99.99%
KP_Goe_828304	ST147	KL64	Perfect	100.00%
KP36	ST15	KL64	High	99.99%
KP5	ST147	KL64	Perfect	100.00%
KPNIH29	ST1518	KL64	Very high	98.22%
MS6671.v1	ST147	KL64	Perfect	100.00%
SKGH01	ST147	KL64	Perfect	100.00%
TGH13	ST147	KL64	Perfect	100.00%
YH43	0	KL67	Very high	96.86%
AR_0049	ST11	KL74	Low	98.88%
ATCC BAA-2146	ST11	KL74	Low	98.88%
Kpn223	ST273	KL74	Very high	98.95%
KP1766	ST307	KL102	High	99.27%
KP1768	ST307	KL102	Very high	99.27%
NR5632	ST307	KL102	Very high	99.27%
HS11286	ST11	KL103	Very high	96.69%
AR_0139	ST37	KL105	High	99.99%
KP38731	ST11	KL105	None	99.94%
KPN1482	ST37	KL105	High	99.97%
32192	ST258	KL106	Good	99.98%
207M1D0-sc	ST258	KL106	Good	99.98%
AR_0098	ST258	KL106	Good	99.97%
BIC-1	ST258	KL106	Good	99.97%
Kp Strain	**ST**	**K-type**	**Match confidence**	**Identity**
CAV1453	ST258	KL106	Good	99.98%

KPNIH24	ST258	KL106	Good	99.97%
KPNIH30	ST258	KL106	Good	99.99%
KPR0928	ST258	KL106	Good	99.99%
KPNIH36	ST258	KL107	High	99.99%
30660/NJST258_1	ST258	KL107	Very high	99.98%
30684/NJST258_2	ST258	KL107	Very high	99.99%
AR_0047	ST258	KL107	Very high	99.99%
AR_0112	ST258	KL107	Very high	99.99%
AR_0113	ST258	KL107	Very high	99.99%
AR_0115	ST258	KL107	Very high	99.99%
AR_0120	ST258	KL107	Very high	99.99%
AR_0125	ST258	KL107	Very high	99.99%
AR_0129	ST258	KL107	Very high	99.99%
BK13043	ST258	KL107	Very high	99.99%
CAV1596	ST258	KL107	Very high	99.98%
CR14	ST258	KL107	Very high	99.97%
H11	ST659	KL107	Very high	99.97%
isolate blood sample 2	ST512	KL107	Very high	99.99%
kp13	ST442	KL107	Very high	99.98%
KPNIH32	ST258	KL107	Very high	99.99%
KPNIH33	ST258	KL107	Very high	99.99%
MNCRE53	ST258	KL107	Very high	99.98%
MNCRE69	ST258	KL107	Very high	99.99%
MNCRE78	ST258	KL107	Very high	99.99%
yzusk-4	ST1665-2LV	KL107	None	82.30%
blaNDM-1	ST395	KL108	Perfect	100.00%
Kp_Goe_62629	ST395	KL108	Perfect	100.00%
BR	ST15	KL112	Very high	98.63%
PMK1	ST15	KL112	None	98.28%
J1	ST111	KL113	High	96.55%
CAV1392	ST11	KL125	High	90.39%
JM45	ST11	KL125	High	90.39%
KP14	ST29-1LV	KL135	None	86.63%
KpN01	ST278	KL136	Very high	99.65%
KpN06	ST278	KL136	Very high	99.65%
FDAARGOS_447	ST152	KL149	Very high	99.98%

Supplementary Table 2: Sequence Type (ST) prediction and allele profiling of the housekeeping genes for all strains analyzed. The number of strains for each ST is also represented.

ST	N^0 of strains	*gapA*	*infB*	*mdh*	*pgi*	*phoE*	*rpoB*	*tonB*
ST258	27	3	3	1	1	1	1	79
ST147	18	3	4	6	1	7	4	38
ST11	15	3	3	1	1	1	1	4
ST340	13	3	3	1	1	1	1	18
ST14	9	1	6	1	1	1	1	1
ST23	9	2	1	1	1	9	4	12
ST15	7	1	1	1	1	1	1	1
ST37	6	2	9	2	1	13	1	16
ST29	4	2	3	2	2	6	4	4
ST101	3	2	6	1	5	4	1	6
ST16	3	2	1	2	1	4	4	4
ST307	3	4	1	2	52	1	1	7
ST323	3	2	1	1	1	9	1	93
ST45	3	2	1	1	6	7	1	12
ST86	3	9	4	2	1	1	1	27
ST1536	2	2	1	2	37	45	4	9
ST278	2	4	1	1	1	12	1	4
ST34	2	2	3	6	1	9	7	4
ST383	2	2	6	1	3	8	1	18
ST392	2	3	4	6	1	7	4	40
ST395	2	3	1	2	4	1	1	4
ST514	2	2	1	1	1	8	1	9
ST749	2	18	23	26	61	11	39	99
ST941	2	6	3	1	20	12	4	4
ST111	1	2	1	5	1	17	4	42
ST1161	1	2	3	2	2	6	4	111
ST146	1	16	24	30	27	36	22	55
ST1518	1	14	1	2	119	21	1	1
ST152	1	2	3	2	1	1	4	56
ST163	1	2	1	1	1	9	1	12
ST1665	1	43	3	5	1	1	4	61
ST17	1	2	1	1	1	4	4	4
ST1941	1	2	102	1	1	9	4	12
ST206	1	16	18	36	40	153*	22	67
ST234	1	2	1	2	1	7	1	24
ST2424	1	2	1	37	1	3	1	56
ST244	1	2	5	1	1	1	1	24
ST273	1	3	4	6	1	7	4	4
ST374	1	2	3	58	37	10	27	9
ST65	1	66	1	65	1	9	11	18
ST659	1	66	1	65	1	9	11	18
ST66	1	2	3	2	1	10	1	13
ST67	1	2	1	9	1	15	5	28

Supplementary Table 2 - Continued								
ST	N^0 of strains	*gapA*	*infB*	*mdh*	*pgi*	*phoE*	*rpoB*	*tonB*
ST700	1	10	1	17	37	12	1	-
ST906	1	16	62	21	27	55	22	75
ST375	1	43	1	2	1	10	4	13
ST38	1	2	1	2	1	2	2	2
ST442	1	2	1	2	1	2	2	2
ST480	1	18	22	55*	16	11	13	51
ST485	1	2	1	1	1	7	1	12
ST505	1	7	1	5	1	1	1	84
ST512	1	54	3	1	1	1	1	79



Supplementary Figure 1: Circular genome map of 44 *K. pneumoniae* genomes and pathogenic island found in Kp B31. The strains were aligned using *K. pneumoniae B31* as a reference. The figure represents the CDS, coding sequences; tRNA, transfer RNA; rRNA, ribosomal RNA; GIs (Genomic Islands) and PIs (Pathogenicity Island).