Data in Brief 34 (2021) 106746

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Characterization of a new multidrug-resistant Brazilian *K. pneumoniae* isolate and 172 *Klebsiella* spp. sequenced strains: Genomic island, multilocus sequence typing and capsule locus dataset

Rodrigo Profeta[a], Núbia Seyffert[a,b], Sandeep Tiwari[a], Marcus V.C. Viana[a], Arun Kumar Jaiswal[a,c], Ana Carolina Caetano[a], Daniel Henrique Bücker[a], Luciana Tavares de Oliveira[d], Roselane Santos[a], Alfonso Gala-Garcia[a], Rodrigo B. Kato[a], Francine F. Padilha[e], Isabel B. Lima-Verde[f], Preetam Ghosh[g], Debmalya Barh[h], Aristóteles Góes-Neto[i], Henrique C.P. Figueiredo[j], Siomar C. Soares[a,c], Roberto Meyer[b], Bertram Brenig[k], Pablo I.P. Ramos[l], Vasco Azevedo[a,†], Thiago L.P. Castro[a,b,†,*]

[a] *Department of Genetics, Ecology and Evolution (GEE), Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[b] *Institute of Health Sciences, Federal University of Bahia, Salvador, BA, Brazil*
[c] *Department of Microbiology, Immunology and Parasitology, Institute of Biological and Natural Sciences, Federal University of Triângulo Mineiro, Uberaba, MG, Brazil*
[d] *Laboratory of Clinical Pathology, Hospital das Clínicas, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[e] *Tiradentes University, Technology and Research Institute, Biomaterials Laboratory, Aracaju, Sergipe, Brazil*
[f] *Swedish University of Agricultural Sciences, Uppsala, Sweden*
[g] *Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia, 23284, USA*
[h] *Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, West Bengal 721137, India*
[i] *Molecular and Computational Biology of Fungi Laboratory, Department of Microbiology, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[j] *AQUACEN – National Reference Laboratory of Aquatic Animal Disease, Ministry of Agriculture, Livestock and Food Supply, Veterinary School, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil*
[k] *Institute of Veterinary Medicine, University of Göttingen, Burckhardtweg 2, Göttingen, Germany*
[l] *Gonçalo Moniz Institute, Oswaldo Cruz Foundation, Salvador, Bahia, Brazil*

https://doi.org/10.1016/j.dib.2021.106746

ARTICLE INFO

Article history:
Received 19 December 2020
Revised 7 January 2021
Accepted 8 January 2021
Available online 14 January 2021

Keywords:
Genomic island
Multilocus sequence typing
Capsular typing
Klebsiella pneumoniae
Antibiotic resistance
Bacterial virulence

ABSTRACT

The genus *Klebsiella* comprises species that cause nosocomial and community-acquired infections. A dataset was created to compile the sequence type (ST) and capsule type (K-locus) information predicted for 172 worldwide isolates of *Klebsiella* spp. whose complete genomes could be retrieved from the GenBank (NCBI) repository. The dataset also includes information related to one multidrug-resistant strain (B31) isolated from a patient who was admitted to an intensive care unit in the Northeast region of Brazil. This strain was phenotypically characterized and submitted to whole-genome sequencing and comparative genomics analysis as we recently reported [1]. The dataset also compiles information on Pathogenicity Islands (PIs), Resistance Islands (RIs) and Miscellaneous Islands (MIS) present in the genome of strain B31. The information provided here may support outbreak prevention policies and future epidemiological studies involving *Klebsiella* spp.

Specifications Table

Subject	Omics: Genomics
Specific subject area	Bacterial genomics
Type of data	Primary data, Tables, Charts
How data were acquired	Whole genome sequence of B31 isolate was obtained using Illumina HiSeq 2500 (Illumina, San Diego, CA, USA). Data from the other 172 strains was gathered using Genbank.
Data format	Raw, Analyzed, Filtered
Parameters for data collection	The complete genomes assigned to 172 *Klebsiella pneumoniae* strains in the publicly available GenBank (NCBI) repository were retrieved for data analysis. In addition, we considered the complete genome of strain *K. pneumoniae* B31, which is a clinical isolate from the Northeast region of Brazil and is characterized by our group.
Description of data collection	The data analyses were conducted using bioinformatics tools. Sequence type (ST) and capsule type (K-locus) information for all *K. pneumoniae* strains were predicted using Kleborate (v2.0). Pathogenicity and antibiotic resistance islands in strain B31 were predicted using GIPSy (Genomic Island Prediction Software).
Data source location	GenBank (NCBI) repository (https://www.ncbi.nlm.nih.gov/nuccore/CP035929) Institutions: Federal University of Minas Gerais and Federal University of Bahia Cities: Belo Horizonte and Salvador Country: Brazil

(*continued on next page*)

DOI of original article: 10.1016/j.gene.2020.145386
* Corresponding author.
E-mail addresses: thiago.luiz@ufba.br, castrotlp@gmail.com (T.L.P. Castro).
† These authors share senior authorship.

