Complete genome sequences and genomic characterization of five plasmids harbored by environmentally persistent Cronobacter sakazakii strains ST83 H322 and ST64 GK1025B obtained from powdered infant formula manufacturing facilities

Flavia J. Negrete1,2, Katie Ko1,2, Hyein Jang1, Maria Hoffmann3, Angelika Lehner4, Roger Stephan4, Séamus Fanning5,6, Ben D. Tall1 and Gopal R. Gopinath1*

Abstract

Background: Cronobacter sakazakii is a foodborne pathogen that causes septicemia, meningitis, and necrotizing enterocolitis in neonates and infants. The current research details the full genome sequences of two extremely persistent C. sakazakii strains (H322 and GK1025B) isolated from powdered infant formula (PIF) manufacturing settings. In addition, the genetic attributes associated with five plasmids, pH322_1, pH322_2, pGK1025B_1, pGK1025B_2, and pGK1025B_3 are described.

Materials and Methods: Using PacBio single-molecule real-time (SMRT®) sequencing technology, whole genome sequence (WGS) assemblies of C. sakazakii H322 [Sequence type (ST)83, clonal complex [CC] 83) and GK1025B (ST64, CC64) were generated. Plasmids, also sequenced, were aligned with phylogenetically related episomes to determine, and identify conserved and missing genomic regions.

Results: A truncated ~13 Kbp type 6 secretion system (T6SS) gene cluster harbored on virulence plasmids pH322_2 and pGK1025B_2, and a second large deletion (~6 Kbp) on pH322_2, which included genes for a tyrosine-type recombinase/integrase, a hypothetical protein, and a phospholipase D was identified. Within the T6SS of pH322_2 and pGK1025B_2, an arsenic resistance operon was identified which is in common with that of plasmids pSP291_1 and pESA3. In addition, PHASTER analysis identified an intact 96.9 Kbp Salmonella SSU5 prophage gene cluster in pH322_1 and pGK1025B_1 and showed that these two plasmids were phylogenetically related to C. sakazakii plasmids: pCS1, pCsa767a, pCsaC757b, pCsaC105731a. Plasmid pGK1025B_3 was identified as a novel conjugative Cronobacter plasmid. Furthermore, WGS analysis identified a ~16.4 Kbp type 4 secretion system gene cluster harbored on pGK1025B_3, which contained a phospholipase D gene, a key virulence factor in several host–pathogen diseases.

*Correspondence: Gopal.Gopinathrao@fda.hhs.gov

© The Author(s) 2022. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
Background

*Cronobacter sakazakii* is an opportunistic foodborne pathogen that causes serious intestinal and extraintestinal systemic infections such as acute gastroenteritis, septicemia, meningitis, urosepsis, osteomyelitis, wound infections, and pneumonia in individuals of all ages [1–5]. Pre-term, low-birth weight, and/or immune compromised neonates and infants are highly susceptible to *C. sakazakii*. Moreover, severe invasive infections such as septicemia, meningitis, and necrotizing enterocolitis are hallmarks of this organism's pathogenicity. Additionally, outcomes from such invasive infantile infections often leave individuals with lifelong debilitating and neurologic impairments such as developmental delays, hydrocephaly, mental retardation, and other chronic neurological sequelae [3, 6, 7]. *C. sakazakii* infections observed in these individuals have been epidemiologically linked to consumption of intrinsically and extrinsically contaminated lots of reconstituted powdered infant (PIF) and follow up formulas; thus, contamination of such products is a challenging task for both infant formula manufacturers and caretakers [7–10]. Another trend that both clinicians and public health scientists must recognize is that unsafe personal hygiene breast-feeding practices, such as the use of contaminated personalized breast pumps, may also lead to infantile infections such as septicemia and meningitis [11–14].

Chase et al. [15] described *C. sakazakii* H322, as a highly persistent sequence type (ST) 83, clonal complex (CC) 83 strain that was obtained from a lot of contaminated PIF manufactured in Europe that was never released to the public. Chase et al. [15] further showed that the persistence of *C. sakazakii* H322 and other phylogenetically related ST83 strains, which were also found within the production environment of this facility, and its presence could be traced back for more than four years. Microarray analysis showed that these strains differed among them by sequence divergence in 5–38 genes [15]. In separate studies, Gopinath et al. [16] and Chase et al. [29] described several malonate-positive ST64, CC64 *C. sakazakii* strains, including GK1025B (a PIF manufacturing environmental isolate), that were found persisting in the environments of another European PIF manufacturing facility. These ST64 strains were phylogenetically related to other strains obtained from sources such as clinical samples, environments of USA dairy powder manufacturing facilities, spices, and mushrooms from the Middle East and China. Draft whole genome sequence (WGS) assemblies of these strains, together with other PIF production environmental-associated strains, confirmed a ST phylogenetic relatedness among them [16]. In the present study, we report the completed genome sequences of these two highly persistent *C. sakazakii* strains, H322 and GK1025B, and describe the genomic characterization of five plasmids harbored by them. The results of this study will facilitate a greater understanding of the survival and persistence of such foodborne pathogens within these “built- environments”.

