Diversity and virulence of the genus *Cronobacter* revealed by multilocus sequence typing (MLST) and comparative genomic analysis

Susan Manju Joseph

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy

July 2013

Experimental work contained in this thesis is original research carried out by the author, unless otherwise stated, in the School of Science and Technology at the Nottingham Trent University. No material contained herein has been submitted for any other degree, or at any other institution.

Susan Manju Joseph

ACKNOWLEDGEMENTS

I would like to express my immense gratitude to my supervisor Prof. Stephen Forsythe for having offered me the opportunity to work on this very exciting project and for having been a motivating and inspiring mentor as well as friend through every stage of this PhD. His constant encouragement and availability for frequent meetings have played a very key role in the progress of this research project. I would also like to thank my co-supervisors, Dr. Alan McNally and Prof. Graham Ball for all the useful advice, guidance and participation they provided during the course of this PhD study.

Special thanks are due to Esin Cetinkaya (Ankara University, Turkey), Arturo Levican and Prof. M. J. Figueras (Rovirai Virgili University, Spain) for all their help and inputs in the speciation studies. I would like to thank the team at Life Technologies (Foster City, CA) for having sequenced the *Cronobacter* spp. genomes used in the latter half of this project, and Prof. Michael McClelland and Dr. Prerak Desai (Vaccine Research Institute, San Diego) for their technical expertise and valuable guidance in the bioinformatic analysis of these genomes.

The journey through this PhD, away from family, would not have been possible for me without the love, laughter and support of some very fine friends I made over the past few years. Many thanks to the Microbiology Prep Room staff, especially Pam and Mike, for their invaluable help at every stage. A huge round of acknowledgement and gratitude goes to all the colleagues and friends (some of whom have now moved on) in the Microbiology Research group as well as Office 104 for their constant support and camaraderie.

Last but definitely not the least; I would like to thank my wonderful family for always showering me with love, blessings and encouragement, even though we have been miles apart. Special thanks are due for my parents for all their sacrifices and hard work that has made me what I am today. Lots of love and gratitude are also due to my wonderful husband, Vishakh, for being my pillar of strength and for patiently enduring the long-distance relationship while I pursued this PhD. I owe all of them a lot.

I would like to dedicate this thesis to my parents and Vishakh, for having always believed in me and supported me unconditionally.

TABLE OF CONTENTS

PUBLICATIONS

- **Joseph, S.,** Forsythe, S.J. 2011. Predominance of *Cronobacter sakazakii* ST4 in neonatal infections. *Emerging Infectious Diseases,* 17, 1713-1715.
- **Joseph, S.,** Cetinkaya, E., Drahovska, H., Levican, A., Figueras, M., Forsythe, S.J. 2012a. *Cronobacter condimenti* sp. nov., isolated from spiced meat and *Cronobacter universalis* sp. nov., a novel species designation for *Cronobacter* sp. genomospecies 1, recovered from a leg infection, water and food ingredients. *International Journal of Systematic and Evolutionary Microbiology*, 62, 1277-83.
- **Joseph, S.,** Sonbol, H., Hariri, S., Desai, P., McClelland, M., Forsythe, S.J. 2012b. Diversity of the *Cronobacter* genus as revealed by multi locus sequence typing. *Journal of Clinical Microbiology*, 50, 3031-3039.
- **Joseph, S.,** Desai, P., Ji, Y., Hamby, S. E., Masood, N., Hariri, S., Sonbol, H., Cumming, C.A., Rico, A., Shih, S. M., Degoricja, L., Brzoska, P., Chuzhanova, N., McClelland, M.., Furtado, M. R., Forsythe, S.J. 2012c. Comparative analysis of genome sequences covering the seven *Cronobacter* species. *PLOS One*, 7, e49455.
- **Joseph, S.,** Forsythe, S.J. 2012. Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Frontiers in Food Microbiology,* 3, 397.
- **Joseph, S.,** Hariri, S., Forsythe, S.J. 2013a. Lack of continuity between *Cronobacter* biotypes and species as determined using multilocus sequence typing. *Molecular and Cellular Probes,* 27, 137-139.
- **Joseph, S.,** Hariri, S., Masood, N., Forsythe, S. 2013b. Sialic acid utilization by *Cronobacter sakazakii*. *Microbial Informatics and Experimentation,* 3, 3.

Contributions during the PhD

- Hamby, S.E., **Joseph, S.,** Forsythe, S.J., Chuzhanova, N. 2011. *In silico* identification of pathogenic strains of *Cronobacter* species from biochemical data reveals association of inositol fermentation with pathogenicity. *BMC Microbiology*, 11, 204.
- Kucerova, E., **Joseph, S.,** Forsythe, S. 2011. The *Cronobacter* genus: ubiquity and diversity. *Quality Assurance and Safety of Crops & Foods*, 3, 104-122.
- Cetinkaya, E., **Joseph, S.,** Ayhan, K., Forsythe, S.J. 2012. Comparison of methods for the microbiological identification and profiling of *Cronobacter* species from ingredients used in the preparation of infant formula. *Molecular and Cellular Probes,* 27, 60-64.
- Hariri, S., **Joseph, S.,** Forsythe, S.J. 2012. Predominance of *Cronobacter sakazakii* ST4 clonal complex strains in *Cronobacter* neonatal meningitis infections in US 2011. *Emerging Infectious Diseases*, 19, 175-177.
- Sonbol, H., **Joseph, S.,** McAuley, C.M., Craven, H.M., Forsythe, S.J. 2012. Multilocus sequence typing of *Cronobacter* spp. from powdered infant formula and milk powder production factories**.** *International Dairy Journal.* 30, 1-7.

Abstract:

Cronobacter spp. (previously known as *Enterobacter sakazakii*) is a diverse bacterial genus consisting of opportunistic food-borne pathogens affecting all age groups, with particularly severe clinical complications such as meningitis and necrotising enterocolitis in neonates and infants. In this study, a multilocus sequence typing (MLST) approach has been established to span the entire *Cronobacter* genus, by employing the alleles of 7 housekeeping genes (*atpD*, *fusA*, *glnS*, *gltB*, *gyrB*, *infB* and *ppsA*, total length 3036 bp). The 325 *Cronobacter* spp. strains used in the study included isolates from the highly publicised *Cronobacter* cases from USA in December 2011. The scheme identified 115 sequence types (ST) across the seven *Cronobacter* species. Multilocus sequence analysis (MLSA) revealed considerable diversity in the genus, with intraspecific variation ranging from low diversity in *C. sakazakii* to extensive diversity within some species such as *C. muytjensii* and *C. dublinensis* including evidence of recombination events between species. An evolutionary analysis revealed the *Cronobacter* genus to have evolved 45-68 million years ago, during the period of evolution of flowering plants. The MLSA was also used in a polyphasic study for the formal recognition of two new species – *C. universalis* and *C. condimenti*. The MLST scheme also revealed the high level of clonality in the species *C. sakazakii* and *C. malonaticus.* ST4 was found to be a highly stable clone of *C. sakazakii*, and a strong association was established between the *C. sakazakii* ST4 clonal complex with neonatal meningitis cases. The curated MLST database is hosted with open access at: http://www.pubmlst.org/cronobacter.

The diversity and virulence study of the organism was then extended to a whole genomic level by analysing eleven high quality draft *Cronobacter* spp. genomes spanning the seven species, including a representative of the *C. sakazakii* ST4 lineage, together with two publicly available genomes. Genome comparison revealed that pair-wise DNA sequence identity varies between 89 and 97% in the seven *Cronobacter* species. The number of annotated genes per genome varied between 3,700 and 4,200. The dataset revealed a pan-genome of more than 6000 genes, 32% of which was found to be conserved across the genus. Genes encoding adhesins, type six secretion systems, metal resistance genes as well as prophages were found in only subsets of genomes and have contributed considerably to the variation of genomic content. Sets of universal core genes and accessory genes unique to each species were identified, which can be used for designing genus/species specific detection assays. The *C. sakazakii* species was found to be unique in the *Cronobacter* genus in encoding genes enabling the utilization of exogenous sialic acid which may have considerable clinical significance. The *C. sakazakii* ST4 genome did not reveal any unique lineage-specific virulence related genes, suggesting the role of selective gene expression and increased host exposure in the pathogenicity of the organism.

