



Draft Genome Sequences of *Cronobacter muytjensii* Cr150, *Cronobacter turicensis* Cr170, and *Cronobacter sakazakii* Cr611

Gal Zizelski Valenci,^a Mor Rubinstein,^a Reuven Afriat,^a Shira Rosencwaig,^a Zeev Dveyrin,^a Efrat Rorman,^a  Israel Nissan^a

^aMinistry of Health, National Public Health Laboratory, Tel Aviv, Israel

ABSTRACT The sequencing and bioinformatics analyses of isolates Cr150, Cr170, and Cr611 from powdered infant formula indicate that the three strains represent new members in the *Cronobacter muytjensii*, *Cronobacter turicensis*, and *Cronobacter sakazakii* groups, respectively.

Cronobacter spp. are a group of opportunistic bacteria associated with severe and potentially life-threatening diseases (1–3). In order to expand our knowledge about this group of pathogens, multiple confirmed *Cronobacter* samples were sequenced and analyzed. One of them is *Cronobacter sakazakii* Cr268 (4). Here, we present the genomes of *Cronobacter muytjensii* Cr150, *Cronobacter turicensis* Cr170, and *Cronobacter sakazakii* Cr611.

The three strains were isolated and identified exactly as indicated in ISO 22964:2006 (5). Matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) analysis (6) indicates that Cr150, Cr170, and Cr611 belong to the *Cronobacter sakazakii* group. The isolates were kept frozen (–80°C) until their revival. Two typical colonies were collected from each isolate using an inoculating loop and suspended in 400 µl of saline followed by two washes and resuspension in 200 µl of saline. DNA was isolated from overnight culture using a MagNA pure compact robot (Roche) following the manufacturer's instructions. Paired-end libraries were generated using an Illumina Nextera DNA Flex DNA library preparation kit according to Illumina protocols. For sequencing, we utilized the Illumina MiSeq platform using a 250-bp paired-end read kit v2.

All bioinformatic analyses were performed using the PATRIC v3.5.36 platform (7) with default parameters, unless otherwise noted. The quality of the raw reads was checked using the FASTQ utilities of PATRIC. The total numbers of reads for Cr150, Cr170, and Cr611 were 2,604,470, 2,086,159, and 2,053,349, respectively. *De novo* assembly of the reads was performed using SPAdes (8) as part of the PATRIC Comprehensive Genome Analysis Service. The assembly and annotation information is detailed in Table 1. Genome annotation was performed with the Rapid Annotations using Subsystems Technology tool kit (RASTtk) (9) and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) at the National Center for Biotechnology Information (NCBI). CRISPR/Cas systems were identified using CRISPRCasFinder (10) (Table 1). In order to identify the closest homologue in the database, as well as the as the closest high-quality representative genome at the NCBI, we utilized the Similar Genome Finder Service, which uses the Mash/MinHash algorithm (11), and the average nucleotide identity (ANI) calculator (12) to estimate average nucleotide identity (Table 1). We also performed taxonomic classification using Kraken v2 (13), confirming strain taxonomy. The multilocus sequence type (MLST) profile was determined using the MLST 2.0 server at the Center for Genomic Epidemiology (genomicepidemiology.org) (14) (Table 1).

Importantly, the MLSTs for these three strains are unknown, with two novel alleles for isolates Cr611 and Cr170 and three novel alleles for isolate Cr150. The closest homologue for each strain identified in the database possesses a low k-mer

Citation Zizelski Valenci G, Rubinstein M, Afriat R, Rosencwaig S, Dveyrin Z, Rorman E, Nissan I. 2020. Draft genome sequences of *Cronobacter muytjensii* Cr150, *Cronobacter turicensis* Cr170, and *Cronobacter sakazakii* Cr611. Microbiol Resour Annuounc 9:e00660-20. <https://doi.org/10.1128/MRA.00660-20>.

Editor Vincent Bruno, University of Maryland School of Medicine

Copyright © 2020 Zizelski Valenci et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Israel Nissan, israel.nissan@moh.gov.il.