Methods

Bacterial strains and DNA isolation

*Cronobacter sakazakii* H322 and GK1025B were grown in 5 ml of Trypticase Soy Broth (TSB, BBL, Becton Dickinson, Franklin Lakes, NJ, USA) supplemented with 1% NaCl (TSBS), and incubated at 37 °C for 18 h with shaking conditions of 160 rpm (Thomas Scientific, Inc., Swedesboro, NJ, USA). Isolation of genomic DNA was performed using a 2 ml aliquot of each culture using the robotic QIACube workstation and the automated Qiagen DNeasy technology (Qiagen, Inc., Germantown, MD, USA) following the manufacturer’s recommendations as described by Jang et al. [17, 18].

Whole genome sequencing, assembly, and annotation:

The single-molecule real-time (SMRT) sequencer technology [19] from PacBio (Pacific Biosciences, Menlo Park, CA, USA) was utilized to create high-quality long-read datasets of *C. sakazakii* strains H322 (SRR8305966) and GK1025B (SRR8305970). The initial processing of long-sequencing reads was carried out using the RS_HGAP_Assembly2 protocol (default parameters) implemented in the Pacific Biosciences SMRT analysis portal (version 2.3.1). Quality filtering was performed automatically during assembly using the

---

**Conclusion:** These data provide high resolution information on *C. sakazakii* genomes and emphasizes the need for furthering surveillance studies to link genotype to phenotype of strains from previous investigations. These results provide baseline data necessary for future in-depth investigations of *C. sakazakii* that colonize PIF manufacturing facility settings and genomic analyses of these two *C. sakazakii* strains and five associated plasmids will contribute to a better understanding of this pathogen's survival and persistence within various “built environments” like PIF manufacturing facilities.

**Keywords:** *Cronobacter sakazakii*, Whole genome sequencing, Plasmids, Built environment, Complete genomes, PHASTER, Phage-plasmids
SMRT Portal P-filter module and using the Hierarchical Genome Assembly Process 3 (HGAP3) pipeline. For generating complete genomes, a hybrid assembly strategy with UniCycler assembly software [20] implemented on the Pathosystems Resource Integration Center (PATRIC) database web-server (https://patricbrc.org/app/Assembly2). Long-read short read archive (SRA) files from PacBio and corresponding WGS datasets of the strains obtained from sequencing runs performed on an Illumina MiSeq platform (Illumina, San Diego, CA, USA) [15, 16, 29] were combined following the instructions on the web-server. The Prokaryotic Genome Annotation Pipeline (PGAP) annotations [21] of these completed genomes, plasmid sequences, and their accession numbers were released under FDA GenomeTrakr Bioproject PRJNA258403, which is part of FDA's foodborne pathogen research comprehensive Bioproject at NCBI (PRJNA186875). The RAST Seedviewer was used to help provide consistent and accurate genome annotations across the genomes and plasmid sequences [22].

Genomic analyses

The PROKSEE server (https://beta.proksee.ca/projects) was used to generate high-quality navigable maps of each circular plasmid as previously described [23]. Each Cronobacter plasmid's sequence was submitted to CGE's Plasmidfinder (https://cge.cbs.dtu.dk/services/PlasmidFinder/) for in silico determination of incompatibility plasmids such as IncF, IncHI1, IncHI2, IncN, and IncI1 plasmids [24]. For prophage sequence identification, C. sakazakii strain FASTA data sets were uploaded to the PHASTER (PHAge Search Tool Enhanced Release) web server and pipeline (https://phaster.ca/, last accessed 8.25.2021, [25, 26]). Mauve, Progressive Mauve, and Geneious suite 12.0 ([27]; https://www.geneious.com/) were used for alignment and visualization as needed. BLAST analysis for the presence of pH322_1 was performed on an in-house database of 683 genomes consisting of GenomeTrakr datasets along with publicly available genomes hosted at NCBI.

Results and discussion

Genome and plasmid characterization

The characteristics of the completed genomes and closed plasmids harbored by C. sakazakii strains H322 and GK1025B are summarized in Table 1. Each genome consisted of a single circular chromosome of 4,350,614 bp and 4,362,605 bp in size, contained a GC content of 56.7% and 56.9%, and 4,146 and 3,693 coding DNA sequences (CDS), respectively. Two plasmids were identified as being harbored by C. sakazakii H322 and three plasmids were identified to be carried by C. sakazakii GK1025B. None of the five plasmids identified by genome sequencing were predicted on CGE's Plasmidfinder [24]. In addition to the closed plasmids generated from long-read sequencing, PHASTER analysis showed that C. sakazakii strain H322 hosted four intact prophage sequences (Additional file 2: Table S2) that were located on the chromosome which included a 47.4 Kbp Salmonella SEN34 (NCBI accession #: NC_028699), a 37.4 Kbp Enterobacteria mEp235 (NC_019708), a 17.7 Kbp Enterobacteria P1 (NC_005856), prophage. An incomplete generalized transducing Salmonella bacteriophage ES18 prophage (NC_006949) prophage was also identified. Three intact Cronobacter prophage sequences were additionally identified by PHASTER analysis on the chromosome of GK1025B: Cronobacter ENT47670 (NC_019927), Cronobacter ESSI_2 (NC_047854), and Cronobacter phiES15 (NC_018454). The complete chromosomal sequences of the two C. sakazakii allowed for detailed annotation and identification of mobilome sequences that could be applied for comparative analysis with other strains of Cronobacter and related organisms.

