ABSTRACT

Title of dissertation: DEVELOPMENT OF A BIOINFORMATICS PLASMID-SEARCH ENGINE FOR CRONOBACTER SPECIES.

Flavia Julissa Negrete Arenas, Master of Science, 2021

Dissertation directed by: Najib El Sayed, Ph. D., Professor, Cell Biology and Molecular Genetics, Bioinformatics and Computational Biology and Director, Integrated Life Sciences

Cronobacter species. are foodborne pathogens that cause serious disease in neonates, infants, and adults. Plasmid classification lays the groundwork for understanding the stable coexistence of various extrachromosomal replicons in a single bacterium, and thus the organization of its genome. This study developed a bioinformatics plasmid-search engine to identify genomic attributes contained on *Cronobacter* plasmids. A database containing 32 *Cronobacter* plasmid sequences from all seven *Cronobacter* species was developed. Another database containing 683 draft and closed plasmids and genomes was also developed. Each strain's plasmid content was sorted into six different categories based on their genetic attributes: virulence, Type-IV, heavy-metal, cryptic, multi-drug resistant, or mobilization. An in-house BLAST+-python script was used to perform a Linux-BLAST analysis to create a formatted %ID output matrix of plasmid genes. This thesis represents the first bioinformatics plasmid-search engine developed for *Cronobacter*. Understanding the role of plasmids in virulence and persistence underpins future mitigation strategies to be developed for controlling this pathogen.

DEVELOPMENT OF A BIOINFORMATICS PLASMID-SEARCH ENGINE FOR *CRONOBACTER* SPECIES.

by

Flavia Julissa Negrete Arenas

Thesis submitted to the Faculty of the Graduate School of the University of Maryland, College Park in partial fulfillment of the requirements for the degree of

Masters of Science

2021

Advisory Committee:

Professor Najib El- Sayed

Professor Vincent Lee

Dr. Ben D. Tall

© Copyright by

Flavia Julissa Negrete Arenas

2021

DEDICATION

To the mentors that have guided my path along this journey, thank you. To my mother, Marybel Arenas, and father, Cesar J. Negrete, who have supported my education through hours of labor, thank you. To Kiko, for being my best friend and the most loyal dog. I would like to express my deepest gratitude to my instructors, Dr. Ben Tall and Dr. Gopal Gopainath, and the principal investigators of the Division of Virulence Assesment, OARSA at U.S. Food and Drug Administration. Drs. Tall and Gopinath have both been instrumnetal for providing insightful knowledge and participating mentorship over the last couple of years. Their professional advice, expertise, and constructive criticism of my work has enabled me to improve and re-analyze my problem statements. The research would not have arrived to its current conclusions without the participation and the knowledge from both to whom I am entirely grateful. Additionally, the thesis could not have been completed without the knowledge and efforts of my friends, Dr. Hyein Jang and Katie Ko who were with me throughout every step of my research preparation. I would like to thank the US F.D.A. and the University of Maryland for sponsoring and offering me this un-imaginable opportunity to complete my studies. I will not forget Dr. El-Sayed whom for without his patience, guidance and understanding, I wouldn't have made it this far. To my advisor Dr. Buchner, who has had the patience to guide me through this program, I thank him. Lastly, I am deeply indebted to my committee for their invaluable patience and insight in this journey.

TABLE OF CONTENTS

DEDICATION ... ii
TABLE OF CONTENTS ... iii
LIST OF TABLES ... v
LIST OF FIGURES ... vi
INTRODUCTION ... 1
BACKGROUND ... 1
Natural History and Taxonomy of Cronobacter/Public Health Concerns ... 1
Pathology of Cronobacter ... 4
Multilocus sequence typing (MLST) analysis of Cronobacter species ... 6
Bacterial plasmids ... 11
History of bacterial plasmid classification ... 13
Introduction to Cronobacter plasmids ... 14
VIRULENCE GENETIC TRAITS ... 18
Metal acquisition and resistance ... 18
Cronobacter plasminogen activator (cpa) ... 20
Type I-VI Secretion Systems ... 23
Antimicrobial Resistance ... 25
MATERIAL AND METHODS ... 27
Bacterial Isolates ... 27
Plasmid sequences ... 29
Plasmidotyping Tool ... 30
RESULTS ... 37
Proposed Cronobacter plasmid classification scheme ... 37
Virulence Plasmids ... 40
Heavy Metal resistance plasmids ... 53
Multi-drug resistance plasmids ... 62
Conjugation-Type IV plasmids ... 68
Mobilization plasmids ... 73
Cryptic plasmids ... 77
CONCLUSION ... 85
CONFLICT OF INTEREST SHARED ... 88

ACKNOWLEDGEMENTS 88
FUNDING 88
SUPPORTING INFORMATION 89
REFERENCES 98
CURRICULUM VITAE 116

LIST OF TABLES

Table 1. Characteristics of known plasmids possessed by Cronobacter species used in this study.[a] ____ 29
Table 2. Summary of Cronobacter plasmids identified using the PlasmidFinder 2.1 tool located at https://cge.cbs.dtu.dk/services/PlasmidFinder/ .[a] ____ 35
Table 3. Proposed plasmid classification scheme of 32 Cronobacter plasmids across the seven Cronobacter species as developed for this study. [a] ____ 39
Table 4. Drug resistance genes shared between multi-drug resistance plasmids, pGW1 and p505108MDR. ____ 66

Supplemental Table S1. Functional annotations for C. sakazakii virulence plasmids conserved domains. ____ 89
Supplemental Table S2. Comprehensive information of Cronobacter strains evaluated in this study. ____ 95
Supplemental Table S3. Cronobacter isolates evaluated by species for presence of Cronobacter plasmids. ____ 95
Supplemental Table S4. Cronobacter isolates evaluated by ST for presence of Cronobacter plasmids. ____ 96

LIST OF FIGURES

Figure 1. Evolutionary distances and ST groupings for concatenated sequence of seven MLST loci for the genus Cronobacter using maximum likelihood tree adapted from Joseph et al.[20] ____ 11
Figure 2. The Plasmidotyper Reference Model specifies a framework for workflow systems attempting to characterize the presence of plasmid genes in Cronobacter species. ____ 32
Figure 3 A) Franco et al., represented repA in a phylogenetic cluster analysis. B) Phylogenetic cluster analysis of the virulence plasmid rep gene used in the study. ____ 43
Figure 4. Mauve progressive algorithm was used to align orthologous and xenologous regions among C. sakazakii virulence plasmid genome sequences shows both local and large-scale changes. ____ 47
Figure 5. Sequence alignment of virulence plasmid class members produced on the PROKSEE Server from the Stothard Research Group that uses BLAST analysis to illustrate conserved and missing genomic sequences (available online: https://beta.proksee.ca/tools).[93] ____ 49
Figure 6. Mauve progressive algorithm used to align orthologous and xenologous regions among C. sakazakii heavy metal plasmid sequences to show both local and large-scale changes. ____ 56
Figure 7. A) The Stothard Research Group's PROKSEE Server, which uses BLAST analysis to show conserved and missing plasmid sequences (available online: https://beta.proksee.ca/tools)[93]*, created a sequence alignment of heavy metal plasmid class members. B) Removal of pCS1 and pH322_1 from heavy metal class.* ____ 59
Figure 8. Phylogenetic cluster analysis of origin of replication gene of heavy metal plasmids members. ____ 59
Figure 9. Phylogenetic cluster analysis of the origin of replication genes (rep) of the multidrug resistance plasmids. ____ 64
Figure 10. Sequence alignment of multidrug resistance plasmid class members produced on the PROKSEE Server from the Stothard Research Group that uses BLAST analysis to illustrate conserved and missing genomic sequences (available online: https://beta.proksee.ca/tools).[93] ____ 68
Figure 11. Phylogenetic cluster analysis of members of the origin of replication family (rep) gene of conjugation type-IV plasmids. ____ 70
Figure 12. The Stothard Research Group's PROKSEE Server, which uses BLAST analysis to show conserved and missing genomic sequences (available online: https://beta.proksee.ca/tools), created a sequence alignment of conjugative-type IV plasmid class members.[93] ____ 73
Figure 13. Alignment of members of the mobilization plasmid class produced on the PROKSEE Server by the Stothard Research Group using BLAST analysis to highlight conserved and missing genomic sequences (available online at https://beta.proksee.ca/tools).[93] ____ 75
Figure 14. Phylogenetic cluster analysis of members of the origin of replication family (rep) gene of the mobilization plasmid class. ____ 76
Figure 15. Mauve progressive algorithm used to align orthologous and xenologous regions among C. sakazakii cryptic plasmid genome sequences to show both local and large-scale changes. ____ 79
Figure 16. Phylogenetic cluster analysis of plasmid origin of replication gene of each member of the cryptic plasmid class. ____ 81

Supplementary Figure 1. Geneious 9.1 map to reference toolkit was used to map ST8 cpa^pos. isolates against ST8 cpa^neg. ______ 94
Supplementary Figure 2. Plasmid for Conjugative Type IV species in Cronobacter species. The PROKSEE Server from the Stothard Research Group uses BLAST analysis to illustrate conserved and missing genomic sequences (available online: https://beta.proksee.ca/tools). ______ 94

LIST OF ABBREVIATIONS

Abbreviation	Meaning	Page
PIF	Powdered Infant Formula	12,14,15,18,24
NEC	Necrotizing Enterocolitis	12,14,15
CNS	Central Nervous System	14
CC	Clonal Complex	14,35
ST	Sequence Type	14
LPS	Lipopolysaccharide	16
MLST	Multilocus Sequence Typing	17
CGH	Comparative Genomic Hybridization	24
IS	Insertion Elements	22,27
MDR	Multidrug Resistance	22,33
INC	Incompatibility	23
T4SS	Type 4 Secretion System	23,33,48,83,92
WGS	Whole Genome Sequencing	24,11
T6SS	Type 6 Secretion System	25,33,38,44,48
ABC	ABC Iron Transporters	26,32
SIL	Silver Resistance	28
CPA	Plasminogen Activator	18-20,36,54,57,49-50
T1SS	Type 1 Secretion System	33
T2SS	Type 2 Secretion System	33
T3SS	Type 3 Secretion System	33
T5SS	Type 5 Secretion System	33
ESBL	Extended-Spectrum Β-Lactamase	33
TETA	Tetracycline	34,73
DFR	Trimethoprim	34,73
RCN	Nickel/Cobalt Resistance	34,73
QNR	Quinolone	34,73
MER	Mercury	34,74
ORFs	Open Reading Frames	34
DFR	Trimethoprim	34,73
NCBI	National Center for Biotechnology Information	35,36,39,46,47
RAST	The Rapid Annotation of Microbial Genomes Using Subsystems Technology	35,50,58
CFSAN	Center for Food Science and Applied Nutrition	35,36
KB	Kilobases	38,48,71
PMID	PubMed Identifier	38,48
BP	Base Pairs	44
CGE	Center for Genomic Epidemiology	45
REP	Origin of Replication	49,62,72,84,89
FHA	Filamentous Hemagglutinin	56
HIPA	Toxin-Antitoxin Gene	59
TA	Toxin-Antitoxin Systems	69
HGT	Horizontal Gene Transfer	76
CTNS	Conjugative Transposons	85
JIFSAN	Joint Institute for Food Safety and Applied Nutrition	95

INTRODUCTION

BACKGROUND

Natural History and Taxonomy of Cronobacter/Public Health Concerns

Cronobacter species are considered an opportunistic group of foodborne Gram-negative pathogens capable of causing intestinal, extra-intestinal, and systemic human disease.[1] These environmentally associated bacteria have been isolated from a variety of foods such as powdered infant and follow up formulae, dried milk protein products, cheese, licorice, candies, dried spices, teas, nuts, herbs, ready to eat foods such as pastas, fresh and frozen vegetables, as well as filth and stable flies, and environments of milk and cheese protein and powdered infant formula (PIF) manufacturing facilities.[1,2] The age groups at highest risk are neonates and infants, and adults over 80 years of age.[1] Neonatal and infantile *Cronobacter* infections may result in necrotizing enterocolitis (NEC), septicemia, and/or meningitis and have often been linked to the consumption of reconstituted, temperature-abused, intrinsically-or-extrinsically-contaminated PIF.[3,4] The mortality rate of *Cronobacter* infections is around ~27% often leaving those who survive with severe neurological and developmental disorders.[2] Adult infections may result in gastroenteritis, septicemia, pneumonia, urinary tract and wound infections; but linking the epidemiology of adult cases to consumption of PIF or infant foods, in general, is difficult to defend, thereby suggesting that there may still be unknown sources.[1]

The first isolate of *Cronobacter* was obtained from dried powder milk in 1950 (*C. sakazakii* NCTC 9238). Later, in 1953, the first medical cases of neonatal meningitis linked to *Cronobacter* were reported by Urmenyi and White-Franklin in 1961, when the organism was then still classified as a yellow pigmented strain of *Enterobacter cloacae.*[3] Later, DNA-DNA hybridization studies in combination with phenotypic studies reported by Farmer *et al.,* proposed classifying these organisms as a single species named *Enterobacter sakazakii,* after bacteriologist Ricchi Sakazakii.[4-6] Farmer *et al.,* described 15 biotypes or biogroups based on phenotypic results.[6] Iversen *et al.,* later reclassified these organisms, including a 16th biogroup into a novel genus, she named *Cronobacter* by utilizing DNA-DNA hybridization, phenotypic, and phylogenetic data to define the original six species, *Cronobacter sakazakii*, *Cronobacter malonaticus*, *Cronobacter turicensis*, *Cronobacter muytjensii*, *Cronobacter dublinensis,* and a genomospecies.[3] Joseph *et al.,* later proposed reclassifying the genomospecies as *Cronobacter universalis* and described *Cronobacter condimenti.*[3,5] Subsequently, a publication by Brady *et al.,* suggested that three non-pathogenic *Enterobacter* species (*Enterobacter pulveris, Enterobacter helveticus,* and *Enterobacter turicensis*) be included in the genus.[7] In rebuttal to this proposal, Stephan *et al.,* showed that these species were not related to either *Enterobacter* or *Cronobacter* and proposed reclassifying them into two new novel genera, *Franconibacter* and *Siccibacter,* respectively.[5] The genus *Cronobacter* currently consists of

seven species: *C. sakazakii*, *C. malonaticus, C. turicensis, C. muytjensii, C. dublinensis* (three subspecies: *Cronobacter dublinensis* subsp. *dublinensis, Cronobacter dublinensis* subsp. *lactaridi*, and *Cronobacter dublinensis* subsp. *lausannensis*), *C. universalis,* and *C. condimenti. C. sakazakii, C. malonaticus,* and *C. turicensis* are currently considered to be the primary pathogenic species and can cause infections in all age groups. *C. sakazakii* isolates represent almost 69.3% (n=2128) of the total *Cronobacter* isolates in the open access *Cronobacter* PubMLST database (http://www.pubmlst.org/*Cronobacter*/, last accessed 6/15/2021). In addition to being the most prevalent species associated with clinical infections, this species has also been linked to several fatal infantile NEC and meningitis cases around the world.[8,9] Studies showed that *C. sakazakii* is able to establish a systemic infection by invading and translocating through the intestinal epithelium into the lamina propria.[8-11] Once these tissue layers are breached, *C. sakazakii* can multiply within macrophages and microglial cells and overcome the physical host barriers of both the intestines and the central nervous system (CNS). Additionally, these isolates can attenuate human serum's bactericidal effects. *C. sakazakii* has distinct pathovars which are depicted as clonal complex (CC) lineages containing related sequence types (ST). Of clinical significance is *C. sakazakii* clonal complex 4 (CC4) strains that contain sequence type 4 (ST4) which have been epidemiologically associated with invasive

neonatal meningitis.[8] Another important *C. sakazakii* sequence type is ST12 which are strongly associated with NEC cases.[8-11,13]

Pathology of Cronobacter

It is suspected that *Cronobacter* can get into the human body through consumption of contaminated reconstituted PIF, contaminated expressed breast milk or contaminated foods.[1,12] Previous *in vitro* studies revealed that *Cronobacter* invades and translocate across both human intestinal epithelial and endothelial tissue culture cells, mimicking the blood brain barrier at a frequency equivalent to that of meningitis causing *Escherichia coli*.[8] Risk factors for NEC, according to Hunter *et al.*, include enteral nutrition using formulae, and unregulated bacterial intestinal colonization and multiplication.[13] The highest risk of infection occurs in a susceptible host suffering from an atypia physiological event such as hypoxia, hypothermia, and intestinal ischemia.[2,11] Additionally, a pathogen such as *Cronobacter* grows unrestricted in reconstituted powder infant formula (PIF) because of availability of nutrients and that these foods are devoid of the natural defenses normally present in breast milk.[12] These conditions in a host trigger increased mucosal inflammation, which leads to the development of high levels of host inflammatory factors such as cytokines, nitric oxide, platelet-activating factor, and prostanoids (mediators of pro-inflammatory reactions), which further harm the apical intestinal epithelium.[14,15]

Hunter *et al.,* presented evidence, using a neonatal rat *Cronobacter* infection model, that *Cronobacter* cells adhering to the intestinal epithelium, triggers barrier breakdown and induces a host pro-inflammatory response.[13] The recruitment of proinflammatory cytokines (such as IL-1 and TNF) characterizes this inflammatory response, which draws immune cells to the site. *Cronobacter* may directly enter the bloodstream or may be carried systemically within macrophages after the barrier is broken, causing sepsis and meningitis. When testing rat brain capillary endothelial (rBCEC4) and human macrophage (U937) cells, Hunter *et al.,* showed that *Cronobacter* could be internalized by both cells and that *Cronobacter* could persist within macrophages.[13] They noted that invasion of rat brain capillary endothelial cells may represent the pathogen's potential to cross the blood–brain barrier leading to central nervous system infections such as meningitis. Persistence in macrophages is a hallmark for many enteric pathogens such as *Salmonella* and the capacity of *Cronobacter* to survive in macrophages up to 96 hours can also be an important *Cronobacter* virulence trait. The degree to which other surface structures, including expression of lipopolysaccharide (LPS) and exopolysaccharide capsules may also allow pathogens to avoid serum killing. However, other factors, such as outer membrane proteins have been reported to improve environmental survival and plant commensalism, as well as various surface appendages that may help in attachment and colonization.[14] For instance, it has been found that expression of both

OmpA and OmpX is essential for *C. sakazakii* to invade the gallbladder and to reach other tissues and certain organs, such as the liver and spleen.[13-15] Furthermore it was demonstrated that OmpA and OmpX were needed for epithelial cell invasion by *C. sakazakii*.[14] Similarly, evidence suggests that the OmpA homologue expressed by *Neisseria meningitidis* could be a contributing factor in the development of human umbilical epithelial cell infection and may play a role in the pathogenesis of meningococcal neonatal meningitis.[13] Other factors such as *fkpA* (also known as macrophage infectivity potentiator-like protein) is a virulence factor known to contribute to Cronobacter' s ability to survive and replicate within macrophages and may also contribute to its persistence in high-stress growth environments.[7,17] Pathogen-directed demethylation, dephosphorylation and acetylation of histones have not been fully examined for *Cronobacter* as it has been for *Bacillus anthracis, Escherichia coli* and *Helicobacter pylori* infections.[18] The cataloging of *Cronobacter* epigenetic effects on host target cells is expected to lead to a better understanding of *Cronobacter* pathogenesis and may drive potential clinical treatment regimens, including the production of a vaccine for this significant foodborne human pathogen.

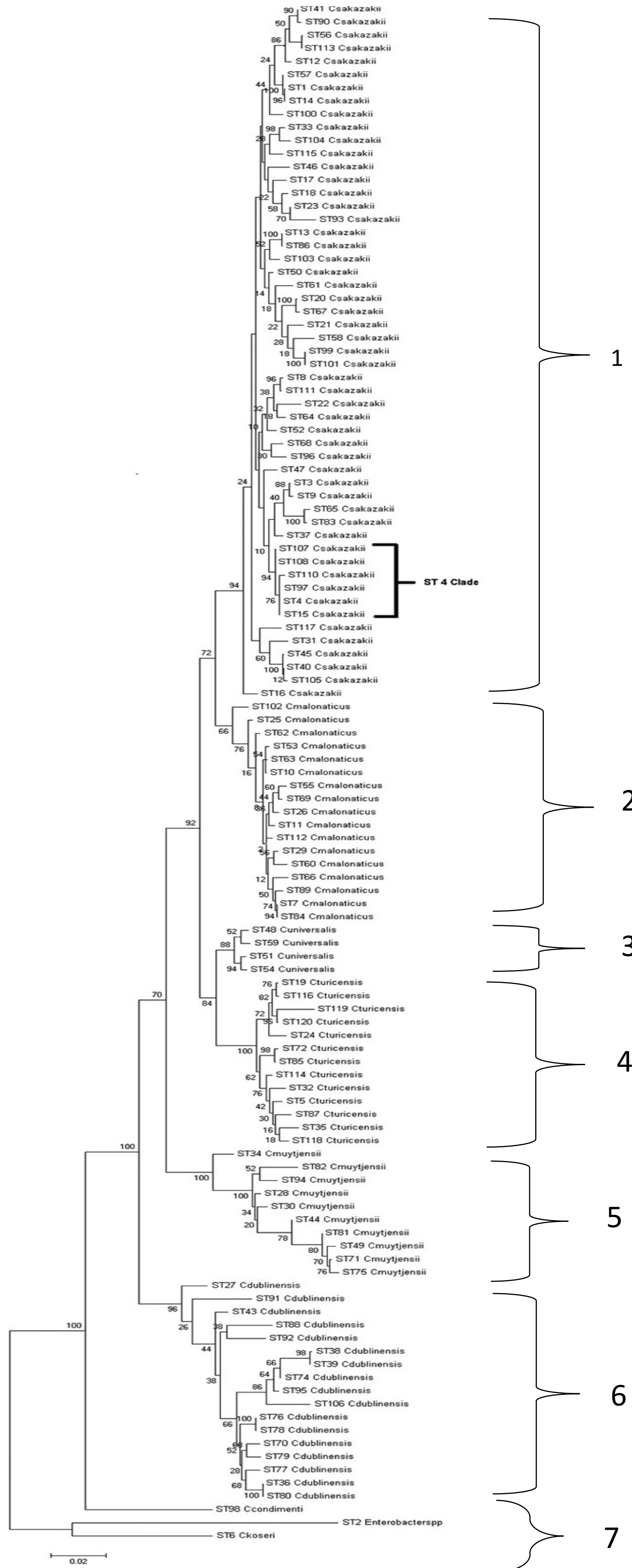
Multilocus sequence typing (MLST) analysis of Cronobacter species

The phylogenetic relationship and diversity of *Cronobacter* species within the *Enterobacteriaceae* have been previously analyzed using a multilocus sequence typing (MLST) approach employing the alleles of

seven genes (*atpD, fusA, glnS, gltB, gyrB, infB,* and *ppsA*; approximate total length 3,036 bp).[19-21] For the entire *Cronobacter* genus, a species-specific multilocus sequence typing (MLST) scheme has been developed and is available online at http://www.pubMLST.org/*Cronobacter*.[19] Based on this series of gene sequences, phylogenetic studies determined that distinct *Cronobacter* species developed during the last 40 million years, with *C. sakazakii* and *C. malonaticus* appearing as distinct species between 11 and 23 million years ago.[18] There are currently 759 sequence types for *Cronobacter* species, some of which appear to be stable clones.[19,21] Additionally, correlations between sequence type and source have been noted.[20] To explore the phylogenetic relatedness of 325 *Cronobacter* strains representative of each species using MLST analysis Joseph *et al.*, described the taxonomic placement within the genus of all seven species (**Figure 1**).[20] Using this combinatorial analytical method, several observations can be made about these findings. *C. universalis* strains ST48 and ST59 grouped among strains of known corresponding ST and CC designations such as ST51 and ST54. The phylogenetic trees illustrate strains causing the formation of species-specific clades. Clades were determined based on the proximity of each strain's phylogeny. Clade 1 included strains of several clinically relevant *C. sakazakii* sequence types including ST1, ST8, and ST4 with ST1 being the most prevalent ST, consisting of 36 strains isolated from around the world over a 25-year span. Most of these clinical isolates were found in human feces, sputum, and

blood samples. *C. sakazakii* ST4 is of note because it seems to have a strong proclivity for causing neonatal meningitis (with 75% of the isolates being linked to meningitis cases over a period of 50 years). This clone seems to be extremely stable, since clinical and non-clinical strains have been isolated from seven countries for more than 50 years. (Clade 4).[20] There have been studies showing that ST110 is connected to ST4 because they share four of the seven alleles.[19] This cluster of STs will be referred to as the " clonal complex 4" as an extension of the prior ST4 group due to the substantial relationship of ST4 and various locus variations (ST15, 97, 107, 108, and 110) with severe cases of meningitis, which is of high clinical relevance. Clade 2 contained *C. malonaticus* strains including ST7 *C. malonaticus* isolates obtained from clinical and PIF sources collected over 30 years. Clinical isolates from the Czech Republic make up STs 84 and 89. Other clades proposed in the study group strains into various *Cronobacter* species clusters. Furthermore, MLST analysis, has more sequence diversity at seven different loci than that seen for a single 16S rRNA gene that has been applied to a variety of bacteria. [19-21] Joseph *et al.,* highlighted the importance of sequence typing and emphasized the efficacy of the MLST scheme when trying to identify *Cronobacter*.[20] We have used a similar approach to study the diversity of this genus. Grim *et al.,* discovered that the *Cronobacter* species had diverged in its bidirectional phylogeny in accordance with species – specific clusters.[24] The C. *dublinensis*-*C. muytjensii* clade (Clade 5 and 6 represented in Figure 1)

evolved to be more adapted to an environmental and plant association niche, whereas the other clade, particularly *C. sakazakii* and *C. malonaticus*, (Clade 1-4 in Figure 1) developed and gained accessory genes that enhanced virulence, host species adaption, and pathogenicity.[23,25-27]



ST41 Csakazakii
ST90 Csakazakii
ST56 Csakazakii
ST113 Csakazakii
ST12 Csakazakii
ST57 Csakazakii
ST1 Csakazakii
ST14 Csakazakii
ST100 Csakazakii
ST33 Csakazakii
ST104 Csakazakii
ST115 Csakazakii
ST46 Csakazakii
ST17 Csakazakii
ST18 Csakazakii
ST23 Csakazakii
ST93 Csakazakii
ST13 Csakazakii
ST86 Csakazakii
ST103 Csakazakii
ST50 Csakazakii
ST61 Csakazakii
ST20 Csakazakii
ST67 Csakazakii
ST21 Csakazakii
ST58 Csakazakii
ST99 Csakazakii
ST101 Csakazakii
ST8 Csakazakii
ST111 Csakazakii
ST22 Csakazakii
ST64 Csakazakii
ST52 Csakazakii
ST68 Csakazakii
ST96 Csakazakii
ST47 Csakazakii
ST3 Csakazakii
ST9 Csakazakii
ST65 Csakazakii
ST83 Csakazakii
ST37 Csakazakii
ST107 Csakazakii
ST108 Csakazakii
ST110 Csakazakii
ST97 Csakazakii
ST4 Csakazakii
ST15 Csakazakii
ST 4 Clade
ST117 Csakazakii
ST31 Csakazakii
ST45 Csakazakii
ST40 Csakazakii
ST105 Csakazakii
ST16 Csakazakii
1
ST102 Cmalonaticus
ST25 Cmalonaticus
ST62 Cmalonaticus
ST53 Cmalonaticus
ST63 Cmalonaticus
ST10 Cmalonaticus
ST55 Cmalonaticus
ST69 Cmalonaticus
ST26 Cmalonaticus
ST11 Cmalonaticus
ST112 Cmalonaticus
ST29 Cmalonaticus
ST60 Cmalonaticus
ST66 Cmalonaticus
ST89 Cmalonaticus
ST7 Cmalonaticus
ST84 Cmalonaticus
2
ST48 Cuniversalis
ST59 Cuniversalis
ST51 Cuniversalis
ST54 Cuniversalis
3
ST19 Cturicensis
ST116 Cturicensis
ST119 Cturicensis
ST120 Cturicensis
ST24 Cturicensis
ST72 Cturicensis
ST85 Cturicensis
ST114 Cturicensis
ST32 Cturicensis
ST5 Cturicensis
ST87 Cturicensis
ST35 Cturicensis
ST118 Cturicensis
4
ST34 Cmuytjensii
ST82 Cmuytjensii
ST94 Cmuytjensii
ST28 Cmuytjensii
ST30 Cmuytjensii
ST44 Cmuytjensii
ST81 Cmuytjensii
ST49 Cmuytjensii
ST71 Cmuytjensii
ST75 Cmuytjensii
5
ST27 Cdublinensis
ST91 Cdublinensis
ST43 Cdublinensis
ST88 Cdublinensis
ST92 Cdublinensis
ST38 Cdublinensis
ST39 Cdublinensis
ST74 Cdublinensis
ST95 Cdublinensis
ST106 Cdublinensis
ST76 Cdublinensis
ST78 Cdublinensis
ST70 Cdublinensis
ST79 Cdublinensis
ST77 Cdublinensis
ST36 Cdublinensis
ST80 Cdublinensis
6
ST98 Ccondimenti
ST2 Enterobacterspp
ST6 Ckoseri
7
0.02

Figure 1. Evolutionary distances and ST groupings for concatenated sequence of seven MLST loci for the genus *Cronobacter* a using maximum likelihood tree adapted from Joseph *et al.*[20] The tree was constructed by Joseph *et al.*[20], using MEGA5 with 1,000 bootstrap replicates. Clade 1 includes sequence types 1 (ST1) and 14 of *C. sakazakii* (ST14). Clade 2 comprises *C. malonaticus* STs 7, 84, and 89 among other stable ST for the genus. Clade 3 includes *C. universalis* ST48 and ST51 strains, as well as two other ST *C. universalis* strains. Cluster 4 is composed of all *C. turicensis* strains which include *C. turicensis* isolates that are extremely stable, since clinical and non-clinical strains have been isolated from seven countries for more than 50 years and include the most well-studied *C. turicensis* sequence type ST32.[14] The remaining clades (Clades 5 and 6) diverge into species-specific clades with *C. muytjensii* in Clade 5 and *C. dublinensis* in Clade 6. *C. condimenti* is a lone outlier. The last cluster (Clade 7) contains enteric sequences that were used as non *Cronobacter* controls in the study.