Received 24 June 2020

Accepted 9 October 2020

Published 29 October 2020

TABLE 1 Assembly, annotation, and MLST details of the *C. sakazakii* Cr611, *C. muytjensii* Cr150, and *C. turicensis* Cr170 genomes

Genomic feature	Data for strain:		
	Cr170	Cr150	Cr611
Assembly details^a			
No. of contigs	27	37	18
<i>N</i> ₅₀ contig length (bp)	534,952	592,056	525,942
GC content (%)	57.42	57.23	56.98
Genome length (bp)	4,498,812	4,613,086	4,443,709
Properly mapped reads (%)	99.84	99.84	99.82
Median base coverage (×)	121	105	105
Contig <i>L</i> ₅₀	3	4	4
Genome annotation			
No. of protein coding sequences	4,281	4,472	4,245
No. of tRNAs	75	77	76
No. of rRNAs	10	10	10
CRISPR			
No. of CRISPR spacers	27	80	19
No. of CRISPR repeats	29	83	21
CRISPR/Cas system	CAS-TypeIE	CAS-TypeIF	CAS-TypeIE
CRISPR/Cas genes	Cas1, Cas2, Cas5, Cas6, Cas7, Cse1, Cse2	Cas1, Cas3-Cas2, Csy1, Csy2, Csy3, Csy6	Cas1, Cas2, Cas5, Cas6, Cas7, Cse1, Cse2
Similar Genome Finder Service (Mash/MinHash)			
Most similar genome in the database	<i>Cronobacter turicensis</i> cro3064B1 (GenBank accession no. NRJM01000000)	<i>Cronobacter muytjensii</i> MOD1-Md1s (GenBank accession no. MSAE00000000)	<i>Cronobacter sakazakii</i> CS-30 (GenBank accession no. QISO01000000)
Most similar genome K-mer count	651/1,000	651/1,000	574/1,000
Average nucleotide identity for most similar genome in the database (%)	98.74	98.96	98.23
Closest high-quality representative genome at NCBI	<i>Cronobacter turicensis</i> z3032 (GenBank accession no. FN543093)	<i>Cronobacter dublinensis</i> 1210 (GenBank accession no. CAKZ00000000)	<i>Cronobacter sakazakii</i> ATCC 29544 (BioProject no. PRJNA276173)
Closest high-quality representative genome K-mer count	517/1,000	134/1,000	529/1,000
MLST			
Locus <i>atpD</i>			
Identity (%)	100	100	100
Coverage (%)	100	100	100
Alignment length (bp)	390	390	390
Allele length (bp)	390	390	390
Allele	<i>atpD_68</i>	<i>atp_35</i>	<i>atpD_3</i>
Locus <i>fusA</i>			
Identity (%)	100	100	100
Coverage (%)	100	100	100
Alignment length (bp)	438	438	438
Allele length (bp)	438	438	438
Allele	<i>fusA_5</i>	<i>fusA_35</i>	<i>fusA_14</i>
Locus <i>glnS</i>			
Identity (%)	100	99.449	99.7245
Coverage (%)	100	100	100
Alignment length (bp)	363	363	363
Allele length (bp)	363	363	363
Allele	<i>glnS_42</i>	<i>glnS_176^b</i>	<i>glnS_9^b</i>
Locus <i>gltB</i>			
Identity (%)	99.6055	99.211	100
Coverage (%)	100	100	100
Alignment length (bp)	507	507	507
Allele length (bp)	507	507	507
Allele	<i>gltB_62^b</i>	<i>gltB_242^b</i>	<i>gltB_109</i>
Locus <i>gyrB</i>			
Identity (%)	100	100	99.32537
Coverage (%)	100	100	100
Alignment length (bp)	402	402	402

(Continued on next page)

TABLE 1 (Continued)

Genomic feature	Data for strain:		
	Cr170	Cr150	Cr611
Allele length (bp)	402	402	402
Allele	<i>gyrB_60</i>	<i>gyrB_45</i>	<i>gyrB_43^b</i>
Locus <i>infB</i>			
Identity (%)	100	100	100
Coverage (%)	100	100	100
Alignment length (bp)	441	441	441
Allele length (bp)	441	441	441
Allele	<i>infB_150</i>	<i>infB_4</i>	<i>infB_56</i>
Locus <i>pps</i>			
Identity (%)	99.596	99.596	100
Coverage (%)	100	100	100
Alignment length (bp)	495	495	495
Allele length (bp)	495	495	495
Allele	<i>pps_156^b</i>	<i>pps_75^b</i>	<i>pps_13</i>

^a The N_{50} length is defined as the shortest sequence length at 50% of the genome, and the L_{50} count is defined as the smallest number of contigs whose length sum produces N_{50} .