Table 1 Characteristics of C. sakazakii H322 and GK1025B complete genomes and plasmids from Bioproject PRJNA258403a

| Strain ID/Plasmid | Genome/plasmid size (bp) | %GC content | Number of CDS | CRISPR arrays | NCBI biosample ID | NCBI genbank ID | NCBI accession number |
|-------------------|--------------------------|-------------|---------------|--------------|------------------|----------------|---------------------|
| H322              | 4,350,614                | 56.7        | 4146          | 1            | SAMN06124518     | CP078110        | MRXM01000000        |
| pH322_1           | 100,741                  | 50.2        | 137           |              |                  | CP078111        |                     |
| pH322_2           | 118,185                  | 56.8        | 118           |              |                  | CP078112        |                     |
| GK1025B           | 4,362,605                | 56.9        | 3693          | 2            | SAMN04329637     | CP078106        | MCOE01000000        |
| pGK1025B_1        | 101,769                  | 51.1        | 141           |              |                  | CP078107        |                     |
| pGK1025B_2        | 120,182                  | 56.6        | 133           |              |                  | CP078108        |                     |
| pGK1025B_3        | 46,238                   | 51.0        | 82            |              |                  | CP078109        |                     |

a Information was obtained from NCBI (https://www.ncbi.nlm.nih.gov/genome/browse/#/prokaryotes/1170/) and summarized.
and confirmed the presence of an intact 96.9 Kb Sal-

BLAST analysis showed that the virulence plas-
C. sakazakii

ST4

pSP291_1 harbored by

virulence plasmid backbones of pSP291_1 and pESA3

mid pH322_2 shares significant homology with the

SEE [23], showed the plasmid to be closely related to

identified by PGAP annotation. Analysis using PROK-

tained a GC content of 56.8%. There were 118 CDS

prophage gene cluster [25] (Additional file 2: Table S2)

SSU5 (NCBI accession number: NC_018843)

tional file 1: Table S1). PHASTER analysis identified

tail fiber identified as side tail fiber protein, Stf (Addi-

prophage-related genes on pH322_1 such as

genes encoding for a phage exonuclease, and a phage

tail fiber identified as side tail fiber protein, Stf (Addi-

GC content of 50.2% and harbored 137 CDS. Unique features con-

on this plasmid included 12 mobile genetic ele-

ments comprising six copies of an Insertion Sequence

family transposase, three exonuclease (3′–5′ exonu-

, SbcCD subunit D, and an unnamed exonu-

), a RecA recombinase, and a site-specific integrase (Additional file 1: Table S1). PGAP analysis also identi-

fied several phage-related genes on pH322_1 such as
genes encoding for a phage exonuclease, and a phage
tail fiber identified as side tail fiber protein, Stf (Addi-

tional file 1: Table S1). PHASTER analysis identified
determined the presence of an intact 96.9 Kb Salmonella SSU5 (NCBI accession number: NC_018843) prophage gene cluster [25] (Additional file 2: Table S2) located on pH322_1.

The second plasmid named pH322_2 and harbored by C. sakazakii H322 was 118,185 bp in size and con-
tained a GC content of 56.8%. There were 118 CDS identified by PGAP annotation. Analysis using PROK-

SEE [23], showed the plasmid to be closely related to C. sakazakii virulence plasmid pSP291_1 harbored by

ST4 C. sakazakii SP291 as described by Power et al. [30]. BLAST analysis showed that the virulence plas-

pH322_2 shares significant homology with the

virulence plasmid backbones of pSP291_1 and pESA3 (data not shown). They share conserved features like

the origin of replication gene, repA, two iron acquisi-
tion systems, an aerobactin-like siderophore (named Cronobactin, iucABCD/iutA), and an ABC ferric-iron transporter gene cluster (eitCBAD) as described by Franco et al. [31].

Description of H322 plasmids: pH322_1 and pH322_2

A 100,741 bp pCS1-like closed plasmid, named pH322_1 was identified to be similar to the plasmid pseudomolecule initially predicted from H322 draft whole genome contig sequences by Chase et al. [15]. The sequence relatedness of pH322_1 to pCS1 har-
bored by C. sakazakii NCIMB 8272 (alias NCTC 8155) after PROKSEE analysis using the β-version of CGView Server [23] is shown in Fig. 1A. It had a GC content slightly larger than the prophage gene cluster present in pH322_1 (96.9 Kb). As was the case for pH322_1, the prophage gene cluster contained genes encoding for prophage structural proteins including terminase, capsid and tail proteins. Genes encoding for a lysoz, an integrase, and a recombinase protein, and possessed a GC content of 51.1% were also noted.