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ABC	ATP binding cassette	DNA	Deoxyribonucleic acid
ACT	Artemis comparison tool	DR	Direct repeats
ANI	Average nucleotide identity	EDTA	Ethylenediamine tetra-acetic acid
ANN	Artificial neural networks	EE	Enterobacteriaceae enrichment
AP-PCR	Arbitrarily primed polymerase chain reaction	EGA	Estimated gestation age
ATP	Adenosine triphosphate	ESIA	*Enterobacter sakazakii* isolation agar
BBB	Blood brain barrier	f-AFLP	Fluorescent amplified fragment length polymorphism
BLAST	Basic local alignment search tool	FAO	Food and Agriculture Organization of the UN
BPW	Buffered peptone water	FDA	Food and Drug Administration
BRIG	BLAST ring image generator	FUF	Follow up formula
CC	Clonal complex	HBMEC	Human brain microvascular epithelial cells
CDC	Centers for disease control and prevention	I_A	Index of association
CFU	Colony forming units	ISFET	Ion-sensitive field-effect transistors
CGH	Comparative genomic hybridization	ISO	International Organization for Standardization
CHO	Chinese hamster ovary	ITS	Internal transcribed spacer
CMOS	Complementary metal-oxide semiconductor	KEGG	Kyoto Encyclopedia of Genes and Genomes
CNS	Central nervous system	MEGA	Molecular Evolutionary Genetics Analysis
COSHH	Control of substances hazardous to health	MLEE	Multilocus enzyme electrophoresis
CRISPR	Clustered regularly interspaced short palindromic repeats	MLSA	Multilocus sequence analysis

CSF	Cerebrospinal fluid	RAPD	Random amplification of polymorphic DNA
DDH	DNA-DNA hybridization	RAT	Recombination analysis tool
DFI	Druggan-Forsythe-Iversen	RFLP	Restriction fragment length polymorphism
DLV	Double locus variant	SDS	Sodium dodecyl sulphate
MYA	Million years ago	SLV	Single locus variant
NCBI	National Centre for Biotechnology Information	SOLiD	Sequencing by oligonucleotide ligation and detection
NCIMB	National collection of industrial and marine bacteria	ST	Sequence type
NCTC	National Collection of Type Cultures	START	Sequence type analysis and recombination tool
NEC	Necrotising enterocolitis	MLST	Multilocus sequence typing
NGS	Next generation sequencing	mLST	Modified lauryl sulfate
NICU	Neonatal intensive care unit	MLVA	Multi-locus variable-number tandem-repeat analysis
NTU	Nottingham Trent University	MPN	Most probable number
O-LPS	Oligo-lipopolysaccharide	TAE	Tris-acetate-EDTA
OM	Outer membrane	TLV	Triple locus variant
ORF	Open reading frame	TSA	Trypticase soy agar
PBRT	PCR-based replicon typing	TSB	Trypticase soy broth
PBS	Phosphate buffered saline	UV	Ultraviolet
PCR	Polymerase chain reaction	VRBGA	Violet red bile glucose agar
PFGE	Pulsed field gel electrophoresis	WHO	World Health Organization
PGM	Personal genome machine	XαGlc	5-bromo-4-chloro-3-indolyl-α,d-glucopyranoside
PHAST	Phage Search Tool		
PIF	Powdered infant formula		
PTS	Phosphotransferase system		

CHAPTER 1

GENERAL INTRODUCTION AND AIMS

1.1 The genus *Cronobacter*: An Introduction

Cronobacter (previously known as *Enterobacter sakazakii*) is a Gram-negative bacterial genus belonging to the family *Enterobacteriaceae.* The genus consists of opportunistic pathogens which are motile, peritrichous, facultatively anaerobic, non-spore forming rods. The genus currently consists of seven species: *C. sakazakii, C. malonaticus, C. turicensis, C. muytjensii, C. dublinensis, C. universalis* and *C. condimenti,* the latter two being a direct result of this PhD study (detailed in Chapter 4).

1.1.1 History of *Cronobacter* taxonomy to 2008

The genus *Cronobacter* as we know it today has undergone a number of taxonomic changes over the last 30 years. Initially identified as "yellow pigmented *E. cloacae*", it was given the status of an independent species called *Enterobacter sakazakii* by Farmer *et al.* (1980), based on a study comprising DNA-DNA hybridization (DDH) and phenotypic characterization.

The taxonomic study conducted by Farmer *et al.* (1980) for the definition of the novel *E. sakazakii* species was carried out using 57 bacterial strains. The DDH results showed the organism to be only 41% and 54% related to *Citrobacter freundii* and *Enterobacter cloacae* respectively. Because of the greater genetic relatedness to *E. cloacae* as well as similarity in biochemical reactions, the organism was assigned to the *Enterobacter* genus. The reclassification also comprised the biochemical characterization of the strains using a wide range of tests. Of them, different combinations of ten variable tests were used for defining fifteen biogroups for *E. sakazakii*. Biogroup 1 was found to be the most frequent (42%) and was considered to be the "wild type". The ten defining traits used were: motility, acid production from inositol, ornithine utilization, malonate utilization, indole production, gas production from glucose, nitrate reduction, acid production from dulcitol, methyl-red and Voges-Proskauer. Biogroup 15 was also the only group with an exception to be negative for the reaction with α-methyl-D-glucoside, and hence the test was not included in the list for defining all the biogroups. The study also reported that the *E. sakazakii* strains could be differentiated from the other related members of the *Enterobacter* genus by the following tests: Yellow pigmentation at 25^{o}C, DNase activity, acid production from D-sorbitol and utilisation of lysine, arginine and ornithine. Apart from the DDH and biochemical traits, the strains were also tested for their antibiotic susceptibilities, laboratory growth patterns on enteric media and lactose fermentation. The *E. sakazakii* species was named in honour of the Japanese microbiologist, Richi Sakazakii,

for his contribution to the field of enteric bacteriology. At that point of time, Farmer *et al.* (1980) had suggested the possibility of the species being reclassified as a genus in the future, because of the diversity observed within some of the isolates.

More than two decades later, Iversen *et al*. (2004) investigated the phylogenetic relationships of *E. sakazakii* with the closely related *Citrobacter* species, using 16S rDNA and *hsp60* gene sequences. This study revealed four distinct clusters among the *E. sakazakii* strains, each representing potential novel species. In a follow up study, a correlation was established between this genotyping clustering and the *E. sakazakii* biogroups identified by Farmer *et al*. (1980). In this study, a new biogroup no. 16 was also characterised (Iversen *et al*. 2006b). Table 1.1 provides a summary of the ten biochemical tests used to define the 16 biogroups.

A polyphasic study comprising of fluorescent – Amplified Fragment Length Polymorphisms (f-AFLP), ribotyping, 16S rDNA sequence analysis, DDH and phenotypic characterization was conducted by Iversen *et al*. (2007 & 2008). This study finally resulted in the taxonomic reclassification of *E. sakazakii* into the genus, *Cronobacter*, belonging to the family *Enterobacteriaceae.* The genus name was derived from the mythical Greek god 'Cronus', who ate his own off-springs, in reference to the pathogenicity of the organism (discussed in Section 1.1.5). The genus was comprised of facultatively anaerobic, Gram-negative, oxidase negative, catalase positive, non-spore-forming rods which are generally motile, able to reduce nitrate to nitrite, produce acetoin (Voges-Proskauer test) and are negative for the methyl red test (Iversen *et al*. 2008). The *Cronobacter* genus at that stage was defined to include 5 species, which were differentiated according to the division of the 16 *E. sakazakii* biogroups (Farmer *et al*. 1980; Iversen *et al*. 2006b) as *Cronobacter sakazakii* (biogroups 1-4, 7, 8, 11 and 13), *Cronobacter malonaticus* (biogroups 5, 9 and 14), *Cronobacter turicensis* (biogroups 16, 16a and 16b), *Cronobacter muytjensii* (biogroup 15), and *Cronobacter dublinensis* (biogroups 6, 10 and 12) (Iversen *et al*. 2007& 2008). A set of 14 biochemical tests were chosen to determine the differentiation of the species, as described in Table 1.2. Since insufficient strains and biochemical tests were available to define strain NCTC 9529^{T} (sole member of biogroup 16c), as a separate species, it was called *Cronobacter* genomospecies 1 (Iversen *et al*. 2008). *C. malonaticus* was initially proposed to be a sub-species of *C. sakazakii* as *C. sakazakii* subsp. *malonaticus,* but later classified as an independent species based on the biochemical test for utilization of malonate. *C. dublinensis* was also further divided into three sub-species: *Cronobacter dublinensis* subsp. *dublinensis*, *Cronobacter dublinensis* subsp. *lausannensis* and *Cronobacter dublinensis* subsp. *lactaridii.* Based on DDH experiments, these three sub-species showed >70% similarity and therefore appeared to be the same species. However, they differed from each other in 6 out of the 14 biochemical traits used for the purpose of speciation (Table 1.2).

The taxonomy of the genus *Cronobacter* has been especially complex due to the high inter-species similarity of the 16S rDNA gene sequences, which ranges from 97.8% to 99.7% and the overlap of biochemical profiles (Dauga & Breuwer. 2008). The existence of microheterogeneities in the 16S rDNA gene sequence is another factor that can generate misidentifications as has been reported for certain strains of *C. sakazakii* and *C. malonaticus* that could not be differentiated based on these sequences. The latter was initially even proposed to be a sub-species of *C. sakazakii*, before being grouped as an independent species (Iversen *et al*. 2007). Furthermore, confusions with other members of the *Enterobacteriaceae* have been reported including a number of *E. cloacae* and *E. hormaechei* strains isolated from human infections which have been misidentified as *Cronobacter* spp. using phenotypic tests (Caubilla-Barron *et al*. 2007; Townsend *et al*. 2008). The *Cronobacter* genus also includes strains which were originally deposited in culture collections as *Prashechia flavescens* belonging to *Azotobacteriaceae* that included a clinical isolate from a nine year old suffering pleuropneumonia (Weisglass & Krznaric-Sucic 1979). In addition, the former *E. sakazakii* Preceptrol^{TM} (quality control) strain ATCC 51329 was reclassified as the *C. muytjensii* type strain. Unfortunately, possibly due to the frequent laboratory use of this strain, it has been referred to as *C. sakazakii* in some published literature. There is consequently uncertainty regarding the relevance of those studies, since *C. muytjensii* is rarely associated with any clinical cases of *Cronobacter* infection.