Bacterial plasmids

In 1947, E. L. Tatum and Joshua Lederberg, described their work involving genetic recombination and organization of the genetic material of *E. coli*, another type of Gram-negative foodborne pathogen and bacteria in general.[29] Six years later, they defined "plasmid" as an extrachromosomal nucleic acid composed structure capable of reproducing autonomously from that of the bacterial chromosome. These studies were followed by those focused on plasmids containing antimicrobial resistance genes, later termed "R-factors".[30-33] From these studies, it was found that plasmids encoded much more than just antimicrobial resistance genes, but also

possessed genes for virulence and the metabolism of rare substrates and that they possessed certain properties including autonomous replication, mobility, host range, and incompatibility.[32] Horizontal gene transfer studies of *E. coli* that followed, broadened our understanding of plasmid introduction, transference, and replication among bacterial species.[34,35] In specific studies on F-factor, a fertility factor that allowed plasmid transfer and replication, provided insights on how plasmid transference to a recipient bacterial cell occurs, and led to the discovery that plasmids contain "replicons" responsible for the control of their own replication.[33-35] Plasmids were also shown to be involved in the transfer of multidrug resistance (MDR) genes which has led to a boost in the interest in plasmid genetics and improved understanding of bacterial conjugation now termed type four secretion system (T4SS).[35] From these results, Cohen *et al.,* demonstrated that plasmids can serve as vectors or as vehicles for DNA cloning.[36]

It was later discovered that properties possessed by plasmids such as autonomous replication, transmissibility, stability, and drug resistance originated from individual genetic modules within the plasmid. For example, the discovery of insertion sequences (IS) and other mobile genomic elements harbored by bacterial plasmids suggested that transposition events mediating deletions or introduction of specific gene clusters could result in the genetic remodeling of a plasmid and subsequently impart new phenotypes to the bacterium harboring them.[37]

These fundamental concepts of bacterial plasmid biology have paved the way to develop the molecular tools now needed to understand the emergence of pathogens and antibiotic resistance which has been the main stalwart of the One Health Initiative.[38] This initiative pivots by accelerating biomedical research breakthroughs, improving public health efficacy, rapidly extending the scientific knowledge base, to improve medical and veterinary medicine and clinical care, the continued synergy of which is to progress health care for millions into the twenty-first century.[38]

History of bacterial plasmid classification

In 1960, Watanabe *et al.,* reported the first attempt to classify plasmids based on the presence of a fertility inhibition factor, which inhibits the transfer of similar conjugative plasmids.[39] Because of its limitations, a new classification system replaced it by the late 1970s. This new classification system separated plasmids based on whether they could coexist inside the same host cell. Incompatible plasmids (*Inc*) refer to the inability of two plasmids to coexist stably over several generations within the same bacterial cell.[40] Commonly, closely related plasmids tend to be incompatible, while unrelated plasmids tend to be compatible. The most important reason for two plasmids being incompatible is that they both possess a replicon (an origin of replication gene) with the same specificity of a Rep protein or controlling elements. Since the plasmid replicon gene determines the Inc group, the terms "Inc" and "Rep" have been used synonymously to characterize plasmid types, such as IncFIIB, IncFIB, and

IncH1.[41] Further development of this classification system introduced the formation of unique DNA probes based on replicon-based regions (origin of replication genes), unique to each plasmid type.[40] Most virulence-associated plasmids in *E. coli* belong to the F incompatibility group.[41] Unlike that for *E. coli*, *Cronobacter* has no plasmid classification scheme yet.

Introduction to Cronobacter plasmids

Like many enteric pathogens, *Cronobacter* species also possess plasmids of varying size. Plasmids are thought to contribute to an organism's genomic plasticity, bacterial virulence, and survival.[2,42-43] Muytjens *et al.,* was the first group to characterize plasmids possessed by clinical *Cronobacter* strains (reported as *E. sakazakii*).[44] The strains were acquired during a six-year surveillance study of nosocomial neonatal meningitis and septicemia cases. Other researchers replicated this study by isolating *Cronobacter* strains from infected infants and comparing them to isolates from utensils and PIF formulations used to prepare baby formulas ingested by the infants. The results indicated that the strains (dubbed *E. sakazakii*) contained plasmids of various sizes..[45] Since then, whole genome sequencing (WGS) has allowed for the study of plasmids at the genomic level and during 2010-2011, the first closed *Cronobacter* genomes represented by *C. sakazakii* strain BAA-894 and *C. turicensis* strain LMG23827^T were made available.[43] Franco *et al.,* used this information to characterize two virulence plasmids, pESA3 and pCTU1 harbored by these

strains.[2] **Table 1** summarizes the plasmids currently found (as of August 2020) in the seven *Cronobacter* species. Discrimination of *Cronobacter* isolates based on NGS data, including plasmid sequences, will augment source-tracking investigations involved in the study of foodborne outbreaks caused by *Cronobacter* species.

In 2010, Kucerova *et al.,* revealed the first fully sequenced or closed genome of *C. sakazakii* ATCC BAA-894.[43] Additionally, this study revealed the sequence of the first two plasmids, pESA2 and pESA3, for the species. pESA2 and pESA3, were found to be 31 kB, (51% GC) and 131-kB, (56% GC) in size, respectively. This strain was isolated from a PIF sample reported to be used in a neonatal intensive care unit that was responsible for a 2001 outbreak of septicemia and meningitis in Tennessee, USA and was found to be indistinguishable by pulsed field gel electrophoresis with that of the clinical outbreak isolate.[43]

Later studies using comparative genomic hybridization (CGH) techniques, revealed that there exists a total of 4382 unique, annotated genes associated with the chromosome and plasmid genomes. Furthermore, only 54.9% of genes were common to all *C. sakazakii* with 43.3% being common across all *Cronobacter* species.[43] Other *Cronobacter* species, such as *C. turicensis*, also had their genomes fully sequenced (strain LMG 23827) in early 2011 by Stephan *et al.*[46] The sequenced strain was isolated from the blood of a post-mortem neonate with a diagnosis of meningitis. The genome size is 4.4 Mb (57% GC) and contains three plasmids of

approximate sizes 138-kB (pCTU1, 57%), 22-kB pCTU2, (49%), and 54-kB pCTU3 (50% GC).[46] Two hundred and twenty-three genes were annotated as virulence- and disease-related proteins, with 413 of 4455 encoded proteins were of unknown function. Ultimately, conserved hypothetical proteins were found that spanned across all *Cronobacter* species, but had not been functionally characterized.

In 2011, Franco *et.al.*, reported the analysis of two shared RepFIB plasmids harbored by *C. sakazakii* BAA-894 (pESA3) and *C. turicensis* LMG 23827 (pCTU1).[2] In addition to the shared RepFIB replicon gene, *repA*, pESA3 possessed a gene encoding for a plasminogen activator- (now called *Cronobacter* plasminogen activator, *cpa*), and a type 6 secretion system (T6SS) gene cluster. Using whole genome sequencing, Tall *et al.,* showed that plasmids pESA3 (131 kB) in *C. sakazakii* BAA-894, pSP291-1 (118 kB) in *C. sakazakii* SP291, and pCTU1 (138 kB) in *C. turicensis* LMG 23827 (alias z3032) are highly homologous in that these plasmid's origin of replication gene, *repA* share significant homology with one another.[45] Furthermore, *in silico* analysis by Franco *et al.*, revealed that both plasmids encode similar groups of genes or gene clusters which comprises a " plasmid backbone" consisting of *repA*, and two iron acquisition systems, an aerobactin-like siderophore (named cronobactin, *iucABCD/iutA*), and an ABC ferric-iron transporter gene cluster (*eitCBAD*), as well as several species-specific virulence gene determinants.[2,45] These plasmids have previously been grouped into a single plasmid category

called “Plasmid group 1”.[45] This plasmid backbone differentiates “Plasmid group 1” from other *Cronobacter* plasmids. Notable, Franco *et al.,* in collaboration with Grim *et al.*, discovered that the *iucABCD/iutA* gene cluster is the only functional siderophore.[2,47] Iron acquisition is an essential aspect of cell physiology for bacteria and for eukaryotes alike. Past studies have shown siderophore-dependent iron acquisition pathways are present in a wide range of prokaryotic and eukaryotic microbes (and even higher plants) and exhibit a great deal of variation in the structure and function of the components involved.[44] The recurring theme is that bacterial cells produce one or more siderophores during times of iron deficiency such as when intracellular iron concentration falls below a critical value (of approximately 10^{-6} M for microbial growth). Other proteins such as the ATP-binding cassette (ABC) transporters are specifically involved in heme/iron transport across membranes. Grim *et.al.,* described in total 14 different iron acquisition systems for *Cronobacter.*[47] The binding protein-dependent type (a subfamily of ABC transporters or traffic ATPases) is represented by three of the transport systems. They are made up of: [1] one or more extracellular (periplasmic) substrate binding proteins (BPs) that are anchored as lipoprotein(s) in the cytoplasmic membrane in Gram-positive cells and some Gram-negative cells, [2] one or two different (homodimer, heterodimer, or pseudoheterodimer) polytopic integral membrane proteins (IMPs), and [3] one or two different ATP hydrolases that face the cytoplasm and supply the system with energy. As a result, microbes have

evolved a variety of strategies for utilizing iron, a critical element for most organisms but not always readily available to them in the environment.[47,51]

VIRULENCE GENETIC TRAITS

The exchange of plasmids among bacteria transfers virulence and other relevant traits to host bacteria, promoting their rapid evolution and adaptation to various growth environments. Thus, plasmids are important "vehicles" for the transmission of virulence factors among bacteria. *Cronobacter* plasmids are strikingly interesting in that they are composed of mobile and exogenously acquired DNA that comprises several conjugative and composite transposons, a pathogenicity island, integrated plasmid genes and phage regions, and many insertions sequence (IS) elements. The next section discusses some of the changes in genomic organization and virulence factor research that has happened since the genome was fully sequenced. These virulence plasmids are thought to help *Cronobacter* species cause systemic disease, survive and persist in food manufacturing environments, and carry antibiotic resistance genes, all of which are serious problems for human health.

Metal acquisition and resistance

Arsenic resistance systems are primarily based on membrane transport pathways (termed efflux) that expel toxic substances from the cytoplasm of a cell.[52] The *ars* operon, associated with bacterial R-factors and discovered almost 50 years ago, are the most prevalent microbial arsenic resistance system.[52] Numerous *ars* operons, each containing a

diverse set of genes in various combinations, populate prokaryotic genomes, as do their accessory plasmids, transposons, and genomic islands.[52] *arsRDABC* gene clusters have been reported in arsenic-resistant prokaryotes.[52,53] Arsenite efflux can be energized in two ways: by ATP hydrolysis catalyzed by ArsA encoded by *arsA* in complex operons, or by the membrane potential in *arsRDABC* operons, with ArsB being guided by the protein motive force.[52,54] Other genes, such as *arsR*, encode for a transcriptional repressor protein that binds to the *ars* operon promoter site. ArsA is an ATPase that interacts with ArsB to form an ATP-activated arsenite efflux pump. Lastly, ArsC are arsenate reductase enzymes, and ArsD is a regulatory repressor for the *ars* operon.[53] Arsenic resistance plasmids have been reported to encode different variations of the *arsRDABC* operon, with some plasmids showing truncated motifs. Toxicity of arsenate is due to its ability to compete for transport and energetic roles with phosphate oxyanions; therefore, the main toxic effects of arsenate result from its transition to arsenite.[52]

Even though copper is needed for biological processes, excessive amounts are highly toxic to bacterial cells. Numerous studies have shown that heavy metal resistance genes, such as *pco* genes, are not always carried on plasmids but can also be present on the chromosomes of *Enterobacteriaceae* species such as *Salmonella enterica* subspecies *enterica* and *E. coli*.[53-55] As previously shown, *pco* operons can be identified as part of a Tn7-like arrangement on the chromosome or on

plasmids as described in *E. coli* strains possessing a silver resistance (*sil*) operon.[54] The *sil* operon comprises nine open reading frames (ORFs), seven of which are apparently structural genes (*silE, silC, silF, silB, silA,* ORF105 and *silP*) and two of which (*silR* and *silS*) encode a putative two-component regulatory circuit. Similarly, *cusCBA* confers tolerance to copper, silver ions and other metals such as zinc, cadmium, and cobalt.[55]

Cronobacter plasminogen activator (cpa)

After the discovery of one of the first plasmids in *Cronobacter*, pESA3, sequence analysis of this plasmid revealed the presence of an outer membrane protease containing an amino acid sequence homologous to members of the omptin membrane protein family.[45] Members of this subfamily act as housekeeping proteases that coat the outer membrane of various *Enterobacteriaceae*.[56] Besides being active in proteolysis, proteolytically-inactive omptins can carry functions such as adhesion and invasion. *Cronobacter* OmptinT-like proteases, now named *Cronobacter* plasminogen activator (Cpa), enhance bacterial colonization by degrading cationic antimicrobial peptides, and several serum proteins providing protection from complement-dependent serum killing.[57] Jang *et al.*, described the phylogenetic relatedness among the OmptinT family of proteins possessed by *C. sakazakii* and *C. universalis* (carried on its virulence plasmid pCUNV1).[58] Interestingly, Cpa shares significant identity with the plasminogen activators Pla of *Yersinia pestis* and PgtE of

Salmonella enterica. PlaA is another example of an omptin-like protease found in *Erwinia* species.[57]

Cpa's ability to degrade several host serum proteins including circulating complement components is thought to cause the rapid activation of plasminogen and inactivation of a plasmin inhibitor, alpha 2-antiplasmin (α2-AP).[58-59] The observed plasminogen activation by Cpa is carried out by proteolytically cleaving critical proteins of the complement cascade: C3, C3a, and C4b. The disruption of complement activation would prevent lysis of a Gram-negative bacterium. Thus, proteolytic inactivation of C3b, C4b, and C5 is likely to increase the systemic survival of a bacterium within a host. Additionally, studies have suggested that the uncontrolled degradation of fibrin clots and extracellular matrix proteins found in *C. sakazakii* infections could be attributed to unretained plasmin production.[2,59]

pESA3 possesses conserved *cpa*-flanking regions in pCTU1 and the other virulence plasmids possessed by all *Cronobacter*; however, pESA3 shows a *cpa* locus of 1427 bp, while pCTU1 shows a collapsed region of 37 bp.[45,58] On pESA3, *cpa* (pESA3p05434) is flanked upstream by an MFS-1 homologue and downstream by genes encoding for proteins associated with carbapenem resistance. A palindromic inverted repeat present in the *cpa* locus of both plasmids of 10 to 13 nucleotides, may act as a transposon attachment site.[2,58] Interestingly, the region of pCTU1 which contains the conserved flanking regions instead of the *cpa* locus and

upstream of this, is a palindromic inverted repeat, while the *cpa* locus found on pESA3, this region is located downstream. The notion of an inverted repeat functioning as a transposon site would clarify the acquisition, evolution, and existence of *cpa* on pESA3 and pCUNV1. Furthermore, a transposon site may suggest a mechanism for the microevolution of *cpa* encoded on pESA3-like plasmids from a prototypical plasmid backbone through the co-integration or deletion of virulence determinants in each *Cronobacter* species. PCR analysis using oligonucleotides designed to detect *cpa* showed that 98% of *C. sakazakii* harbor the *cpa* gene.[58] Moreover, only two other strains of *C. universalis* harbored the *cpa* gene while all other *Cronobacter* species were PCR-negative for *cpa*, suggesting that the presence of *cpa* may be species-specific. Collectively, these results revealed a substantial degree of polynucleotide sequence heterogeneity in virulence plasmids. Furthermore, evolutionary evidence that suggests an ongoing process of sequence rearrangements, microevolution, or co-integration or deletion is occurring in these plasmids and possibly other replicons which occur more frequently than has hitherto been appreciated.[58]

When evaluating the role of *cpa* in a Zebrafish infection model, a *cpa*-deficient mutant (*C. sakazakii* ATCC BAA-894△*cpa*) exhibited a 10% mortality rate in comparison with 80% mortality rate with the wild-type parental strain (ATCC BAA-894).[60] Furthermore, evaluating the complemented strain (*C. sakazakii* ATCC BAA-894△*cpa*/pQE30::*cpa*) in

the infection model demonstrated a 40% restoration of lethality.[60] In addition to a decrease in virulence, mutants showed a slower growth rate after 24 hours of infection. Cpa remains an important virulence factor that *C. sakazakii* and *C. universalis* may use as a means to counteract serum killing and promoting host systemic spread. However, because there was not an absolute abolishment of virulence in the Zebrafish infection model with the *cpa* mutant, these findings also suggest that other virulence factors are involved. [2,60]

Type I-VI Secretion Systems

Many well-known bacterial pathogens possess secretion systems which are protein transport nanomachines used by numerous Gram-negative bacterial pathogens to secrete several critical virulence factors. In Gram-negative bacteria, the dedicated secretion systems are designated Type I through Type VI, with each system transporting a distinct subset of proteins. A key distinction amidst these systems is the existence of a *sec-tat* secretion pathway that effectively secretes a folded protein through the cell wall. Distinctly, T2SS T5SS and TISSs lack this mechanism, while the other three classes are *sec-tat* dependent.

T1SSs are structurally like a wide family of ATP-binding cassette (ABC) transporters that are involved in the export of small molecules such as antibiotics and toxins from cells.[61-62] T2SSs secrete enzymes that aid in metabolism and their adaptation within their habitat, which may include plant and animal hosts.[63] The T2SS is needed for the secretion of several

proteins (generally enzymes) in some bacteria but is not required for the secretion of a single protein in others.[63-64] Type III secretion systems (T3SSs) are evolutionarily related to the flagellar apparatus, consisting of over 20 proteins and a sequence of ring structures found in the bacterial inner and outer membranes that allows the delivery of partly unfolded virulence effector proteins into a host cell or the environment.[66,67] As with T3SSs, T4SSs can span an additional host cell membrane, allowing substrates to be directly transferred into the recipient cell's cytoplasm.[67,68] T4SSs can perform a number of functions due to their ability to transfer both DNA and proteins, including conjugative DNA transfer, DNA uptake and release, and direct translocation of effector proteins or DNA/protein complexes into recipient cells.[68] Unlike other secreted substrates, which cross the bacterial membrane through a dedicated secretion apparatus or membrane channel, Type V secretion system (T5SS) substrates secrete themselves, this is why they are sometimes referred to as autotransporters.[69] These classes of proteins each contain repetitive β-barrel domains that insert into the outer membrane to create channels to funnel protein transport in and out of the cell .[69] Unlike many other Gram-negative secretion systems, T6SSs are capable of transmitting effector proteins between bacteria in a contact-dependent manner, which is thought to contribute to bacterial-bacterial communication and interactions in the environment.[70,71]

Antimicrobial Resistance

Antimicrobial resistance has been identified as a critical public health and social problem of grave importance. The issue is particularly severe where strains resistant to several (multidrug-resistant [MDR]) or nearly all available agents (extensively drug-resistant) are emerging.[72-76] The first antimicrobial agents with Gram-negative activity were identified in the 1940s by routine sampling of soil *Actinobacteria*. The spread of extended-spectrum β-lactamase (ESBL)-producing *Enterobacteriaceae* in the 1990s was especially troubling, as was the appearance and rapid spread of carbapenemase-producing species.[77] The three main groups of antimicrobial agents with Gram-negative activity are β-lactams (particularly β-lactam-β-lactamase inhibitor hybrids, later-generation cephalosporins, and carbapenems), fluoroquinolones, and aminoglycosides.[77-79]

Zeng *et al.,*[80] identified a *C. sakazakii* isolate ("505108"), obtained from a female neonate with hyperpyrexia in Nanjing City, China in 2015, possessing three plasmids: [1] p505108-Multi-drug resistance (MDR), [2] p505108- New Delhi metallo-β-lactamase 1 (NDM), and [3] p505108-T6SS.[80] These plasmids had circularly closed DNA sequences of 312,880, 53,793, and 139,553 bp in length, with mean G+C contents of 47.7, 49.0, and 56.4 %, respectively, and contained 359, 62, and 126 predicted open reading frames (ORFs), respectively. Shi *et al.,* performed *in silico* analysis of the genomic landscape of p505108-MDR which revealed multiple

regions of multi-drug resistance.[80] Two regions, labeled MDR-1 and MDR-2, acquired several accessory modules which carry an extremely large number of resistance genes such as resistance to carbapenems, aminoglycoside, tetracyclines, and phenicols and sulphonamide/trimethoprim.[80,81] The MDR-1 region is a 25-kB conserved domain that holds antibiotic resistance gene cassettes for trimethoprim (*dfr*), aminoglycoside (*aph3*, *strAB*), and nickel/cobalt (*rcn*) resistance. MDR-2 is 50.2 kB in length and confers resistance for beta-lactam (bla_{SHV-12}, bla_{DHA-1}), aminoglycoside (*aacC3*, *aacA27*), quinolone (*qnr*), tetracycline (*tetA*), phenicol (*catA2*), sulphoamide (*sul1*), rifampicin (*arr7*), and mercury (*mer*).[81]

The aim of this study is to develop a bioinformatics plasmid-search engine to identify genomic attributes contained on *Cronobacter* plasmids. Only a few complete genomes of *Cronobacter* plasmids are available in public databases. Therefore, an effort was made to analyze plasmid sequences and those of closed genomes of *Cronobacter* species that can be used in NGS analysis to support detailed source tracking investigations. This research project involves the development of the first bioinformatics plasmid search engine for *Cronobacter*. Understanding the function of plasmids in virulence and survival lays the groundwork for developing potential preventive strategies for combating this foodborne pathogen.

MATERIAL AND METHODS

Bacterial Isolates

Six hundred and eighty three *Cronobacter* genome sequences encompassing the entire collection of sequence strains housed by the U.S. Food and Drug Adminsitration for *Cronobacter* were used in this study (see Supplemental **Table 1**). All strains were isolated from various surveillance studies reported by Restaino *et al.,* Jaradat *et al.,* Chon *et al.,* Yan *et al*., Gopinath *et al.,* Chase *et al*., and Jang *et al.*[82-88] Every strain present in the collection was identifed as *Cronobacter* using species-specific (*rpoB* and *cgcA*) PCR assays, except for those genomes obtained from the National Center for Biotechnology Information (NCBI).[82-88] Genome assemblies (FASTA) were submmited to the *Cronobacter* MLST website (http://www.pubmlst.org/*Cronobacter*) for determination of allelic, sequence type (ST), and clonal complex (CC) profiles.[12,16,89] In addition, MLST was confirmed by submitting the genomes to the Center for Food Safety and Applied Nutrition (CFSAN) Galaxy GenomeTrakr website's MLST tool (https://galaxytrakr.org/root/) for analysis.[89] For those genomes missing annotation, FASTA files were uploaded to the Rapid Annotation Subsystem Technology (RAST) server (online annotation; http://rast.theseed.org).[90]

Reference genomes were obtained from NCBI. Nucleotide sequences of the strains were deposited into NCBI's GenBank and were released to the public through submission to NCBI under the FDA-CFSAN

bioproject *Cronobacter* GenomeTrkr Project (PRJNA258403), which is part of the FDA's Center for Food Safety and Applied Nutrition (CFSAN) foodborne pathogen research umbrella project at NCBI (PRJNA186875) project. As part of this study the genome for *C. sakazakii* strain H322 was closed and its chromosome and plasmids, pH322-1 and pH322-2 assemblies, along with the PGAP annotations were released under FDA GenomeTrakr Bioproject on NCBI (PRJNA258403) as described above.[91] *C. sakazakii* H322, a strain that was originally described by Chase *et al.*[87], represented a persistent strain that was repeatedly isolated over a period of five years from the environment of a Swiss powdered infant formula manufacturing facility.

The accession numbers of the genomes and plasmids used in the study are shown in **Table 1.** To understand plasmid homology between various STs in *Cronobacter*, we concentrated our efforts to understanding the presence of conserved homology between the isolates and each plasmid. Twenty *Cronobacter* STs indicative of the seven species have been used primarily because these STs represent the most virulent STs. *C. sakazakii* ST8 isolates were an intriguing group because they demonstrated the existence of *cpa* in some, but not in all isolates. To represent the distinction among this ST, we added csak35 and csak295, isolates to reflect the presence and absence of the *cpa* gene in *C. sakazakii* ST8 isolates.

Plasmid sequences

Thirty-two *Cronobacter* plasmid sequences were downloaded from the NCBI website and are shown in **Table 1.**

Table 1. Characteristics of known plasmids possessed by Cronobacter species used in this study.[a]

Cronobacter species	Plasmid Name	RefSeq Record[b]	INSDC Record[c]	Size (Kb)	GC %	CDS	References (PMID)
	CSK29544 _1p	NZ_CP011048	CP011048	93.9	57.02	61	6920972
	pCS2	NZ_CP012255	CP012255	117.8	57.23	103	27013041
	pCsaCS09a	NZ_CP027110	CP027110	131.2	56.85	116	Unpublished
C. sakazakii	pGW2	NZ_CP028976	CP028976	135.3	56.81	117	30334728
	pESA3	NC_009780	CP000785	131.2	56.85	116	3500316
	pSP291-1	NC_020263	CP004092	118.1	57.23	106	23516209
	pH322_2	This study	CP078112	118.1	56.8	118	This study
C. malonaticus	p1	NC_023024	CP006732	126.5	57.29	101	24435860
	pCMA1	NZ_CP013941	CP013941	126.5	57.29	104	27013041
C. dublinensis	pCDU1	NZ_CP012267	CP012267	197.3	56.84	160	27013041
C. turicensis	pCTU1	NC_013283	FN543094	138.3	56.05	109	21037008
C. universalis	pCUNV1	NZ_CP012258	CP012258	129.8	57.03	114	27013041
C. muytjensii	CmuyJZ38 _p2	NZ_CP017664	CP017664	145.8	58.1	113	32218777
C. condimenti	pCCO1	NZ_CP012265	CP012265	151.5	54	125	27013041
C. sakazakii	CSK29544 _2p	NZ_CP011049	CP011049	4.9	54.88	4	6920972
	pCSA2	NC_021293	KC663407	5.1	55.01	6	25332122
	pSP291-3	NC_020262	CP004094	4.4	54.03	6	23516209
	CSK29544 _3p	NZ_CP011050	CP011050	53.5	50.07	58	6920972
	pCS1	NZ_CP012254	CP012254	110.1	50.71	122	27013041
	pCS3	NZ_CP012256	CP012256	53.4	49.25	55	27013041
	pSP291-2	NC_020261	CP004093	52.1	49.16	55	23516209
	pH322_1	This study	CP078111	100.7	50.2	137	This study
C. malonaticus	p2	NC_023025	CP006733	55.9	50.38	59	24435860
	pCMA2	NZ_CP013942	CP013942	52.6	50.25	55	27013041
C. turicensis	pCTU3	NC_013285	FN543096	53.8	50.04	55	21037008
C. sakazakii	pESA2	NC_009779	CP000784	31.2	51.58	36	3500316
	pGW1	NZ_CP028975	CP028975	340.7	48.58	358	30334728
C. turicensis	pCTU2	NC_013284	FN543095	22.4	49.22	27	21037008
C. sakazakii	p505108-MDR	NZ_KY978628	KY978628	312.9	47.7	352	28771073
	p505108-NDM	NZ_KY978629	KY978629	53.8	49	40	28771073
	p505108-T6SS	NZ_KY978630	KY978630	139.6	56.4	125	28771073
C. muytjensii	CmuyJZ38 _p1	NZ_CP017663	CP017663	109	49	142	32218777

[a] Information was obtained from NCBI and then summarized, last accessed 12.31.2020.