pGK1025B_2 was identified as a slightly smaller version (120,182 bp) of the virulence plasmid, pESA3 (131,196 bp) that Franco et al. [31] described for C. saka-
zakii strain BAA-894. It contained 133 CDS, possessed a GC content of 56.6%, and harbored a homolog of repA, the plasmid’s origin of replication gene and a homolog of Cronobacter plasmogen activator, cpA (location: 6338 to 7276 bp). As described earlier, other noted genetic

Description of GK1025B plasmids: pGK1025B_1, pGK1025B_2 and pGK1025B_3

Complete sequences of three plasmids, pGK1025B_1, pGK1025B_2, and pGK1025B_3 was obtained by long-read sequencing and PGAP annotation (Additional file 1: Table S1 [21]). pGK1025B_1 was 101,769 bp in size and contained 141 CDS, of which 70 genes encoded for hypo-
thetical proteins (Additional file 1: Table S1) and had a GC content of 51.1%. An intact 99.4 Kbp gene cluster encoding for a Salmonella SSU5 prophage (NC_018843) was identified using PHASTER and was like pH322_1 (Additional file 2: Table S2) described earlier. This is a slightly smaller sized prophage SSU5 gene cluster than what was reported by Kim et al. (103 Kbp; 2012) but is slightly larger than the prophage gene cluster present in pH322_1 (96.9 Kb). As was the case for pH322_1, the prophage gene cluster contained genes encoding for prophage structural proteins including terminase, capsid and tail proteins. Genes encoding for a lysoz, an integrase, and a recombinase protein, and possessed a GC content of 51.1% were also noted.

pGK1025B_2 was identified as a slightly smaller version (120,182 bp) of the virulence plasmid, pESA3 (131,196 bp) that Franco et al. [31] described for C. sakazakii strain BAA-894. It contained 133 CDS, possessed a GC content of 56.6%, and harbored a homolog of repA, the plasmid’s origin of replication gene and a homolog of Cronobacter plasmogen activator, cpA (location: 6338 to 7276 bp). As described earlier, other noted genetic
features among these virulence plasmids include: a siderophore aerobactin biosynthesis gene cluster (now named Cronobactin/siderophore receptor, iucABCD/iutA), a bicistronic toxin-antitoxin gene complex encoding for HigB/HipA, and a methyl-accepting chemotaxis protein I (serine chemoreceptor protein) gene. pGK1025B_3 is a plasmid of 46,528 bp in size. It possesses a GC content of 51.1% and harbored 50 genes...
encoding for hypothetical proteins; however, it is a conjugative plasmid like pESA2 which is harbored by \textit{C. sakazakii} BAA-894 [31, 33]. A ∼ 16.4 Kbp type 4 secretion system gene cluster was found, and most notably it contains within this gene cluster a copy of a phospholipase D gene (\textit{plD}, located between ∼ 9308 to ∼ 9841 bp, NCBI Locus Tag: AUM97_022060).

**Comparative genomic analysis of the novel phage-plasmids pH322_1 and pGK1025B_1**

Initial sequence analyses described above suggested that pGK1025B_1 and pH322_1 belong to a unique category of plasmids called phage-plasmids (extrachromosomal DNA molecules that host intact prophage sequences and strictly behave like plasmids) that have been known since the 1960s [34, 35]. Sequence comparison of \textit{repA} gene sequences from pH322_1 and pGK1025B_1 (locus tags BTK77_021130 and AUM97_021275, respectively) suggested that these plasmids possessed a mutually exclusive origin of replication and different sub-groups of IncF1B category. It was reported that SSU5 phage bearing plasmids usually belong to IncF1B incompatibility group in other \textit{Enterobacteriaceae} members [35]. Some previously reported \textit{Cronobacter} plasmids belong to this rare category of phage-plasmids along with the two plasmids identified in this study suggesting an expansion of genetic diversity among this emerging foodborne pathogen [15, 36]. We identified sequences containing significant homology to pGK1025B_1 and pH322_1 from among the known \textit{Cronobacter} plasmids by BLAST analysis. The properties of these prophages are summarized in Additional file 2: Table S2 and suggest a prevalence of the known plasmids by BLAST analysis.

Cronobacter homology to pGK1025B_1 and pH322_1 from among [15, 36]. We identified sequences containing significant diversity among this emerging foodborne pathogen identified in this study suggesting an expansion of genetic category of phage-plasmids along with the two plasmids previously reported. A detailed analysis of these plasmids, and their inclusion in plasmid-finding pipelines, would enable the identification of SSU5-like sequences from the growing number of \textit{Cronobacter} WGS datasets.
780 phage–plasmids grouped into eight distinct categories based on sequence features. This study further suggested a role for the phage-plasmids in genetically connecting phages and other mobile (and transducing) genetic elements [35]. *Salmonella* prophage SSU5 represents a different type of lysogenic phage with a circular phage-plasmid that is also very common in other members of the *Enterobacteriales*; however, they are only very rarely annotated as being phage related, much less as prophages [38]. Often a genome may contain an integration hot spot such as that found for *Lactococcus lactis* subsp. *cremoris* which contains 20% of its genome as IS elements [39]. This suggests that a genome can exist in an active evolutionary state, as it can readily accommodate new DNA and/or loose genome regions

Fig. 2 (See legend on previous page.)
as well. Similarly, pH322_1 and pGK1025B_1 with an abundance of mobile genetic elements found in these phage-plasmid sequences (Additional file 1: Table S1) may represent such a genetic element. Furthermore, the fact that C. sakazakii strains H322 and GK1025B contain multiple plasmids may offer selective advantages to a bacterial host which may also reflect their adaptative abilities to persist within the nutrient-rich environment of a built environment, such as that of powdered infant formula manufacturing facilities [40].