Biogroup[a]	Phenotypic Tests[b]										16S rDNA Cluster	Species[c]
	VP	MR	Nit	Orn	Mot	Ino	Dul	Ind	Malo	Gas		
1	+	-	+	+	+	+	-	-	-	+	1	*C. sakazakii*
2	+	-	+	+	+	-	-	-	-	+	1	*C. sakazakii*
2a	+	-	+	+	-	-	-	-	-	+	1	*C. sakazakii*
3	+	-	+	+	-	+	-	-	-	+	1	*C. sakazakii*
4	+	-	+	-	+	+	-	-	-	+	1	*C. sakazakii*
4a	+	-	+	-	-	+	-	-	-	+	1	*C. sakazakii*
5	+	-	+	+	+	+	-	-	+	+	1	*C. malonaticus*
5a	+	-	+	+	-	+	-	-	+	+	1	*C. malonaticus*
6	+	-	+	+	+	+	-	+	-	+	4	*C. dublinensis*
7	+	-	+	+	+	+	-	-	-	-	1	*C. sakazakii*
8	+	-	-	+	+	+	-	-	-	+	1	*C. sakazakii*
8a	+	-	-	+	+	-	-	-	-	+	1	*C. sakazakii*
8b	+	-	-	+	+	+	-	-	+	+	1	*C. sakazakii*
8c	+	-	-	+	+	-	-	-	+	+	1	*C. sakazakii*
9	+	-	+	+	+	-	-	-	+	+	1	*C. malonaticus*
9a	-	-	+	+	+	-	-	-	+	+	1	*C. malonaticus*
10	+	-	+	+	+	-	-	+	-	+	4	*C. dublinensis*
11	+	-	+	+	+	-	+	-	-	+	1	*C. sakazakii*
12	+	-	+	+	+	+	-	+	+	+	4	*C. dublinensis*
13	-	+	+	+	+	+	-	-	-	+	1	*C. sakazakii*
13a	-	+	+	+	-	+	-	-	-	+	1	*C. sakazakii*
13b	-	+	+	-	+	+	-	-	-	+	1	*C. sakazakii*
13c	-	-	+	+	+	+	-	-	-	+	1	*C. sakazakii*
14	+	-	+	-	+	-	-	-	+	+	1	*C. malonaticus*
14a	+	-	+	-	+	-	-	-	-	+	1	*C. malonaticus*
15	+	-	+	+	+	+	+	+	+	+	3	*C. muytjensii*
16	+	-	+	+	+	+	+	-	+	+	2	*C. turicensis*
16a	+	-	+	+	-	+	+	-	+	+	2	*C. turicensis*
16b	+	-	+	+	+	+	+	-	-	+	2	*C. turicensis*
16c	+	-	+	-	-	+	+	-	+	+	2	*C. universalis*

Table 1.1 Description of the biogroups of *Cronobacter* spp. and their distribution across the genus. Adapted from Iversen *et al*. (2006).

a: Defined by Farmer *et al*. (1983) and Iversen *et al*. (2006) b: VP – Vogues Proskauer ; MR – Methyl Red ; Nit – Nitrate reduction ; Orn – Ornithine utilization; Mot – Motility at 37^{o}C; Ino – Acid production from inositol; Dul – Acid production from dulcitol; Ind – Indole production; Malo – Malonate utilization; Gas – Gas production from glucose. c: *Cronobacter* spp. species breakdown as described by Iversen *et al*. (2008).

Characteristic	1	2	3	4	5	6	7	8
Indole Production	-	-	-	-	+	+	+	v
Carbon utilization:								
Dulcitol	-	-	+	+	+	-	-	-
Lactulose	+	+	+	+	+	+	+	-
Malonate	-	+	+	v	+	+	-	-
Maltitol	+	+	+	+	-	+	+	-
Palatinose	+	+	+	+	v	+	+	+
Putrescine	+	v	+	v	+	+	+	v
Melezitose	-	-	+	-	-	+	-	-
Turanose	+	+	+	v	v	+	v	-
***myo*-inositol**	v	v	+	+	+	+	+	-
***cis*-aconitate**	+	+	+	+	v	+	+	+
***trans*-aconitate**	-	+	-	+	v	+	+	+
4-Aminobutyrate	+	+	+	v	+	+	+	+
1-0-Methyl α-D-glucopyranoside (AMG)	+	+	+	+	-	+	+	+

Table 1.2 Tests for biochemical differentiation of species of the *Cronobacter* genus as determined by Iversen *et al*. (2008).

1 - *C. sakazakii;* 2 - *C. malonaticus*; 3 - *C. turicensis*; 4 - *Cronobacter* genomospecies 1; 5 - *C. muytjensii*; 6 - *C. dublinensis* subsp. *dublinensis*; 7 - *C. dublinensis* subsp. *lactaridi;* 8 - *C. dublinensis* subsp. *lausannensis*.

+: 90% positive; V: 20–80% positive; -: >10% positive.

1.1.2 Physiology

Certain aspects of the physiology of *Cronobacter* spp. have been of particular interest because of its survival in powdered infant formula (PIF). One such attribute has been suggested to be the organism's resistance to dry stress. *Cronobacter* spp. have been reported to exhibit better osmotic stress survival than other organisms isolated from PIF such as *Escherichia coli, Salmonella* spp. and *Citrobacter* spp. (Breeuwer *et al*. 2003). A number of possible reasons have been put forward for this stress resistance. Riedel and Lehner (2007) conducted a proteomic study to identify 53 different protein groups involved in the osmotic stress response of *Cronobacter* spp. The ability of *Cronobacter* spp. to synthesise trehalose was put forward as one such candidate since trehalose can act as a molecular chaperone protecting proteins and cellular membranes from denaturation, a property that has been previously proven in *E. coli* (Breeuwer *et al*. 2003; Horlacher and Boos 1993). *Cronobacter* can also produce large amounts of capsular polysaccharides that can protect the cell in conditions of desiccation. Caubilla-Barron and Forsythe (2007) showed through their study using 27 *Enterobacteriaceae* strains that capsulated strains of *Cronobacter* spp. could be recovered from dehydrated PIF even after 2.5 years. The capsulation of *Cronobacter* is also associated with the biofilm formation properties of the organism. Strains of *Cronobacter* spp. have been found to form biofilms in enteral feeding tubes obtained from the neonatal intensive care units (NICU) of hospitals (Hurrel *et al*. 2009a)

Thermotolerance is another trait of the organism that has been studied in much detail. Heat resistance has been found to vary between the strains of *Cronobacter*, though all strains could be inactivated at pasteurization temperatures at 72^{o}C (Nazarowec-White and Farber 1997; Iversen *et al*. 2004b; Breeuwer *et al*. 2003; Nazarowec-White *et al*. 1999). This observation was one which was used to review the guidelines for reconstitution of PIF to reduce the risk of the organism (FAO-WHO 2004, 2006). More recently, a combination of temperature and ultrasound parameters was found to significantly decrease *C. sakazakii* numbers in reconstituted PIF, an aspect that needs to be investigated further (Adekunte *et al*. 2010). *Cronobacter* can survive in reconstituted PIF within a temperature range of 6 to 47^{o}C (Iversen *et al*. 2004). For laboratory purposes, strains of *Cronobacter* are found to grow comfortably in 16-18 hrs at 37^{o}C on different media such as Trypticase Soy Agar (TSA) and MacConkey's agar. Almost 80% of *Cronobacter* strains exhibit yellow pigmentation when grown on TSA at 25^{o}C, a trait which is believed to be temperature dependent, as the numbers of pigmented strains reduce when grown at 37^{o}C (Iversen and Forsythe 2007).

The properties of polysaccharide production, pigmentation and desiccation resistance have led to the hypothesis of a natural plant habitat for *Cronobacter* spp. by Iversen and Forsythe (2003). The polysaccharide aids the organism to attach itself to plant surfaces while

the desiccation resistance enables it to survive the harsh weather conditions of the environment. The pigmentation in turn also helps to protect the organism from oxygen radicals produced by the radiations of the sun. This also offers a plausible explanation for the survival of the organism in starches which are a major component in the manufacture of products such as PIF.

Some *Cronobacter* strains have been observed to be moderately resistant to acidic conditions of upto pH 3.0, this resistance being enhanced by a pre-exposure to non-lethal acidic conditions (Edelson-Mammel *et al*. 2006). Studies have also shown the survival of *Cronobacter* spp. in a number of acidic fruits and vegetables at 25^{o}C for upto 48 hours (Kim and Beuchat 2005).