[b] RefSeq refers to Reference Sequence (RefSeq) collection record

[c] INSDC refers to International Nucleotide Sequence Database Collaboration (INSDC) record.

Plasmidotyping Tool

The plasmid search engine, named "Plasmidotyper," is a nucleotide database of 32 *Cronobacter* plasmids formatted using the National Center for Biotechnology Information's Basic Localization Alignment and Search Tool (BLAST) command line tools. Six hundred thirty-eight *Cronobacter* genomes from FDA's in-house database were analyzed for the presence of all *Cronobacter* plasmids using the Plasmidotyping search engine. Specific BLASTn parameters for each run consisted of the following: number of threads = 12, e- value= 1E-10, score edge =0.05, best hit overhang = 0.25, percent identity=70%. In-house python scripts were written to transform BLAST results from their original format into a datapoint matrix in Excel. The program is composed of three python scripts: the first executes the BLAST command, the second converts the output to .csv format, and the third generates a datapivot matrix for the results. This matrix contains the %ID results of the BLAST run between the nucleotide plasmid database and the 638 *Cronobacter* genome sequences. A total of 2954 protein encoding genes were identified; of these, 794 genes were hypothetical proteins. Each hypothetical gene encoding protein was surveyed for sequence similarity using BLASTX. Using this analysis, 56 hypothetical proteins were better annotated to the highest %ID sequence similarity results. An overview of the

schematic for the program can be seen in **Figure 2.** The proposed workflow is comprised of three elements: 1) Data Format – element; 2) BLAST+ analysis – element; 3) Datapivot-matrix presence activity element. For any one incoming genomic sequence (FASTA), the Plasmidotyper program runs a BLASTn analysis against the built-in nucleotide plasmid database (2-element). The program will then create a datapivot matrix that will use the plasmid scheme to categorize BLAST results against the ST and *Cronobacter* species. Activities used in the workflow were configured with several elements, which makes it possible to reuse the same activity in different applications. For the purposes of this study, we analyzed the presence or absence of plasmids and their attributes harbored by the seven *Cronobacter* species.

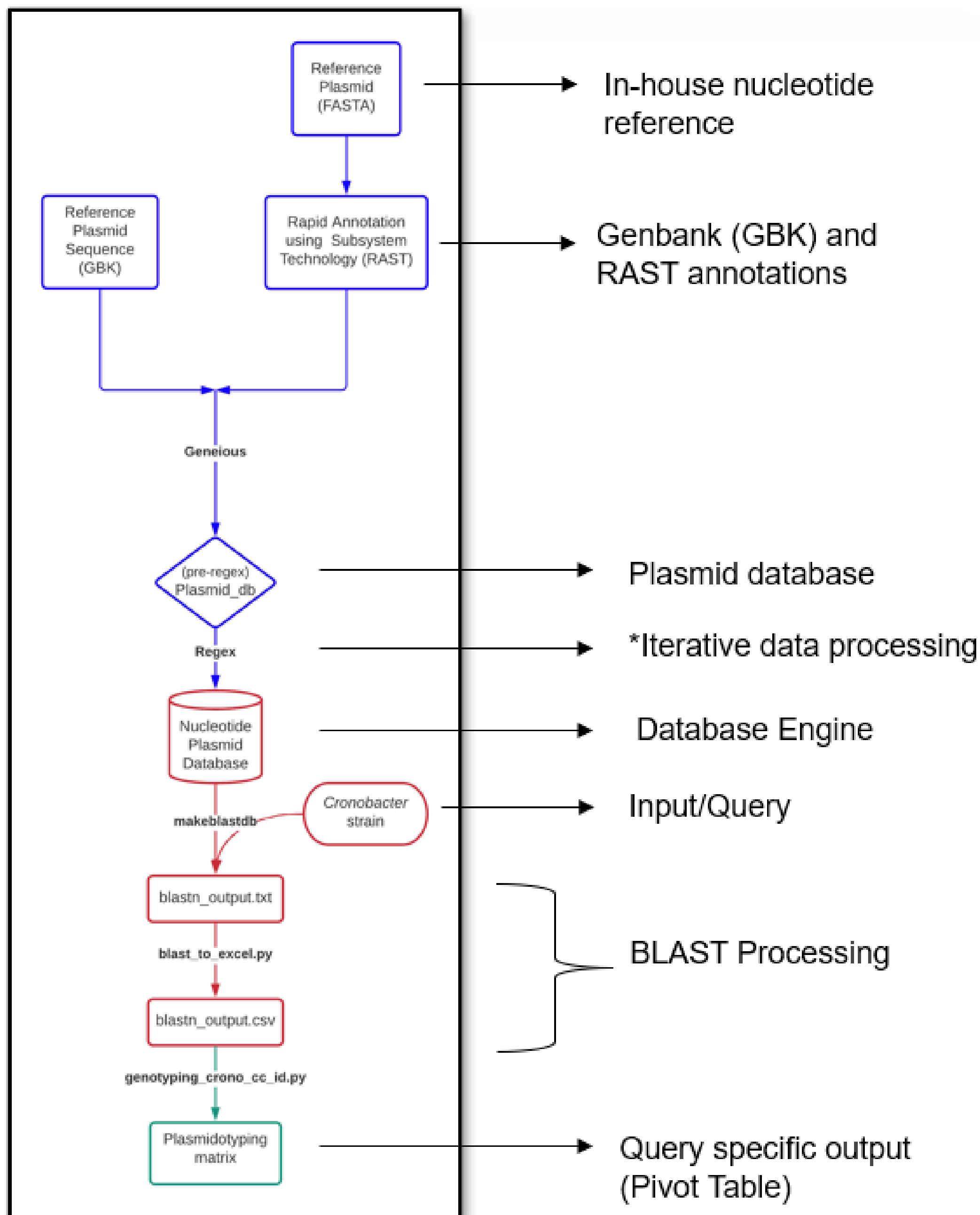


Figure 2. The Plasmidotyper Reference Model specifies a framework for workflow systems which attempts to characterize the presence of plasmid genes in *Cronobacter* species. Plasmidotyper has three main elements: 1) Formatting element (Blue colored section); 2) BLAST+ analysis element (Red colored section); and 3) Datapivot-matrix

presence/absence activity element (Green colored section). GenBank files are formatted into the plasmid database using EditPad Lite 9.1 using regular expression (Formatting Element). Modular architecture of the plasmid database is upheld when a new *Cronobacter* plasmid sequence or plasmid classes may need to be added to the exisiting database. This modular architecture was designed to (1) deliver and adapt the core base to accommodate features needed to address specific challenges (e.g., sequences, filetypes), (2) add input plasmid/draft genomic sequences to the database engine without disruption of the workflow, and (3) offer flexibility in adding new plasmids and or classes to the classification schemes. For example, a new plasmid category can be added to the database easily without having to disrupt the already built-in architecture (Analysis Element). Plasmidotyper generates a datapivot matrix which contains the %ID results of the BLASTn processesses. (Results Element). Curation of quality controls is achieved after the program highlights error messages to which the user can respond accordingly. Iterative data processing occurs after every addition of new plasmid sequences brought into the plasmid engine. After addition of input sequences to the plasmid database, automated classfication of plasmids will occur in all future uses and manual invovlement is no longer required. A seventy percent *Cronobacter* threshold is used for nucleotide diversity which was was based on previous microarray studies.[101] BLAST parameters include: number of threads = 12, e- value= 1E-10, score edge =0.05, best hit overhang = 0.25, percent identity=70%. BLAST parameters give e-value cuts out the possiblity outside of BLAST length criteria.

Comparative Genomic Analysis

Progressive Mauve implementation using the Geneious 9 suite[91] was used for global alignment. Gapped aligner was implemented on using UGENE 38.1

software (available for download at: http://ugene.net/) using MAFFT for with a match seed weight of 15, a minimum LCB score of 30,000 at full alignment.[92,96] After BLAST alignments, the PROKSEE Alignment Tool (https://beta.proksee.ca/projects) was used to generate high-quality navigable maps of the circular genomes.[93] Each *Cronobacter* sequence was submitted to CGE's Plasmidfinder Tool (https://cge.cbs.dtu.dk/services/PlasmidFinder/) for *in silico* determination of incompatibility plasmids such as IncF, IncHI1, IncHI2, IncN, and IncI1 plasmids (at least six plasmid sequences from each incompatibility group with a MLST scheme) to the Web server. [95] Results are presented in **Table 2**. All isolates contained best-matching hits for each replicon sequence was given as output, using a % ID threshold value of 80% nucleotide identity and >96% coverage. Plasmids have been divided into homology groups using the new plasmid nomenclature dependent on incompatibility groups using the PlasmidFinder 2.1 database and analytical tools available at https://cge.cbs.dtu.dk/services/PlasmidFinder/ [Current Software version: 2.0.1 (2020-07-01); Database version: (2021-07-12)].[95]

Table 2. Summary of Cronobacter plasmids identified using the PlasmidFinder 2.1 tool located at https://cge.cbs.dtu.dk/services/PlasmidFinder/ .[a]

Category	***Cronobacter* species**	**Plasmid**	**Plasmidofinder Result**	**Nucleotide position (bp)**	**Accession Number**	**Identity**
Virulence	*sakazakii*	CSK29544_1p	No hit	-	-	-
		pCS2	No hit	-	-	-
		pCsaCS09a	No hit	-	-	-
		pGW2	No hit	-	-	-
		pESA3	No hit	-	-	-
		pSP291-1	No hit	-	-	-
		pH322_2	No hit	-	-	-
		p505108-T6SS	No hit	-	-	-
	malonaticus	p1	No hit	-	-	-
		pCMA1	No hit	-	-	-
	dublinensis	pCDU1	No hit	-	-	-
	turicensis	pCTU1	IncFIB	110412..111220	FN543094	100
	universalis	pCUNV1	No hit	-	-	-
	muytjensii	CmuyJZ38_2	No hit	-	-	-
	condimenti	pCCO1	No hit	-	-	-
Cryptic	*sakazakii*	CSK29544_2p	No hit	-	-	-
		pCSA2	No hit	-	-	-
		pSP291-3	No hit	-	-	-
Heavy Metal Resistance	*sakazakii*	CSK29544_3p	IncFIB	16107..16797	FN543096	99.13
		pCS3	IncFIB	6795..7485	FN543096	99.13
		pSP291-2	IncFIB	32668..33358	FN543096	99.13
	malonaticus	p2	IncFIB	45366..46056	FN543096	99.13
		pCMA2	IncFIB	2185..2875	FN543096	99.28
	turicensis	pCTU3	IncFIB	2655..3347	FN543096	100
Conjugation type-VI	*sakazakii*	pESA2	IncFII; pESA2	17198..17774;2497..3246	FN543095;CP000784	96.71;100

	turicensis	pCTU2	IncFII; pESA2	8327..8903, 2477..3227	FN543095; CP000784	100;97.87
Multi-drug resistance	*sakazakii*	pGW1	IncHI2; IncHI2A	312987..313313; 294290..294815	BX664015; BX664015	100;100
	sakazakii	p505108-MDR	IncHI2; IncHI2A	297899..298225;211..840	BX664015; BX664015	100;100
Mobilization	*muytjensii*	CmuyJZ38_1	No hit	-	-	-
	sakazakii	pCS1	No hit	-	-	-
		pH322_1	No hit	-	-	-
Multi-drug resistance	*sakazakii*	p505108-NDM	IncX3	1..374	JN247852	100

[a] Each *Cronobacter* sequence indicated was submitted to the Center for Genomic Epidemiology's (CGE) Plasmidfinder tool (https://cge.cbs.dtu.dk/services/PlasmidFinder/) for *in silico* detection and best-matching hits in each genome for each replicon sequence were given as output, using a % ID threshold value of 80%, and coverage of >96%, dashes represent no IncFIB origin found for that plasmid. Plasmids have been divided into homologous groups using the new plasmid nomenclature dependent on incompatibility groups using the PlasmidFinder function. Heavy metal plasmid isolates shared presence of IncFIB incompatibility group, while conjugation type-IV plasmids subtyped to a shared coverage of IncFII and a similar coverage of pESA2. All plasmid sequences of the mutli-drug resistance category shared presence of IncHI2 and IncHI2A replicons. Plasmid classes are made to prevent drastic length disparities between plasmid sequences.

RESULTS

Proposed Cronobacter plasmid classification scheme

Results of Plasmidfinder 2.1 analysis (**Table 2**) showed that only 12 of the 32 (35.5%) *Cronobacter* plasmids used in this study were contained in the plasmidfinder database (https://cge.cbs.dtu.dk/services/PlasmidFinder/). Of the 15 virulence plasmids, only pCTU1 was identified. However, all heavy metal, conjugation-Type IV and multidrug resistance plasmids were also identified as IncFIB plasmids, IncFII, IncHI2, and IncX3, respectively. It is hoped that the results of this Thesis will enrich the plasmidfinder 2.1 database. The proposal to classify the various *Cronobacter* plasmids into six classes is based on their specific genomic attributes and nucleotide compatibility to other closed plasmids or their ability to coexist within the same cell. The proposed six classes of plasmids are described in **Table 3**. Fifteen of the plasmids were identified as virulence plasmids and each species harbors a virulence plasmid. The decision to classify them as such was based on the criteria proposed by Franco *et al.*[2] Heavy-metal resistance plasmids were characterized by possessing multiple heavy metal resistance gene clusters (most notably various efflux pump gene clusters). Conjugation-Type IV plasmids contained T4SS gene clusters. Cryptic plasmids were characterized on the basis that they only carried mobilization genes such as *mobABCD* and a methyl-accepting chemotaxis protein gene. These plasmids also possess relaxase genes, such as *mobABCD* and several genes encoding for hypothetical proteins. *mobABCD* genes are thought to be involved in plasmid transfer.[10] Multi-drug resistance plasmids were characterized by harboring

multiple antimicrobial resistance genes and insertion elements, such as genes for β-lactam resistances and other resistance genes for carbapenems, quinolones, aminoglycosides, tetracyclines, phenicols, sulphonamides, trimethoprims, rifampicins, bleomycin, acriflavine, and the New Delhi metallo-β-lactamase 1 (*NDM-1*) gene as described by Shi et al.[77] Lastly, mobilization plasmids possessed many plasmid-borne genes involved in recombination such as transposons, insertion elements, prophages, and DNA methylases. WGS plasmid assemblies were configured using regex to make them compatible with the plasmidotyping python script developed here in this study. Three *C. sakazakii* plasmids were not added to the plasmidotyping database because at the time of development their genomic sizes were considered too large for it to be a plasmid (pCsaCS031a, pCsaCS09b, pCsaCS09c) and that NCBI had not annotated them. p505108-NDM was classified as a multi-drug resistance plasmid even though it shared a low genetic homology with all other plasmid members in the database. Take note that since then NCBI has broken the plasmids into smaller sized assemblies, but these changes were too late to incorporate these plasmids into the present plasmid database, so they too were excluded from the database. The below sections describe each plasmid class in greater detail.

Table 3. Proposed plasmid classification scheme of 32 Cronobacter plasmids across the seven Cronobacter species as developed for this study. [a]

Classification	*Cronobacter* species	Plasmid Name	RefSeq Record[b]	INSDC Record[c]	Size (Kb)	GC%	CDS	References (PMID)
Virulence	*C. sakazakii*	CSK29544_1p	NZ_CP011048	CP011048	93.9	57.02	61	28053670
		pCS2	NZ_CP012255	CP012255	117.8	57.23	103	27013041
		pCsaCS09a	NZ_CP027110	CP027110	131.2	56.85	116	Unpublished
		pGW2	NZ_CP028976	CP028976	135.3	56.81	117	30334728
		pESA3	NC_009780	CP000785	131.2	56.85	116	21245266
		pSP291-1	NC_020263	CP004092	118.1	57.23	106	23516209
		pH322_2	NZ_CP078112	CP078112	118.1	56.8	118	This study
		p505108-T6SS	NZ_KY978630	KY978630	139.6	56.4	125	28771073
	C. malonaticus	p1	NC_023024	CP006732	126.5	57.29	101	24435860
		pCMA1	NZ_CP013941	CP013941	126.5	57.29	104	27013041
	C. dublinensis	pCDU1	NZ_CP012267	CP012267	197.3	56.84	160	27013041
	C. turicensis	pCTU1	NC_013283	FN543094	138.3	56.05	109	21037008
	C. universalis	pCUNV1	NZ_CP012258	CP012258	129.8	57.03	114	27013041
	C. muytjensii	CmuyJZ38_2	NZ_CP017664	CP017664	145.8	58.1	113	32218777
	C. condimenti	pCCO1	NZ_CP012265	CP012265	151.5	54	125	27013041
Cryptic	*C. sakazakii*	CSK29544_2p	NZ_CP011049	CP011049	4.9	54.88	4	28053670
		pCSA2	NC_021293	KC663407	5.1	55.01	6	25332122
		pSP291-3	NC_020262	CP004094	4.4	54.03	6	23516209
Heavy Metal Resistance	*C. sakazakii*	CSK29544_3p	NZ_CP011050	CP011050	53.5	50.07	58	28053670
		pCS3	NZ_CP012256	CP012256	53.4	49.25	55	27013041
		pSP291-2	NC_020261	CP004093	52.1	49.16	55	23516209
		p2	NC_023025	CP006733	55.9	50.38	59	24435860
	C. malonaticus	pCMA2	NZ_CP013942	CP013942	52.6	50.25	55	27013041
	C. turicensis	pCTU3	NC_013285	FN543096	53.842	50.039	55	33374633
Conjugation type-VI	*C. sakazakii*	pESA2	NC_009779	CP000784	31.208	51.5829	36	28694793
	C. turicensis	pCTU2	NC_013284	FN543095	22.448	49.2204	27	21037008
Multi-drug resistance	*C. sakazakii*	pGW1	NZ_CP028975	CP028975	340.723	48.5802	358	30334728
		p505108-MDR	NZ_KY978628	KY978628	312.9	47.7	352	28771073
Mobilization	*C. sakazakii*	pCS1	NZ_CP012254	CP012254	110.1	50.71	122	27013041
		pH322_1	NZ_CP078111	CP078111	100.7	50.2	137	This study
	C. muytjensii	CmuyJZ38_1	NZ_CP017663	CP017663	109	49	142	32218777
Multi-drug resistance	*C. sakazakii*	p505108-NDM	NZ_KY978629	KY978629	53.8	49	40	28771073

All seven *Cronobacter* species are represented accordingly across the 32 plasmids. Also, in this table, *C. sakazakii* p505108-NDM plasmid which had relatively low homology (<20%) to all other *Cronobacter* plasmid sequences did

carry multiple insertion elements and *bla*$_{NDM-1}$, bla$_{SHV-12}$, and *ble*$_{MBL}$ antimicrobial resistance genes.

Virulence Plasmids

As shown in **Table 3** each *Cronobacter* species possesses a virulence plasmid which harbors common genes such as an IncFIB incompatibility class replication of origin gene or plasmid replication initiator family protein (*rep*), two iron acquisition gene clusters (*eitCBAD* and *iucABCD/iutA*) and various type 1 and type II toxin-antitoxin bicistronic gene pairs. Most plasmids contain unique regions encompassing genes for origin of replication initiation proteins that encode functions necessary for replication activation and copy number control. Franco *et al.,* showed that *rep* genes of the then two known closed *Cronobacter* species, together with several plasmid *rep* genes from *Cronobacter* assemblies, phylogenetically clustered together with the RepFIB *rep* genes of pILF82 harbored by pathogenic *E. coli*, pHCM2 of *Salmonella enterica* subspecies *enterica* serovar Typhi, and pMT1 harbored by *Yersinia pestis*. This tree is shown in **Figure 3A**.[2,11] The plasmid replication initiator family protein gene sequences of these *Cronobacter* plasmids grouped separately into two distinct clusters, signifying that they shared homology, but the homology followed species epithets. Clade 1 was comprised of the plasmid *rep* gene from *C. sakazakii* strains BAA-894 and Jor100, *C. malonaticus* strains E763 and LMG 23826, *C. dublinensis* subsp. *lausannensis* strain LMG 23824, *C. turicensis* strains z3032 and E681, and the two *Cronobacter* genomospecies group 1 now named *C. universalis* strains E680 and NCTC 9529. Strains residing in cluster 2 shared

homology with the plasmid *rep* genes from draft assemblies of *C. muytjensii* strains Jor174 and Jor171 and *C. dublinensis* subsp. *dublinensis* strain LMG 23823. Take note that *C. condimenti* had not yet been described when this analysis had been performed. To investigate further the phylogenetic relatedness of the origin of replication genes associated with the virulence plasmids investigated in the current study in comparison with that described by Franco *et al.*, we aligned the origin of replication gene of these plasmids in a similar fashion. Fifteen virulence plasmids, each of which contain the gene encoding for the plasmid rep protein, are represented in each of the seven *Cronobacter* species.[2] Similar to what Franco *et al*., had found, this alignment mirrored a similar phylogenetic species-specific distinction and is shown in **Figure 3B**.[2] In addition to that found by Franco *et al.,* three clades were determined to share sequence homology among the plasmid *rep* genes of the 15 *Cronobacter* virulence plasmids. Clade 1 contained plasmid *rep* genes of *C. malonaticus* pCMA1, *C. universalis* pCUNV1, *C. turicensis'* pCTU1, sharing sequence homology with that of *C. sakazakii* plasmid pESA3, respectively.[2]

A)

Escherichia coli chi7122 pAPEC-1
Salmonella typhimurium G8430 pU302L
Escherichia coli pVM01
Escherichia coli pIP1206
Escherichia coli S88 pECOS88
Salmonella enterica Kentucky CVM29188
Escherichia coli APEC O1 pAPEC-O1-ColBM
Escherichia coli Vir68 pVir68
Escherichia coli pAPEC-O103-ColBM
Escherichia coli pUTI89
Escherichia coli F11
Plasmid F
Escherichia coli plasmid R386
Escherichia coli DU1 pO86A1
Escherichia coli 55989p
Escherichia coli O157:H- 3072/96 pSFO157
Escherichia coli O157:H7 EDL933 pO157
Escherichia coli E110019
Salmonella enterica Choleraesuis SC-B67
Salmonella enterica Choleraesuis pOU7519
Salmonella enteritidis
Salmonella typhimurium LT2 pSLT
Escherichia coli pMAR7
Escherichia coli E2348/69
Escherichia coli O127:H6 E2348/69 pMAR2
Escherichia coli B171
Escherichia coli plasmid EAF
Escherichia coli APEC O1 pAPEC-O1-R
Yersinia pestis Nepal516 pMT
Yersinia pestis CO92 pMT1
Yersinia pestis KIM pMT-1
Yersinia pestis Microtus 91001 pMT1
Yersinia pestis Medievalis 91001
Yersinia pestis CA88-4125 pMT1
Yersinia pestis Antiqua pMT
Yersinia pestis Pestoides F
Salmonella enterica Typhi CT18 pHCM2
Escherichia coli pILF82
Cronobacter muytjensii Jor174 pCMU1
Cronobacter muytjensii Jor171 pCMU1
C. dub. dublinensis LMG23823 pCDD1
Cronobacter sakazakii BAA-894 pESA3
Cronobacter malonaticus E763 pCMA1
Cronobacter sakazakii Jor100 pCSA1
C. dub. lausannensis LMG23824 pCDLs1
Cronobacter malonaticus LMG23826 pCMA1
Cronobacter turicensis z3032 pCTU1
Cronobacter turicensis E681 pCTU1
Cronobacter genomospecies 1 E680 pCGS1
Cronobacter genomosp. 1 NCTC9529 pCGS1
S enterica Schwarzengrund pCVM19633 110
Klebsiella pneumoniae MGH 78578 pKPN3
Escherichia coli E22
Klebsiella pneumoniae CG43 pLVPK
Klebsiella pneumoniae NTUH-K2044 pK2044
Escherichia coli A2363 pAPEC-02-ColV

***repA* Clade 2**

***repA* Clade 1**

0.05

B)

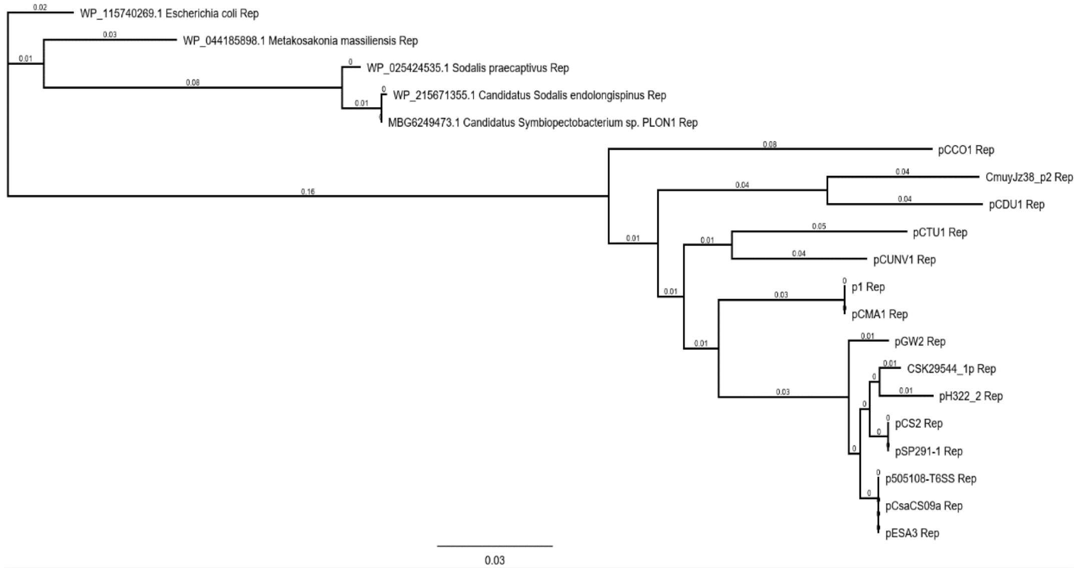


Figure 3 **A)** Franco *et al.,* represented *rep* in a phylogenetic cluster analysis. The neighbor-joining approach was used to infer the evolutionary history.[2] Next to the branches are the percentages of duplicate trees in which the related taxa clustered together in the bootstrap test (1,000 repetitions). The branch lengths are in the same units as the evolutionary distances used to estimate the phylogenetic tree, and the tree is drawn to scale. The first, second, and third codon positions, as well as noncoding sites, were included. Gaps and missing data were removed from all points in the data set (complete deletion option). The total number of positions in the final data set was 780. MEGA4 was used to perform phylogenetic analysis. **B)** Phylogenetic cluster analysis of virulence plasmid *rep* genes used in the current study. The evolutionary history was inferred by aligning nucleotide sequences for the 15 virulence plasmid *rep* genes along with neighboring enteric sequences used as outliers. The annotation of each *rep* gene is from NCBI. The evolutionary history was inferred by first aligning the *rep* nucleotide sequences using the MAFFT algorithm, then identifying phylogeny using the maximum

likelihood method.[95-96] The percentages of replicate trees in which the associated taxa clustered together (bootstrap test 1,000 replicates) are shown next to the branches (64). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the maximum composite likelihood method, and data shown are the number of base substitutions per site. Bar marker represents 0.03 nucleotide divergence. Tree was constructed using UGENE 38.1 software (available for download at: http://ugene.net/).[93]

Further analysis in determining the distribution and prevalence of the origin of replication gene suggest that 89% (488 of 550) of *C. sakazakii* isolates contain the origin of replication gene for virulence plasmids: p505108-T6SS, pCsaCS09a and pESA3. Due to their high degree of similarity, all three origin of replication genes from these three plasmids were discovered to be shared by *C. sakazakii* ST1, 4, 8, 64, and 83 isolates (see **Supplemental Table S4**). Ninety-three percent of *C. sakazakii* ST4 isolates shared the origin replication gene for all three virulence plasmids. In C. *dublinensis* isolates, only one virulence plasmid was discovered to contain its origin of replication gene: pCDU1. More precisely, 95% of C. *dublinensis* isolates possessed the *rep* gene from pCDU1. At 79 %, C. *malonaticus* isolates shared the origin of replication gene for p1 and pCMA1. Six of the nine *C. muytjensii* isolates shared homology for the *rep* gene from CmuyJZ38 p2, including *C. muytjensii* ST546 and ST294. This is a unique finding for the type species strain ATCC51329 does not contain any plasmids. Eighty nine percent of *C. turicensis* isolates shared the pCTU1origin of the replication gene. Among which are *C. turicensis* ST19, ST32, and ST569 isolates. All C. *universalis* ST54 and ST622 isolates shared the origin of replication gene for pCUNV1, the species' sole origin of replication gene. Similarly, the *C. condimenti* ST98 isolate possessed a plasmid *rep* gene that shared homology with the *rep* gene harbored on pCCO1.