Genomic analysis of the virulence plasmids, pH322_2 and pGK1025B_2

The shared genome backbone (Additional file 1: Table S1) of pH322_2 and pGK1025B_2 with that of virulence plasmids, pESA3 [31] and pSP291_1 [30] is shown in Fig. 1B. Both plasmids harbored a Cronobacter plasminogen activator (cpa) encoding Protease VII or Ompitin precursor (EC 3.4.23.49) (Additional file 1: Table S1) homologous to the Salmonella outer membrane protease, PgtE [1, 31, 32]. Both the plasmids contained a truncated ~13 kbp type six secretion system (T6SS) gene cluster which shares homology with a similar region harbored by pESA3, and pSP291_1. PROKSEE analysis showed that a similar truncated T6SS with a large deletion in the region for SP291_1 and pH322_2 compared to that of pESA3. In addition to the deletion within the T6SS gene cluster, pH322_2 also had a second large deletion of ~6 Kbp, which includes genes for a tyrosine-type recombinase/integrase, a hypothetical protein, and a phospholipase D. These results correlate with those reported by Franco et al. [31]; Tall et al. [32], Chase et al. [15], and Jang et al. [1] who had described the presence of a virulence plasmid like pH322_2, pGK1025B_2, pESA3 and pSP291_1 in a high percentage of C. sakazakii strains (629 of 652, 96%). Two functional T6SS clusters were reported by Wang et al., 2018 in the C. sakazakii strain ATCC12868 although the genome sequences are not available on NCBI for comparison [42]. In contrast, truncated T6SS segments on pESA3-like virulence plasmids reported by
Franco et al. [31] and others had not been characterized in vivo or share sequence homology with the chromosomal clusters rendering their use just as a possible 'signature sequence' for this category of plasmids. A Cobalt ABC transporter gene cluster encoding for an ATP-binding protein (CbtL), permease protein (CbtK), and two copies of a substrate-binding protein gene (CbtJ) were found on pGK1025B_2, but not on pH322_2 Additional file 4: Figure S1.

Sequence analysis of conjugative plasmid pGK1025B_3 compared with pESA2 and other Enterobacteriaceae plasmids

Sequence alignment of newly described pGK1025B_3 compared with other conjugative class members produced on the PROKSEE server (Fig. 3) suggest that this plasmid represents a new conjugative plasmid that only has marginal sequence homology with pEAS2 from C. sakazakii BAA-894 [1, 33]. Interestingly, a known virulence gene coding a phospholipase D (PLD) was
identified within a complete T4SS cluster harbored on pGK1025B_3 (Additional file 5: Figure S2). Results of a BLASTn analysis using the rep (CP078109.1) gene from pGK1025B_3 queried against Enterobacteriaceae and related endosymbionts (NCBI taxid:91347), showed a shared homology with many related rep genes. Alignment of these gene sequences, shown in Fig. 4 revealed that the rep gene of pGK1025B_3 clustered distinctly separated from a larger cluster of 91 rep genes of related plasmids of members of the Enterobacteriaceae. These results suggest that the rep gene of pGK1025B_3 may represent a novel Cronobacter origin of replication gene carried by a previously uncharacterized Cronobacter conjugative plasmid that harbors within its gene cluster a phospholipase D gene. Future surveillance studies to identify the prevalence of pGK1025B_3 like plasmids as well as functional genetic studies are needed. Phospholipase D (PLD) represents a heterogeneous group of lipolytic esterases, which are either secreted into the extracellular milieu, or directly injected into the host cell cytosol by a wide variety of Gram-positive and Gram-negative bacteria through Type 6 and Type 4 secretion systems [41]. It plays an important role in several host–pathogen physiological interactions involved in bacterial pathogenesis, including cell invasion, evasion of the host immune response through escape of or maturation avoidance within phagosomes, establishment of tissue colonization, and systemic spread. The contribution of the Cronobacter version of PLD in the pathogenicity of this organism needs to be further studied.