Acids are key factors to food preservation and in recent times, studies have focussed more on the use of the bacteriostatic properties of these acids as antimicrobial agents against *Cronobacter* spp., especially in PIF and baby food products. The organic acids, propionic and acetic acid have shown high inhibitory effects against the growth and survival of *Cronobacter* strains in liquids such as fruit juices and baby food (Back *et al*. 2009). Red muscadine (*Vitis rotundifolia*) juices have also exhibited antimicrobial activity against *C. sakazakii*, attributed to the combined activity of malic, tartaric and tannic acids in their contents (Kim *et al*. 2010). The combination of lactic acid and copper at sub-lethal concentrations has also shown inactivation of *Cronobacter* spp. in infant formula (Al-Holy *et al*. 2010).

1.1.3 Reservoirs

Cronobacter spp. are ubiquitous organisms, having been isolated from a very wide range of sources. The organism has often been isolated from different types of herbs, spices, cereals as well as ready-to-eat products such as salads and confectionery, cheese products, meat and vegetables (Iversen and Forsythe 2004; Friedemann 2007; Baumgartner *et al*. 2009). Many of these products also correlate with the proposed natural plant habitat for *Cronobacter* spp. as mentioned earlier, with the raw materials acting as a source for contamination of the organism.

From the environment, the organism has been isolated from freshwater, the isolate being the species type strain of *C. universalis* (previously *C.* genomospecies 1) NCTC 9529 (Iversen *et al*. 2007). A strain of *Cronobacter* spp. has also been isolated from a marine environment in Barcelona, Spain, in a study conducted to examine the bacterial population in the sea surface microlayer (Agogue *et al*. 2006). Other natural environmental sources of *Cronobacter* have included soil, grass silages, and thermal mineral water springs among others (Neelam *et al*. 1987; Mosso *et al*. 1994; Van Os *et al*. 1996). In domestic environments, *Cronobacter* was isolated from 5 out of 16 tested vacuum cleaner bags in a study conducted by Kandhai *et al*. (2004). *Cronobacter* spp. has also been isolated from locations such as the floor,

roofs, tanker bays, drying towers, roller dryers, conveyors and air filters in the manufacturing, processing as well as non-processing environment of a number of milk powder factories (Hein *et al*. 2009; Craven *et al*. 2010; Jacobs *et al*. 2011). One of the strains included in the initial study for the definition of the species *E. sakazakii* was NCIMB 8272, now identified as *C. sakazakii*. This strain had been isolated from dried milk in 1950, thus showing the presence of the presence of the organism in milk products over many decades.

Among animals, a *Cronobacter* strain was recently isolated from the nostril of a stable horse (Holy *et al*. 2011). The organism has also been reported to be isolated from wild house flies (*Musca domestica*), filth flies, larvae of the stable fly, *Stomoxys calcitrans* and the guts of Mexican fruit fly, *Anastrepha ludens* (Kuzina *et al*. 2001; Mramba *et al*. 2006; Butler *et al*. 2010; Pava-Ripoll *et al*. 2012). These reports hint at the possibility that flies could serve as a vector for *Cronobacter* spp.

Cronobacter spp. have been isolated from a number of clinical samples such as blood, cerebrospinal fuid (CSF), sputum, bone marrow, urine, faeces, wound infections (Farmer *et al*. 1980; Muytjens *et al*. 1983; Gallagher and Ball 1991; Iversen *et al*. 2006; Caubilla-Barron *et al*. 2007). The type strain of the species *C. malonaticus*, LMG 23826, is an isolate from a breast abscess. As has already been mentioned before, the organism has also been isolated from neonatal enteral feeding tubes in intensive care units (Hurrell *et al*. 2009b).

However, the one isolation source that the organism has been especially infamous for has been its presence in infant related food products such as follow up formula (FUF) and weaning foods, and most significantly PIF (Chap *et al*. 2009). Powdered infant formula is often used as a substitute for breast milk for infants. As a product, PIF is found to be non-sterile and therefore has often been a source for members of the *Enterobacteriaceae* family, especially *Cronobacter*. The organism has been isolated from PIF products from across the world in multiple surveys conducted over the years (Muytjens *et al*. 1988; Nazarowec-White and Farber 1997; Iversen and Forsythe 2004). In the survey conducted by Muytjens *et al*. (1988), the organism (then *E. sakazakii*) was found to be the third most frequent *Enterobacteriaceae* (after *E. agglomerans* and *E. cloacae*) and had been isolated from 20 of the 141 infant formula samples tested from across the world, at a concentration of not more than 1 CFU/g in any of the samples. Nazarowec-White and Farber (1997) found *Cronobacter* spp. in 6.7% of the 120 infant formula brands in Canada that they tested.

The bacterial contamination of PIF could be either intrinsic, having occurred during the manufacturing and packaging process, or extrinsic, occurring during the process of reconstitution of the formula using non-sterile or contaminated utensils or spoons. A number of *Cronobacter* infections and outbreaks have been traced to contaminated PIF (Muytjens *et al* 1983; Van Acker *et al*. 2001; Himelright *et al*. 2002). A major issue in some of these cases has

also been the temperature abuse of the reconstituted or prepared formula, discussed in Section 1.1.9.

Some of the outbreaks or cases associated with formula-fed infants have also resulted in market recalls of the PIF products citing *Cronobacter* contamination. The first such recall had occurred as a result of the Tennessee outbreak in 2001 (Himelright *et al*. 2002), while the latest one occurred as recently as December 2011, when the supermarket chain Walmart recalled a PIF product following the death of an infant in Missouri, USA (CDC 2011).

1.1.4 Epidemiology

Cronobacter spp. have been implicated as the causative agents in various life threatening diseases such as meningitis, necrotising enterocolitis (NEC), septicaemia and pneumonia, affecting a wide range of age groups (Caubilla-Barron *et al*. 2007, Muytjens *et al*. 1983).

Cronobacter spp. cases have been detected and reported among infants, and even among the infant population, the high risk group has been pre-term, low birth weight infants, <2500 g and less than 28 days of age (Lai *et al*. 2007). The symptoms that have been observed among infants suffering from *Cronobacter* infections include abscess, bacteraemia or sepsis, conjunctivitis, digestive problems, necrotising enterocolitis, meningitis and tonsillitis. Among these, the meningitis cases have been reported to be the most frequent followed by bacteraemia and necrotising enterocolitis (Healy *et al*. 2009). Based on a survey conducted by the Foodborne Diseases Active Surveillance Network (FoodNet) in 2002, the incidence rate of *Cronobacter* meningitis or bacteraemia was calculated to be approximately 1 case for every 100,000 infants per year (Bowen and Braden 2008). According to reports, nearly 40-80% of the meningitic *Cronobacter* infections among infants result in fatalities while *Cronobacter* NEC has been reported to have a fatality rate of 10-55% among infants (Iversen and Forsythe 2003; Lai *et al*. 2007).

The very first reported *Cronobacter* case was that of neonatal meningitis in England by what was at that time identified only as pigmented *E. cloacae* by Urmenyi and Franklin (1961). This was followed by yet another neonatal meningitic case in Denmark by Joker *et al*. (1965). As mentioned earlier, a number of cases of neonatal *Cronobacter* spp. infections have been associated with PIF contamination (Muytjens *et al* 1983; Van Acker *et al*. 2001; Himelright *et al*. 2002).

Bowen and Braden (2006) had surveyed and analyzed possible risk factors of 46 neonatal *Cronobacter* spp. (then *E. sakazakii*) cases, of which 45 were cases of either meningitis or NEC. They observed that in contrast to patients suffering bacteraemia, the meningitic infants had all been of higher gestational age as well as birthweight. Also, the meningitic neonates were

reported to be infected by the organism at a much earlier stage of <1 week age. Bowen and Braden have associated this division in the population to the differences in the formula-feeding regimes of the infants, since the full-term, healthy infants were more likely to be fed non-sterile infant formula at an earlier age.

Cronobacter may be more famous as a causative agent of neonatal infections through its presence in PIF, but the fact that it affects individuals of all age groups is a reality. It has been frequently found to affect elderly or immunocompromised adults, having been isolated in cases of bacteraemia, sepsis, pneumonia as well as wound infections. The majority of these adults were reported to be older than 50 years (Lai 2001). However, these adult cases are generally not as severe as the ones in the younger age group. Also, because of the higher sensitivity associated with the neonatal age group, the cases among adults and older children do not seem to be reported or documented as often. Among adults, greater incidences of only colonization have been reported rather than infection. Some adult *Cronobacter* infections have also arisen as secondary infections, due to other underlying problems such as malignancies. To date, meningitis cases caused by *Cronobacter* spp. have not been reported in any adults (Friedemann 2009).

However, the truth remains that the currently available numbers for cases of *Cronobacter* infections could very well be an underestimate. Improved surveillance and identification methods need to be in place in order to enable a more accurate enumeration of the incidences of *Cronobacter* infections (FAO/WHO 2008).

