DATA NOTE

Draft genome sequence of *Enterobacter chengduensis* ECC445, isolated from fresh water in the West Indies

Matthieu Pot[1]*, Célia Ducat[1], Yann Reynaud[1], David Couvin[1], Séverine Ferdinand[1], Sébastien Breurec[1,2,3,4], Antoine Talarmin[1] and Stéphanie Guyomard-Rabenirina[1]

Abstract

Objectives The *Enterobacter cloacae* complex is considered an important opportunistic pathogen. It comprises many members that remain difficult to delineate by phenotypic approaches. Despite its importance in human infection, there is a lack of information on associated members in other compartments. Here we report the first de novo assembled and annotated whole-genome sequence of a *E. chengduensis* strain isolated from the environment.

Data description ECC445 specimen was isolated in 2018 from a drinking water catchment point in Guadeloupe. It was clearly related to *E. chengduensis* species according to hsp60 typing and genomic comparison. Its whole-genome sequence is 5,211,280-bp long divided into 68 contigs, and presents a G + C content of 55.78%. This genome and associated datasets provided here will serve as a useful resource for further analyses of this rarely reported *Enterobacter* species.

Keywords Caribbean, CRISPR spacer, *Enterobacter cloacae* complex, Environment, hsp60, Phylogeny, Typing

Objective

Members of the *E. cloacae* complex (ECC) are ubiquitous *Enterobacteriaceae* involved in various opportunistic infections in animals, humans and plants [1]. The taxonomy of this bacterial complex has been re-evaluated regularly [1–3]. There are to date 24 species and 22 *Enterobacter* genomospecies in this bacterial complex [3]. In parallel to whole-genome sequencing, different typing approaches have been proposed to help in identify them, such as the amplification of the partial gene coding for the heat-shock protein 60 (hsp60) [4].

The objective of this study is to report a draft genome sequence of an environmentally isolated *E. chengduensis*. We also provided a global overview of this species which is rarely described in the scientific literature. ECC445 strain is associated with a broader analysis of *Enterobacter* population diversity and was isolated from a water sample [5]. This sample was collected in June 2018 during prospective control of a catchment point in Petit-Bourg, Guadeloupe (French overseas territory; 16.188009 N, 61.659017 W). After serial dilutions, 100 mL were filtered (0.45 µm, Millipore, Guyancourt, France). The membrane was placed in 9 mL of buffered peptone water solution for overnight pre-enrichment at 37 °C. Then, 100 µL of broth were cultured aerobically onto chromogenic agar supplemented with ceftriaxone at 4 mg/L (CCA, CHRO-Magar, Paris, France).

*Correspondence:
Matthieu Pot
matthieu.pot@ird.fr
[1] Transmission, Reservoir and Diversity of Pathogens Unit, Pasteur Institute of Guadeloupe, Pointe-À-Pitre, France
[2] Faculty of Medicine Hyacinthe Bastaraud, University of the Antilles, Pointe-À-Pitre, France
[3] Centre for Clinical Investigation 1424, INSERM, Pointe-À-Pitre/Les Abymes, France
[4] Laboratory of Clinical Microbiology, University Hospital Centre of Guadeloupe, Pointe-À-Pitre/Les Abymes, France