Data accessibility	The raw sequencing data of strain B31 have been deposited in the NCBI SRA database. The data are accessible via the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA521748 Dataset is available in the Mendeley repository via the link: https://data.mendeley.com/datasets/fskkcpkwk2 Available files are: Sequence type and capsular type of 173 Klebsiella spp. strains.pdf MIS_list.xlsx (Miscellaneous Islands predicted in B31) PI_list.xlsx (Pathogenicity Islands predicted in B31) RI_list.xlsx (Resistance Islands predicted in B31)
Related research article	R. Profeta, N. Seyffert, S. Tiwari, M.V.C. Viana, A.K. Jaiswal, A.C. Caetano, D.H. Bücker, L.T. de Oliveira, R. Santos, A. Gala-Garcia, R.B. Kato, F.F. Padilha, I.B. Lima-Verde, P. Ghosh, D. Barh, A. Góes-Neto, H.C.P. Figueiredo, T.L.P. Castro, S.C. Soares, R. Meyer, B. Brenig, P.I.P. Ramos, V. Azevedo, Comparative genomics with a multidrug-resistant Klebsiella pneumoniae isolate reveals the panorama of unexplored diversity in Northeast Brazil, Gene. (2020). https://doi.org/10.1016/j.gene.2020.145386.

Value of the Data

- The dataset comprises relevant information on *Klebsiella* spp. isolates of prominent public health concern due to the involvement in pathogenicity and antimicrobial resistance spread.
- The dataset provides direct access to the sequence type (ST) and capsule type (K-locus) information of 173 worldwide isolates of *Klebsiella* spp.
- The dataset may be expanded or used in new characterization studies on *Klebsiella* spp. For example, it is possible to explore resistance and virulence determinants and make associations with bacterial types.

1. Data Description

K. pneumoniae and other phylogenetically related species cause community-acquired and nosocomial infections [2]. The importance of these pathogens as a public health concern has recently raised along with the high incidence of antibiotic-resistant strains [3]. The characterization of clinical *Klebsiella* spp. as to their phenotypic diversity and gene constitution is extremely important to assist the treatment of patients and explore epidemiologic aspects such as the distribution of multidrug-resistant strains. In addition, valuable information is provided to understand disease severity and outbreaks, supporting prevention policies of nosocomial infections caused by *Klebsiella spp*.

This dataset is an extension of our recent study on the comparative genomics analysis with a multidrug-resistant *Klebsiella pneumoniae* isolate (B31) from the Northeast of Brazil and other 172 worldwide *Klebsiella* spp. strains [1]. Contributing to the genomic characterization of *K. pneumoniae*, the study included genomic DNA sequencing, bacterial typing, and screening of virulence and drug resistance determinants. The complete dataset can be found in the Mendeley Data repository, via the link https://data.mendeley.com/ datasets/fskkcpkwk2.

Among all 173 strains analysed, 52 STs and 42 K-loci were reported (see section Data accessibility in the Specifications Table). ST258, ST147, ST11, and ST340 were the most prevalent STs found in the dataset. Allele identification for each ST is likewise described.

Table 1). The capsular typing revealed KL64, KL107, KL2, and KL15 as the most abundant K-loci in the dataset (Fig. 1).

Pathogenicity Islands (PI), Resistance Islands (RIs), and Miscellaneous Islands (MIS) in the genome of *K. pneumoniae* B31 were predicted (see section Data accessibility in the Specifications Table). B31 presented 11 RIs, 14 PIs, and 8 MISs containing both antibiotic resistance and pathogenicity related genes (Fig. 2). The number of genes found in every genomic island of B31 is depicted in Fig. 3.