Conclusions

The mechanisms related to the persistence of Cronobacter strains within the built environment such as that of powdered infant formula manufacturing facilities are currently unknown. The use of whole-genome sequencing of Cronobacter isolates obtained from the “built environment” as part of a routine surveillance strategy is only in its infancy but is a first step in determining the relationships of Cronobacter species that possess a long-term persistence phenotype in food manufacturing facilities. WGS analyses demonstrated that these two persistent C. sakazakii strains possess five plasmids of which fall into three different plasmid classes, such as the virulence plasmids pSP291_1 and pESA3 originally characterized by Power et al. [30] and Franco et al. [31], a prophage bearing pCS1-like plasmid originally described by Chase et al. [15], and an uncharacterized conjugative plasmid like that of pGK1025B_3 that possesses a phospholipase D gene within its T4SS gene cluster. The genomic information about these two highly persistent C. sakazakii strains H322 and GKI025B provides insights to design further in-depth investigations of a facility’s microbiota profile.

This information could also be used in future studies to develop basic differences between non-pathogenic and pathogenic microorganisms found within these food manufacturing environments. Finally, future analysis of the genome sequences of wild-type C. sakazakii strains will shed more light on the importance of plasmids and phage-plasmids and their role in survival and persistence in PIF manufacturing environments, and as causative agents of severe-invasive human infectious diseases. This study highlights the increased discriminatory power of WGS analysis and emphasizes the need for furthering extended surveillance studies and provides insights linking the genotype–phenotype of C. sakazakii from previously published longitudinal surveillance investigations.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13099-022-00500-5.

Additional file 1: Table S1. PGAP annotation of genes carried on pH322_1, pH322_2, pGK1025B_1, pGK1025B_2, and pGK1025B_3 that are harbored by C. sakazakii H322 and GKI025B. A

Additional file 2: Table S2. Results of PHASTER analysis for various Cronobacter plasmids including pH322_1, pH322_2, pGK1025B_1, pGK1025B_2, and pGK1025B_3. A

Additional file 3: Table S3. BLAST analysis of 683 Cronobacter genomes housed in a local database for the presence of phage-plasmid pH322_1.

Additional file 4: Figure S1. Multiple alignment analysis of the Cronobacter arsenic operon within the T6SS of virulence plasmids, pSP291_1, pH322_2, pGK1025B_2, and pESA3, as displayed by using Geneious suite. The black horizontal bar indicates the consensus sequence. The blue line indicates sequence coverage; the green represents percent identity with red presenting little homology; and green representing high homology. The arsenic operon consists of three genes: arsenate reductase (arsC), glutaredoxin, arsenic transporter, and a gene encoding a metalloregulator ArsR/SmtB family transcription factor. The operon is flanked by genes encoding for a TRh family potassium uptake protein and dihydrodipicolinate synthase family protein.

Additional file 5: Figure S2. Cronobacter phospholipase D family protein within the T4SS of pGK1025B_3 as displayed by using Geneious suite. The phospholipase D family protein is flanked by genes encoding for two hypothetical proteins and a conjugal relaxase and Vbr871 (a member of the superfamily of traffic ATPases). Other adjacent genes include Vbr810, which has a role in regulating substrate transfer to the extracellular space, and Vbr89 which encode for a channel protein that forms heterodimers with Vbr87. Vbr87 is localized at the outer membrane and plays a stabilizing role with the other Vbr8 proteins during assembly of the T4SS pilus.

Acknowledgements

This manuscript is being submitted by FN in partial fulfillment of the requirements for a Master of Science degree in Biological Sciences, University of Maryland Biological Science’s Graduate Program, University of Maryland, College Park. We thank Dr. Felix Reich (Institute for Food Quality and Safety, University of Veterinary Medicine Hannover, Bischofsholer Damm 15, 30173 Hannover, Germany) for providing the strain GKI025B and for helpful suggestions and discussions. Trade names of commercial products mentioned in this publication do not imply any recommendation or endorsement by the Food and Drug Administration.

Author contributions

FN, GG, BDT, and AL developed the concept for this paper. FN, HJ, GG, AL, SF, RS, IK, and BDT designed the experiments and contributed to the initial
drafts of the paper. BDT and GG completed the final draft and revisions. MH carried out PacBio sequencing and initial SMRT pipeline processing. GG performed hybrid assembly, annotations, and genomic data submission. GG, FN, and KK carried out the genome analyses and illustrations. KK carried out PHASTER analysis, and KK and FN performed the PROKSEE analyses. All authors contributed to the article and approved the submitted version. All authors read and approved the final manuscript.

**Funding**

Funds supporting this work were obtained internally through U.S. FDA appropriations, and the University of Maryland, Joint Institute for Food Safety and Applied Nutrition (JIFSAN) that supported FN and KK through a cooperative agreement with the FDA (#FDU001418). Moreover, funding for research fellow HJ was provided by Oak Ridge Institute for Science and Education of Oak Ridge, Tennessee.

**Availability of data and materials**

Not applicable.

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Yes.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1. Jang H, Gopinath GR, Eshwar A, Srikumar S, Nguyen S, Gangiredla J, Patel IR, Finkelstein SB, Negrete F, Woo J, Lee Y, Fanning S, Stephan R, Tall BD, Lehner A. The secretion of toxins and other exoproducts of Cronobacter: role in virulence, adaption, and persistence. Microorganisms. 2020;8(2):229. https://doi.org/10.3390/microorganisms8020229.