Some of the major outbreaks in *Cronobacter* research have been summarized below. The strains recovered in these cases have been included for further detailed characterization in this PhD study.

- Muytjens *et al*. (1983) reported *Cronobacter* cases in eight infants in The Netherlands, five of which were from the same hospital and the remaining three from three different hospitals. The infants were all suffering from neonatal meningitis, of which six cases resulted in fatalities. This was possibly the first reported *Cronobacter* outbreak case which was associated with contamination of infant formula, though a direct implication was not concluded. The investigation had resulted in isolation of the organism also from the hospital environment such as powdered formula tin, dishwashing brush and a stirring spoon, even though plasmid profiling had eliminated these strains to be causative agents.

- Twelve infants in an NICU in Belgium were reported to be victims of a *Cronobacter* NEC outbreak by Van Acker *et al*. (2001). All the affected were found to be pre-term, low birth weight, formula fed infants, of which two eventually died

because of the NEC. The strains from the outbreak were analysed by arbitrarily primed polymerase chain reaction (AP-PCR) to reveal three different profiles among the patient isolates. This outbreak happened to be the first one to show a strong association between *Cronobacter* colonisation and NEC cases.

- Block *et al*. (2002) reported the isolation of the organism from five infants in a hospital in Jerusalem. Three of them were infected showing symptoms of bacteraemia and meningitis, while the others were colonizations. These strains were unique in that they were all found to be negative for nitrite reduction. Another interesting fact in this outbreak was about a *Cronobacter* strain that had been isolated from a blender in the hospital in which the formula had been processed. This strain was found to be the same pulsetype as the ones from the infants when tested by pulsed field gel electrophoresis (PFGE). Also, the organism was found to be extremely persistent in the blender, having been isolated for upto 5 months despite repeated decontaminations. This suggested the establishment of an environmental niche by *Cronobacter* spp. in the environment of the blender.

- An outbreak of *Cronobacter* spp. in a hospital in Tennessee in 2001 was reported by Himelright *et al*. (2002). A meningitic infant's unfortunate death led to the testing of 49 hospitalized infants, among which eight more were found to be positive. Of these, two had respiratory illnesses while six were found to be asymptomatic. Further investigations implicated a formula product in this outbreak, leading to the eventual recall of the product by the manufacturers. As a result, this turned out to be a path breaking case in *Cronobacter* research, both for the formula industry as well as regulatory authorities. Ironically, the formula fed to the infants in this outbreak was a commercial formula product not intended for use by neonates, fed on the instructions of the neonatologists. However, this is an issue which has often been overlooked by researchers and never been pursued.

- A *Cronobacter* spp. outbreak had occurred in the NICU of a French hospital in 1994 over a 3-month period. Caubilla-Barron *et al*. (2007) had conducted a detailed phenotypic and genotypic analysis of the 31 *Cronobacter* spp. strains isolated from various body sites of 16 infected infants, of which 3 did not survive, as well as prepared and unprepared infant formula tins. PFGE analysis grouped the strains into four different pulsetypes. The strains from all three infants who died belonged to pulsetype 2. Pulsetypes 1 to 3 were strains obtained from the infants and opened formula, while pulsetype 4 comprised strains from the unopened formula. Two

isolates from the fatalities also showed extended spectrum β-lactamase (ESBL) activity.

- More recently, in December 2011, the Centers for Disease Control and Prevention (CDC) reported cases of *Cronobacter* spp. infections in four different US states (Florida, Illinois, Missouri and Oklahoma). Of these, three of the infants were meningitic, two of whom died (in Missouri and Florida). All the infants had been fed PIF, though all of them were found to be unlinked manufacturers and brands. Following one of the deaths, the supermarket giant Walmart had recalled from their shelves the infant formula used. In Missouri, the organism was also found in an opened PIF tin, prepared PIF and opened nursery water bottle used to prepare the PIF; while in Illinois, the organism was also isolated from an opened nursery water bottle. PFGE analysis of the strains from Illinois and Missouri were also conducted by CDC, and the cases were found to be completely unlinked (CDC 2011).

1.1.5 Pathogenicity and virulence factors

Even though *Cronobacter* spp. infections are not very frequent, they are of great significance mainly because of the severity of infections the organism causes as well as the sensitive age group of the neonates that are affected by them. The pathogenesis of *Cronobacter* spp., as yet, remains unknown and there is a lot of ongoing research trying to identify the exact mechanism and factors associated with the virulence of this organism. It is important to note that not all *Cronobacter* species have been linked with infections, and the severity of virulence has also been found to vary among the different strains. *C. sakazakii* is the type species of the genus and this species has been the most dominant among clinical cases, followed by *C. malonaticus* and *C. turicensis.*

Among infants, the more frequently reported cases have been those of meningitis and NEC. Unlike virtually all other bacterial or viral meningitis cases which result in the inflammation of the meninges of the brain, a meningitic *Cronobacter* infection leads to liquefaction of the cerebral content. This sets it apart from other meningitic bacteria such as *Neisseria meningitidis* and *E. coli* K1. The only other bacterium causing such clinical symptoms in meningitis is the closely related bacterium *Citrobacter koseri* (Kline 1988). Other resulting complications include cerebral haemorrhage, necrosis, infarction and cyst formation. Some patients have also exhibited intractable seizures. This brain damage due to *Cronobacter* spp. meningitis eventually results in fatality in 40-80% of the cases. In the case of survivors, they often exhibit severe permanent neurological damage, including quadriplegia and developmental delays. In a number of the meningitic cases, *Cronobacter* isolates have often been obtained

from blood and CSF samples. The spread of the organism within the body of the host is believed to be blood-borne (Bowen and Braden 2006 & 2008).

Unlike *Cronobacter* spp. meningitis, bacteraemia has been observed in patients of all age groups. Typical symptoms have included high temperature fever, altered mental status and hypotension. In some infants, bacteraemia has been found to occur in tandem with meningitis, and is possibly also a causative factor of the meningitis (Bowen and Braden 2008).

Like various other members of the *Enterobacteriaceae*, *Cronobacter* too has been isolated from faecal samples of patients. They generally have been found to suffer from bloody diarrhoea or NEC (Van Acker *et al*. 2001; Bowen and Braden 2008). NEC is a frequently observed gastrointestinal infection among infants, often resulting in fatalities. The causative factors for the intestinal destruction in NEC have been attributed to prematurity among infants (and therefore an immature intestinal barrier), enteral formula feeding regimes, abnormal bacterial colonization of the intestine with the help of host-pathogen interactions, mucosal injury during ischaemia and increase in the protein levels in the gastrointestinal lumen (associated with infant formula consumption) (Lin and Stoll 2006).

Earlier theories have suggested vaginal infection during birth as a causative agent of *Cronobacter* spp. infections among infants. However, this has been disproved after infections were found to affect infants born via caesarean section as well (Muytjens *et al*. 1983; Caubilla-Barron *et al*. 2007). The exact mechanism of *Cronobacter* spp. pathogenicity has not been completely understood. However, it is highly plausible that the organism enters the central nervous system (CNS) from the blood via the choroid plexus which is the site of production of the CSF. Following this, the organism may permeate through the blood-brain barrier (BBB) aided by secretory factors such as endotoxins, glycopeptides and collagenases. The BBB translocation may happen by paracellular or transcellular methods and in some cases even by destruction of the brain microvascular endothelium of the BBB (Iversen and Forsythe 2003; Huang *et al*. 2000; Kim 2008). Investigations into such pathogenesis mechanisms of bacterial meningitis have included *in vivo* and *in vitro* model studies of meningitic pathogens such as *E. coli* K1, *Citrobacter* spp. (Wang and Kim 2002; Badger *et al*. 1999).

For *Cronobacter* spp. too, a number of *in vitro* studies have been carried out to investigate the bacterium's pathogenicity mechanisms. Studies have proven the ability of the organism to translocate across the human blood brain barrier as a mode of infection for meningitis. *Cronobacter* spp. strains isolated from a fatal NICU outbreak in France were characterized and found to invade Caco-2 cells and brain capillary endothelial cells, as well as survive in macrophages. Greater invasiveness was observed for the strains isolated from the fatal meningitis and NEC cases, with rates comparable to a meningitic *E. coli* K1 strain (Townsend *et al*. 2007; Townsend *et al*. 2008b). Mange *et al*. (2006) conducted *in vitro* studies using HeLa and Caco-2 epithelial cells, as well as human brain microvascular epithelial cell

lines (HBMEC) to characterize the adhesive properties of *Cronobacter* spp. Greater adhesion was observed during the late exponential phase of growth and it was also found based on mannose-sensitivity tests that the adhesion of the organism could be multi-factorial and not solely fimbrial based. Laboratory studies using Caco-2 epithelial cells have shown the role of actin filaments, microtubule structures and tight junction disruption in the adhesion and invasion properties of the organism (Kim and Loessner 2008). Recent *in vitro* studies have demonstrated the role of dendritic cells in tissue damage during NEC (Emami *et al*. 2011).