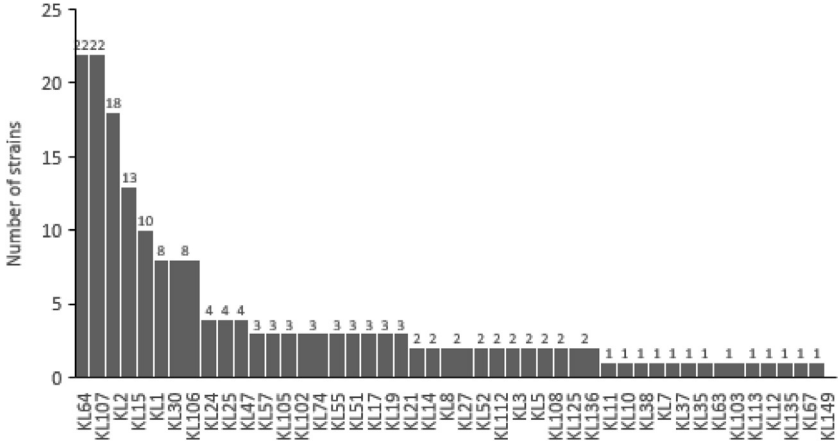

Fig. 1. Chart of number of strains for every capsule locus (KL) found in the dataset.

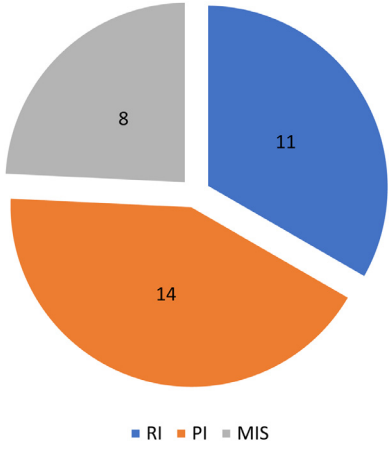

Fig. 2. Total number of genomic islands predicted in the genome of strain *K. pneumoniae* B31.

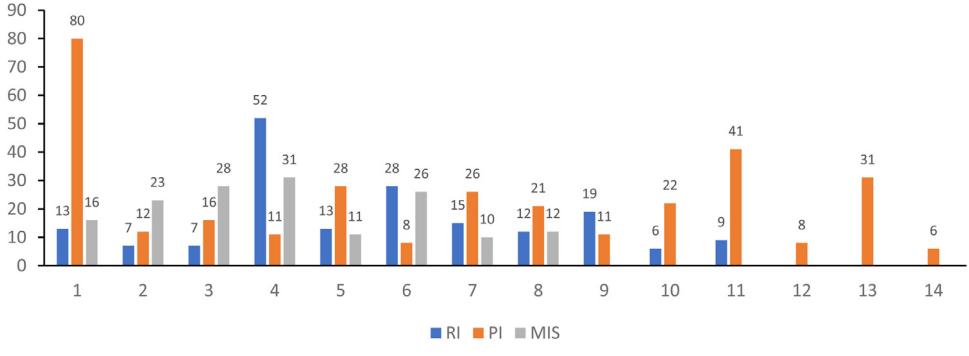

Fig. 3. Number of genes predicted in every genomic island of strain B31.

Table 1
Allele profiling and sequence type (ST) prediction for the 173 *Klebsiella* spp. isolates.

		Allele identification						
Sequence type	Number 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
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

2. Experimental Design, Materials and Methods

2.1. Selection of genomes

The complete genome of one multidrug resistant *K. pneumoniae* strain (B31), isolated from a patient admitted to an intensive care unit in the Northeast of Brazil, was sequenced using

Illumina Inc. technology and assembled as described in our recent study on the comparative genomics with *Klebsiella* spp. isolates. The publicly available 172 complete chromosome sequences of strains assigned to *K. pneumoniae* were retrieved from the GenBank (NCBI) repository.

2.2. Genomic island prediction

Resistance and pathogenicity islands in B31 were predicted using GIPSy (Genomic Island Prediction Software) [4]. The genome of the non-pathogenic strain *K. oxytoca* AR380 (GenBank assembly accession: GCA_003073975.1) was used as the reference for the prediction of genomic island features present exclusively in the genome of the pathogenic bacterium. Criteria for the selection of a non-pathogenic strain included classification as *Klebsiella* spp. and public availability of a quality complete genome sequence (the *K. oxytoca* AR380 genome was previously obtained using both Pacbio® and Illumina Inc. technologies). Miscellaneous Islands (MIS) included genes related to the Resistance and Pathogenicity Islands (RI and PI, respectively).