Besides sharing significant plasmid *rep* gene sequence homology, these virulence plasmids also harbored species-specific virulence factors in addition to the “plasmid backbone” genes and such shared homology is shown in **Figure 4**.

For example, the pESA3-like virulence plasmids of *C. sakazakii* (pESA3) and *C. universalis* (pCUNV1) strains BAA-894 and NCTC 9529 harbored *cpa* or *Cronobacter* plasminogen activator gene. Whereas the virulence plasmids possessed by *C. malonaticus* (pCMA1, p1) and pCTU1 of *C. turicensis* have a 27 kb gene cluster encoding for a *Bordetella*-like filamentous hemagglutinin gene cluster. Five additional virulence plasmids from *C. sakazakii* (CSK29544_1p (93 kB), pCS2 (117 kB), pCsaCS09a (131 kB), pGW2 (135 kB), and pH322_2 (118 kB)) were added to the virulence plasmid category. Using an implementation of the ProgressiveMauve algorithm in Geneious suite 9.1, six *C. sakazakii* genomes revealed that there are seven core regions conserved among all virulence plasmids, which account for an average of 94.2% of each plasmid sequence (see **Figure 4**). The average plasmid length, GC% content, and number of proteins in this class is of 133 kB, 56.8%, and 112, respectively.

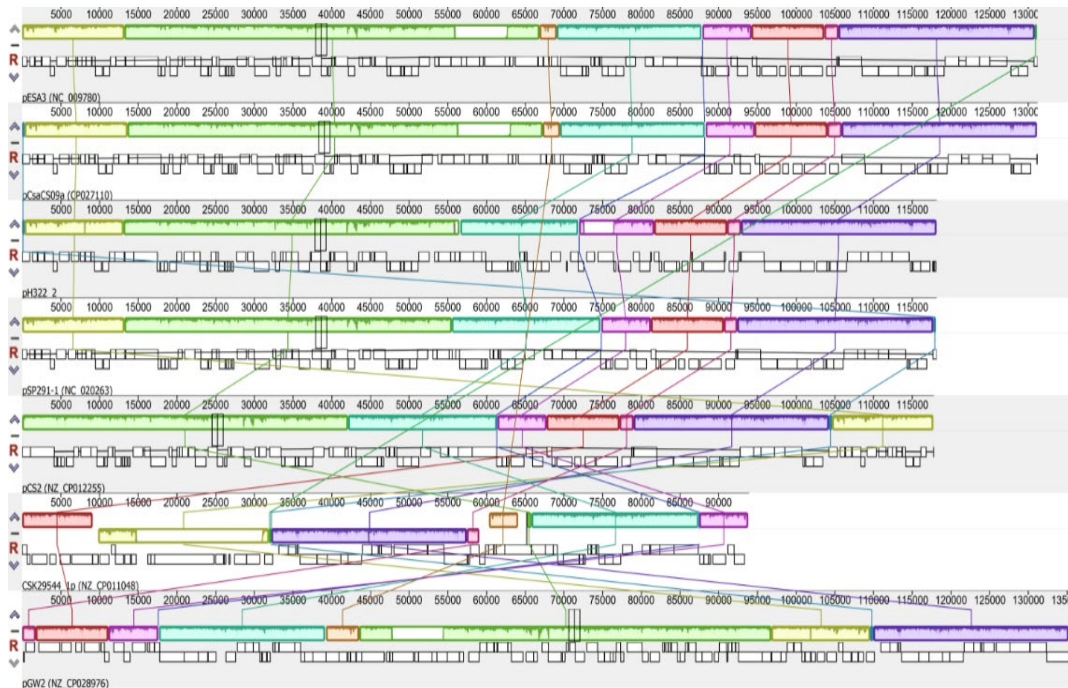


Figure 4. The Mauve progressive algorithm was used to align closed *C. sakazakii* virulence plasmid genome sequences to show both local and large-scale changes in phylogeny. Included in this alignment are *C. sakazakii* virulence class plasmid members (top to bottom): 1) pESA3; 2) pCsCs09a; 3) pH322_2; 4) pSP921-1; 5) pCS2; 6) CSK29544_1p; 7) and pGW2. Genomes are divided into horizontal panels with a single black horizontal center line, the name of the genome sequence, and a scale giving the sequence coordinates for that genome. Colored block outlines surround a portion of the genome sequence that corresponded to another genome and is presumed homologous and free of genomic rearrangement on the inside. Outside these blocks, there is no discernible homology among the input genomes. A similarity profile of the genomic sequence is displayed within each block. The average level of conservation in that area of the genome sequence correlates to the height of the similarity profile. Regions that are completely white were not aligned and are likely to include sequence fragments unique to a particular plasmid sequence. For functional annotations of each conserved region shown above,

please see **Supporting Table 1.** progressiveMauve Algorithmic Tool 9.1 was used to perform alignment.[91]

In addition to *eitCBAD* and *iucABCD iutA* gene clusters, pESA3 and pSP291-1 contain a ~17 kB region which house a type 6 secretion system (T6SS). pCTU1 and pCTU1-like plasmids, which were first discovered in *C. turicensis* z3032 and *C. malonaticus*, contain instead an ~27-kB region which encodes for *fhaB* (a *Bordetella* pertussis-like filamentous hemagglutinin gene cluster), as well as the transporter gene (*fhaC*) and the FHA locus (encoding for five putative adhesins).[2,36,37]

Alignment of all members of this class using BLAST+ on PROKSEE online bioinformatic tool is shown in **Figure 5**. p505108-MDR was removed from the virulence plasmid group and placed in the multi-drug resistance class after their alignment produced <20% homology with plasmid members of the virulence class. Interestingly, sequence alignment of pESA3, pCsaCs09a, p505108-T6SS, and pGW2 show homology for the type VI secretion system (T6SS) gene region.

ParB/RepB/Spo0J family partition protein
hypothetical protein
acetyltransferase
SidA/IucD/PvdA family monooxygenase
AAA family ATPase
omptin family outer membrane protease
TonB-dependent siderophore receptor
MFS transporter
iucC
YafY family transcriptional regulator
aerobactin synthase IucA
efflux RND transporter periplasmic adaptor subunit
MFS transporter
two-component sensor histidine kinase
type II toxin-antitoxin system HipA family toxin
efflux RND transporter permease subunit
treZ
NUDIX domain-containing protein
treY
ABC transporter six-transmembrane domain-containing protein
glgX
MFS transporter
LysR family transcriptional regulator
autotransporter domain-containing protein
TetR/AcrR family transcriptional regulator
metallophosphoesterase
hypothetical protein
cupin domain-containing protein
diguanylate phosphodiesterase
mgtA
MerR family transcriptional regulator
HAMP domain-containing protein
AAA family ATPase
AAA family ATPase
GNAT family N-acetyltransferase
methyl-accepting chemotaxis protein
winged helix-turn-helix transcriptional regulator
alpha/beta hydrolase
glutathione S-transferase family protein
PTS transporter subunit EIIC
guaD
2-dehydro-3-deoxy-6-phosphogalactonate aldolase
MBL fold metallo-hydrolase
GntR family transcriptional regulator
DUF4440 domain-containing protein
MCP four helix bundle domain-containing protein
sensor domain-containing diguanylate cyclase
hypothetical protein
hypothetical protein
SDR family oxidoreductase
hypothetical protein
helicase
potassium transporter TrkG
arsenic transporter
phospholipase D family protein
hypothetical protein
dihydrodipicolinate synthase family protein
tyrosine-type recombinase/integrase
PLP-dependent aminotransferase family protein
MFS transporter
tssC
iron ABC transporter permease
ABC transporter substrate-binding protein
OmpA family protein
RepB family plasmid replication initiator protein
tssK
tssH
thioredoxin family protein
vgrG
130 kbp
10 kbp
120 kbp
20 kbp
110 kbp
30 kbp
100 kbp
40 kbp
90 kbp
50 kbp
80 kbp
70 kbp
60 kbp
pESA3
pCsaCS09a
p505108-T6SS
pGW2
pCS2
pSP291-1
pH322_2
pCUNV1
CSK29544_1p
pCTU1
pCMA1
p1
pCDU1
CmuyJz38_p2
pCCO1

Figure 5. Sequence alignment of virulence plasmid class members produced on the PROKSEE Server from the Stothard Research Group that uses BLAST analysis to illustrate conserved and missing genomic sequences (available online: https://beta.proksee.ca/tools).[93] The annotation of each gene is from NCBI. The inner circle represents the sequence clockwise and the scale marks are selection markers detailing position. GenBank annotations of the reference pESA3 (CDS in blue arranged outside ring) were downloaded as a GFF file for analysis using the default configuration on the PROKSEE server. Across the circular genomes, selected genes or regions of interest are shown as follows: Franco *et al.*, adapted siderophore loci with Cronobactin gene, Iron ABC transporter genes, (T6SS), *parAB* genes, and the *cpa* gene. Missing regions identified by the BLAST analysis on the CGView server's PROKSEE software are shown as 'gaps' on each of the circular genomes.

Using the Plasmidotyper program we were able to obtain results for comparison of the plasmids in this group. Genes encoding for cytoplasmic protein

anchoring complexes of the T6SS, e.g., TssK are absent from many virulence plasmids possessed by *C. turicensis* and *C. malonaticus* strains. Using PROKSEE visual alignment tool, the locus of the siderophore genes was found across all plasmid backbones around nucleotide positions 117025….117909.

ABC iron transporters specifically involved in heme/iron transport across virulence factors was a common gene cluster across twelve of the virulence plasmid sequences. No copies of ABC iron transporters were found in any virulence plasmid in *C. sakazakii* ST226 or *C. dublinensis* ST162 strains. In contrast to all other species, *C. malonaticus* strains had the highest percentage of ABC iron transporters genes among all virulence plasmids. CmuyJZ38 p2 had the largest number of ABC iron transporter copies. The locus of the ABC iron transporter was identified using the PROKSEE visual alignment method across all plasmid backbones at nucleotide positions 74943…75701.

Except for CSK29544 1p and pGW2, all virulence plasmids contained a tyrosine-type recombinase or integrase. *C. malonaticus* ST60 and ST7 strains shared the highest number of for tyrosine-type recombinase genes as these were found in all *C. sakazakii* and *C. malonaticus* plasmids. Other virulence plasmids lacked a similar presence of tyrosine-type recombinase genes, occurring only once per strain. Using PROKSEE analysis (see **Figure 5**) a tyrosine-type recombinase locus was detected at nucleotide positions 80518…81492 by visual alignment across all plasmid backbones.

Despite their importance as a virulence factor, T6SS shared gene homology in only six of the fifteen plasmids. The C. *muytjensii* ST129 isolates

was the only *C. muytjensii* sequence type to possess a T6SS. *C. sakazakii* ST64 isolates shared the highest number of T6SS genes among all *Cronobacter* isolates. Interestingly, even though T6SS genes were found on pCDU1, *C. dublinensis* ST162 strains did not share homology with any T6SS genes from any plasmid. Also, *C. condimenti* did not contain T6SS genes from their respective species plasmid, pCCO1, but rather shared a T6SS gene cluster from *C. sakazakii*-derived plasmids pCS2, pCsaCs09, pESA3, and pGW2. For example at nucleotide positions 80518…81492, a T6SS gene cluster was identified for plasmids derived from *C. sakazakii*.

Similarly, plasmid-encoded toxin-antitoxin (*hipAB*) genes for *C. turicensis* isolates were found on plasmid pCTU1. The existence of a type-II toxin-antitoxin genes in a *C. dublinensis* pCDU1 plasmid was shared by the *C. condimenti* isolate. *C. malonaticus* ST7 strains shared the highest number of genes for type-II toxin-antitoxin systems across all isolates. More intriguingly, the type-II toxin-antitoxin bicistronic gene cluster is present and shared in *C. sakazakii* ST4 isolates with that of four *C. sakazakii* plasmids (pCsaCS09a, pESA3, pGW2, and pSP291-1).

Except for CSK29544 1p and pCsaCS09a plasmids, arsenic efflux systems were found in all other *C. sakazakii* virulence plasmids. Heavy metal efflux gene clusters, such as those for arsenic, copper, and silver were discovered on plasmids other than those involved in virulence, namely heavy metal resistance plasmids such as plasmid pCTU3. Heavy metal plasmids contain a complete arsenic locus that accounts for 100% of the arsenic operon (*arsA, arsB, arsC, arsD, arsR,*

permease*)*. Genes encoding bacterial exchangers, Acr3 and ArsB were identified at nucleotide position 47151…47564 on the plasmid backbone using PROKSEE alignment viewer. *C. muytjensii* ST546 isolates was the only ST for *C. muytjensii* that lacked presence of a plasmidborne arsenic efflux gene system.

Despite their importance as a virulence factor, T6SS shared gene homology in only six of the fifteen plasmids. The C. *muytjensii* ST129 isolates was the only *C. muytjensii* sequence type to possess a T6SS. *C. sakazakii* ST64 isolates shared the highest number of T6SS genes among all *Cronobacter* isolates. Interestingly, even though T6SS genes were found on pCDU1, *C. dublinensis* ST162 strains did not share homology with any T6SS genes from any plasmid. Also, *C. condimenti* did not contain the T6SS genes from their respective species plasmid, pCCO1, but rather shared a T6SS gene cluster from *C. sakazakii*-derived plasmids pCS2, pCsaCs09, pESA3, and pGW2. A T6SS gene cluster was identified at nucleotide positions 80518…81492, in plasmids derived from *C. sakazakii*.

Results showed that all *C. sakazakii* virulence plasmids except CSK29544 1p contained an omptin membrane protease (*cpa*) gene. Notably, the *cpa* gene was not consistently shown to share homology in plasmids of several *C. sakazakii* ST8 isolates, and a distinction for the presence of *cpa* for this ST was made. In addition, *C. sakazakii* ST8 cpa^{pos} isolates showed shared homology with the gene from species-specific *C. sakazakii* plasmids (pCS2, pCsaCS09a, pESA3, pGW2, pH322_2, pSP291-1). *C. sakazakii* isolates have the highest level of homologous presence for *cpa* compared to other organisms, in specific ST64 and ST143

strains sharing copies of *cpa* from pCS2, pCsaCS09a, pESA3, pGW2, pH322_2, and pSP291-1. When a ST8 *cpa*^{neg} strain (csak29544) is mapped to pESA3 plasmid, it does not show presence of a shared *cpa* gene between 5978 ...47564 bp on pESA3. This gap region, named "g1" (gap 1) could explain the absence of *cpa* present in some *C. sakazakii* ST8 isolates. (**see Supplemental Figure 3**).

Heavy Metal resistance plasmids

It is thought that heavy metal plasmids offer growth stability genetic elements to cells through enhancing host cell persistence during times of stressful growth conditions. Plasmid members of this class encode genes that impart resistance to heavy metals such as silver, copper zinc, cadmium, cobalt, and other metals by periplasmic detoxification through the actions of various efflux pumps. *C. sakazakii* plasmids included in this class are: (CSK2954 3p (53.5 kB), pCS3 (53.4 kB), pSP291-2 (52.1 kB), p2 (55.9 kB), pCMA2 (52.6 kB) and *C. turicensis* pCTU3 (53.84 kB). The heavy metal plasmid sequence has an average BLAST query coverage of 95% and a nucleotide identity of 99.3% with the reference IncFIB plasmid (**see Table 2)**. The average size, GC% content, and number of genes encoding proteins of heavy metal plasmids are 53.6 kB, 50.0%, and 56, respectively. Originally, pCS1 and pH322_1 were considered members of this class; however, after Progressive Mauve Alignment 9.1 (**Figure 6)** and PROKSEE BLAST-based analysis **(Figure 7A)**, these two *C. sakazakii* plasmids were removed. The final consensus of members for this class was decided to be 1) pCTU3; 2) pCMA2; 3) pCS3; 4) p2; 5) pSP291-2; 6) CSK29544_3p and their alignment using PROKSEE can be seen in **Figure 7B.**

Phylogenetically, the sequence homology of the plasmid replication initiator family protein (*rep*) or its origin of replication gene, of the heavy metal resistance plasmids possessed by six *Cronobacter* strains compared to that of other enteric strains is shown in **Figure 8**. pCTU3, the heavy metal prototypic plasmid possessed by *C. turicensis* strain z3032 shared sequence homology with the *rep* genes from the other *Cronobacter* heavy metal plasmids. Interestingly, the *rep* gene from *C. sakazakii* 29544_3p, *C. malonaticus* pCMA2, and *C. malonaticus* p2 also shared a high level of sequence homology. These results suggest that the *Cronobacter* plasmids pCS3, and pSP291_3 grouped in Clade 1 may have shared a common ancestry. A similar speculation may be offered for the *Cronobacter* plasmids: *C. sakazakii* 29544_3p, *C. malonaticus* pCMA2 and *C. malonaticus* p2 plasmids. Clades 2 and 3 represent *rep* genes from other phylogenetically-related heavy metal plasmids.

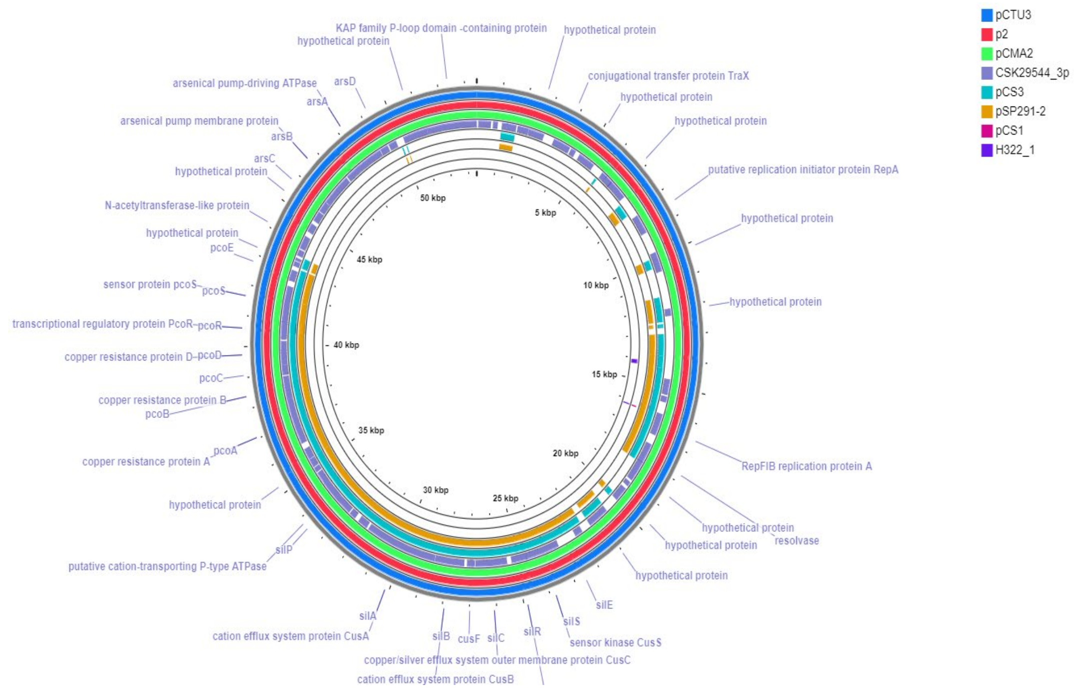
One hundred and sixty-two of the 550 *C. sakazakii* isolates examined shared homology for the origin of replication gene possessed by heavy metal plasmids. Furthermore, *C. sakazakii* ST4 isolates had the highest shared homology for the origin of replication gene among all the plasmids in the study. Similarly, 25% of total *C. dublinensis*, C. *malonaticus*, and C. *muytjensii* isolates shared the presence of a heavy metal plasmid origin of replication gene. Nine of the twenty *C. malonaticus* ST7 isolates possessed this gene. *C. malonaticus* ST7, 60, and 129 all shared homology for the origin of the replication gene of the heavy metal class plasmids. Forty seven percent of *C. turicensis* isolates shared the origin of replication gene for all heavy metal plasmids, however no C.

universalis or *C. condimenti* isolates possessed a heavy metal plasmid *rep* gene (see **Supplemental Table S3 and S4**). Three clades were determined to share sequence homology among the *rep* genes of the six *Cronobacter* heavy metal plasmids starting with Clade 1 which contained the *rep* genes of *C. turicensis* pCTU3, *C. sakazakii* pCS3 and pSP291-2. The addition of other enterics added to the analysis, which included plasmids from *E. coli, Salmonella enterica*, formed Clade 2, with no shared sequence homology with any *Cronobacter* plasmid *rep* gene. Clade 3 contained the *rep* gene from *C. sakazakii* CSK29544_3p, *C. malonaticus* pCMA2 and p2, respectively (see **Figure 8**).[2]

Figure 6. The Mauve progressive algorithm was used to align orthologous and xenologous regions among *C. sakazakii* heavy metal plasmid sequences to show both local and large-scale changes. Included in this alignment are heavy metal class plasmid members (read top to bottom): 1) pCS1; 2) p_H322_1; 3) pCTU3; 4) pCMA2; 5) pCS3; 6) p2; 7) pSP291-2; 8) CSK29544_3p. This analysis determined that pCS1 and pH322_1 share no conserved domains with the other members in this plasmid class. Genomes are separated into horizontal panels by a single black horizontal center line, the name of the genome sequence, and a scale that displays the genome's sequence coordinates. On the inside, colored block outlines surround a region of the genome sequence that matched another genome and is deemed to be homologous and free of genomic rearrangement. There is no visible commonality among the input genomes outside of these blocks. Within each block, a similarity profile of the genomic sequence is displayed. The height

of the similarity profile is proportional to the average level of conservation in that region of the genome sequence. Completely white areas were not aligned, and thus are likely to include sequence fragments exclusive to a genome. progressiveMauve Algorithmic Tool 9.1 was used to perform alignment.[91]

A)



B)

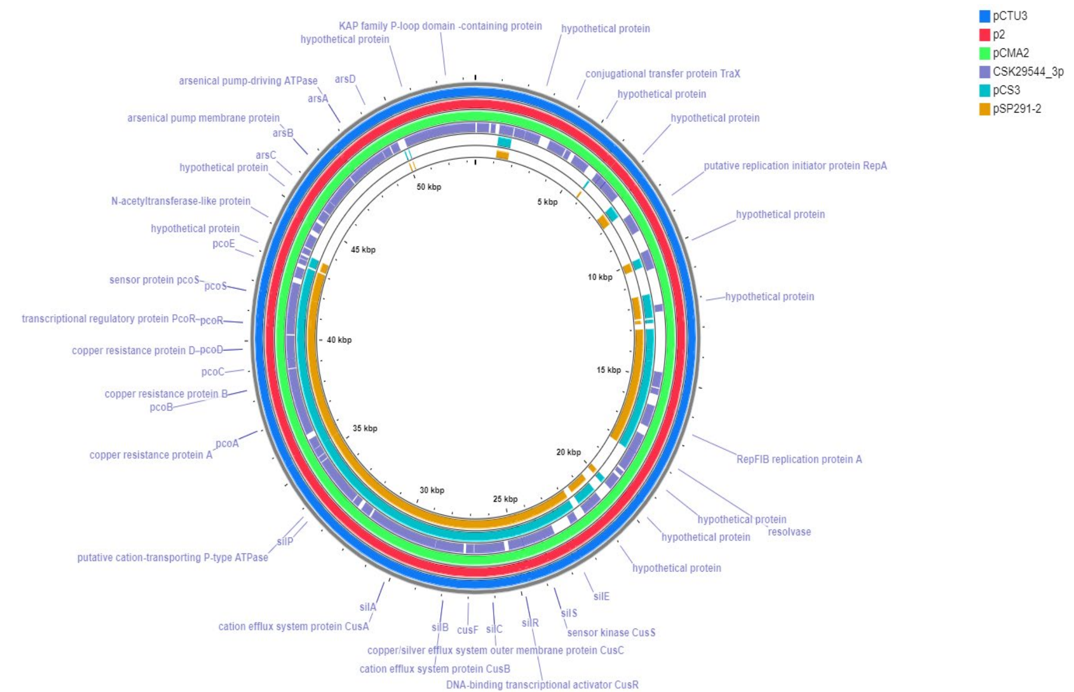


Figure 7. **A)** The Stothard Research Group's PROKSEE Server, which uses BLAST analysis to show conserved and missing plasmid sequences (available online: https://beta.proksee.ca/tools)[93], created a sequence alignment of heavy metal plasmid class members. The annotation of each gene is from NCBI. **B)** Removal of pCS1 and pH322_1 from the heavy metal class. After PROKSEE analysis of heavy metal plasmids class members, plasmids pCS1 and pH322_1 have, if any, have no significant genomic homology to other members of the heavy metal plasmids class. Thus, these plasmids sequences were removed, and the remaining heavy metal plasmid class included: 1) pCTU3; 2) p2; 3) pCMA2; 4) CSK29544_3p; 5) pCS3; 6) pSP291-2. The sequence clockwise is represented by the inner circle, while the scale marks, which include position and selection markers, are shown by the outside circle.

CSK29544_3p Rep
pCS3 Rep
pSP291-2 Rep
p2 Rep
pCMA2 Rep
pCTU3 Rep
0.3
WP_204231772.1_Leclercia_adecarboxylata_Rep
WP_114388507.1 Leclercia_sp._W17_Rep
WP_058663139.1_Enterobacteriaceae_MULTISPECIES_Rep
WP_214599088.1_Enterobacter_mori_Rep
ECO7189163.1_Salmonella_enterica_Rep
EFJ7594911.1_Escherichia_coli_Rep
WP_048029504.1_Leclercia_Rep
0.04

Figure 8. The origin of the replication gene (*rep*) of heavy metal plasmids members was studied using phylogenetic cluster analysis. Six nucleotide sequences for the *rep* genes from each member of the heavy metal plasmid class, as well as nearby phylogenetically related sequences used as outliers, were aligned to determine the evolutionary history. The evolutionary history was deduced by first utilizing the MAFFT algorithm to align the

nucleotide sequences, then using the maximum likelihood technique to determine phylogeny.[95-96] Next to the branches are the percentages of replicate trees where the relevant taxa clustered together (bootstrap test 1,000 repetitions). The phylogenetic tree is displayed to scale, with branch lengths in the same units as the evolutionary distances used to build the tree. The statistics given are the number of base substitutions per site, and the evolutionary distances were calculated using the maximum likelihood composite approach. The bar indicates a nucleotide difference of 0.04 nucleotide divergence. NCBI was used to annotate each *rep* gene. UGENE 38.1 software (available for download at http://ugene.net/) was used to create the tree. [93]

Heavy metal genes were identified and classified as structural genes like *silE, silC, silF, silB, silA*, ORF105 (25395...26780) or transcriptional regulators like *silR* and *silS*, which were found in the *sil* operons at nucleotide positions 23057...24532 and represented a shared plasmid backbone similar to that found for the virulence plasmids. The presence of the *sil* operon in the same nucleotide location across all plasmids belonging to this group suggests the presence of a conserved gene cluster. The *pco* (copper resistance) and *sil* (silver resistance) operons were found concurrently in all screened isolates, and the plasmidotyping tool showed that they were distributed among a diversity of species-specific *Cronobacter* strains. *C. muytjensii* and *C. condimenti* lacked the presence of the *pco* locus, followed by *C. universalis* which showed presence of the *pco* locus that was shared with *C sakazakii* pCS3 locus in ST622 isolates. Despite having a high levels of shared gene homology for the *sil/pco* locus in *C. sakazakii* isolates, ST1 isolates seemed not to possess many genes of this plasmid. Only *pcoD* and *pcoA* genes were found *in C. sakazakii* ST1 isolates, and shared comparable genes

for the *pco* locus found in *C. turicensis* and *C. universalis* isolates. While *pco* loci from *C. turicensis* pCTU3 and *C. sakazakii* pSP291-2 plasmids was found in *C. malonaticus* ST60 and ST129 isolates. *C. sakazakii ST83* were the only sequence type within the species to share the *pco* locus from any heavy metal resistance plasmid.