A number of studies have been carried out showing the association of the outer membrane protein A (OmpA) with the virulence of *Cronobacter* spp. OmpA along with cellular microtubules and microfilaments was shown to play a role in the invasion of intestinal epithelial cells by Nair and Venkitanarayanan (2007). Singamsetty *et al*. (2008) have implicated the role of OmpA expression in *Cronobacter* invasion of HBMEC cells. Mittal *et al*. (2009) carried out studies using *ompA*-mutants of *Cronobacter* spp. to prove the role of the protein in causing meningitis by multiplication in blood and traversal of the BBB. OmpA was also shown to be the major fibronectin-binding protein in *Cronobacter* spp., thus responsible for initiating the invasion of brain microvascular endothelial cells of the BBB (Nair *et al*. 2009).

Kothary *et al*. (2007) have characterized a zinc-containing metalloprotease (Zpx) unique to *Cronobacter* spp. which was found to cause rounding of Chinese hamster ovary (CHO) cells. They have speculated the role of this enzyme in the necrosis and cellular destruction occurring during the NEC in infants. In a study to characterize the fimbrial haemagglutinins in the *Enterobacter* genus, 50% of the *Cronobacter* spp. strains (then *E. sakazakii*) tested were found to be positive for fimbrial mannose sensitivity, haemagglutinins and fimbrial production (Adegbola and Old 1983). A plasmid borne outer membrane protease termed *Cronobacter* plasminogen activator (Cpa), was identified in mutant studies with *C. sakazakii* as playing a role in the organism's serum resistance activity. It was found to be highly similar to the Pla subfamily of omptins, identified as virulence factors in a number of *Enterobacteriaceae* members (Franco *et al*. 2011). Using microarray analysis, Townsend *et al* (2007) reported about the possible role of the *sodA* gene in macrophage survival of the organism by exhibiting resistance to oxidative stress conditions. Despite being a constitutive trait, the gene was found to be variably present in the strains tested in the study.

Transposon mutagenesis studies have identified two candidate genes to play a role in the biofilm formation abilities of *C. sakazakii* (Hartmann *et al*. 2010). The biofilm formation may be linked to the adhesive properties of the organism, and therefore linked to virulence potential. *Cronobacter* spp. have been reported to form biofilms on a wide range of surfaces such as glass, stainless steel, polyvinyl chloride as well as enteral feeding tubes obtained from NICU (Iversen *et al*. 2004b; Kim *et al*. 2006; Hurrell *et al*. 2009).

Pagotto *et al*. (2003) had used a mouse model to show the enterotoxin production in some strains of the *Cronobacter* spp. In this study, they also determined a minimum lethal dose of the organism. They found all the 18 *Cronobacter* spp. strains tested in the study to be lethal to the suckling mice at a dose of 10^8 CFU per mouse, though the minimum dose was found to vary, with some strains being found to be lethal at a 1000-fold less dose too.

Investigation of the virulence of the organism has moved on to the genomic level ever since the *C. sakazakii* BAA-894 and *C. turicensis* z3032 genomes were published (Kucerova *et al*. 2010; Stephan *et al*. 2010). Researchers have identified various putative virulence factors related to Type VI secretion system, iron acquisition, blood-brain barrier penetration, enterobactin and aerobactin synthesis among others (Kucerova *et al*. 2011; Franco *et al*. 2011).

The epidemiology and pathogenicity of the organism raise important questions regarding the possible factors that trigger the mode of virulence in this organism. It is highly pertinent to find out what sets apart the virulent *Cronobacter* strains from the non-virulent ones, especially with the variation in pathogenicity observed between the species too. This is an answer which hopefully can be answered in the present age of genomic studies. The diversity and significance of a number of these virulence factors in the genomes of the *Cronobacter* genus have been discussed in detail in Chapter 7.

1.1.6 Culture and phenotypic identification methods

The diversity in sources, coupled with the clinical cases, has necessitated the development of robust detection methods for the organism. The U.S Food and Drug Administration (FDA) have approved standard protocols for the isolation of the bacterium from PIF, based on enrichment methods and biochemical characterization. The earlier methods were similar to the protocol for isolation of *Salmonella* spp. from milk powder samples (FDA 2002). These included a pre-enrichment step in distilled water or buffered peptone water (BPW), followed by enrichment in the *Enterobacteriaceae* enrichment (EE) broth at 36^{o}C overnight. The sample is then streaked or spread onto Violet Red Bile Glucose (VRBG) agar plates and incubated overnight at 36^{o}C, to then observe for typical colony morphology of *Enterobacteriaceae*. The organism appears on the plate as purple colonies surrounded by bile acids precipitated in the form of a purple halo. Five such presumptively positive colonies are then picked and grown on TSA plates at 25^{o}C to select for yellow pigmented colonies. The identification of the colonies is then confirmed using API20E strips and a Most Probable Number (MPN) is also determined. However, there were a number of limitations associated with this method. The protocol did not select for *Cronobacter* spp. initially, while all *Enterobacteriaceae* are enriched in the EE broth to form pink-purple colonies on the VRBGA

plates, and hence could overgrow *Cronobacter* spp. Also, since only 80% *Cronobacter* spp. strains are reported to be yellow pigmented, there are possibilities of false negative identifications when picking colonies off the TSA plates. The overall procedure was also laborious and long, requiring nearly 5 days for completion. These issues were overcome in a revised FDA protocol released in 2012, which also took into account the taxonomic revision of the organism as the genus *Cronobacter* into the then six species (Chen *et al*. 2012). In brief, the procedure involves a pre-enrichment step in sterile BPW for 24 hrs. The samples are then centrifuged and the pellets are aliquoted in phosphate buffered saline (PBS). These aliquots are then cultured and screened for the presence of *Cronobacter* spp. by using a combination of chromogenic media - Druggan-Forsythe-Iversen (DFI) agar and R&F agar, biochemical tests (Rapid ID32E or VITEK 2.0) and real-time PCR assays. The entire procedure until the final confirmed identification of *Cronobacter* spp. takes 2 days and has been graphically presented in Fig. 1.1. This revised protocol has been validated in a number of studies in recent times (Chen *et al*. 2010; Chen *et al*. 2011a).

The above method is only valid for the US, while the rest of the world follows the validation of the International Organization for Standardization (ISO) method for the detection of *Cronobacter* spp. in PIF samples. Briefly, in the ISO method, the pre-enrichment step with BPW is followed by the inoculation of 100 µl of the culture to 10 ml of a selective enrichment medium called the modified lauryl sulphate (mLST) broth and incubation at 44^{o}C for 24 hours. This is then followed by streaking for growth on the chromogenic *Enterobacter sakazakii* isolation agar (ESIA) at 44^{o}C, and subsequent sub-culture of typical colonies on TSA plates at 25^{o}C for yellow pigmentation. The yellow pigmented colonies are then picked for biochemical characterisation (Anonymous 2006).

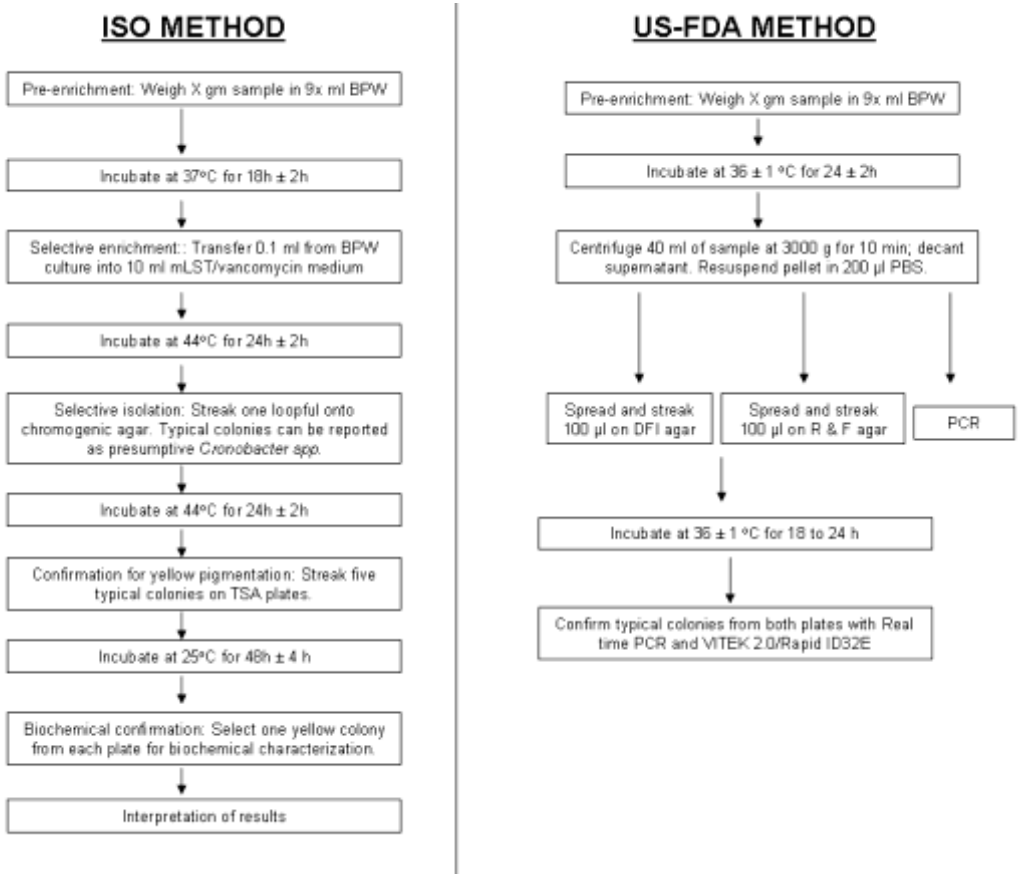


Fig 1.1 Flowcharts of the US-FDA and ISO approved protocols for the isolation of *Cronobacter* spp. from PIF. Adapted from Anonymous (2006) & Chen *et al*. (2012)