Pot *et al. BMC Genomic Data*   (2023) 24:16

Page 2 of 5

Table 1 Overview of datafiles/datasets

Label	Name of DataFile /data set	File types (file extension)	Data repository and identifier (DOI or accession number)
DataFile 1	Table S1, ECC445 – Antibiotic resistance profile	MS Excel file (.xlsx)	Figshare, https://doi.org/10.6084/m9.figshare.21294888 [8]
DataFile 2	Script S1, Analysis script of *E. chengduensis*	Bash file (.bash)	Figshare, https://doi.org/10.6084/m9.figshare.22041197 [9]
DataFile 3	Table S2, ECC445 – Quast and short BUSCO summary	Normal text file (.txt)	Figshare, https://doi.org/10.6084/m9.figshare.21294882 [14]
DataFile 4	Table S3, *E. chengduensis* metadata from NCBI and quality control criteria	MS Excel file (.xlsx)	Figshare, https://doi.org/10.6084/m9.figshare.21294903 [15]
DataFile 5	Alignment S1, Roary – *E. chengduensis* core-gene alignment	Fasta file (.fasta)	Figshare, https://doi.org/10.6084/m9.figshare.22041257 [16]
DataFile 6	MLTree S1, IQ-TREE – Maximum likelihood phylogenetic tree of *E. chengduensis*	Newick tree format (.nhx)	Figshare, https://doi.org/10.6084/m9.figshare.22041293 [17]
DataFile 7	Figure S1, Maximum likelihood phylogenetic tree of *E. chengduensis*	Portable Document Format file (.pdf)	Figshare, https://doi.org/10.6084/m9.figshare.21294879 [18]
DataFile 8	Table S4, ECC445 – PathogenFinder summary	Normal text file (.txt)	Figshare, https://doi.org/10.6084/m9.figshare.21294915 [19]
DataFile 9	Table S5, ECC445 – CrisprCasFinder summary	Tab-separated values (.tsv)	Figshare, https://doi.org/10.6084/m9.figshare.21294924 [20]
DataFile 10	Table S6, CRISPR spacers comparison	MS Excel file (.xlsx)	Figshare, https://doi.org/10.6084/m9.figshare.21294933 [21]
DataSet 1	Partial hsp60 coding sequence of the ECC445 strain	Fasta file(.fasta)	NCBI Nucleotide, https://identifiers.org/nucleotide:OM687483 [10]
DataSet 2	Illumina reads of the ECC445 strain	SRA file (.sra)	NCBI Sequence Read Archive, https://identifiers.org/ncbi/insdc.sra:SRR16640525 [22]
DataSet 3	Annotated assembly of the ECC445 strain	Fasta file (.fasta)	NCBI Nucleotide, https://identifiers.org/nucleotide:JAKLRZ000000000 [23]

ECC445 strain was isolated from this selective plate after overnight growth at 37 ℃. It was initially identified as *E. cloacae* (97.7% ID) with the API-20E system (Bio-Mérieux, Marcy-l'Étoile, France), and as *E. hormaechei* (99.9% ID) by matrix assisted laser desorption/ionization time–of–flight mass spectrometry (VITEK MS, BioMérieux, Marcy-l'Étoile, France) [6]. This strain was resistant to third-generation cephalosporins by chromosomal overproduction of AmpC (CA-SFM/EUCAST guideline 2018, disc diffusion method; DataFile 1; Table 1) [5, 7, 8].

Data description
Upon isolation and to better identify this bacterium, DNA analysis was conducted. Unless otherwise stated on the detailed methodology file, default parameters were used for all software tools (DataFile 2) [9]. DNA was extracted from a pure isolate with the QIAamp DNA minikit (Qiagen, Hilden, Germany). We conducted the hsp60 typing approach and obtained a 273-bp amplicon that was sequenced at Eurofins (Eurofins Genomic SAS, Les Ullis, France; DataSet 1) [4, 5, 10]. BLASTn submission to the NCBI nucleotide collection database indicated 100% coverage and a maximum percentage identity with the *E. chengduensis* reference (v2.11.0; 99.27%; CP043318) [11–13].

The second sequencing step was performed with a NextSeq 500 system (Illumina; 150-bp paired-end configuration; Nextera XT Kit; DataSet 2) [22]. Raw reads ($n=7,941,390$) were trimmed and filtered with Alien-Trimmer v0.4.0 [24]. De novo assembly and annotation were performed with SPAdes (v3.12.0; –careful) and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP v5.3; DataSet 3) [23, 25, 26]. Whole-genome sequence of ECC445 strain has a 95-fold coverage and BUSCO single-copy completeness of 98.60% (v5.3.2) [27]. It is 5,211,280-bp long divided onto 68 contigs, and presents a G + C content of 55.78% (QUAST v5.0.2; DataFile 3) [14, 28].