The *cusCBA* efflux gene cluster, which confers tolerance to zinc, cadmium, cobalt, and other metals, is encoded within the *sil* operon's conserved region (23057...31725) on the plasmid backbone of all the members of the heavy metal resistance class. These genes were interchangeably considered to be a central and shared gene cluster associated with the heavy metal resistance plasmids. Further examination of the heavy metal resistance plasmid members' core backbone revealed the existence a toxin-antitoxin (TA) system. MazF toxin and its antitoxin MazE were identified in all members of the heavy metal class at nucleotide positions 18140...18370. The *mazEF* operon was detected in the lone *C. condimenti* isolate and not in any other plasmids in the heavy metal plasmid class, including non-species-specific plasmids.[93,93] Neither *C. sakazakii* nor *C. malonaticus* isolates possessed this toxin-antitoxin. All *C. turicensis* isolates contained one copy of the *mazEF* operon, which seemed to share homology with a *C. malonaticus* pCMA2 plasmid rather than a *C. turicensis* plasmid. In heavy metal resistance plasmids, type II toxin-antitoxin operons are presented as either *RelE/ParE* or *vapBC*. *C. turicensis* and *C. muytjensii* ST94 isolates possessed the *RelE/ParE* family toxin-antitoxin system. A shared bicistronic TA operon was identified as a parE-relE/parD-relB type II gene clusters, where the antitoxin gene

preceded the toxin gene and was located between base position 985190...985426. Type II toxin-antitoxin *vapBC* systems were found in all *C. turicensis*, *C. muytjensii* ST546, *C. universalis* isolates. In comparison to other species, *C. turicensis* isolates shared the highest number of type I and type II toxin-antitoxin gene clusters. Except for CSK29544_3p, all heavy metal plasmids contained the *hok/sok* locus type I toxin/antitoxin operon. *C. sakazakii* ST8, *C. dublinensis* ST106, and all other strains except *C. universalis* ST54 and *C. malonaticus* ST7 possessed the *hok* type I TA operon that seemed to be derived from *C. malonaticus* plasmid pCMA2. Isolates possessing the *hok* type I TA operon derived from *C. malonaticus* pCMA2 included *C. sakazakii* ST8, *C. dublinensis* ST106, and all other strains except for *C. universalis* ST54 and *C. malonaticus* ST7. Multiple copies of the *hok/sok* locus type I toxin/antitoxin operon were also identified in *C. turicensis* ST19 strains.

Conjugal transfer *traX* genes were present in all plasmids from the heavy metal resistance class except CSK29544_3p.[95] The only copy of the gene found for its genus in *C. muytjentsii* ST294 was shared with the *traX* gene of *C. malonaticus* plasmid p2. Similarly, the conjugal transfer *traX* gene was also found in *C. sakazakii* ST83 strains. Most notable was that *C. universalis* and *C. condimenti* isolates lacked any conjugal transfer genes that were associated with heavy metal plasmids.

Multi-drug resistance plasmids

Cronobacter isolates are typically susceptible to the most widely used antimicrobial agents used in clinical practice, but resistance to one or more

antimicrobials including cephalothin, streptomycin, gentamicin, and tetracycline has emerged in a few *Cronobacter* isolates. Plasmids of the multi-drug resistance class have acquired several accessory gene modules encoding many resistance genes, especially those encoding resistance to carbapenem, aminoglycosides, tetracycline, and phenicol, as well as sulphonamide/trimethoprim resistance. *C. sakazakii* p505108-MDR (312.9kB) and pGW1 (340.7kB) compose the multidrug resistance plasmid class. Both the pGW1 and p505108-MDR backbones shared the key IncHI2A origin of replication gene markers, including *repHI2A* and *repHI2B* for replication initiation, and had 95 percent BLAST query coverage and 100 percent nucleotide identity to the reference IncHI2 plasmid (**see Table 2)**. The plasmids in this group have an average size (kB), GC % content, and number of proteins of 326.1, 48.1%, and 355, respectively.[77]

Six percent of both C. *sakazakii* (30 of 550) and *C. malonaticus* (3 of 48) isolates possessed the origin of replication gene for p505108-MDR, compared to roughly 1% (5 of 550) of C. *sakazakii* and 2% (1 of 48) of *C. malonaticus* isolates to that of pGW1's. No other *Cronobacter* species was found to contain the origin of replication gene for any of the two plasmids that comprise the multi-drug resistance class. These STs contained the following *C. sakazakii* isolates that were discovered to contain the *rep* gene for p505108-MDR: ST1, ST4, ST8, ST64, and ST83. Fourteen *C. sakazakii* ST4 isolates shared gene homology with the *rep* gene of p505108-MDR, representing 50% (14 of 28) of *Cronobacter* isolates sharing gene homology with the origin of replication gene for p505108-MDR. Three C. *sakazakii* ST1, ST64, and ST83 isolates, as well as two C. *malonaticus*

ST 7 isolates, were considered to share homology with the origin of replication gene from pGW1. Phylogenetic analysis identified two clades that shared sequence homology among the *rep* genes of the two *Cronobacter* multidrug resistance plasmid class. Of interest, both members of the class were in Clade 2 sharing high homology to neighboring enteric sequence's origin of replication gene (see **Figure 9**).

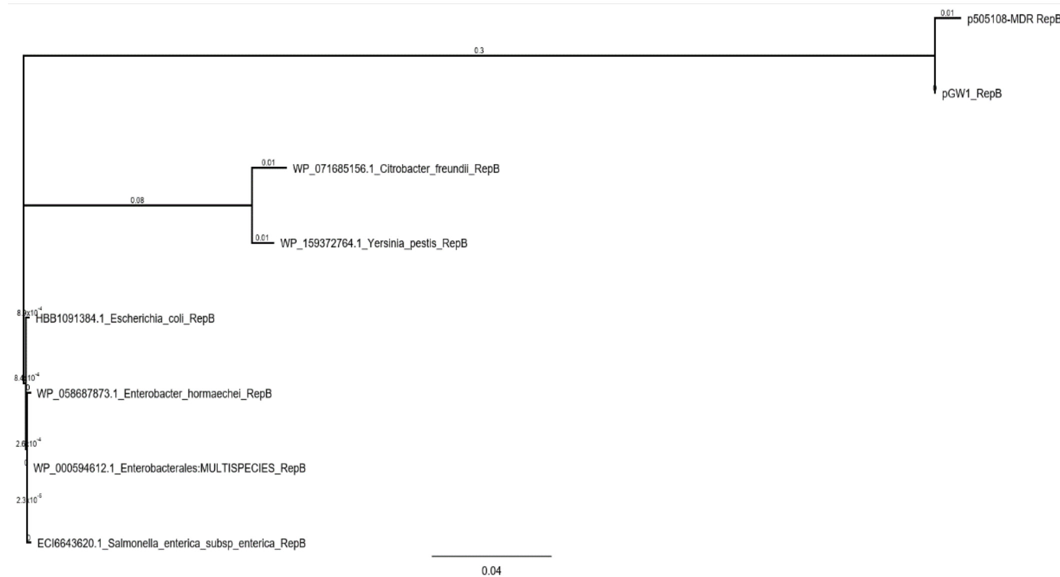


Figure 9. Phylogenetic cluster analysis of the origin of replication genes (*rep*) of the multidrug resistance plasmids. The annotation of each gene is from NCBI. The evolutionary history of the heavy metal plasmid class was estimated by aligning the two *rep* gene sequences with nearby phylogenetically related sequences used as outliers. The evolutionary history was determined by first aligning the rep sequences using the MAFFT algorithm and then using the maximum likelihood technique to determine phylogeny. Next to the branches are the percentages of replicate trees in which the related taxa clustered together (bootstrap test 1,000 repetitions).[95-96] The tree is drawn to scale, with branch lengths equal to the evolutionary distances used to determine the

phylogenetic tree's branch lengths. The evolutionary distances were calculated using the maximum composite likelihood approach, and the statistics displayed represent the number of substitutions per site. The bar indicates a nucleotide difference of 0.04 nucleotides. The tree was created with the UGENE 38.1 software (which may be downloaded at: http://ugene.net/). [93]

On the plasmid backbone of p505108-MDR, the tellurite resistance operon (*terZABCDE*) is located between nucleotide positions 62591...82538, and 216882...226136. When compared to the operon present in p505108-MDR (19947 bp), the *ter* operon identified in pGW1 (9254 bp) is truncated by approximately 10.6 kB. More intriguingly, no *Cronobacter* isolate had all the *ter* genes present in the *ter* locus of p505108-MDR. Indeed, only the *C. sakazakii* ST1, ST83, and *C. malonaticus* ST7 isolates possessed *terC* and structural genes (*terC, terD, terE, terF*). Most prevalent were some *C. sakazakii* ST83 isolates which exhibited shared homology for the complete *ter* locus of pGW1. Arsenic resistance is shared by both plasmids in this class, between nucleotide positions 156203...159087 kB. p505108 (*arsB, arsC, arsH, arsR*) and pGW1 (*arsB, arsC, arsH, arsR*) show different distributions of *ars* genes in this class.[95] The *ars* gene module from multidrug plasmids, on the other hand, was found in *Cronobacter* isolates with a high degree of specificity; for instance, the *ars* gene module was identified only in specific isolates: *C. dublinensis* ST79, *C. muytjensii* ST546, *C. universalis* ST622, and *C. condimenti* ST98. Only in *C. dublinensis* ST106 isolates was the *ars* gene module from both multidrug resistance plasmids identified **(Table 4).**

The MDR-1 region of p505108-MDR was conserved in pGW1 (124249…143642), containing antibiotic resistance cassettes for trimethoprim (*dfr*), aminoglycoside (*aph3*, *strAB*), and nickel/cobalt (*rcn*). Antibiotic resistance cassettes of the MDR-2 module of p505108-MDR consisting of β-lactam ($bla_{SHV\text{-}12}$, $bla_{DHA\text{-}1}$), aminoglycoside (*aacC3*, *aacA27*), quinolone (*qnr*), tetracycline (*tetA*), phenicol (*catA2*), sulphoamide (*sul1*), rifampicin (*arr7*), and mercuric (*mer*) were equally conserved between the two members of this class. The MDR-2 region (142132…265807) was absent in all *C. sakazakii* ST1 and *C. universalis* ST54 isolates, with only a transient sharing of the β-lactam and sulphonamide resistance genes from p505108-MDR (*blaSHV-12, sul1*) being identified **(Table 4).** Interestingly, apart from β-lactam resistance, *C. sakazakii* ST64 isolates possessed an entire MDR-2 region.

Table 4. ***Drug resistance genes shared between multi-drug resistance plasmids, pGW1 and p505108MDR.***

Plasmids[a]	Resistance Marker[b]	Resistance phenotype	Nucleotide position
The MDR-1 region	The *ter* locus	Tellurium resistance	62591..82538
	The *ars* locus	Arsenic resistance	156203..159087
	dfrA18	Trimethoprim resistance	124249..124818
	strAB	Aminoglycoside resistance	126592.. 128231
	The *rcn* locus	Nickel/cobalt resistance	142132..143642
The MDR-2 region	*bla*SHV-12	β-lactam resistance	239239..240099
	*bla*DHA-1	β-lactam resistance	247503..248642
	aacC3	Aminoglycoside resistance	265953..266762
	aacA27	Aminoglycoside resistance	268847..269428
	aacA4cr	Quinolone and aminoglycoside resistance	235221..235775
	qnrB4	Quinolone resistance	252763..253410
	tetA(D)	Tetracycline resistance	224408..225592
	catA2	Phenicol resistance	220639..221280
	sul1	Sulphonamide resistance	245087..245926
	sul1	Sulphonamide resistance	262128..262967

arr7	Rifampicin resistance	265412..265807
The *mer* locus	Mercuric resistance	102624..106878
aphA1a	Aminoglycoside resistance	164776..165591

[a] *C. sakazakii* plasmids pGW1 and p505108-MDR.

[b] Genotype markers for resistance phenotype.

These two plasmids harbored similar multi-drug resistance profiles against both AmpC β-lactamase and cephalosporin resistance (*bla*$_{SHV-12}$ and *bla*$_{DHA-1}$). Other drug-resistance antimicrobial profiles discovered in both plasmids include — azithromycin resistance (*mph(A)*), sulphonamide resistance (*sul1*), aminoglycoside resistance (*aadA2, aac(3)-Ild*), Trimethoprim resistance *dfrA12.* Interestingly both plasmids include *aac(6')-Ib-cr*, a mutant aminoglycoside acetyltransferase capable of reducing ciprofloxacin operation. Alignment of both pGW1 and p505108-MDR using PROKSEE software can be seen in **Figure 10**.

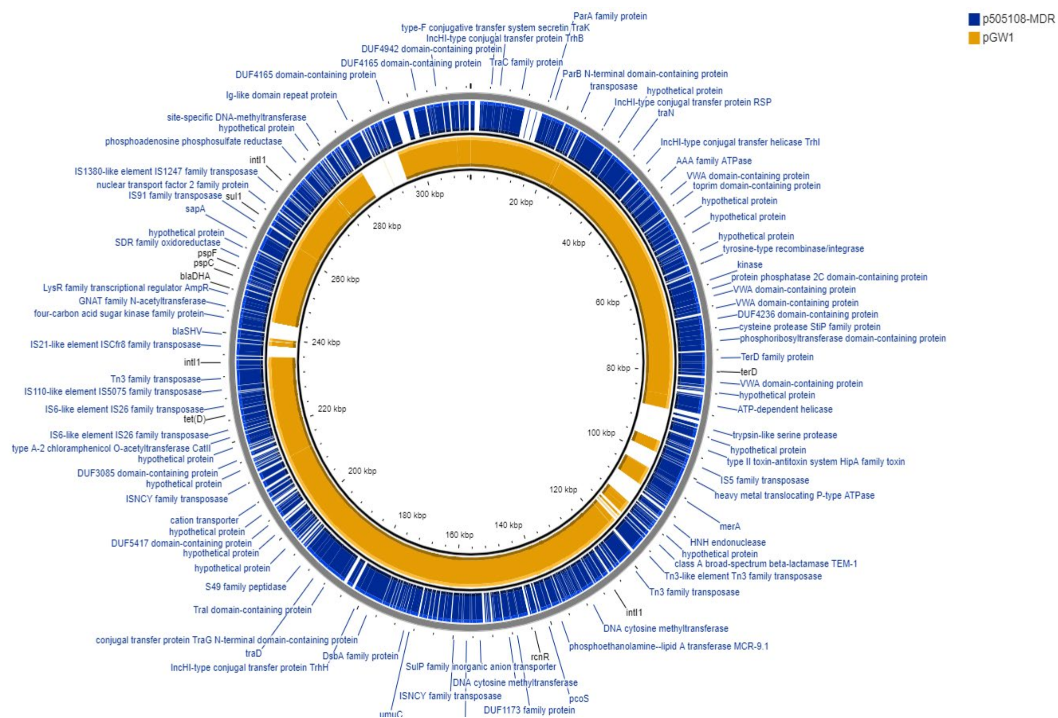


Figure 10. Sequence alignment of multidrug resistance plasmid class members produced on the PROKSEE Server from the Stothard Research Group that uses BLAST analysis to illustrate conserved and missing genomic sequences (available online: https://beta.proksee.ca/tools).[93] The annotation of each gene is from NCBI. The GenBank annotations for the reference p505108-MDR (CDS in blue arranged outside ring) were downloaded as a GFF file and analyzed on the PROKSEE server using the default settings. The sequence clockwise is represented by the inner circle, while the scale marks, which include position and selection markers, are shown by the outside circle.

Conjugation-Type IV plasmids

Conjugative plasmids play a significant role in horizontal gene transfer (HGT), which helps bacteria adapt faster to their everchanging environments by transmitting ecologically significant features among strains and species. Because these plasmids contain genes that govern their own replication and transmission,

they must be subject to natural selection in their own right, with fitness goals that aren't always matched with those of their bacterial hosts. Conjugative plasmids are a collection of (mainly) circularized DNA molecules that are extrachromosomal entities and along with prophage make up an organism's mobilome. Plasmids are made up of a backbone of essential genes that control basic plasmid processes and a set of non-essential accessory genes, such as the *virB* transportal complex which is known to mediate the transfer of DNA and proteins into plant cells. Members of this plasmid type possess bacterial *virB/*D transport complex, cag I and II loci, and lastly T4SS genes and examples of such plasmids are *C. sakazakii* pESA2 (31.2 kB) and *C. turicensis* pCTU2 (22.5 kB). BLAST query coverage for pESA2 was >99 %, and nucleotide identity to the IncFII plasmid was 96.71 %. pCTU2 has >97.8% BLAST query coverage against pESA2 and 100% nucleotide coverage against the IncFII plasmid (see **Table 2**). The average size, GC% content, and number of proteins for members in this class is 28.6 Kb, 50.4%, and 31.

Of the conjugation plasmid members, the origin of replication gene of C. *turicensis* pCTU2 was found to be present in three C. *turicensis* (16%) and 14 *C. sakazakii* (3%) isolates. On the other hand, the origin of replication gene of pESA2, was not identified in any C. *turicensis* isolates. More specifically, the origin of replication gene found on pCTU2 was found in C. *turicensis* ST19 isolates. (see **Supplemental Table S3** and **S4**). Of interest, a phylogenetic tree was created using the origin of replication gene of the two plasmid members of the conjugative class and compared to the origin of replication gene from

plasmids of other phylogenetically related enteric species and is shown in **Figure 11**. Two clusters formed, of which *C. turicensis*' pCTU2's *rep* gene was more phylogenetically related to the *rep* gene of *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* UCICRE 14 plasmid than the other *rep* gene. Clade 2 held pESA2's *rep* gene among four other enteric sequence's plasmids (see **Figure 11**).

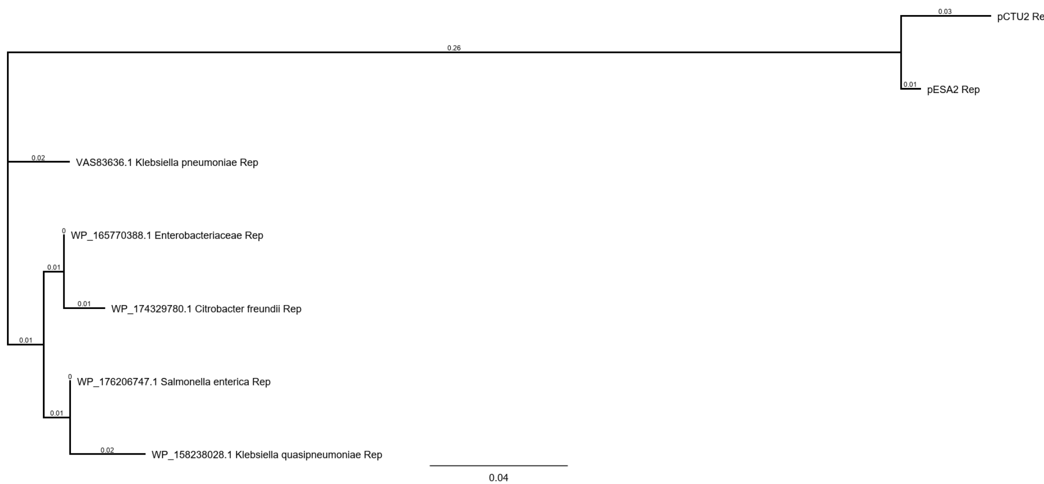


Figure 11. Phylogenetic cluster analysis of members of the origin of replication family (*rep*) gene of conjugation type-IV plasmids. The evolutionary history of the conjugative-type IV class was estimated by aligning *rep* gene sequences from each member of the class with nearby phylogenetically related *rep* sequences used as outliers. The annotation of each gene is from NCBI. The evolutionary history was determined by first aligning the nucleotide sequences using the MAFFT algorithm and then using the maximum likelihood technique to determine phylogeny. Next to the branches are the percentages of replicate trees in which the related taxa clustered together (bootstrap test 1,000 repetitions).[95-96] The tree is drawn to scale, with branch lengths equal to the evolutionary distances used to determine the phylogenetic tree's branch lengths. The evolutionary

distances were calculated using the maximum composite likelihood approach, and the statistics displayed represent the number of base substitutions per site. The bar indicates a nucleotide difference of 0.04 nucleotides. The tree was created with the UGENE 38.1 software (which may be downloaded at: http://ugene.net/).[93]

Multiple copies of T4SS protein encoding genes are found on both plasmids between 24639…28358 bp across the plasmid backbone (see **Figure 12**). *C. turicensis* isolates possessed the highest levels of T4SS genes. Notably, *C. turicensis* ST19 isolates uniquely possessed all T4SS genes from both plasmids. *C. sakazakii*, *C. dublinensis* and *C. muytjensii* isolates shared homologies with T4SS and T6SS from pCsaCs09a and pESA3. The presence of 12 *cag* PI genes from both *cag* I and *cag* II loci, were also found within the *virB/D* complex (*virB4, virB7, virB8, virB9, virB10, virB11*, and *virD4*) that were located on both plasmids. The *virB/D* complex was shared at a low level in all *Cronobacter* strains with minor conserved gene homology, and no one *Cronobacter* isolate possessed the full *virB/D* complex after visual alignment. Lastly, although pGW1 was first thought to be a member of the class of plasmid, it was later removed as the alignment using BLASTn shared little to any significant sequence similarity with the other two members of the Conjugative-T4SS class members, pESA2 and pCTU2, (see **Supplemental Figure 1).**

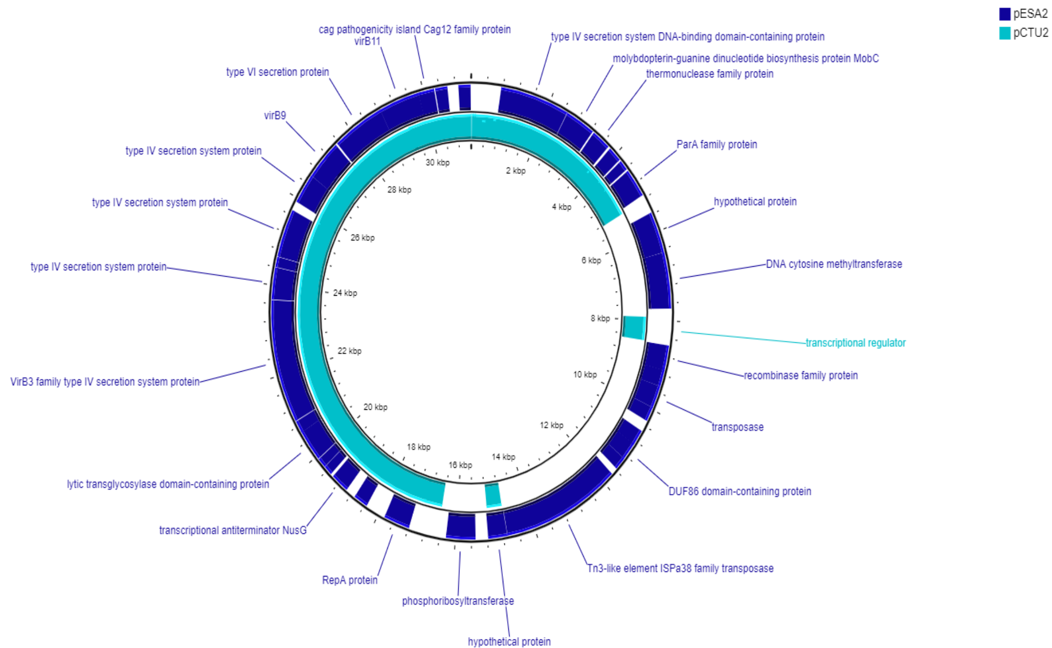


Figure 12. The Stothard Research Group's PROKSEE Server, which uses BLAST analysis to show conserved and missing genomic sequences (available online: https://beta.proksee.ca/tools), created a sequence alignment of conjugative-type IV plasmid class members.[93] The annotation of each gene is from NCBI. The GenBank annotations for the reference pESA2 (CDS in blue placed outside ring) were downloaded as a GFF file and analyzed using the default parameters on the PROKSEE service. The inside circle represents the clockwise sequence, while the outside circle represents the scale marks, which comprise location and selection markers.

Mobilization plasmids

The plasmids included in the mobilization plasmid class revealed a mosaic structure consisting of mobile genetic elements originating from bacteriophage, other plasmids and transposable elements. The PROKSEE image reveals a shared plasmid backbone composed of a cohort of gene complexes related to DNA methylation, Type I restriction modification, insertion elements, and phage-like

integrated genes be harbored by members of the mobilization plasmid class (see **Figure 13)**. The average size, GC% content, and number of genes encoding proteins of mobilization plasmids is 106.6 kb, 49.97%, and 133 genes, respectively. The mobilization plasmid class includes *C. muytjensii*'s CMuyJz38_p1, *C. sakazakii*'s pCS1, and pH322_1. No plasmid sequence from *C. turicensis*, *C. universalis*, and *C. condimenti* showed the presence of a DNA methylase gene. More interestingly, only one sequence type (ST) from *C. sakazakii* (ST83) and *C. malonaticus* (ST649) and *C. muytjensii* (ST129) share the presence of DNA methylase. However, in diverse sequence types of *C. sakazakii* isolates, BLAST coverage for DNA methylase was >95.6 percent. Interestingly, DNA methylase was present in *cpa*-^{pos} *C. sakazakii* ST8 isolates, but not in *cpa*-^{neg} *C. sakazakii* ST8 isolates. These plasmids contain exonuclease, methyltransferase, and methylase genes (functional genes), as well as exonuclease, methyltransferase, and methylase, and (structural) phage-linked genes. Both plasmids encode a phage-like integrase that has been shown to mediate site-specific integration into the 5' end of the *E. coli* chromosome's *prfC* gene. Additionally, the chromosomal association of these elements makes detection difficult, it is possible that such elements are representatives of a much larger and important group of mobile genetic elements, potentially acting as vectors of antibiotic resistances and other genes related to persistence. Two types of methylation classes were identified. Type I restriction modification systems and DNA methylase, were found in *C. sakazakii* pH322_1 and C. *muytjensii* CmuyJz38_p1.