1. Holy O, Petřelová J, Hanulík V, Chromá M, Matsušková M, Forsythe SJ. Epidemiology of Cronobacter spp. isolates from patients admitted to the Olomouc University Hospital (Czech Republic). Epidemiol Mikrobiol Imunol. 2014;63:69–72.

1. Patrick ME, Mahon BE, Greene SA, Rounds J, Cronquist A, Wymore K, et al. Incidence of Cronobacter spp. infections, United States, 2003–2009. Emerg Infect Dis. 2014;20:1520–5.

1. Alsonosi A, Harri S, Kajskl M, Onelisková M, Hanulk V, Rodero V, et al. The speciation and genotyping of Cronobacter isolates from hospitalised patients. Eur J Clin Microbiol Infect Dis. 2015;34:1979–88.

1. Yong W, Guo B, Shi X, Cheng T, Chen T, Jiang X, et al. An investigation of an acute gastroenteritis outbreak: Cronobacter sakazakii, a potential cause of food-borne illness. Front Microbiol. 2018;9:549.

1. Friedemann M. Epidemiology of invasive neonatal Cronobacter (Enterobacter sakazakii) infections. Eur J Clin Microbiol Infect Dis. 2009;28:1297–304.

1. Strysko J, Cope JR, Martin H, Tarr C, Hise K, Collier S, et al. Food safety and invasive Cronobacter infections during early infancy, 1961–2018. Emerg Infect Dis. 2020;26:857–65. https://doi.org/10.3201/eid2605.190858.

1. Nonjeav FR, Kotloff KL, Martin MA, Schwalbe RS. Nosocomial bacteremia caused by Enterobacter sakazakii and leuconostoc mesenteroides resulting from extrinsic contamination of infant formula. Pediatr Infect Dis J. 1990;9:447–9.

1. Himelright I, Harris E, Lorch V, Anderson M, Jones T, Craig A, et al. Enterobacter sakazakii infections associated with the use of powdered infant formula—tennessee, 2001. Morb Mortal Wkly Rep. 2002;51:297–300.

1. Iversen C, Forsythe SJ. Risk profile of Enterobacter sakazakii, an emergent pathogen associated with infant milk formula. Trends Food Sci Technol. 2003;20(4):443–54.

1. Friedemann M. Enterobacter sakazakii in food and beverages (other than infant formula and milk powder). Int J Food Microbiol. 2007;116:1–10.

1. Jason J. Prevention of invasive Cronobacter infections in young infants fed powdered infant formulas. Pediatrics. 2012;130:e1076–84. https://doi.org/10.1542/peds.2011-3855.

1. Bowen A, Wiesenfeld HC, Klesz J, Pasuelle AW, Nowalk AJ, Brink L, Elliott E, Martin H, Tarr CL. Notes from the field: Cronobacter sakazakii infection associated with feeding extrinsically contaminated expressed human milk to a premature infant—pennsylvania. Morb Mortal Wkly Rep. 2017;66:761–2.

1. McMullan R, Mennon V, Beukers AG, Jensen SO, van Hai SJ, Davis R. Cronobacter sakazakii infection from expressed breast milk. Australia Emerg Infect Dis. 2018;24:393–4.

1. Chase JR, Gopinath GR, Eshwar KA, Stoller A, Fricker-Feer C, Gangiredla J, et al. Comparative genomic characterization of the highly persistent and potentially virulent Cronobacter sakazakii ST83, CC65 strain H322 and other ST83 strains. Front Microbiol. 2017;8:1136.

1. Gopinath GR, Chase HR, Gangiredla J, Eshwar A, Jiang H, Patel I, Negrete F, Finkelstein S, Park E, Jung T, Lee Y, Park J, Choi H, Kowar B, Kim M, Lee C, Jeong H, Fanning S, Stephan R, Iversen C, Reich F, Klein G, Lehner A, Tall BD. Genomic characterization of malonate positive Cronobacter sakazakii serotype O:2, sequence type 64 strains, isolated from clinical, food, and environment samples. Gut Pathog. 2018. https://doi.org/10.1186/s13909-018-0238-9.

1. Jiang H, Chase HR, Gangiredla J, Grimm CJ, Patel IR, Kotary MH, et al. Analysis of the molecular diversity among Cronobacter species isolated from fifth flies using targeted PCR, pan genomic DNA microarray, and whole genome sequencing analyses. Front Microbiol. 2020;11:561:204.

1. Jiang H, Woo J, Lee Y, Negrete F, Finkelstein S, Chase HR, Addy N, Ewing L, Beaubrun JIG, Patel I, Gangiredla J, Eshwar A, Jaradat ZW, Seo K, Shabarint S, Fanning S, Stephan R, Lehner A, Tall BD, Gopinath GR. Draft genomes of Cronobacter sakazakii strains isolated from dried spices bring unique insights into the diversity of plant-associated species. Stand Genomic Sci. 2018;13:35. https://doi.org/10.1186/s40793-018-0335-6.