Whole-genome sequencing allowed accurate species identification using FastANI v1.3 and Genome-to-Genome Distance Calculator (GGDC; v3.0) against last *Enterobacter* references panel provided by Feng Y. and colleagues [3, 29–31]. Results confirmed that ECC445 strain is closely related to *E. chengduensis* with the novel Sequence Type ST1533 (Average Nucleotide Identity – ANI: 98.48%; in silico DNA-DNA Hybridization – isDDH *formula 2*: 88.20%) [5, 11, 29, 32, 33].

Pot *et al. BMC Genomic Data* (2023) 24:16

Page 3 of 5

On June 29, 2022, a total of 19 *E. chengduensis* with complete metadata was retrieved from the NCBI database [13]. All these whole-genome sequences have passed our quality control (DataFile 4) [15]. To place ECC445 isolate in a global context, a core-genome alignment with Roary v3.13.0 was performed using reference sequences of *Enterobacter*, including all previous *E. chengduensis* (DataFile 5) [3, 13, 16, 34]. If only fastqs were available, de novo assembly was performed with SPAdes [25]. All additional sequences were annotated with PGAP as described above [26]. Maximum-likelihood phylogenetic reconstructions were performed by IQ-TREE (v2.1.2) and associated dependencies (*i.e.* ModelFinder and UFBoot2; −merit BIC: GTR + F + R8 −ufboot 1,000 −bnni; DataFile 6) [17, 35–37]. The tree was drawn with iTOL v6.4.2 (DataFile 7) [18, 38].

To date, ECC445 strain is the first *E. chengduensis* described in the natural environment and is a probable human pathogen (81.9%; PathogenFinder v1.1; DataFile 8) [19, 39]. The remaining 19 *E. chengduensis* isolates were recovered from human samples (*i.e.* ST414 $n = 18$; ST1535 $n = 1$). Most of them were described from China (10/19, 52.6%). Unlike ECC445, many of these strains had acquired resistance genes and plasmids (ResFinder v4.1, PlasmidFinder v2.1; details in the DataFile 4) [15, 40, 41].

The CRISPRCasFinder tool v1.1.2 identified a CRISPR/Cas system in the strain ECC445 genome (3 array sequences; Cas type I-F with evidence level of 4; DataFile 9) [20, 42]. As shown in DataFile 10 [21], none of the 46 CRISPR spacer sequences from strain ECC445 were found in the other *E. chengduensis* isolates using the CRISPR Comparison ToolKit (CCTK; v0.8.4) [43]. In contrast, some spacer arrays were identical between ST1535 and ST414 strains. These observations are consistent with the evolution of ECC445 ST1533 isolate in a different ecosystem than other *E. chengduensis* STs.

Limitations
This Data note was limited to a single draft genome sequence. To overcome this limitation and place this strain in a global context, an extensive analysis of *E. chengduensis* strains was performed. Its correct identification by the hsp60 approach was confirmed here by whole-genome sequencing. However, further analysis on a larger collection is needed to appreciate the robustness of this technique to identify *E. chengduensis* among other ECC members.

Abbreviations
ANI	Average Nucleotide Identity
CCTK	CRISPR Comparison ToolKit
ECC	*Enterobacter cloacae* Complex
GGDC	Genome-to-Genome Distance Calculator
hsp60	Heat-shock protein 60
isDDH	in silico DNA-DNA Hybridization
PGAP	Prokaryotic Genome Annotation Pipeline
ST	Sequence Type