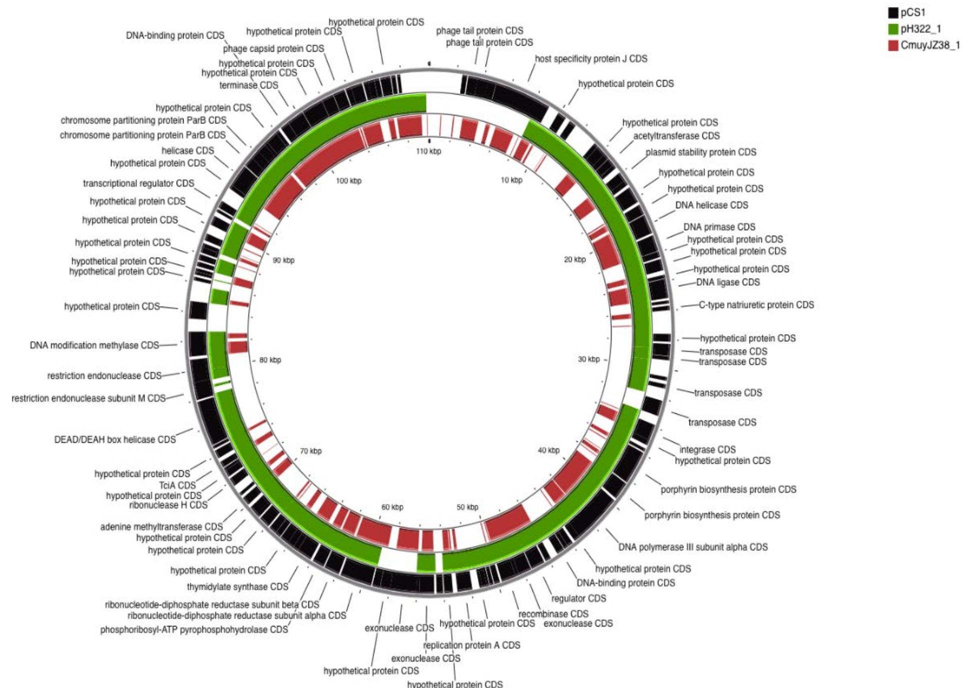


Figure 13. Alignment of members of the mobilization plasmid class produced on the PROKSEE Server by the Stothard Research Group using BLAST analysis to highlight conserved and missing genomic sequences (available online at https://beta.proksee.ca/tools).[93] The annotation of each gene is from NCBI. The inner circle denotes the clockwise sequence, while the scale marks include position and selection indicators. The GenBank annotations for the reference pCS1 (CDS in black, placed outside the ring) were downloaded as a GFF file and analyzed using the PROKSEE server's default setup. The circular genomes contain the following genes or regions of interest: DNA methylation, the Type I restriction modification system, and phage-like integration genes. Missing sections detected by the PROKSEE server's BLAST analysis are depicted as 'gaps' on each of the circular genomes.

Additionally, the origin of replication gene of plasmid CmuyJZ38_p1 was not shared by any other *Cronobacter* species. *Cronobacter* plasmids pCS1 and

pH322_1 shared their origin of replication gene with 140 of the 550 *C. sakazakii* (25%) and 14 C. *malonaticus* (6%) isolates, respectively. Among the *C. sakazakii* isolates, the origin of replication origin gene for pCS1 was identified in 63 of 140 ST4 *C. sakazakii* strains. Finally, it was demonstrated that both C. *malonaticus* ST7 and ST129 strains also possess the origin of replication genes from pCS1 and pH322_1 (see **Supplemental Tables S3** and **S4**). Phylogenetic analysis of the origin of replication gene across the members of the mobilization class produced two clades with a clear outlier from the *rep* gene of CmuyJZ38_p1. The latter clade contained the *rep* gene of *Cronobacter* mobilization plasmid class members, pH322_1 and pCS1 that also shared significant homology with that of a common *Enterobacteriaceae* and an *Enterobacter hormaechei* plasmid (see **Figure 14**). These data support the prevalence findings.

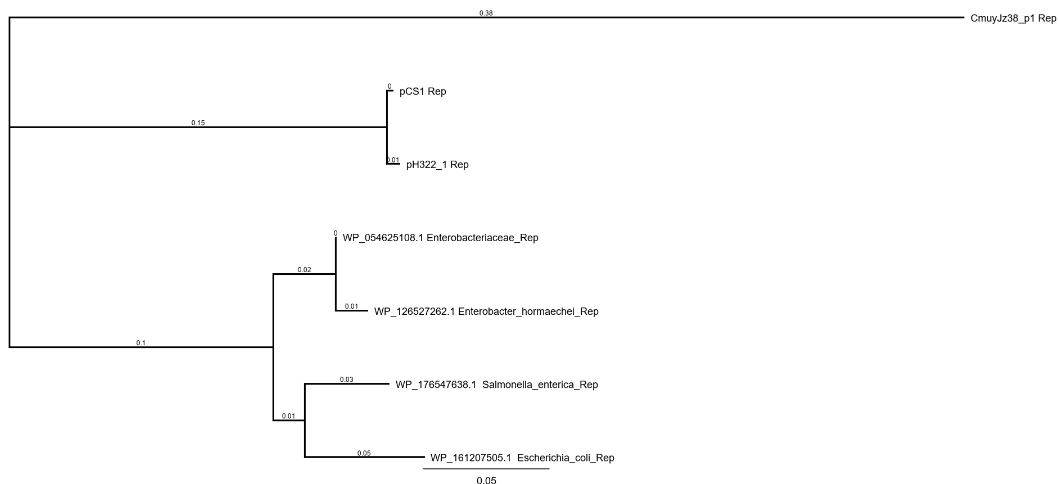


Figure 14. Phylogenetic cluster analysis of members of the origin of replication family (rep) gene of the mobilization plasmid class. The evolutionary history of the mobilization class was estimated by aligning rep gene sequences from each member of the class with nearby phylogenetically related sequences used as outliers. The annotation of each gene is from NCBI. The evolutionary history was determined by first aligning the nucleotide

sequences using the MAFFT algorithm and then using the maximum likelihood technique to determine phylogeny. Next to the branches are the percentages of replicate trees in which the related taxa clustered together (bootstrap test 1,000 repetitions).[95-96] The tree is drawn to scale, with branch lengths equal to the evolutionary distances used to determine the phylogenetic tree's branch lengths. The evolutionary distances were calculated using the maximum composite likelihood approach, and the statistics displayed represent the number of base substitutions per site. The bar indicates a nucleotide difference of 0.05 nucleotides. The tree was created with the UGENE 38.1 software (which may be downloaded at: http://ugene.net/).[93]

There were thirteen phage-associated genes found in *C. sakazakii* ST4 and 64 isolates, by far the isolates reported with the highest number of mobilization class plasmid-borne genes. *C. turicensis* isolates harbored all five type I restriction modification system genes of both plasmids. *C. sakazakii* ST64 and *C. malonaticus* ST60 were the only isolates to harbor a phage-like integrase from pH322_1 and pCS1. *C. sakazakii* ST8 *cpa*+ isolates harbor two DNA methylase genes from different plasmid origins. Similarly, *C. malonaticus* ST60 and ST129 harbor DNA methylase genes from pH322_1 and CmuyJz38_p1.

Cryptic plasmids

Although these plasmids include mobile genetic elements such as conjugative transposons (CTns), their peculiar limited genomic size (average= 4.8kB) distinguishes them from the mobilization plasmid class. Conjugative transposons can move from one bacterial cell to another by a process requiring cell-to-cell contact usually through T4SS conjugation pili. Such elements have been found in many bacterial genera but are particularly common among the

Gram-positive *Streptococci* and *Enterococci* and many Gram-negative organisms as well. This plasmid class includes three *C. sakazakii* plasmids: CSAK29544 2p (4.9 kB), pSP291-3 (4.4 kB), and pCSA2 (5.1 kB). A three-gene operon, *mobABC*, and the transfer origin gene, *oriT*, were identified in all three plasmids. The Mob proteins were predicted to mediate the formation of the relaxosome complex, nick DNA at the oriT, and shuttle the DNA/protein complex to the mating-pore apparatus. The *mobC* gene from CSK29544 2p was found in *C. sakazakii* ST64 and ST83, *C. malonaticus* ST7, and ST60 isolates. *C. dublinensis* ST79, were found to harbor the *mobC* gene from pCSA2. The *mcp* gene, shown to regulate motility of *C. sakazakii*, was identified in pCSA2 only. *C. muytjensii* ST294 isolates were the only other species to show presence of *mcp* from pCSA2. Progressive Mauve Alignment of cryptic plasmid members showed conservation of two blocks of genomic DNA between all three members. Interestingly, an inversion of the two conserved genomic regions was found in pSP291-3 (**see Figure 15**).

Shared insertion sequence elements in *C. sakazakii* plasmids H322_1 and pCS1 comprised most of the insertion elements found in the study (6 of 9 total insertional elements). The other two harbored insertion elements were found on *C. malonaticus* virulence plasmid p2 and *C. sakazakii* cryptic plasmid pSP291-3.

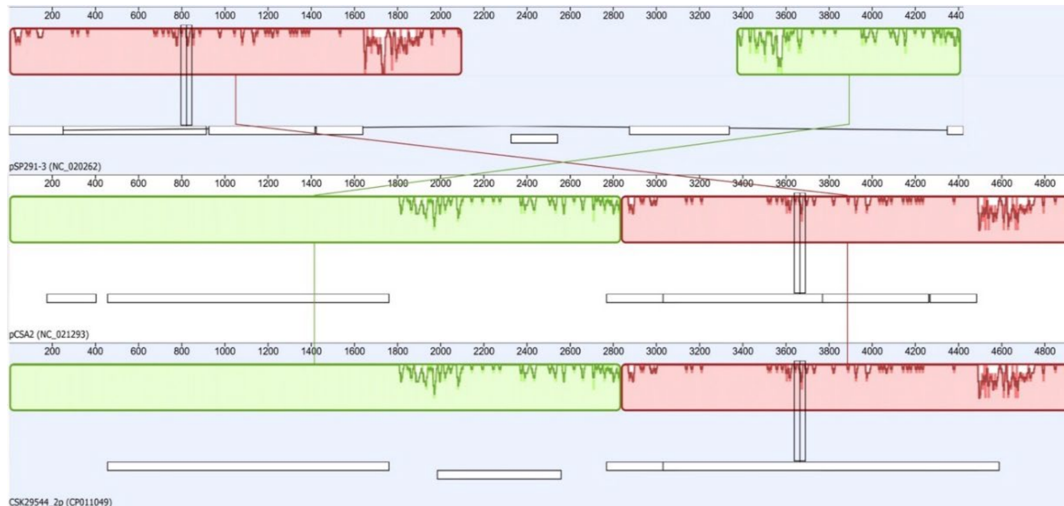


Figure 15**.** Mauve progressive algorithm used to align orthologous and xenologous regions among *C. sakazakii* cryptic plasmid genome sequences to show both local and large-scale changes. Included in this alignment are heavy metal class plasmid members (read top to bottom): 1) pSP291-3; 2) pCSA2; 3) CSK29544_2p. pSP291-3 contains an inverse region of the two conserved domains shared by pCSA2 and CSK29544 2p, as well as a deletion spanning the region 2105...3389 bp. Genomes are separated into horizontal panels by a single black horizontal center line, the name of the genome sequence, and a scale that displays the genome's sequence coordinates. On the inside, colored block outlines denote regions of the genome sequence that matched another and were deemed to be homologous and free of genomic rearrangement. Outside of these blocks, there is no discernible similarity between the input genomes. A similarity profile of the genomic sequence is displayed within each block. The profile's height is proportional to the average degree of conservation in that region of the genome sequence. Since all white areas were not aligned, they are likely to contain sequence fragments unique to a genome. The progressiveMauve Algorithm Tool was used to perform this alignment.[91]

Due to its small size, the plasmid initiator gene was used to mark the origin of replication. The most prevalent shared homology of the origin of replication gene, as a marker and observed in *C. sakazakii* isolates for the cryptic plasmids was found in ST4 isolates, with 110 of the 192 isolates in the database sharing the presence of the origin of replication gene for CSAK29544 3p. Interestingly, pCSA2 origin of replication gene was identified in 114 *C.sakazakii* isolates. pSP291-3, in comparison to the other two plasmids in this class, was seen in only 61 *C. sakazakii* isolates (11%) of which the majority were ST 4. 10% of isolates carried the origin of replication gene for both pCSA2 and CSAK29544 3p, but not for pSP291-3. The absence of shared homology for the origin of replication gene of pSP291-3 in comparison to CSAK29544 3p and pCSA2 was similarly observed in *C. malonaticus*, with just 8% of strains sharing the pSP291-3 *rep* gene, compared to 27% and 29% for CSAK29544 3p and pCSA2, respectively. The origin of the CSAK29544 3p *rep* gene was mostly detected in *C. malonaticus* ST7, with nine of the twenty isolates sharing this gene. The origin of replication gene of pSP291-3 was not identified in any of the isolates from *C. turicensis*. While only 11% of the isolates from *C. muytjensii* shared homology with the pCSA2's origin of replication gene, *C. universalis* and *C. condimenti* isolates did not share any cryptic plasmid origin of replication genes (see **Supplemental Tables S3** and S**4**). Thus, using the shared homology of a *C. sakazakii* cryptic plasmid’s *rep* gene we were able to identify cryptic plasmids in both *C. malonaticus* and *C. muytjensii*. Phylogenetic analysis showed that the plasmid replication initiator gene present in CSAK29544_2p and pCSA2 were

paralogous in nature in comparison to that of pSP291-3. The latter clustered similarly to the plasmid initiator protein of other phylogenetically related enteric sequences, rather than that of CSAK29544_2p and pCSA2 (see **Figure 16**).

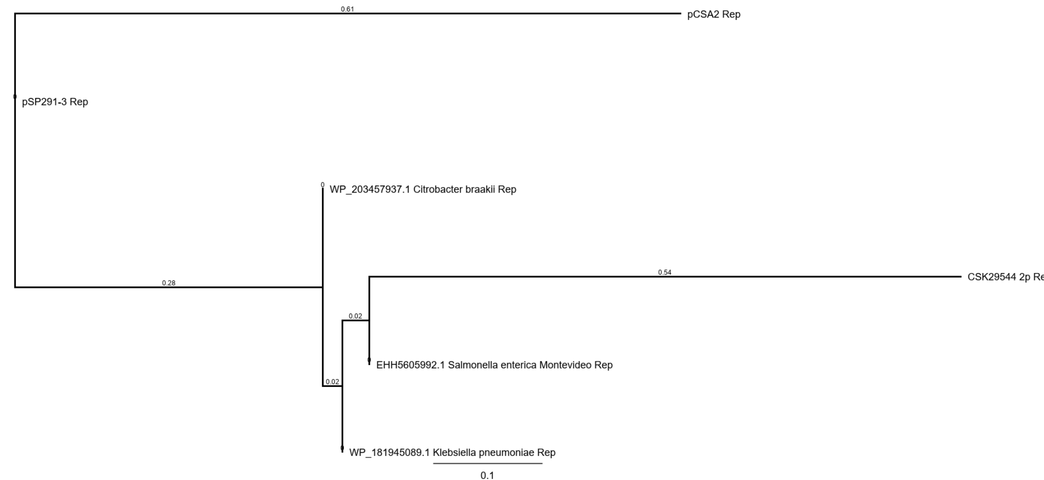


Figure 16. The plasmid replication gene (rep) of each cryptic plasmid member was analyzed using phylogenetic clustering. NCBI was used to annotate each gene. By comparing rep gene sequences from each cryptic plasmid class member with phylogenetically related sequences as outliers, the evolutionary history of the cryptic class was estimated. The evolutionary history was determined by first utilizing the MAFFT method to align the nucleotide sequences and then determining phylogeny using the maximum likelihood technique. The percentages of duplicate trees where related taxa clustered together are displayed next to the branches (bootstrap test 1,000 repetitions).[95-96] The tree is shown to scale, with branch lengths proportional to the evolutionary distances used to determine the branch lengths of the phylogenetic tree. The evolutionary distances were computed using the maximum composite probability method, and the data shown are the number of base substitutions per site. A difference of 0.1 nucleotides is indicated by the bar. The tree was made with the UGENE 38.1 software (available at http://ugene.net/). [93]

DISCUSSION

The taxonomy of *Cronobacter*, like that of many other bacterial genera, has evolved and expanded as more sensitive molecular and sequence-based methods have become available. In this study, we expanded the characterization of plasmid virulence genes on plasmids harbored by *Cronobacter* first described by Franco *et al.*[2].To do so, we performed genome-scale sequence analyses to discern the taxonomic relationship of extant *Cronobacter* plasmid sequences and genome-scale alignment and phylogenetic reconstruction of syntenic, orthologous plasmid genes such as the plasmid replication initiator family protein (*rep*) genes. We discovered that this gene, which appears to be ubiquitous and highly conserved among plasmids carried by many *Enterobacteriaceae*, is a part of the *Cronobacter* core genome in this study. We found that the results from this study were more meaningful in discerning relationships between pairs of *Cronobacter* plasmids that have not been previously studied. Additionally, using the plasmidotyping scheme, we analyzed for the presence of 2954 virulence genes in 32 *Cronobacter* plasmids in 654 genome sequences of a diverse collection of *Cronobacter* strains that were obtained from several surveillance studies.[77-82] Plasmids were divided into six classes based on homology, with each class containing numerous plasmids. The classification of *Cronobacter* plasmids into six groups is based on genomic characteristics and nucleotide compatibility with other plasmids. The compatibility of these plasmids, or their ability to dwell within the same cell, and the specific features possessed by each plasmid type was used to classify them. Fifteen of the plasmids were determined to be virulence

plasmids, based on the criteria established by Franco *et al.* (**see Table 3**).[2] Plasmids with several heavy metal resistance gene clusters were identified as heavy metal resistance plasmids (most notably were the presence of various efflux pump genes). T4SS gene clusters were found in conjugation-type 4 plasmids. Cryptic plasmids were identified by the lack of transfer genes and solely carrying mobilization genes like *mobABCD* and a methyl-accepting chemotaxis protein gene. Multiple antimicrobial resistance genes and insertion elements, such as genes for β-lactam resistance and other resistance genes for carbapenems, quinolones, aminoglycosides, tetracyclines, phenicols, sulphonamides, trimethoprims, rifampicins, bleomycin, and acriflavine, were found on two multi-drug resistance plasmids. Finally, the highest mobilization class plasmid-borne genes such as transposons and DNA methylase were found in mobilization plasmids. Using comparative genomics, we were able to define the syntenic *Cronobacter* shared attributes between the genomes of 15 different plasmids of the virulence plasmid class, including an IncFIB incompatibility class replication of origin gene, *rep*, and two iron acquisition gene clusters (*eitCBAD* and *iucABCD iutA*), as well as various type 1 and type II toxin-antitoxin bicistronic gene pairs.

Recombination frequently rearranges the genome, through horizontal gene transfer which brings new sequences into bacterial chromosomes and plasmids, while deletions eliminate parts of the genome. As a result, each genome has a mosaic of lineage-specific segments, shared entities with a subset of other genomes, and conserved portions across all genomes. Consequently, each genome

that was removed from its original plasmid class post-alignment analysis of genomic DNA may have shared a subset of other genomes at some point prior to events of rearrangements and horizontal transfer. Such plasmids as pGW1 which was originally considered a member of the conjugative plasmid class presented a challenge when sequence comparisons did not show high %ID between itself and other members in the conjugative plasmids class. Conversely, pGW1 and p505108-MDR became its own class of plasmids identified by the abundance of multidrug resistance genes and mobile genetic elements (e.g., IS elements, transposons, and integrases) present. Our comparative phylogenetic examination of these 32 plasmids demonstrated how powerful HGT has been in the microevolution scrupling of *Cronobacter* species mobilome itself including gene clusters and prophages that were associated with these plasmids. This has led the widespread distribution and sharing of genes, and gene clusters across all the plasmids. For example, *C. condimenti* has a Type-6 secretion system that shared a higher homology with that found on *C. sakazakii* plasmids pCS2, pCsaCs09, pESA3, and pGW2.   Similarly, the filamentous hemagglutinin from *C. malonaticus*, *C. dublinensis*, and other non-*C. condimenti* plasmids like pCTU1 were found on the *C. condimenti* plasmid. This occurrence repeats itself in the heavy metal class as well as *mazEF* toxin-antitoxin system observed in *C. turicensis* isolates. These classes may share some common ancestry with those found in *C. malonaticus* pCMA2 plasmid. Furthermore, like plasmidborne TAs, chromosomal TAs might play a role in stabilizing genomic, pathogenicity islands and prophages, which are similarly obtained through horizontal gene transfer

mechanisms. We propose that HGT between plasmids allowed acquisition of virulence factor genes, antibiotic resistance genes, and other gene clusters among various plasmids possessed by *Cronobacter* species that have given *Cronobacter* species advantages in survival and persistence. One of these more studied virulence factors analyzed was that of *cpa*, a gene encoding for a plasminogen activator that plays an important role in *Cronobacter* serum resistance.[2] Although ST143 and ST64 exhibited the highest mean shared homology, these STs are not known to have the strongest proclivity for neonatal meningitis [8,14] and related *Cronobacter* pathology.

When taken together, the benefits of the plasmidotyper bioinformatic toolset for Cronobacter species studies assist to highlight previously unknown molecular properties shared among *Cronobacter* strains and species plasmids. Certain gene contributions to phytobeneficial functions associated with bacterial-growth promotion, virulence and persistence were also discovered. For example, the putrescine export system permease protein was exclusively discovered in ST4 and ST83 *C. sakazakii* in p505108-MDR. Other examples include the *ars* module only identified in specific isolates: *C. dublinensis* ST79, *C. muytjensii* ST546, *C. universalis* ST622, and *C. condimenti* ST98 strains. DNA methylase was found in *cpa*^{pos} *C. sakazakii* ST8 isolates but not in *cpa*^{neg} ST8 *C. sakazakii* isolates, which was a surprising finding that needs further study.

CONCLUSION

In bioinformatics, *in silico* studies entail the coordinated use of computational tools and data sources. A growing number of these resources are

being made available as Web services enabling programmatic access. As part of their investigations, bioinformatics experts will need to organize various Web services and in-house python programs in workflows. As consequence of this goal, classification using a combination of various Web services to promote a new program to characterize the presence of virulence genes in *Cronobacter* species was conducted. *Cronobacter* are emerging pathogens that can cause meningitis, sepsis and necrotizing enterocolitis in neonates and infants, and cause septicemia, pneumonia, gastroenteritis, and wound and urinary tract infections in adults. The genus *Cronobacter* consists of seven species: *C. sakazakii, C. malonaticus, C. muytjensii, C. turicensis, C. condimenti, C. dublinensis* which contains several plasmids.[45]

Plasmid classification lays the framework for deciphering the steady coexistence of diverse extrachromosomal replicons in a single cell, and consequently the genome's arrangement. *Cronobacter* plasmids are strikingly distinctive in that they are composed of mobile and exogenously acquired DNA that comprises several conjugative and composite transposons, a pathogenicity island, integrated plasmid genes and phage regions, and many insertions sequence (IS) elements. Mobile genetic elements can be freely shared among a wide range of bacteria, contributing to bacterial genome flexibility. A number of previous studies have demonstrated that plasmid carriage increases bacterial fitness in the presence of selection for specific functions such as antibiotic resistance. [97-99] Bacterium possessing plasmid DNA have shown the ability to transfer during plasmid-mediated bioaugmentation at low moisture content, a common clinical

origin for *Cronobacter* isolates (PIF facilities).[100] Moreover, plasmids, transposons and other mobile elements allow the bacterium to constantly re-assort virulence factor genes complicating our efforts to group Cronobacter into well-defined genomic pathotypes. Hence, understanding the role of plasmids in virulence and persistence is critical for developing future mitigation measures to combat this bacterium. The goal of this project was to create a bioinformatics plasmid-search engine to find which Cronobacter isolates share specific plasmid DNA genomic properties between Cronobacter plasmids and isolates.

For these reasons, a bioinformatic plasmid search engine created using in-house python scripts ran a comparative search using information derived from two databases. The first database consisted of 32 *Cronobacter* plasmid sequences from all seven *Cronobacter* species, while the second featured 683 draft and closed genomes. Additionally, each plasmid sequence was placed into six categories in accordance to their respective genomic attributes: virulence, T4SS-Conjugation, heavy-metal, cryptic, multi-drug resistant, and mobilization. The resulting datapivot matrix displayed recorded BLAST results for the potential shared homology between *Cronobacter* isolates and *Cronobacter* plasmids. Taken together, the advantages of the plasmidotyper bioinformatic toolkit for *Cronobacter* species research help to emphasize molecular characteristics of *Cronobacter* species strains by ST or subspecies that have not previously been shown. As a result, the plasmidotyper workflow developed in this study may be built with many parts, allowing the same activity to be reused across multiple applications. The objective of this research was to create a bioinformatics

plasmid-search engine that will help researchers find genomic features on *Cronobacter* plasmids. Only a few *Cronobacter* plasmids are available in public databases. As a result, an effort was made to evaluate these sequences and compare these to *Cronobacter* plasmids using various tools. As these strains will be used for NGS analysis to aid source tracking investigations associated with strains causing foodborne outbreaks. Understanding the role of plasmids in virulence and survival would pave the way for the development of novel antimicrobial therapies to combat this pathogen.

CONFLICT OF INTEREST SHARED

Nothing declared.

ACKNOWLEDGEMENTS

We thank the University of Maryland at College Park, Joint Institute for Food Safety and Applied Nutrition (JIFSAN) for supporting JIFSAN graduate student Flavia Negrete.

FUNDING

Financial funds sponsoring this work were obtained internally through U.S. FDA appropriations and this research was also funded in part by the University of Maryland JIFSAN Program through a cooperative agreement with the FDA. #FDU001418.