Apart from these methods, there are a number of commercially available media and broths selective for *Cronobacter* spp. The organism can be identified selectively in the laboratory on chromogenic media such as DFI agar and *Enterobacter sakazakii* Isolation Agar (ESIA) media, which are primarily based on the α-glucosidase activity of the organism (Iversen *et al*. 2004a). Earlier, Muytjens *et al*. (1984) had reported about the α-glucosidase activity of the organism that could be used for differentiating it from *E. cloacae*. The *Cronobacter* spp. are detected on the DFI agar plates based on their α-glucosidase activity that is detected using the chromogen 5-bromo-4-chloro-3-indolyl-α,d-glucopyranoside (XαGlc). This reaction results in the formation of blue-green colonies selective for *Cronobacter* spp. in a mixed *Enterobacteriaceae* culture (Iversen *et al*. 2004). The use of these chromogenic media has been a major improvement in the detection process of *Cronobacter* spp. in laboratories. The osmotic resistance property of *Cronobacter* spp. has also been used in the design of a number of enrichment broths, such as the mLST broth with 0.5 M NaCl, *Cronobacter* screening broth with 10% sucrose and the *E. sakazakii* enrichment broth with sodium deoxycholate. A study carried out with 177 *Cronobacter* spp. strains (then *E. sakazakii*) to analyse the efficiency levels of some of these broths concluded that there is a constant need to re-evaluate the amount of selective agents used, in order to improve the accuracy of *Cronobacter* spp. detection (Iversen and Forsythe 2007).

Biochemical tests have been traditionally used for microbial identification even at the species level. For *Cronobacter* spp. too, a range of tests have been chosen which can, in combination, help to identify the individual species. This biochemical typing scheme was also used in the taxonomic reclassification study for the genus *Cronobacter* as described in Section 1.1.1. There are also a number of commercial kits available for the identification of *Cronobacter* spp. strains based on a set of selective biochemical reactions. Some of the more popular ones are API20E, ID32E and VITEK 2.0 (Biomerieux). However, a major drawback to some of these phenotypic kits has been the outdated nomenclature in the databases, with some still identifying the organism as *E. sakazakii.*

Since these phenotypic identification methods have often exhibited discrepancies resulting in false-positive and false-negative detections of *Cronobacter* spp. (Iversen *et al.* 2004a), there is a need to constantly evaluate and improve them for improved accuracy.

1.1.7 Molecular typing and identification methods

Research on *Cronobacter* has considerably progressed in recent years, with a number of molecular based typing methods being developed for improved strain identification and characterisation.

Like most members of *Enterobacteriaceae*, for *Cronobacter* too, the 16S rDNA region has been a popular target for molecular identification studies. A number of probes have been designed over the years by various researchers specifically targeting different regions of the gene either for the purpose of end-detection or for identification of the organism (Keyser *et al*. 2003; Lehner *et al*. 2004; Iversen *et al*. 2007), which also included a TaqMan real-time PCR assay (Malorny & Wagner. 2005). Iversen *et al*. (2004) had analysed the phylogenetic relationships and diversity of *Cronobacter* spp. (then *E. sakazakii*) within the *Enterobacteriaceae* family using the 16S rDNA and hsp60 sequences which had grouped the strains into four distinct phylogenetic clusters, giving strong indications of the existence of the *Cronobacter* genus. In a separate study using Artificial Neural Networks (ANN), Iversen *et al*. (2006) had also identified *E. sakazakii*-specific target regions in the partial 16S rDNA sequence. This is a 528 bp long region which can be used for strain identification. However, for the purpose of phylogenetic analysis full length 16S rDNA sequences (>1300 bp) are used. 16S rDNA sequence diversity has traditionally been used to define genus (5%) and species (3%) boundaries for prokaryotic populations. However, with respect to very closely related organisms, 16S rDNA sequence analysis has limitations due to minimal diversity in the sequences. Also, most bacteria possess multiple copies of the 16S rDNA operon, introducing discrepancies because of the sequence diversity between these multiple versions (Acinas *et al*. 2004). This has been an issue for *Cronobacter* spp. too, especially for the differentiation between *C. sakazakii* and *C. malonaticus*, which show 99.7% similarity for the full length 16S rDNA sequence, even though DDH studies revealed them to be <70% related. During the taxonomic reclassification from the *E. sakazakii* species to *Cronobacter* genus status, *C. malonaticus* was initially proposed as a subspecies of *C. sakazakii* (Iversen *et al*. 2007), before being classified as an independent species (Iversen *et al*. 2008).

The internal transcribed spacer (ITS) region is the nucleotide sequence located between the 16S and 23S rDNA genes in bacteria. Liu *et al*. (2006a & 2006b) have conducted multiple studies involving oligonucleotide arrays as well as TaqMan Real-time PCR probes targeting these ITS regions for the detection of the organism in infant formula. A separate real-time PCR assay for infant formula detection of the organism was also designed targeting the partial macromolecular synthesis operon, specifically the rpsU and the primase (*dnaG*) genes (Seo & Brackett. 2005). However, in all these mentioned instances, the probes were designed using a limited number of *E. sakazakii* strains, and therefore have not been adequately validated for use after the taxonomic revision to the *Cronobacter* genus.

The taxonomic study that resulted in the reclassification of *E. sakazakii* as the genus *Cronobacter* employed a number of molecular techniques for clarifying the diversity within the strains. Apart from the mandatory biochemical tests required to confirm the phenotypic characteristics of the species, other techniques such as DDH, f-AFLP using *EcoRI* and *TaqI*,

ribotyping and full length 16S rDNA sequencing (>1300 bp) were used for defining the genus *Cronobacter* (Iversen *et al*. 2007 & 2008).

PFGE has traditionally been used as a "gold standard" for outbreak analysis, source tracking and surveillance of the organism (Caubilla-Barron *et al*. 2007; Nazarowec-White & Farber. 1997; Craven *et al*. 2010; CDC. 2011). The methodology follows a Pulsenet approved protocol for *Salmonella* (Swaminathan *et al*. 2001), with the most frequently used restriction enzyme being *XbaI*. PFGE has successfully exhibited a high genomic level discrimination of *Cronobacter* spp. strains in various studies. However, it does pose certain disadvantages of high labour and time required for the process and the lack of portability of results. Recently, a modified PFGE protocol was published for *Cronobacter* isolates, validated to work for all the recognised species of the genus, with further confirmation of results by using an additional enzyme *SpeI* (Brengi *et al*. 2012). Studies have reported the use of PFGE in the profiling of *Cronobacter* spp. isolated from milk powder and infant formula processing plants, in order to trace the distribution of clones within a factory or manufacturing unit (Craven *et al*. 2010, Jacobs *et al*. 2011). Another important PFGE analysis study of 30 strains isolated from a fatal NICU outbreak in France (Caubilla-Barron *et al*. 2007) revealed that strains of multiple pulsetypes had been isolated from the same infant, stressing the need to analyse multiple colonies from the same sample.

Nazarowec-White and Farber (1999) had conducted a study using 17 *Cronobacter* spp. (then *E. sakazakii*) isolates to evaluate phenotypic and genotypic methods for characterizing the organism. The genotypic methods used in the study included ribotyping, random amplification of polymorphic DNA (RAPD) and PFGE. The results showed RAPD and PFGE to be the most discriminatory methods of the three for analysing the bacterium.

A multi-locus variable-number tandem-repeat analysis (MLVA) scheme was proposed for the organism by Mullane *et al*. (2008a) based on the genome sequence of *C. sakazakii* BAA-894, and was tested on 112 isolates of the organism. This study was conducted during the same time as when the taxonomic revision of *E. sakazakii* to the genus *Cronobacter* had been proposed (Iversen *et al*. 2007) and the results showed considerable diversity that correlated with the same. However, the scheme was not pursued for re-evaluation later after the genus *Cronobacter* and species had been defined.