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Davin-Regli A, Lavigne JP, Pagès JM. *Enterobacter* spp.: update on taxonomy, clinical aspects, and emerging antimicrobial resistance. Clin Microbiol Rev. 2019;32:e00002-19. https://doi.org/10.1128/CMR.00002-19.
2. Sutton GG, Brinkac LM, Clarke TH, Fouts DE. *Enterobacter hormaechei* subsp. *hoffmannii* subsp. nov., *Enterobacter hormaechei* subsp. *xiangfangensis* comb. nov., *Enterobacter roggenkampii* sp. nov., and *Enterobacter muelleri* is a later heterotypic synonym of *Enterobacter asburiae* based on computational analysis of sequenced *Enterobacter* genomes. F1000Res. 2018;7:521. https://doi.org/10.12688/f1000research.14566.2.
3. Feng Y, Hu Y, Zong Z. Reexamining the Association of AmpC variants with *Enterobacter* species in the context of updated taxonomy. Antimicrob Agents Chemother. 2021;65(12):e0159621. https://doi.org/10.1128/AAC.01596-21.

Pot *et al. BMC Genomic Data*   (2023) 24:16

Page 4 of 5

4. Hoffmann H, Roggenkamp A. Population genetics of the nomenspecies *Enterobacter cloacae*. Appl Environ Microbiol. 2003;69:5306–18. https://doi.org/10.1128/AEM.69.9.5306-5318.2003.

5. Pot M, Reynaud Y, Couvin D, Ducat C, Ferdinand S, Gravey F, Gruel G, Guérin F, Malpote E, Breurec S, Talarmin A, Guyomard-Rabenirina S. Wide distribution and specific resistance pattern to third-generation cephalosporins of *Enterobacter cloacae* complex members in humans and in the environment in Guadeloupe (French West Indies). Front Microbiol. 2021;12:628058. https://doi.org/10.3389/fmicb.2021.628058

6. Patel R. Matrix-assisted laser desorption ionization–time of flight mass spectrometry in clinical microbiology. Clin Infect Dis. 2013;57:564–72. https://doi.org/10.1093/cid/cit247.

7. Bonnet R, Bru JP, Cacon F, Cattoen C, Cattoir V, Courvalin P, Dubreuil L, Jarlier V, Lefort A, Merens A, Plesiat P, Ploy MC, Soussy CJ, Varon E, Weber P. CA-SFM / EUCAST: Comité de l'antibiogramme de la Société Française de Microbiologie -Recommandation 2018, V1.0 February ed. France; 2018. p. 1–132.

8. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. Table S1, ECC445 - Antibiotic resistance profile. 2022. https://doi.org/10.6084/m9.figshare.21294888. figshare.

9. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. Script S1 Analysis script of *E. chengduensis*. 2023. https://doi.org/10.6084/m9.figshare.22041197. figshare.

10. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. *Enterobacter chengduensis* strain ECC445 heat-shock protein 60 (*groL*) gene, partial cds. Genbank. https://identifiers.org/nucleotide:OM687483. 2022.

11. Wu W, Feng Y, Zong Z. Characterization of a strain representing a new *Enterobacter* species, *Enterobacter chengduensis* sp. nov. Antonie Van Leeuwenhoek. 2019;112(4):491–500. https://doi.org/10.1007/s10482-018-1180-z.

12. Oren A, Garrity GM. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol. 2019;69(7):1844–6. https://doi.org/10.1099/ijsem.0.003452.

13. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2016;44(D1):D7-19. https://doi.org/10.1093/nar/gkv1290.

14. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. Table S2, ECC445 - Quast and short BUSCO summary. 2022. https://doi.org/10.6084/m9.figshare.21294882. figshare.

15. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. Table S3, *E. chengduensis* metadata from NCBI and quality control criteria. 2022. https://doi.org/10.6084/m9.figshare.21294903. figshare.

16. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. Alignment S1, Roary – *E. chengduensis* core-gene alignment. 2023. https://doi.org/10.6084/m9.figshare.22041257. figshare.

17. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. MLTree S1, IQ-TREE – Maximum likelihood phylogenetic tree of *E. chengduensis*. 2023. https://doi.org/10.6084/m9.figshare.22041293. figshare.

18. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. Figure S1, Maximum likelihood phylogenetic tree of *E.* chengduensis. 2022. https://doi.org/10.6084/m9.figshare.21294879. figshare.

19. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. Table S4, ECC445 - PathogenFinder summary. 2022. https://doi.org/10.6084/m9.figshare.21294915. figshare.

20. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. Table S5, ECC445 – CrisprCasFinder summary. 2022. https://doi.org/10.6084/m9.figshare.21294924. figshare.

21. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. Table S6, CRISPR spacers comparison. 2022. https://doi.org/10.6084/m9.figshare.21294933. figshare.

22. NCBI Sequence Read Archive https://identifiers.org/ncbi/insdc.sra:SRR16640525. 2021.

23. Pot M, Ducat C, Reynaud Y, Couvin D, Ferdinand S, Breurec S, Talarmin A, Guyomard-Rabenirina S. ECC445 Whole genome shotgun sequencing project. Genbank. https://identifiers.org/nucleotide:JAKLRZ000000000. 2022.

24. Criscuolo A, Brisse S. AlienTrimmer: a tool to quickly and accurately trim off multipleshort contaminant sequences from high-throughput sequencing reads. Genomics. 2013;102(5-6):500-6. https://doi.org/10.1016/j.ygeno.2013.07.011.

25. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes *de novo* assembler. Curr Protoc Bioinformatics. 2020;70(1):e102. https://doi.org/10.1002/cpbi.102.

26. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res. 2021;49(D1):D1020–8. https://doi.org/10.1093/nar/gkaa1105.

27. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2. https://doi.org/10.1093/bioinformatics/btv351.

28. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5. https://doi.org/10.1093/bioinformatics/btt086.

29. Wu W, Feng Y, Zong Z. Precise species identification for *Enterobacter*: a genome sequence-based study with reporting of two novel species, *Enterobacter quasiroggenkampii* sp. nov. and *Enterobacter quasimori* sp. nov. mSystems. 2020;5:e00527-20. https://doi.org/10.1128/mSystems.00527-20.

30. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res. 2022;50(D1):D801–7. https://doi.org/10.1093/nar/gkab902.

31. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9(1):5114. https://doi.org/10.1038/s41467-018-07641-9.

32. Miyoshi-Akiyama T, Hayakawa K, Ohmagari N, Shimojima M, Kirikae T. Multilocus sequence typing (MLST) for characterization of *Enterobacter cloacae*. PLoS One. 2013;8(6):e66358. https://doi.org/10.1371/journal.pone.0066358.

33. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124. https://doi.org/10.12688/wellcomeopenres.14826.1.

34. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3. https://doi.org/10.1093/bioinformatics/btv421.

35. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74. https://doi.org/10.1093/molbev/msu300.

36. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22. https://doi.org/10.1093/molbev/msx281.

37. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9. https://doi.org/10.1038/nmeth.4285.

38. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6. https://doi.org/10.1093/nar/gkab301.

39. Cosentino S, Larsen MV, Aarestrup FM, Lund O. Correction: Pathogen-Finder- distinguishing friend from foe using bacterial whole genome sequence data. PLoS ONE. 2013;8:1–11. https://doi.org/10.1371/annotation/b84e1af7-c127-45c3-be22-76abd977600f.

40. Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. Microb Genom. 2022;8(1):000748. https://doi.org/10.1099/mgen.0.000748

41. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing.

Antimicrob Agents Chemother. 2014;58(7):3895–903. https://doi.org/10.1128/AAC.02412-14.

42. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46(W1):W246–51. https://doi.org/10.1093/nar/gky425.

43. Collins AJ, Whitaker RJ. CRISPR comparison toolkit (CCTK): Rapid identification, visualization, and analysis of CRISPR array diversity. bioRxiv [preprint]. 2022;502198. https://doi.org/10.1101/2022.07.31.502198.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.