SUPPORTING INFORMATION

Supplementary Table 1. Functional annotations for C. sakazakii virulence plasmids conserved domains.[a]

Region	Length (referenced to pCsaCS09a)	Genes	Length
Yellow	382…13645	CpmJ protein CDS	543
		CpmK protein CDS	492
		DNA-binding response regulator CDS	693
		drug:proton antiporter CDS	384
		histidine phosphatase family protein CDS	618
		hypothetical protein CDS	1,059
		hypothetical protein CDS	768
		MFS transporter CDS	1,209
		MipA/OmpV family protein CDS	771
		N-acetyltransferase CDS	531
		omptin family outer membrane protease CDS	939
		RepB family plasmid replication initiator protein CDS	>588
		transcriptional regulator CDS	891
		two-component sensor histidine kinase CDS	>1164
		YafY family transcriptional regulator CDS	696
Light Green	13646…67283	AcrB/AcrD/AcrF family protein CDS	3,051
		alpha/beta hydrolase CDS	822
		amino acid-binding protein CDS	399
		antibiotic biosynthesis monooxygenase CDS	324
		arsC CDS	414
		arsenical efflux pump membrane protein ArsB CDS	1,284
		ArsR family transcriptional regulator CDS	726

cell envelope biogenesis protein OmpA CDS	1,704
clpV CDS	2,661
darcynin CDS	321
diguanylate cyclase CDS	486
diguanylate phosphodiesterase CDS	1,218
dihydrodipicolinate synthase family protein CDS	888
EamA/RhaT family transporter CDS	879
efflux RND transporter periplasmic adaptor subunit CDS	1,149
four-helix bundle copper-binding protein CDS	327
GGDEF domain-containing protein CDS	1,464
glyoxalase CDS	387
GntR family transcriptional regulator CDS	720
Hcp1 family type VI secretion system effector CDS	483
helix-turn-helix-type transcriptional regulator CDS	744
hypothetical protein CDS	939
hypothetical protein CDS	930
hypothetical protein CDS	816
hypothetical protein CDS	432
hypothetical protein CDS	312
hypothetical protein CDS	300
hypothetical protein CDS	>292
hypothetical protein CDS	246
LysR family transcriptional regulator CDS	900
metallophosphoesterase CDS	741
methyl-accepting chemotaxis protein CDS	1,911
MFS transporter CDS	1,308
molybdenum ABC transporter substrate-binding protein CDS	807

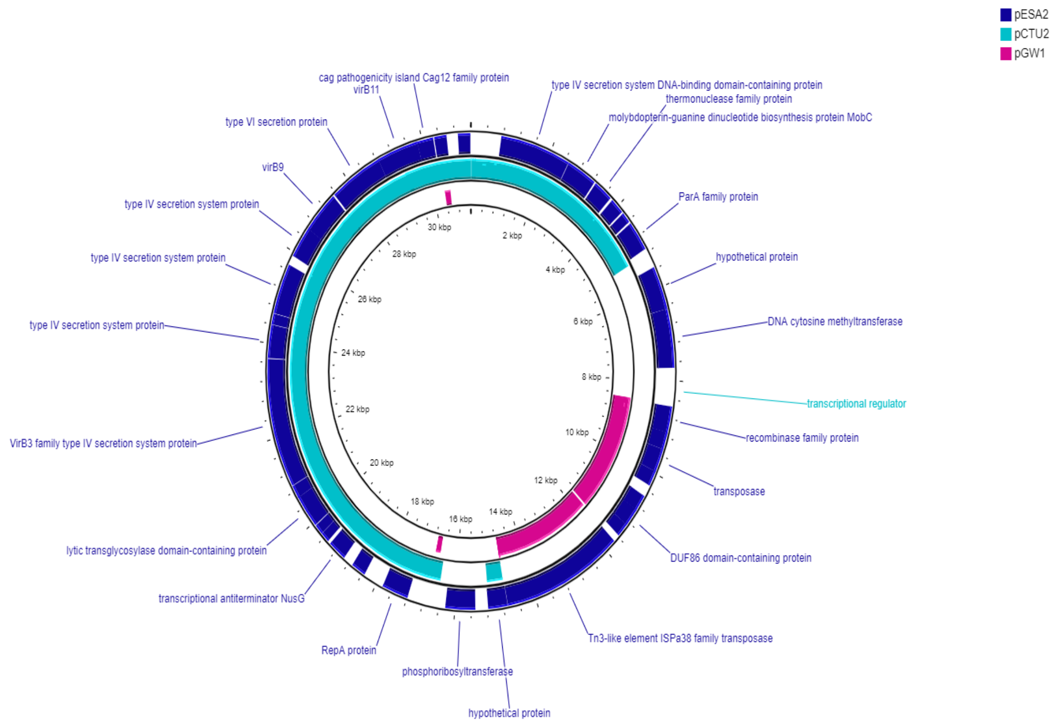
		N-acetyltransferase CDS	516
		NINE protein CDS	294
		nuclear transport factor 2 family protein CDS	357
		nuclear transport factor 2 family protein CDS	348
		NUDIX domain-containing protein CDS	444
		PLP-dependent aminotransferase family protein CDS	1,335
		potassium transporter TrkG CDS	1,476
		PTS sugar transporter CDS	1,377
		starvation-sensing protein RspA CDS	1,200
		TetR family transcriptional regulator CDS	558
		TetR/AcrR family transcriptional regulator CDS	591
		transcriptional regulator CDS	336
		tssB CDS	498
		tssC CDS	1,536
		tssK CDS	1,341
		two-component sensor histidine kinase CDS	>15
		two-component-system connector protein AriR CDS	279
		two-component-system connector protein YcgZ CDS	240
		type IV secretion protein Rhs CDS	554
		type VI secretion system protein ImpK CDS	654
		type VI secretion system tip protein VgrG CDS	2,910
		type VI secretion system tube protein Hcp CDS	492
Orange	69542…67284	DUF2931 domain-containing protein CDS	648
		hypothetical protein CDS	789
		hypothetical protein CDS	>389

		PAAR domain-containing protein CDS	267
Magenta	69543...88209	ABC transporter ATP-binding protein CDS	759
		ABC transporter substrate-binding protein CDS	984
		copper-binding protein CDS	459
		DUF255 domain-containing protein CDS	2,022
		DUF2534 domain-containing protein CDS	261
		DUF2543 domain-containing protein CDS	246
		helicase CDS	3,162
		hypothetical protein CDS	1,818
		hypothetical protein CDS	1,224
		hypothetical protein CDS	390
		hypothetical protein CDS	198
		integrase CDS	975
		iron ABC transporter permease CDS	1,041
		MFS transporter CDS	1,236
		SDR family oxidoreductase CDS	>48
		SMI1/KNR4 family protein CDS	522
		thiol:disulfide interchange protein DsbG CDS	750
		tssJ CDS	390
Red	94689...104067	AAA family ATPase CDS	1,185
		glutathione S-transferase family protein CDS	597
		hypothetical protein CDS	771
		hypothetical protein CDS	327
		methyl-accepting chemotaxis protein CDS	1,947
		mgtA CDS	2,715
Light Pink	104068...105920	autotransporter domain-containing protein CDS	>13
		cupin domain-containing protein CDS	1,122

Dark Purple	105921...131195	aerobactin synthase IucA CDS	1,749
		autotransporter domain-containing protein CDS	>2927
		glgX CDS	2,076
		hypothetical protein CDS	195
		IucA/IucC family siderophore biosynthesis protein CDS	1,746
		lysine 6-monooxygenase CDS	1,323
		MFS transporter CDS	1,212
		N-acetyltransferase CDS	948
		ParA family protein CDS	1,200
		ParB/RepB/Spo0J family partition protein CDS	969
		RepB family plasmid replication initiator protein CDS	87
		siderophore-interacting protein CDS	885
		TonB-dependent siderophore receptor CDS	2,196
		transcriptional regulator CDS	282
		treY CDS	2,532
		treZ CDS	1,785
		type II toxin-antitoxin system HipA family toxin CDS	1,320

[a] Conserved domain for plasmids pictured in Figure 4.

Supplementary Figure 1. Geneious 9.1 map to reference toolkit was used to map ST8 cpa^{+} isolates against ST8 cpa^{-}. pESA3 plasmid sequence (1…65314 bp) was used as a positive control, and CSAK29544_1p plasmid sequence as a negative control. Because the latter plasmid sequence does not share homology with flanking regions of *cpa* gene cluster, then it did not map and is not shown in the alignment. Minimum mapping quality was set to 20 meaning a 99% minimum confidence that the mapping is correct. Minimum support for structural variant discovery for reads to use during the next iteration was set to 2. Blue bar (top) signifies % coverage between mapped regions and reference sequence (pESA3).



Supplementary Figure 2. *Cronobacter* Conjugative Type IV plasmids as analyzed using the PROKSEE Server from the Stothard Research Group. PROKSEE uses BLAST analysis to illustrate conserved and missing genomic sequences (available online: https://beta.proksee.ca/tools). Sequence alignment of Conjugative-Type IV plasmid class

members produced on the PROKSEE Server from the Stothard Research Group that uses BLAST analysis to illustrate conserved and missing genomic sequences (available online: https://beta.proksee.ca/tools). The reference pESA2 (CDS in dark blue arranged outside ring) was downloaded as a GFF file and analyzed on the PROKSEE server using the default settings. If pGW1 shares any significant sequence similarity with the other two members of the Conjugative-Type VI class members, pESA2 and pCTU2, it is from 8527...11677 bp (%ID=99.6). On each of the circular genomes, missing sections detected by the BLAST analysis on the CGView server are displayed as ‘gaps.'

Supplementary Table 2. *Cronobacter strains evaluated in this study.*[a]

[See Attachment]

Supplementary Table 3. *Cronobacter isolates evaluated by species for presence of Cronobacter plasmids.* [a]

	Number of Strains by Species with Origin of Replication Gene for Plasmid						
Plasmid	*C. sakazakii*	*C. dublinensis*	*C. malonaticus*	*C. muytjensii*	*C. turicensis*	*C. universalis*	*C. condimenti*
pCCO1	1	0	0	0	0	0	2
pCDU1	3	19	0	0	0	0	0
p1	0	0	38	0	0	0	0
p2	162	5	12	2	9	0	0
pCMA1	0	0	38	0	0	0	0
pCMA2	162	5	12	2	9	0	0
CmuyJZ38_p1	0	0	0	1	0	0	0
CmuyJZ38_p2	0	0	0	6	0	0	0
CSAK29544_1p	464	0	5	0	0	0	0
CSAK29544_2p	206	2	13	0	1	0	0
CSAK29544_3p	162	5	12	2	9	0	0
p505108-MDR	31	0	3	0	0	0	0
p505108-NDM	1	0	0	0	0	0	0
p505108-T6SS	488	0	5	0	0	0	0
pCS1	140	0	3	0	0	0	0
pCS2	486	0	5	0	0	0	0
pCS3	162	5	12	2	9	0	0
pCSA2	224	2	14	1	2	0	0
pCsaCS09a	488	0	5	0	0	0	0
pESA2	11	0	0	0	0	0	0
pESA3	488	0	5	0	0	0	0
pGW1	5	0	1	0	0	0	0
pGW2	200	0	0	0	0	0	0
pH322_1	140	0	3	0	0	0	0
pH322_2	113	0	4	0	0	0	0
pSP291-1	486	0	5	0	0	0	0
pSP291-2	162	5	12	2	9	0	0

pSP291-3	61	0	4	0	0	0	0
pCTU1	1	0	0	1	17	0	0
pCTU2	14	0	0	0	3	0	0
pCTU3	161	5	12	2	9	0	0
pCUNV1	0	0	0	0	1	3	0

[a] Origin of replication gene, *repA* or *repB* of each plasmid located in various *Cronobacter* isolates, sorted by species.

Supplementary Table 4. *Cronobacter isolates evaluated by ST for presence of Cronobacter plasmids.*[a,b]

	Number of Strains by Sequence Type with Origin of Replication Gene for Plasmid																	
Species	*C. sakazakii*					*C. dublinensis*		*C. malonaticus*			*C. muytjensii*			*C. turicensis*			*C. universalis*		*C. condimenti*
Sequence Type *(Total)*	1 *(148)*	4 *(192)*	8 *(28)*	64 *(31)*	83 *(25)*	106 *(4)*	79 *(3)*	60 *(5)*	7 *(20)*	129 *(5)*	546 *(3)*	81 *(1)*	294 *(1)*	19 *(5)*	32 *(4)*	569 *(2)*	54 *(3)*	622 *(1)*	98 *(1)*
pCCO1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CmuyJZ38_p2	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0
pCDU1	0	0	0	0	0	4	3	0	0	0	0	0	0	0	0	0	0	0	0
CSAK29544_1p	134	179	25	29	1	0	0	0	4	1	0	0	0	0	0	0	0	0	0
p2	2	110	21	2	0	0	0	0	9	3	1	0	1	3	0	2	0	0	0
pCMA2	2	110	21	2	0	0	0	0	9	3	1	0	1	3	0	2	0	0	0
p1	0	0	0	0	0	0	0	4	12	4	0	0	0	0	0	0	0	0	0
pCMA1	0	0	0	0	0	0	0	4	12	4	0	0	0	0	0	0	0	0	0
CmuyJZ38_p1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pESA2	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pGW1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CSAK29544_2p	37	93	22	3	9	0	0	4	4	3	0	0	0	0	0	0	0	0	0
pCSA2	40	105	22	3	9	0	0	4	4	3	1	0	0	0	0	0	0	0	0
pSP291-3	18	24	1	0	6	0	0	3	0	0	0	0	0	0	0	0	0	0	0
p505108-MDR	7	14	3	1	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0
p505108-NDM	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p505108-T6SS	134	179	25	29	1	0	0	0	4	1	0	0	0	0	0	0	0	0	0
CSAK29544_3p	2	110	21	2	0	0	0	0	9	3	1	0	1	3	0	2	0	0	0
pCS1	12	63	0	13	8	0	0	0	1	2	0	0	0	0	0	0	0	0	0
pCS3	2	110	21	2	0	0	0	0	9	3	1	0	1	3	0	2	0	0	0
pH322_1	12	63	0	13	8	0	0	0	1	2	0	0	0	0	0	0	0	0	0
pSP291-2	2	110	21	2	0	0	0	0	9	3	1	0	1	3	0	2	0	0	0
pCS2	134	179	25	29	1	0	0	0	4	1	0	0	0	0	0	0	0	0	0
pCsaCS09a	134	179	25	29	1	0	0	0	4	1	0	0	0	0	0	0	0	0	0
pESA3	134	179	25	29	1	0	0	0	4	1	0	0	0	0	0	0	0	0	0
pCUNV1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0
pGW2	130	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

pH322_2	3	3	0	29	23	0	0	0	4	0	0	0	0	0	0	0	0	0	0
pSP291-1	134	179	25	29	1	0	0	0	4	1	0	0	0	0	0	0	0	0	0
pCTU2	3	2	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
pCTU3	2	110	21	2	0	0	0	0	9	3	1	0	1	3	0	2	0	0	0
pCTU1	0	0	0	0	0	0	0	0	0	0	1	0	0	3	4	2	0	0	0

[a] Origin of replication gene, *repA* or *repB* of each plasmid located in various *Cronobacter* isolates, sorted by sequence type.

[b] Prevalence-based *Cronobacter* ST isolates were analyzed.

REFERENCES

1. Parra-Flores, J., Aguirre, J., Juneja, V., Jackson, E. E., Cruz-Córdova, A., Silva-Sanchez, J., & Forsythe, S. (2018). Virulence and antibiotic resistance profiles of *Cronobacter* sakazakii and *Enterobacter* species involved in the diarrheic hemorrhagic outbreak in Mexico. *Frontiers in Microbiology*, *9*, 2206.
2. Franco, A. A., Hu, L., Grim, C. J., Gopinath, G., Sathyamoorthy, V., Jarvis, K. G., ... & Tall, B. D. (2011). Characterization of putative virulence genes on the related RepFIB plasmids harbored by *Cronobacter* species *Applied and Environmental Microbiology*, *77*(10), 3255-3267.
3. Urmenyi, A. M. C., & Franklin, A. W. (1961). Neonatal death from pigmented coliform infection. *Lancet*, 313-15.
4. Iversen, C., Mullane, N., McCardell, B., Tall, B. D., Lehner, A., Fanning, S., ... & 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* muytjensii 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. *International Journal of Systematic and Evolutionary Microbiology*, *58*(6), 1442-1447.

5. Stephan, R., Grim, C. J., Gopinath, G. R., Mammel, M. K., Sathyamoorthy, V., Trach, L. H., ... & Tall, B. D. (2014). Re-examination of the taxonomic status of *Enterobacter helveticus*, *Enterobacter pulveris* and *Enterobacter turicensis* as members of the genus *Cronobacter* and their reclassification in the genera *Franconibacter* gen. nov. and *Siccibacter* gen. nov. as *Franconibacter helveticus* comb. nov., *Franconibacter pulveris* comb. nov. and *Siccibacter turicensis* comb. nov., respectively. *International Journal of Systematic and Evolutionary Microbiology*, 64(Pt 10), 3402.
6. Farmer III, J. J. (2015). My 40-year history with *Cronobacter/Enterobacter sakazakii*–lessons learned, myths debunked, and recommendations. *Frontiers in Pediatrics*, *3*, 84.
7. Brady, C., Cleenwerck, I., Venter, S., Coutinho, T., & De Vos, P. (2013). Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA): proposal to reclassify *E. nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov. as *Lelliottia nimipressuralis* comb. nov. and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into *Pluralibacter* gen. nov. as *Pluralibacter gergoviae* comb. nov. and *Pluralibacter pyrinus* comb. nov., respectively, *E. cowanii*, *E. radicincitans*, *E. oryzae* and *E. arachidis* into *Kosakonia* gen. nov. as .... *Systematic and Applied Microbiology, 36(5), 309-319.*

8. Hariri, S., Joseph, S., & Forsythe, S. J. (2013). Cronobacter sakazakii ST4 strains and neonatal meningitis, United States. *Emerging Infectious Diseases*, *19*(1), 175.

9. Centers for Disease Control and Prevention (CDC. (2009). *Cronobacter* species isolation in two infants-New Mexico, 2008. *MMWR. Morbidity and Mortality Weekly Report*, *58*(42), 1179-1183.

10. Bar-Oz, B., Preminger, A., Peleg, O., Block, C., & Arad, I. (2001). *Enterobacter sakazakii* infection in the newborn. *Acta Paediatrica*, *90*(3), 356-358.

11. Giri, C. P., Shima, K., Tall, B. D., Curtis, S., Sathyamoorthy, V., Hanisch, B., ... & Kopecko, D. J. (2012). *Cronobacter* species (previously *Enterobacter sakazakii*) invade and translocate across both cultured human intestinal epithelial cells and human brain microvascular endothelial cells. *Microbial Pathogenesis*, *52*(2), 140-147.

12. Wood, R. A., Sicherer, S. H., Vickery, B. P., Jones, S. M., Liu, A. H., Fleischer, D. M., ... & Sampson, H. A. (2013). The natural history of milk allergy in an observational cohort. *Journal of Allergy and Clinical Immunology*, *131*(3), 805-812.

13. Hunter, C. J., & Bean, J. F. (2013). *Cronobacter*: an emerging opportunistic pathogen associated with neonatal meningitis, sepsis and necrotizing enterocolitis. *Journal of Perinatology*, *33*(8), 581-585.

14. Liu, Q., Mittal, R., Emami, C. N., Iversen, C., Ford, H. R., & Prasadarao, N. V. (2012). Human isolates of *Cronobacter sakazakii* bind efficiently

to intestinal epithelial cells in vitro to induce monolayer permeability and apoptosis. *Journal of Surgical Research*, *176*(2), 437-447.

15. Kim, K., Kim, K. P., Choi, J., Lim, J. A., Lee, J., Hwang, S., & Ryu, S. (2010). Outer membrane proteins A (OmpA) and X (OmpX) are essential for basolateral invasion of *Cronobacter* sakazakii. *Applied and Environmental Microbiology*, *76*(15), 5188-5198.
16. Joseph, S., Cetinkaya, E., Drahovska, H., Levican, A., Figueras, M. J., & Forsythe, S. J. (2012). *Cronobacter condimenti* sp. nov., isolated from spiced meat, and Cronobacter universalis sp. nov., a species designation for *Cronobacter* sp. genomospecies 1, recovered from a leg infection, water and food ingredients. *International Journal of Systematic and Evolutionary Microbiology*, *62*(6), 1277-1283.
17. Holý, O., & Forsythe, S. (2014). *Cronobacter* species as emerging causes of healthcare-associated infection. *Journal of Hospital Infection*, *86*(3), 169-177.
18. Craven, H. M., McAuley, C. M., Duffy, L. L., & Fegan, N. (2010). Distribution, prevalence and persistence of *Cronobacter* (*Enterobacter sakazakii*) in the nonprocessing and processing environments of five milk powder factories. *Journal of Applied Microbiology*, *109*(3), 1044-1052.
19. Baldwin, A., Loughlin, M., Caubilla-Barron, J., Kucerova, E., Manning, G., Dowson, C., & Forsythe, S. (2009). Multilocus sequence typing of *Cronobacter* sakazakii and *Cronobacter malonaticus* reveals stable

clonal structures with clinical significance which do not correlate with biotypes. *BMC Microbiology*, *9*(1), 1-9.

20. Joseph S, Sonbol H, Hariri S, Desai P, McClelland M, *et al.,et al.,* (2012) Diversity of the *Cronobacter* genus as revealed by multi locus sequence typing. *J. Clin. Microbiol.* 50: 3031–3039.
21. Joseph, S., Desai, P., Ji, Y., Cummings, C. A., Shih, R., Degoricija, L., ... & Forsythe, S. J. (2012). Comparative analysis of genome sequences covering the seven *Cronobacter* species. *PloS One*, *7*(11), e49455.
22. Alsonosi, A., Hariri, S., Kajsík, M., Oriešková, M., Hanulík, V., Röderová, M., ... & Holý, O. (2015). The speciation and genotyping of *Cronobacter* isolates from hospitalised patients. *European Journal of Clinical Microbiology & Infectious Diseases*, *34*(10), 1979-1988.
23. Brandão, M. L. L., Umeda, N. S., Jackson, E., Forsythe, S. J., & de Filippis, I. (2017). Isolation, molecular and phenotypic characterization, and antibiotic susceptibility of *Cronobacter* species from Brazilian retail foods. *Food Microbiology*, *63*, 129-138.
24. Grim, C. J., Kotewicz, M. L., Power, K. A., Gopinath, G., Franco, A. A., Jarvis, K. G., ... & Tall, B. D. (2013). Pan-genome analysis of the emerging foodborne pathogen *Cronobacter* spp. suggests a species-level bidirectional divergence driven by niche adaptation*. BMC Genomics*, 14(1), 1-16.
25. Jaradat, Z. W., Al Mousa, W., Elbetieha, A., Al Nabulsi, A., & Tall, B. D. (2014). *Cronobacter* species–opportunistic food-borne pathogens. A

review of their virulence and environmental-adaptive traits. *Journal of Medical Microbiology*, *63*(8), 1023-1037.

26. Vojkovska, H., Karpiskova, R., Orieskova, M., & Drahovska, H. (2016). Characterization of *Cronobacter* species isolated from food of plant origin and environmental samples collected from farms and from supermarkets in the Czech Republic. *International Journal of Food Microbiology*, *217*, 130-136.
27. Costa, P. V., Vasconcellos, L., da Silva, I. C., de Mello Medeiros, V., Forsythe, S. J., & Brandão, M. L. L. (2020). Multi-locus sequence typing and antimicrobial susceptibility profile of *Cronobacter sakazakii* and *Cronobacter malonaticus* isolated from corn-based farinaceous foods commercialized in Brazil. *Food Research International*, *129*, 108805.
28. Fei, P., Man, C., Lou, B., Forsythe, S. J., Chai, Y., Li, R., ... & Jiang, Y. (2015). Genotyping and source tracking of *Cronobacter sakazakii* and *C. malonaticus* isolates from powdered infant formula and an infant formula production factory in China. *Applied and Environmental Microbiology*, *81*(16), 5430-5439.
29. Johnson, T. J., & Nolan, L. K. (2009). Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiology and Molecular Biology Reviews*, *73*(4), 750-774.
30. Lederberg, J., & Tatum, E. L. (1946). Gene recombination in *Escherichia coli*. *Nature, 158(4016), 558.*

31. Johnson, T. J., & Nolan, L. K. (2009). Pathogenomics of the virulence plasmids of Escherichia coli. *Microbiology and Molecular Biology Reviews, 73*(4), 750-774.

32. Smith, H. W., & Halls, S. (1967). Studies on Escherichia coli enterotoxin. *Journal of Pathology and Bacteriology*, 93(2), 531-43.

33. Datta, N., & Hedges, R. W. (1971). Compatibility groups among fi− R factors. *Nature*, 234(5326), 222-223.

34. Juhas M, Crook DW, Hood DW. Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. Cell Microbiol. 2008;10:2377-2386.

35. Mellata, M., Touchman, J. W., & Curtiss III, R. (2009). Full Sequence and Comparative Analysis of the Plasmid pAPEC-1 of Avian Pathogenic E. coli χ7122 (O78∶ K80∶ H9). *PloS one*, 4(1), e4232.

36. Cohen, S. N., & Chang, A. C. (1973). Recircularization and autonomous replication of a sheared R-factor DNA segment in *Escherichia coli* transformants. *Proceedings of the National Academy of Sciences, 70(5),* 1293-1297.

37. Siguier, P., Filée, J., & Chandler, M. (2006). Insertion sequences in prokaryotic genomes. *Current opinion in microbiology*, *9*(5), 526-531.

38. Piffaretti, J. C. (2016). Antibiotic resistance: the emergence of plasmid-mediated colistin resistance enhances the need of a proactive one-health approach. *FEMS microbiology letters*, *363*(5). Piffaretti, J. C. (2016). Antibiotic resistance: the emergence of plasmid-mediated colistin

resistance enhances the need of a proactive one-health approach. *FEMS microbiology letters*, *363*(5).

39. Watanabe, T., & Ogata, C. (1966). Episome-Mediated Transfer of Drug Resistance in Enterobacteriaceae IX. Recombination of an R Factor with F. *Journal of Bacteriology, 91(1), 43-50.*
40. Johnson, T. J., & Nolan, L. K. (2009). Plasmid replicon typing. In *Molecular Epidemiology of Microorganisms* (pp. 27-35). Humana Press, Totowa, NJ.
41. Hedges, R. W., & Baumberg, S. (1973). Resistance to arsenic compounds conferred by a plasmid transmissible between strains of Escherichia coli. *Journal of Bacteriology*, *115*(1), 459-460.
42. Yan, Q. Q., Condell, O., Power, K., Butler, F., Tall, B. D., & Fanning, S. (2012). *Cronobacter* species (formerly known as *Enterobacter sakazakii*) in powdered infant formula: a review of our current understanding of the biology of this bacterium. *Journal of Applied Microbiology*, *113*(1), 1-15.
43. Kucerova, E., Clifton, S. W., Xia, X. Q., Long, F., Porwollik, S., Fulton, L., ... & Forsythe, S. J. (2010). Genome sequence of *Cronobacter sakazakii* BAA-894 and comparative genomic hybridization analysis with other *Cronobacter* species. *PloS One*, *5*(3), e9556.
44. Muytjens, H. L., Zanen, H. C., Sonderkamp, H. J., Kollée, L. A., Wachsmuth, I. K., & Farmer 3rd, J. J. (1983). Analysis of eight cases of neonatal meningitis and sepsis due to Enterobacter sakazakii. *Journal of Clinical Microbiology, 18*(1), 115-120.

45. Tall, B. D., Chen, Y., Yan, Q., Gopinath, G. R., Grim, C. J., Jarvis, K. G., ... & Lampel, K. A. (2014). *Cronobacter*: an emergent pathogen causing meningitis to neonates through their feeds. *Science Progress*, *97*(2), 154-172.

46. Stephan, R., Lehner, A., Tischler, P., & Rattei, T. (2011). Complete genome sequence of Cronobacter turicensis LMG 23827, a food-borne pathogen causing deaths in neonates. *Journal of bacteriology*, *193*(1), 309-310.

47. Grim, C. J., Kothary, M. H., Gopinath, G., Jarvis, K. G., Beaubrun, J. J. G., McClelland, M., ... & Franco, A. A. (2012). Identification and characterization of *Cronobacter* iron acquisition systems. *Applied and Environmental Microbiology*, *78*(17), 6035-6050.

48. Miethke, M., & Marahiel, M. A. (2007). Siderophore-based iron acquisition and pathogen control. *Microbiology and Molecular Biology Reviews*, *71*(3), 413-451.

49. Köster, W. (2001). ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. *Research in Microbiology*, *152*(3-4), 291-301.

50. Beretta, G. L., Cassinelli, G., Pennati, M., Zuco, V., & Gatti, L. (2017). Overcoming ABC transporter-mediated multidrug resistance: the dual role of tyrosine kinase inhibitors as multitargeting agents. *European Journal of Medicinal Chemistry*, *142*, 271-289.

51. Li, W., Zhang, H., Assaraf, Y. G., Zhao, K., Xu, X., Xie, J., ... & Chen, Z. S. (2016). Overcoming ABC transporter-mediated multidrug resistance: molecular mechanisms and novel therapeutic drug strategies. *Drug Resistance Updates*, *27*, 14-29.

52. Ben Fekih, I., Zhang, C., Li, Y. P., Zhao, Y., Alwathnani, H. A., Saquib, Q., ... & Cervantes, C. (2018). Distribution of arsenic resistance genes in prokaryotes. *Frontiers in Microbiology*, *9*, 2473.

53. Fang, L., Li, X., Li, L., Li, S., Liao, X., Sun, J., & Liu, Y. (2016). Co-spread of metal and antibiotic resistance within ST3-IncHI2 plasmids from *E. coli* isolates of food-producing animals. *Scientific Reports*, *6*(1), 1-8.

54. Mourao, J., Marçal, S., Ramos, P., Campos, J., Machado, J., Peixe, L., ... & Antunes, P. (2016). Tolerance to multiple metal stressors in emerging non-typhoidal MDR *Salmonella* serotypes: a relevant role for copper in anaerobic conditions. *Journal of Antimicrobial Chemotherapy*, *71*(8), 2147-2157.

55. Staehlin, B. M., Gibbons, J. G., Rokas, A., O'Halloran, T. V., & Slot, J. C. (2016). Evolution of a heavy metal homeostasis/resistance island reflects increasing copper stress in enterobacteria. *Genome Biology and Evolution*, *8*(3), 811-826.

56. Haiko, J., Suomalainen, M., Ojala, T., Lähteenmäki, K., & Korhonen, T. K. (2009). Invited review: breaking barriers—attack on innate immune

defences by omptin surface proteases of enterobacterial pathogens. *Innate Immunity*, *15*(2), 67-80.