1. Eid J, Fehr A, Gray J, Luong K, Kyle L, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibilbo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner P, Harnbol P, Heiner C, Hester K, Holden D, Kears G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra S, Shen G, Sorenson J, Tomanay A, Travers K, Trudson M, Vecelli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8.

1. Wick RR, Judd LM, Gorrie CL, and Holt KE Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6): e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

1. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O’Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Than K, Yamashita RA, Cheng Z, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Prutt KD. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res. 2018;46(D1):D851–60. https://doi.org/10.1093/nar/gky1068.

1. Overbeek R, Olson R, Puscheck GD, Olsen GJ, Davidsen JTF, et al. The SEED and the rapid annotation of microbial genomes using subsystems
technology (RAST). Nucleic Acids Res. 2014;42:D206–14. https://doi.org/10.1093/nar/gkt1226.
23. Stothard P, Grant JR, Van Domselaar G. Visualizing and comparing circular genomes using the CGView family of tools. Brief Bioinform. 2019;2019(20):1576–82.
24. Caratelli A, Hasman H. PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). In: de la Cruz F, editor. Horizontal gene transfer. New York: Humana, 2020. p. 285–94.
25. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44(W1):W16-21. https://doi.org/10.1093/nar/gkw387.
26. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39:W347–52.
27. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14:1394–403. https://doi.org/10.1101/gr.2389704.
28. Darling AE, Mau B, Perna NT. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 2010;5: e11147. https://doi.org/10.1371/journal.pone.0011147.
29. Chase HR, Gopinath GR, Gangireddy J, Patel IR, Kothary MH, Carter L, Sathymaurothy V, Lee B, Park E, Yoo YJ, Chung TJ, Choi H, Jun S, Park J, Jeong S, Kim M, Reich F, Klein C, Tall BD. Genome sequences of multiple-positiv e-cronobacter sakazakii serogroup O:2, sequence Type 64 strains CDC 1121-73 and GK1025, isolated from human bronchial wash and a powder ed infant formula manufact uring plant. Genome Announc. 2016;4(6):e01072-1116. https://doi.org/10.1128/genomeA.00082-13.
30. Franco AA, Hu L, Grim CJ, Gopinath G, Sathymaurothy V, Jarvis KG, Lee C, Sadowski J, Kim J, Kothary MH, McCardell BA, Tall BD. Characterization of putative virulence genes on the related repFIB plasmids harbored by Cronobacter spp. Appl Environ Microbiol. 2011;77:2355–67.
31. Tall BD, Gopinath GR, Gangireddy J, Patel IR, Fanning S, Lehrer A. Food microbiology fundamentals and frontiers. In: Doyle MP, Diez-Gonzalez F, Hill C, editors. Chapter 14. Cronobacter species. 5th ed. Washington, DC: ASM Press, 2019. p. 389–414.
32. Kucerova E, Clifton SW, Xia XQ, Long F, Porwollik S, Fulton L, Fronick C, Minx P, Kyung K, Warren W, Fulton R, Feng D, Wollam A, Shah N, Bhonagiri V, Nash WE, Hallsworth-Pepin K, Wilson RK, McClelland M, Forsythe SJ. Genome sequence of Cronobacter sakazakii BAA-894 and comparative genomic hybridization analysis with other Cronobacter species. PLoS ONE. 2010;5(3):9556. https://doi.org/10.1371/journal.pone.0009556.
33. Ikeda H, Tomizawa J. Prophage P1, and extrachromosomal replication unit Cold Spring Harb. Symp Quant Biol. 1968;33(791):798. https://doi.org/10.1101/sqb.1968.033.01.091.
34. Pfeiffer E, Moura de Sousa JA, Touchon M, and Rocha EPC. Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. Nucleic Acids Res. 2021;49:2655–73. https://doi.org/10.1093/nar/gkb064.
35. Eida AA, Bougouffa S, L’Haridon F, Alam I, Weisskopf L, Bajic VB, Saad MM, Hirt H. Genome insights of the plant-growth promoting bacterium Cronobacter muytjensii Z38 with volatile-mediated antagonistic activity against Phytophthora infestans. Front Microbiol. 2020;11:369. https://doi.org/10.3389/fmicb.2020.00369.
36. Srikumar S, Cao Y, Yan Q, Van Hoorde K, Nguyen S, Cooney S, Gopinath GR, Tall BD, Sivasankaran SK, Lehner A, Stephan R, Fanning S. RNA Sequencing-based transcriptional overview of xerotolerance in Cronobacter sakazakii SP291. Appl Environ Microbiol. 2019. https://doi.org/10.1128/AEM.01993-18.
37. Flores-Diaz M, Monturiol-Gross L, Naylor C, Alape-Girón A, Flieger A. Bacterial sphingomyelinases and phospholipases as virulence factors. Microbiol Mol Biol Rev. 2016;80:597–628. https://doi.org/10.1128/MMBR.00082-15.
38. Wang M, Cao H, Wang Q, Xu T, Guo X, Liu B. The roles of two type VI secretion systems in Cronobacter sakazakii ATCC 12868. Front Microbiol. 2018;22(9):2499.

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