2.3. Sequence type and capsular type information

To predict the sequence type (ST) of all 173 *Klebsiella* spp. strains, allelic profiling for seven housekeeping genes (*gapA, infB, mdh, pgi, phoE, rpoB*, and *tonB*) was performed using BIGSdb (Bacterial Isolate Genome Sequence Database, accessed via the link http://bigsdb.pasteur.fr/Klebsiella/Klebsiella.html).

Table 1). These genes encode for glyceraldehyde 3-phosphate dehydrogenase, translation initiation factor IF-2, malate dehydrogenase, phosphoglucose isomerase, phosphorin E, DNA-directed RNA polymerase subunit beta, and periplasmic energy transducer, respectively. Kleborate (v2.0) [5] and the Kaptive [6] were used to assign the best matching capsule types for the *K. pneumoniae* strains.

Credit Author Statement

Wrote the manuscript; **(R.P., N.S., S.T., A.C.C., R.S.):** designed the study; **(R.P., S.T., P.I.P.R., V.A., and T.L.P.C.):** investigation; **(B.B., N.S., D.H.B., L.T.O. and R.S.):** conducted in silico analyses and generated the results; **(R.P., M.V.C.V., A.K.J., R.B.K. and P.I.P.R.):** critically reviewed and revised the manuscript; **(N.S., A.G.G., A.G.N., F.F.P., I.B.L., H.C.P.F., S.C.S., D.B., P.G., R.M., P.I.P.R., and T.L.P.C.):** supervised the study; **(S.T, V.A., and T.L.P.C.):** Resources; **(H.C.P.F., P.G. and B.B.)**.

Ethics Statement

Written informed consent was obtained from the patient, with protocols approved by and in accordance with the Research Ethics Committee (COEP 5149 / 2016) in Federal University of Minas Gerais, Belo Horizonte, Brazil.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Acknowledgments

The authors acknowledge the collaboration and assistance of all team members and the Brazilian funding agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) for providing financial support. TC and NS were supported by CAPES grant numbers 88881.068000/2014-01 and 88887.091518/2014-00, respectively. Currently, TC and NS are supported by CNPq grant number 433859/2018-0 and CAPES grant number 8887.508856/2020-00, respectively.

References

[1] R. Profeta, N. Seyffert, S. Tiwari, M.V.C. Viana, A.K. Jaiswal, A.C. Caetano, D.H. Bücker, L.T. de Oliveira, R. Santos, A. Gala-Garcia, R.B. Kato, F.F. Padilha, I.B. Lima-Verde, P. Ghosh, D. Barh, A. Góes-Neto, H.C.P. Figueiredo, T.L.P. Castro, S.C. Soares, R. Meyer, B. Brenig, P.I.P. Ramos, V. Azevedo, Comparative genomics with a multidrug-resistant Klebsiella pneumoniae isolate reveals the panorama of unexplored diversity in Northeast Brazil, Gene (2020), doi:10.1016/j.gene.2020.145386.

[2] J.E. Choby, J. Howard-Anderson, D.S. Weiss, Hypervirulent Klebsiella pneumoniae – clinical and molecular perspectives, J. Intern. Med. (2019) 0–3, doi:10.1111/joim.13007.

[3] K.L. Wyres, K.E. Holt, Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria, Curr. Opin. Microbiol. 45 (2018) 131–139, doi:10.7287/peerj.preprints.26852v1.

[4] S.C. Soares, H. Geyik, R.T.J. Ramos, P.H.C.G. de Sá, E.G.V. Barbosa, J. Baumbach, H.C.P. Figueiredo, A. Miyoshi, A. Tauch, A. Silva, V. Azevedo, GIPSy: Genomic island prediction software, J. Biotechnol. 232 (2016) 2–11, doi:10.1016/j.jbiotec.2015.09.008.

[5] M.M.C. Lam, R.R. Wick, K.L. Wyres, C. Gorrie, M. Judd, S. Brisse, A. Jenney, K.E. Holt, Frequent emergence of pathogenic lineages of Klebsiella pneumoniae via mobilisation of yersiniabactin and colibactin, BioRxiv (2017), doi:10.1101/098178.

[6] K.L. Wyres, R.R. Wick, C. Gorrie, A. Jenney, R. Follador, N.R. Thomson, K.E. Holt, Identification of Klebsiella capsule synthesis loci from whole genome data, Microb. Genomics. 2 (2016) 1–15, doi:10.1099/mgen.0.000102.