57. Haiko, J., Laakkonen, L., Juuti, K., Kalkkinen, N., & Korhonen, T. K. (2010). The omptins of *Yersinia pestis* and *Salmonella enterica* cleave the reactive center loop of plasminogen activator inhibitor 1. *Journal of Bacteriology*, *192*(18), 4553-4561.
58. Jang, H., Gopinath, G. R., Eshwar, A., Srikumar, S., Nguyen, S., Gangiredla, J., ... & Lehner, A. (2020). The secretion of toxins and other exoproteins of *Cronobacter*: role in virulence, adaption, and persistence. *Microorganisms*, *8*(2), 229.
59. Karatan, E., & Watnick, P. (2009). Signals, regulatory networks, and materials that build and break bacterial biofilms. *Microbiology and Molecular Biology Reviews*, *73*(2), 310-347.
60. Zogaj, X., Bokranz, W., Nimtz, M., & Römling, U. (2003). Production of cellulose and curli fimbriae by members of the family *Enterobacteriaceae* isolated from the human gastrointestinal tract. *Infection and immunity*, *71*(7), 4151-4158.
61. Morgan, J. L., Acheson, J. F., & Zimmer, J. (2017). Structure of a type-1 secretion system ABC transporter. *Structure*, *25*(3), 522-529.
62. Masi, M., & Wandersman, C. (2010). Multiple signals direct the assembly and function of a type 1 secretion system. *Journal of Bacteriology*, *192*(15), 3861-3869.

63. Thomassin, J. L., Santos Moreno, J., Guilvout, I., Tran Van Nhieu, G., & Francetic, O. (2017). The trans-envelope architecture and function of the type 2 secretion system: New insights raising new questions. *Molecular Microbiology*, *105*(2), 211-226.

64. Abendroth, J., Mitchell, D. D., Korotkov, K. V., Johnson, T. L., Kreger, A., Sandkvist, M., & Hol, W. G. (2009). The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of *Vibrio cholerae*. *Journal of Structural Biology*, *166*(3), 303-315.

65. Rêgo, A. T., Chandran, V., & Waksman, G. (2010). Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis. *Biochemical Journal*, *425*(3), 475-488.

66. Charro, N., & Mota, L. J. (2015). Approaches targeting the type III secretion system to treat or prevent bacterial infections. *Expert Opinion on Drug Discovery*, *10*(4), 373-387.

67. Ugalde-Silva, P., Gonzalez-Lugo, O., & Navarro-Garcia, F. (2016). Tight junction disruption induced by type 3 secretion system effectors injected by enteropathogenic and enterohemorrhagic *Escherichia coli*. *Frontiers in Cellular and Infection Microbiology*, *6*, 87.

68. Fronzes, R., Schäfer, E., Wang, L., Saibil, H. R., Orlova, E. V., & Waksman, G. (2009). Structure of a type IV secretion system core complex. *science*, *323*(5911), 266-268.

69. Meuskens, I., Saragliadis, A., Leo, J. C., & Linke, D. (2019). Type V secretion systems: an overview of passenger domain functions. *Frontiers in microbiology*, *10*, 1163.
70. Silverman, J. M., Brunet, Y. R., Cascales, E., & Mougous, J. D. (2012). Structure and regulation of the type VI secretion system. *Annual Review of Microbiology*, *66*, 453-472.
71. Russell, A. B., Peterson, S. B., & Mougous, J. D. (2014). Type VI secretion system effectors: poisons with a purpose. *Nature Reviews Microbiology*, *12*(2), 137-148.
72. Wright, G. D., & Sutherland, A. D. (2007). New strategies for combating multidrug-resistant bacteria. *Trends in Molecular Medicine*, *13*(6), 260-267.
73. Magiorakos, A. P., Srinivasan, A., Carey, R. T., Carmeli, Y., Falagas, M. T., Giske, C. T., ... & Monnet, D. T. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clinical Microbiology and Infection*, *18*(3), 268-281.
74. Van Duin, D., & Paterson, D. L. (2016). Multidrug-resistant bacteria in the community: trends and lessons learned. *Infectious Disease Clinics*, *30*(2), 377-390.
75. Souli, M., Galani, I., & Giamarellou, H. (2008). Emergence of extensively drug-resistant and pandrug-resistant Gram-negative bacilli in Europe. *Eurosurveillance*, *13*(47), 19045.

76. Piddock, L. J., Walters, R. N., Jin, Y. F., Turner, H. L., Gascoyne-Binzi, D. M., & Hawkey, P. M. (1997). Prevalence and mechanism of resistance to 'third-generation' cephalosporins in clinically relevant isolates of *Enterobacteriaceae* from 43 hospitals in the UK, 1990-1991. *The Journal of Antimicrobial Chemotherapy*, *39*(2), 177-187.

77. Bonnet, R. (2004). Growing group of extended-spectrum β-lactamases: the CTX-M enzymes. *Antimicrobial Agents and Chemotherapy*, *48*(1), 1-14.

78. Lautenbach, E., Patel, J. B., Bilker, W. B., Edelstein, P. H., & Fishman, N. O. (2001). Extended-spectrum β-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*: risk factors for infection and impact of resistance on outcomes. *Clinical Infectious Diseases*, *32*(8), 1162-1171.

79. Vaidya, V. K. (2011). Horizontal transfer of antimicrobial resistance by extended-spectrum β lactamase-producing *Enterobacteriaceae*. *Journal of Laboratory Physicians*, *3*(1), 37.

80. Zeng, H., Lei, T., He, W., Zhang, J., Liang, B., Li, C., ... & Wu, Q. (2018). Novel multidrug-resistant *Cronobacter sakazakii* causing meningitis in neonate, China, 2015. *Emerging Infectious Diseases*, *24*(11), 2121.

81. Shi, L., Liang, Q., Zhan, Z., Feng, J., Zhao, Y., Chen, Y., ... & Zhou, D. (2018). Co-occurrence of 3 different resistance plasmids in a multi-drug resistant *Cronobacter sakazakii* isolate causing neonatal infections. *Virulence*, *9*(1), 110-120.

82. Restaino, L., Frampton, E. W., Lionberg, W. C., & Becker, R. J. (2006). A chromogenic plating medium for the isolation and identification of *Enterobacter sakazakii* from foods, food ingredients, and environmental sources. *Journal of Food Protection*, *69*(2), 315-322.

83. Jaradat, Z. W., Ababneh, Q. O., Saadoun, I. M., Samara, N. A., & Rashdan, A. M. (2009). Isolation of *Cronobacter* species (formerly *Enterobacter sakazakii*) from infant food, herbs and environmental samples and the subsequent identification and confirmation of the isolates using biochemical, chromogenic assays, PCR and 16S rRNA sequencing. *BMC Microbiology*, *9*(1), 1-11.

84. Chon, J. W., Song, K. Y., Kim, S. Y., Hyeon, J. Y., & Seo, K. H. (2012). Isolation and characterization of *Cronobacter* from desiccated foods in Korea. *Journal of Food Science*, *77*(7), M354-M358.

85. Yan, Q., Wang, J., Gangiredla, J., Cao, Y., Martins, M., Gopinath, G. R., ... & Fanning, S. (2015). Comparative genotypic and phenotypic analysis of *Cronobacter* species cultured from four powdered infant formula production facilities: indication of pathoadaptation along the food chain. *Applied and Environmental Microbiology*, *81*(13), 4388-4402.

86. Gopinath, G. R., Chase, H. R., Gangiredla, J., Eshwar, A., Jang, H., Patel, I., ... & Tall, B. D. (2018). Genomic characterization of malonate positive *Cronobacter sakazakii* serotype O: 2, sequence type 64 strains, isolated from clinical, food, and environment samples. *Gut Pathogens*, *10*(1), 1-22.

87. Chase, H. R., Gopinath, G. R., Eshwar, A. K., Stoller, A., Fricker-Feer, C., Gangiredla, J., ... & Lehner, A. (2017). Comparative genomic characterization of the highly persistent and potentially virulent *Cronobacter sakazakii* ST83, CC65 strain H322 and other ST83 strains. *Frontiers in Microbiology*, *8*, 1136.

88. Jang, H., Woo, J., Lee, Y., Negrete, F., Finkelstein, S., Chase, H. R., ... & Gopinath, G. R. (2018). Draft genomes of *Cronobacter sakazakii* strains isolated from dried spices bring unique insights into the diversity of plant-associated strains. *Standards in Genomic Sciences*, *13*(1), 1-16.

89. Gangiredla, J., Rand, H., Benisatto, D., Payne, J., Strittmatter, C., Sanders, J., ... & Strain, E. (2021). GalaxyTrakr: a distributed analysis tool for public health whole genome sequence data accessible to non-bioinformaticians. *BMC Genomics*, *22*(1), 1-11.

90. Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., ... & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(D1), D206-D214.

91. Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., ... & Pruitt, K. D. (2018). RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic acids research*, *46*(D1), D851-D860.

92. Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss, and rearrangement. *PloS One*, *5*(6), e11147.

93. Okonechnikov, K., Golosova, O., & Fursov, M. (2012). team U.(2012). *Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics*, *28*(8), 1166-1167.

94. Stothard, P., Grant, J. R., & Van Domselaar, G. (2019). Visualizing and comparing circular genomes using the CGView family of tools. *Briefings in Bioinformatics*, *20*(4), 1576-1582.

95. Carattoli, A., & Hasman, H. (2020). PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). In *Horizontal Gene Transfer* (pp. 285-294). Humana, New York, NY.S

96. Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*(4), 772-780.

97. Dykhuizen, D. E., & Hartl, D. L. (1983). Selection in chemostats. *Microbiological reviews*, *47*(2), 150-168.

98. Lenski, R. E., & Nguyen, T. T. (1988). Stability of recombinant DNA and its effects on fitness. Pages S18-20 in J. Hodgson and AM Sugden, editors. Planned release of genetically engineered organisms. *Trends in Biotechnology/Trends in Ecology and Evolution Special Publication. Elsevier, Cambridge, England*.

99. Cooper, N. S., Brown, M. E., & Caulcott, C. A. (1987). A mathematical method for analysing plasmid stability in micro-organisms. *Microbiology*, *133*(7), 1871-1880.

100. Gao, C., Jin, X., Ren, J., Fang, H., & Yu, Y. (2015). Bioaugmentation of DDT-contaminated soil by dissemination of the catabolic plasmid pDOD. *Journal of Environmental Sciences*, *27*, 42-50.

101. Tall, B. D., Gangiredla, J. Gopinath, G. R., Yan, Q. Q., Chase, H. R., Lee, B., Hwang, S., Trach, L., Park, E., Yoo, Y.J., Chung, T., Jackson, S. A., Patel, I. R., Sathyamoorthy, V., Pava-Ripoll, M., M. L. Kotewicz, M. L., Carter L., Iversen, C., Pagotto, F., Stephan, R., Lehner, A., Fanning, S., and Grim, C. J. (2015). Development of a custom-designed, pan genomic DNA microarray to characterize strain-level diversity among *Cronobacter* spp. *Frontiers in Pediatrics 3*:36.

CURRICULUM VITAE

Book Chapters:

1. Jang H., Gopinath GR, **Negrete F**, Weinstein L, Lehner A, and Fanning S, and Tall BD. (2021). 5th edition of Foodborne Infections and Intoxications. (15), 265-283, *Cronobacter* species. Elsevier, Inc. (ISBN 978-0-12-819519-2).
2. Jang H, Gangiredla J, Patel IR, **Negrete F**, Weinstein LM, Ko K, Tall BD, and Gopinath GR. (2021). Diagnosis of Pathogenic Microorganisms Causing Infectious Diseases. *Cronobacter.* Science Publishers, Inc.

Manuscript:

1. Carter L, Chase HR, Choi H, Jun S, Park J, Jeong S, Kim M, Han K, Lee C, Jeong H, Finkelstein S, **Negrete F**, Cinar HN, Tall BD, Gopinath GR. (2017). Draft genome sequences of enterotoxigenic *Bacillus cereus* strains obtained from powdered infant formula. Genome Announc. 0:e01644-16. https://doi.org/10.1128/genomeA.01644-16.
2. Kothary MH, Gopinath GR, Gangiredla J, Rallabhandi PV, Harrison LM, Yan QQ, Chase HR, Lee B, Park E, Yoo Y, Chung T, Finkelstein SB, **Negrete FJ**, Patel IR, Carter L, Sathyamoorthy V, Fanning S and Tall BD. (2017). Analysis and Characterization of Proteins Associated with Outer Membrane Vesicles Secreted by *Cronobacter* spp. *Front Microbiol.*(9), 8-134. https://doi:10.3389/fmicb.2017.00134.

3. Chase HR, Gopinath GR, Eshwar AK, Stoller A, Fricker-Feer C, Gangiredla J, Patel IR, Cinar HN, Jeong H, Lee C, Negrete F, Finkelstein S, Stephan R, Tall BD and Lehner A. (2017). Comparative Genomic Characterization of the Highly Persistent and Potentially Virulent *Cronobacter sakazakii* ST83, CC83 Strain H322 and Other ST83 Strains. *Front. Microbiol.* 8-1136. Https://doi:10.3389/fmicb.2017.01136.
4. Hindermann D, Gopinathrao G, Chase H, Negrete F, Althaus D, Zurfluh K, Tall BD, Stephan R and Nüesch-Inderbinen. (2017). Salmonella enterica serovar Infantis from Food and Human Infections, Switzerland, 2010–2015: Poultry-Related Multidrug Resistant Clones and an Emerging ESBL Producing Clonal Lineage. *Front. Microbiol.* 8-1322. https://doi:10.3389/fmicb.2017.01322.
5. Chase, HR, Eberl L, Stephan R, Jeong H, Lee C, Finkelstein S, **Negrete F**, Gangiredla J, Patel I, Tall BD, Gopinath GR, and Lehner A. (2017). Draft genome sequence of *Cronobacter sakazakii* GP 1999, sequence type 145, an epiphytic isolate obtained from the tomato's rhizoplane/rhizosphere continuum. *Genome Announcements*. 5:e00723-17. https://doi.org/10.1128/genomeA.00723-17.
6. Gopinath GR, Chase HR, Gangiredla J, Eshwar A, Jang H, Patel I, **Negrete F**, Finkelstein S, Park E, Chung T, Yoo Y, Woo J, Lee Y, Park J, Choi H, Jeong S, Jun S, Kim M, Lee C, Jeong H, Fanning S, Stephan R, Iversen C, Reich F, Klein G, Lehner A, and Tall BD. (2018). Genomic characterization of malonate positive *Cronobacter sakazakii* serotype O:2,

sequence type 64 strains, isolated from clinical, food, and environment samples. *Gut Pathogens* 10-11. https://doi.org/10.1186/s13099-018-0238-9.

7. Jang H, Addy N, Ewing L, Beaubrun JJG, Lee Y, Woo J, **Negrete F**, Finkelstein S, Tall BD, Lehner A, Eshwar A, and Gopinath GR. (2018). Whole-genome sequences of *Cronobacter sakazakii* isolates obtained from foods of plant origin and dried-food manufacturing environments. *Genome Announcements*. 6:e00223-18. https://doi.org/10.1128/genomeA.00223-18.
8. Jang H, Gopinath GR, Chase HR, Gangiredla J, Eshwar A, Patel I, Addy N, Ewing L, Beaubrun JJG, **Negrete F**, Finkelstein S, Woo J, Lee Y, Jaradat ZW, Seo K, Fanning S, Stephan R, Lehner A, and Tall BD. (2018). Draft genomes of *Cronobacter sakazakii* strains isolated from dried spices bring unique insights into the epidemiology of plant-associated strains. *Std. Genomics Sciences.* 13-35 https://doi.org/10.1186/s40793-018-0339-6.
9. Negrete F, Gangiredla J, Jang H, Chase HR,Woo J, Lee Y, Patel I, Finkelstein SB, Tall BD, and Gopinath GR. (2019). Prevalence and distribution of efflux pump complexes genes in *Cronobacter sakazakii* using whole genome and pan-genomic datasets. *Current Opinion in Food Science*. (30)32-42. https://dol.org/10.1016/j.cofs2018.11.005.
10. Finkelstein S, Negrete F, Jang H, Gangiredla J, Mammel M, Patel IR, Chase HR, Woo J, Lee Y-Y, Wang CZ, Weinstein L, Tall BD, and Gopinath GR. (2019). Prevalence, distribution, and phylogeny of type two

toxin-antitoxin genes possessed by *Cronobacter* species where *C. sakazakii* homologs follow sequence type lineages. Special issue on Toxins-Antitoxins, *Microorganisms 7*-554.

11. 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, and Lehner A. (2020). The secretion of toxins and other exoproteins of *Cronobacter*: Role in virulence, adaption, and persistence. *Microorganisms,* 8- 229; http://doi:10.3390/microorganisms8020219.
12. Jang H, Chase HR, Gangiredla J, Grim CJ, Patel IR, Kothary MH, Jackson SA, Mammel MK, Carter L, **Negrete F**, Finkelstein S, Weinstein L, Yan Q, Iversen C, Pagotto F, Stephan R, Lehner A, Eshwar AK, Fanning S, Farber J, Gopinath GR, Tall BD and Pava-Ripoll M. (2020). Analysis of the Molecular Diversity Among *Cronobacter* Species Isolated From Filth Flies Using Targeted PCR, Pan Genomic DNA Microarray, and Whole Genome Sequencing Analyses. *Front. Microbiol*. 11:561204. http://doi:10.3389/fmicb.2020.561204.
13. Jang H, Beaubrun JJG, Gangiredla J, Müller M, Tamber S, Patel IR, Ewing L, Weinstein L, Wang CZ, **Negrete F**, Finkelstein SB, Sockett DC, Pagotto F, Stephan R, Tall BD, and Gopinath GR. (2020). Parallel Phylogenetic and MLST Analyses Using Core Genes Showed That Clinical and Food *Salmonella enterica* ssp. *enterica* Serovar Bovismorbificans Polyphyletic Strains Clustered into Three Highly Related Groups. *Applied and Environmental Microbiology.* (Under

review).

14. Jang H, Gopinath GR, Gangiredla J, Patel IR, Beaubrun JJG, Chase HR, Addy N, Ewing L, **Negrete F**, Finkelstein SB, Ko K, Woo J, Lee Y, Wang CZ, Fanning S, Stephan R, Lehner A, and Tall BD. (2020). Characterization of plant-associated *Cronobacter sakazakii* strains using molecular, whole genome, and pan-genome sequence analyses, and Zebrafish infectivity studies identifies clinically relevant and virulent sequence types. (In Prep).
15. Jang H, Negrete F, Weinstein LM, Gangiredla J, Patel IR, Ko K, Chase HR, Finkelstein SB, Chen Y, Stephan R, Lehner A, Eshwar A, Fanning S, Hayman M, Tall BD, Reich F, Klein G, and Gopinath GR. (2021). Comparative Genomic Characterization of *Cronobacter* Species Obtained from a German Powdered Infant Formula Production Facility with Other Strains from Europe, Asia, and the United States. (In Prep).
16. Jang H, Gopinath GR, Weinstein L, **Negrete F**, Finkelstein S, Ko K, Gangiredla J, Patel IR, Edelson-Mammel SG, Carter PJ, Chen Y, Metz M, Sheehan JF, Thompson CJ, Smoot LA, Hayman MM, and Tall BD, (2020). Genomic Characterization of *Cronobacter* Isolates Obtained from Milk Powder Manufacturing Facilities in the United States (In Prep).
17. Gopinath GR, Jang H, Mammel MK, Tamber S, Patel IR, Gangiredla J, Müller A, Wang CZ, Beaubrun JJG, Ewing L, Finkelstein SB, **Negrete F**, Sockett DC, Pagotto F, BD, and Roger Stephan. (2020). Identification of new polyphyletic lineages in *Salmonella enterica* subsp. *enterica* using

novel k-mer binning and minimal core-gene genome-typing methods. (In Prep).

18. Jang H, **Negrete F**, Weinstein LM, Ko K, Gangiredla J, Patel IR, Tall BD, Gopinath GR, and Morávková M. (2021). Phylogenomic characterization of *Cronobacter* species isolated from fresh produce, frozen vegetables, and farm environments in the Czech Republic. (In Prep.).
19. Negrete **FJ**, Hoffman M, Ko K, Jang H, Gopinath GR, Fanning S, Reich F, Lehner A5, and Tall BD. (2021). Complete genome sequences and genomic characterization of five plasmids harbored by environmentally persistent *C. sakazakii* strains ST83 H322 and ST64 GK1025B obtained from powdered infant formula manufacturing facilities. Gut Pathogens. (In Prep).

Abstracts:

1. Finkelstein, S., J. Gangiredla, **F. Negrete**, H. R. Chase, C. Lee, H. Jeong, I.R. Patel[1], G.R. Gopinath, and B.D. Tall. (2017). Phylogenomic analyses of type II toxin-antitoxin genes in the foodborne pathogen *Cronobacter* spp. using sequenced-based bioinformatics combined with DNA microarray analysis demonstrates an evolutionary shared species-specific line of evolution. IAFP Abstract #:P2-130.
2. Gopinath, G. R. H.R. Chase, J. Gangiredla, I. Patel, N. Addy, J.J-G. Beaubrun, B. Li, C. A. Elkins, **F. Negrete**, S. Finkelstein, H. Jeong, C Lee, K. Han, S. Zhao, G.Tyson, H.Tate, D.H. Tilley, Jr., M. P. Simons A.J. McCoy, R. Meza, A.C. Brown, C.R. Friedman and B.D. Tall. (2017).

Comparative Genomics of *bla*$_{CTX\text{-}M\text{-}65}$-resistant clinical strains of *Salmonella enterica* Serovar Infantis from Peru and resistant strains from chicken, cattle and humans. IAFP Abstract #:P2-145.

3. Lee, C., H. Jeong[1], H. Kwon, K. Han, S. Choi, S. Kim, J. Lee, J.Y. Do, S. Finkelstein, **F. Negrete**, H.R. Chase, I.R. Patel, J. Gangiredla, G.R. Gopinath, and B.D. Tall. (2017). Prevalence, distribution, and comparative genomics of a hemolysin III gene (COG1272) and related hemolysin genes among *Cronobacter* spp. IAFP Abstract #: P2-156.
4. **Negrete, F.** J. Gangiredla, S. Finkelstein, I. Patel, H.R. Chase, C. Lee, H. Jeong, B.D. Tall, and G. R. Gopinath. (2017). Phylogenomic analyses of efflux pump complexes in the foodborne pathogen *Cronobacter* spp. using DNA microarray analysis combined with sequenced-based bioinformatics demonstrates presence of species-specific orthologues. IAFP Abstract #:P2-157.
5. Chase, H.R., G.R. Gopinath, J. Gangiredla, I.Patel, Q.Q.Yan, V. Sathyamoorthy, M.H. Kothary, H.N. Cinar, L. Carter, **F. Negrete**, S. Finkelstein, H. Jeong, C. Lee, K. Han, S. Fanning, C. Iversen, R. Stephan, A. Eshwar, A. Lehner, G. Klein, F. Pagotto, M. Pava-Ripoll, and B.D. Tall. (2017). Next-Generation Sequence-Based Typing of *Cronobacter* Spp. For Food and Environmental Safety. Microbe Abstract #: Friday-723.
6. Gopal R. Gopinath, Hannah R. Chase, Jayanthi Gangiredla, Isha Patel, Baoguang Li, Christopher A. Elkins, **Flavia Negrete**, Samantha

Finklestein, HyeJin Jeong, Chaeyoon Lee, Drake H Tilley, Jr, Mark P. Simons, Andrea J. McCoy, Rina Meza, Denise Hindermann, Denise Althaus, Katrin Zurfluh, Roger Stephan, Magdalena Nüesch-Inderbinen and Ben D. Tall. (2017). Resistomics: Tracking the emergence and widening occurrence of *bla*$_{CTX\text{-}M\text{-}65}$ -resistance associated with *Salmonella enterica* Serovar Infantis by NGS and comparative genomics of clinical isolates from Peru. ASM-NGS Meet. Abstract #: Tuesday-116.

7. **Flavia Negrete**, Jayanthi Gangiredla, Samantha Finkelstein, Hyein Jang, Hannah R. Chase, JungHa Woo, YouYoung Lee, Isha Patel, Hannah R. Chase, Ben D. Tall, and Gopal R. Gopinath. (2018). Prevalence and distribution of efflux pump complexes genes in *Cronobacter sakazakii* using whole genome and pan-genomic datasets. IAFP Abstract#: P2-200.
8. Hyein Jang, Athmanya Eshwar, Gopal R. Gopinath, Jayanthi Gangiredla, Isha R. Patel, Junia Jean Gilles Beaubrun, Hannah R. Chase, Nicole Addy, Laura Ewing, **Flavia Negrete**, Samantha Finkelstein, JungHa Woo, YouYoung Lee, Séamus Fanning, Roger Stephan, Angelika Lehner and Ben D. Tall. (2018). Characterization of plant-associated *Cronobacter sakazakii* strains using molecular, whole genome, and pan-genome sequence analyses, and Zebrafish infectivity studies identifies clinically relevant and virulent sequence types. IAFP Abstract#: P2-201.
9. Samantha Finkelstein, Hyein Jang, Gopal R. Gopinath, Jayanthi Gangiredla, Isha R. Patel, **Flavia Negrete**, Hannah R. Chase, JungHa Woo, YouYoung Lee, and Ben D. Tall. (2018). Phylogenomic analyses of

type II toxin-antitoxin genes in the foodborne pathogen *Cronobacter sakazakii* using sequence-based bioinformatics. IAFP Abstract#: P2-202

10. H. Jang, J. Gangiredla, A. Müller, S. Tamber, J. J. G. Beaubrun, I. R. Patel, L. Ewing, S. Finkelstein, **F. Negrete**, D. C. Sockett, F. Pagotto, R. Stephan, B. D. Tall, and G. R. Gopinath. (2019). Parallel Phylogenetic and MLST Analyses Using Core Genes Showed That Clinical and Food *Salmonella enterica* ssp. *enterica* Serovar Bovismorbificans Strains Clustered into Three Highly Related Groups. MICROBE2019 Abstract#: SATURDAY-AAR-594.
11. G. R. Gopinath, H. Jang, M. Mammel, S. Tamber, I.R. Patel, J. Gangiredla, A. Müller, C. Wang, J. J.G. Beaubrun, L. Ewing, S. Finkelstein, **F. Negrete**, D. C. Sockett, F. Pagotto, B.D. Tall, and R. Stephan. (2019). Identification of new polyphyletic lineages in *Salmonella enterica* subsp. *enterica* using novel k-mer binning and minimal core-gene genome-typing methods. MICROBE2019 Abstract#: FRIDAY - EEB-510.
12. H. Jang, J. Gangiredla, I. R. Patel**, F. Negrete**, S. B. Finkelstein, L. M. Weinstein, C. Z. Wang, M. K. Mammel, J. J. G. Beaubrun, N. Addy, G. R. Gopinath, and B.D. Tall**.** Phylogenomic Characterization of Plant-associated *Cronobacter.* (2019).Abstract-AMI-2019 Annual Conference and International Symposium on "Microbial Technologies in Sustainable Development of Energy, Environment, Agriculture and Health at Central University of Haryana during November 15-18, 2019.
13. Jang H, Weinstein LM, Gopinath GR, **Negrete F**, Gangiredla J, Tall BD,

Moravkova M. (2020). Phylogenomic characterization of *Cronobacter* species isolated from fresh produce, frozen vegetables, and farm environments in the Czech Republic. IAFP Ann. Meeting Abstract: Paper ID# 23493.

14. Weinstein LM, Jang H, Gopinath GR, **Negrete F**, Gangiredla J, Patel IR, Tall BD. (2020). Genomic Characterization of Prophage Encoding Regions in *Cronobacter sakazakii.* IAFP Ann. Meeting Abstract: Paper ID# 23494.
15. Hyein Jang, **Flavia Negrete**, Leah M. Weinstein, Jayanthi Gangiredla, Isha R. Patel, Katie Ko, Hannah R. Chase, Samantha B. Finkelstein, Yi Chen, Roger Stephan, Angelika Lehner, Athmanya Eshwar, Seamus Fanning, Melinda Hayman, Ben D. Tall, Felix Reich, Günter Klein, and Gopal R. Gopinath. (2021). Comparative Genomic Characterization of *Cronobacter* Species Obtained from a German Powdered Infant Formula Production Facility with Other Strains from Europe, Asia, and the United States. IAFP Ann. Meeting Abstract: Paper ID#:25735
16. **Flavia Negrete**, Hyein Jang, Katie Ko, Leah M. Weinstein, Jayanthi Gangiredla, Isha R. Patel, Hannah R. Chase, Samantha B. Finkelstein, Mark K. Mammel, David W. Lacher, Roger Stephan, Angelika Lehner, Athmanya Eshwar, Seamus Fanning, Ben D. Tall, Najib El-Sayed, and Gopal R. Gopinath. (2021). Development of a bioinformatics plasmid-search engine for *Cronobacter* species. IAFP Ann. Meeting Abstract: Paper ID#:25683.