^b Novel allele.

similarity (15) (Table 1). Notably, although Cr150 was classified using Kraken v2 as *C. mytjensii*, the Mash/MinHash algorithm identified *Cronobacter dublinensis* 1210 (GenBank accession number [CAKZ000000000](https://www.ncbi.nlm.nih.gov/GenBank/CAKZ000000000)) as the closest high-quality representative genome in NCBI, indicating that the NCBI database does not contain a high-quality representative genome for *C. mytjensii*.

In conclusion, strains Cr150, Cr170, and Cr611 represent new members in the *C. mytjensii*, *C. turicensis*, and *C. sakazakii* groups, respectively.

Data availability. The sequences are now available to the public at the NCBI (BioProject number [PRJNA636126](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA636126)).

REFERENCES

- Blackwood BP, Hunter CJ. 2016. *Cronobacter* spp. Microbiol Spectr 4: E110-0002-2015. <https://doi.org/10.1128/microbiolspec.E110-0002-2015>.
- Holy O, Forsythe S. 2014. *Cronobacter* spp. as emerging causes of healthcare-associated infection. J Hosp Infect 86:169–177. <https://doi.org/10.1016/j.jhin.2013.09.011>.
- Iversen C, Mullane N, McCardell B, Tall BD, Lehner A, Fanning S, Stephan R, Joosten H. 2008. *Cronobacter* gen. nov., a new genus to accommodate the biogroups of *Enterobacter sakazakii*, and proposal of *Cronobacter sakazakii* gen. nov., comb. nov., *Cronobacter malonaticus* sp. nov., *Cronobacter turicensis* sp. nov., *Cronobacter mytjensii* sp. nov., *Cronobacter dublinensis* sp. nov., *Cronobacter genomospecies* 1, and of three subspecies, *Cronobacter dublinensis* subsp. *dublinensis* subsp. nov., *Cronobacter dublinensis* subsp. *lausannensis* subsp. nov. and *Cronobacter dublinensis* subsp. *lactaridi* subsp. nov. Int J Syst Evol Microbiol 58:1442–1447. <https://doi.org/10.1099/ijs.0.65577-0>.
- Zizelski Valenci G, Rubinstein M, Afriat R, Rosencwaig S, Dveyrin Z, Rorman E, Nissan I. 2019. Draft genome sequence of *Cronobacter sakazakii* Cr268, isolated from infant powder formula. Microbiol Resour Announc 8:e01067-19. <https://doi.org/10.1128/MRA.01067-19>.
- International Organization for Standardization. 2006. ISO-IDF 22964|IDF/RM 210: specifies a method for the detection of *Enterobacter sakazakii* in milk and milk products. <https://www.iso.org/standard/41258.html>.
- Stephan R, Ziegler D, Pfluger V, Vogel G, Lehner A. 2010. Rapid genus- and species-specific identification of *Cronobacter* spp. by matrix-assisted laser desorption/ionization-time of flight mass spectrometry. J Clin Microbiol 48:2846–2851. <https://doi.org/10.1128/JCM.00156-10>.
- Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res 45:D535–D542. <https://doi.org/10.1093/nar/gkw1017>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, III, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. <https://doi.org/10.1038/srep08365>.
- Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Neron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 46:W246–W251. <https://doi.org/10.1093/nar/gky425>.
- Ondov BD, Treangen TJ, Mellsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132. <https://doi.org/10.1186/s13059-016-0997-x>.
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. <https://doi.org/10.1099/ijs.0.64483-0>.
- Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. <https://doi.org/10.1186/gb-2014-15-3-r46>.
- Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Ponten T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361. <https://doi.org/10.1128/JCM.06094-11>.
- Compeau PE, Pevzner PA, Tesler G. 2011. How to apply de Bruijn graphs to genome assembly. Nat Biotechnol 29:987–991. <https://doi.org/10.1038/nbt.2023>.