The O-antigen is a very important component of the cell walls of Gram-negative bacteria, its variations responsible for the various serotypes in bacterial species. The genes involved in the synthesis of this O-antigen region are located in *Enterobacteriaceae* at the *rfb* locus (encoding dTDP-D-glucose 4,6 dehydratase), between the *galF* and *gnd* genes which encode for UDP-glucose pyrophosphorylase and gluconate-6-phosphate dehydrogenase respectively. Owing to their diversity and variation, these genes have been popular targets for typing in bacteria and such PCR based O-serotyping schemes have also been designed for the

Cronobacter spp. (Mullane *et al*. 2008b; Jarvis *et al*. 2011; Sun *et al*. 2011). Initially, the O1 and O2 serotypes for *Cronobacter* spp. (then *E. sakazakii*) were identified by Mullane *et al*. (2008b) with PCR probes targeting the *wehC* and *wehI* genes respectively. These were later reclassified as *C. sakazakii* serotypes O1 and O2 by Jarvis *et al*. (2011). Subsequently, this study has been further elaborated to cover the diversity of the genus *Cronobacter* and ten more serotypes have been identified for the genus– four more for *C. sakazakii* (serotypes O3, O4, O5, O6 and O7), two for *C. malonaticus* (serotypes O1 and O2), two for *C. turicensis* (serotype O1 and O2) and one for *C. muytjensii* (serotype O1). The PCR identification schemes for these serotypes have been designed with primers targeting either the *wzx* or *wzy* genes (Jarvis *et al*. 2011; Sun *et al*. 2011, 2012a & 2012b). Though the serotyping scheme is useful for identifying the serotypes in the *Cronobacter* genus and has revealed the high level of diversity of the lipopolysaccharide variations in the organism, there have been a few practical issues associated with it. The fact that every serotype requires a different set of primers makes the process fairly complicated, especially since a prior knowledge of the species identification would be required to ease the process of selecting from a large number of PCR primer pair combinations. In the study by Jarvis *et al*. (2011), the serotypes of nearly 20% of the strains tested could not be determined using the current range of primers, suggesting a greater diversity and variation in the regions. Hence, there is a need to re-evaluate the scheme to encompass the diversity with fewer primers to target the all the variations.

A number of PCR probes have been designed for the organism based on individual housekeeping genes such as *gyrB* and *rpoB* (Dauga & Breeuwer. 2008; Stoop *et al*. 2009). The *rpoB* scheme has recently been expanded to incorporate the latest taxonomic revisions of the genus, to encompass all seven *Cronobacter* species. However, since the scheme involves a separate set of primers (with varying amplification conditions) for each individual species, the convenience of using a single target gene for typing the genus can get overcome in the process (Lehner *et al*. 2012). The virulence related outer membrane protein gene, *ompA* has also been targeted for the development of a PCR probe for detecting the organism in infant formula (Nair & Venkitanarayanan. 2006). Proudy *et al*. (2008) had conducted a study using 27 *Cronobacter* spp. strains to evaluate a BOX-PCR technique, a PCR-RFLP sequencing scheme for the flagellin gene, *fliC*, and PFGE. Their results showed a very comparable discriminatory power of isolates using BOX-PCR and PFGE, though the *fliC* gene did not prove to be a good target to detect strain variability. More recently, Carter *et al*. (2012) have designed a multiplexed PCR assay for *Cronobacter* spp. specific for the di-guanylate cyclase encoding gene, *cdgc1*, for the purpose of species-level identification. However, it has been validated only for six out of the seven *Cronobacter* species.

Analysis based on the sequences of multiple housekeeping genes (multilocus sequence analysis, MLSA) has previously proven to be a useful tool for members of *Enterobacteriaceae*

(Lacher *et al*. 2007, Ibarz Pavón & Maiden. 2009). A MLSA study was conducted by Kuhnert *et al*. (2009) based on the three genes *recN*, *rpoA* and *thdF* comprising the genus *Cronobacter* and also related members of *Enterobacteriaceae.* The scheme was successful in explaining the phylogenetic relationships between the various species. However, this scheme too involved the use of multiple primers for each gene with varying amplification parameters.

A multi-locus sequence typing (MLST) scheme developed by Pagotto and Farber was first reported for *Cronobacter* spp. This scheme was based on the ten loci *fusA, gyrB, ileS, lepA, pyrG, recA, recG, rplB* and *rpoB* (Fanning & Forsythe 2008). However, no further details of primers, sequences or results have been published to date.

Recently, a 7-loci MLST scheme was developed for the identification and characterisation of *C. sakazakii* and *C. malonaticus* (Baldwin *et al*. 2009). This scheme required the partial sequence analysis of seven housekeeping genes: *atpD, fusA, glnS, gltB, gyrB, infB* and *ppsA*. When concatenated together these sequences provided 3036 nucleotides for analysis. This scheme was effectively able to demonstrate a robust phylogenetic frame to separate the two species which was previously found to be quite difficult using techniques such as 16S rDNA sequencing. The study also reported that some of the previous confusion between the two species may have been due to incorrect speciation of some of the biotype index strains, hinting at the possible flaws in the initial use of phenotyping to define the species. The *Cronobacter* MLST scheme is hosted with open access at http://www.pubmlst.org/cronobacter.

It is important to point out that some of the detection and typing methods that have been referred to above were developed during the initial phase of concern over the microbiological safety of PIF, while some others were based on poorly characterized bacterial strain collections of *E. sakazakii* that did not truly represent the diversity of the organism. Subsequently, not all methods are applicable for present day *Cronobacter* spp. detection as the probes have not been adequately validated for use after the taxonomic revisions. Since the current international regulations require the absence of *Cronobacter* spp. in 10g of PIF, the end-detection method must be able to cover all formally recognised species. Therefore the suitability of many of these tests is questionable and need to be re-evaluated.

1.1.8 Genome studies

Genomic studies of *Cronobacter* spp. have made considerable progress in the last few years. The very first genome of the genus was that of *C. sakazakii* strain BAA-894, sequenced by the Genome Centre at Washington University, USA, using the whole genome shotgun method, supplemented with end sequencing of a fosmid library. The strain was an isolate from a formula tin associated with a *Cronobacter* NICU outbreak in Tennessee, USA (Details in

Section 1.1.4). The genome comprised of a 4.4 Mb chromosome and two plasmids, pESA2 and pESA3, which were 31 kb and 131 kb in size, respectively (Accesion No. NC_009778 – 80). This genome was published by Kucerova *et al*. (2010) in a whole-genome microarray study of the diversity of the *Cronobacter* genus. The genomic sequence of *C. sakazakii* BAA-894 genome was used for the construction of a 384,030 probe oligonucleotide tiling DNA microarray, comprising of 10 strains belonging to the five species of the genus – *C. sakazakii, C. malonaticus, C. turicensis, C. muytjensii* and *C. dublinensis*. The results of the microarray study showed a 55% core genome for *C. sakazakii* and 43% core genome for the *Cronobacter* genus, and identified mobile elements including prophage regions to be responsible for a large amount of the diversity. Along with a follow-up publication (Kucerova *et al*. 2011), the study reported the presence of a number of virulence factors in the genome such as iron acquisition systems, type six secretion systems and fimbrial clusters.

Not long after this study, a genome of *C. turicensis* strain z3032 was sequenced and announced by the Institute for Food Safety and Hygiene, Zurich, Switzerland (Stephan *et al*. 2011). The strain was a 2005 isolate from a fatal case in a Children's Hospital in Zurich. The chromosome of this genome was of a similar size at 4.38 kb, with three plasmids sized 22 kb, 53 kb and 138 kb (Accession No. NC_013282 – 85).

Franco *et al*. (2011) conducted a study which included a comparison of the larger plasmids of the two genomes (pESA3 and pCTU1; 131 kb and 138 kb). They reported both plasmids to be largely similar in content and identified them as virulence plasmids harbouring key regions such as type six secretion systems, iron acquisition systems, plasminogen activator and filamentous haemagglutinin genes.

More recently, during the course of this project, the whole genome shotgun assembly sequence of *C. sakazakii* strain E899 (Accesion No. AFMO00000000) was announced on Genbank and in print by Chen *et al*. (2011b). No plasmid sequences were available for this genome. There have been no studies published to date linked to the comparative analysis of any of these genomes.

1.1.9 Public health significance and future directions

Even though *Cronobacter* infections are not very frequent, there is a great need to clarify the risks associated with them to the general public, so that adequate measures can be implemented to control the growth of the organism. Since the target population of the pathogen has often been infants, with PIF being implicated as a causative agent in a number of infections, extra caution and awareness needs to be raised to maintain the microbiological safety of infant formula and related products. This caution needs to be practised both at the manufacturing stage

as well as the PIF handling and reconstitution stage. Since manufactured PIF is not a sterile product, it is the responsibility of manufacturers to ensure least possibilities of contamination of the product in the processing environment.

The incidences of *Cronobacter* infections have eventually led to the changes implemented by Food and Agricultural Organization/World Health Organization (FAO/WHO) to the microbiological criteria of manufactured PIF during their risk assessment meetings. These changes resulted in the recognition of *Cronobacter* spp. along with *Salmonella enterica* as a major bacterial risk factor to infants being fed reconstituted PIF and thereby stressed on the significance of good practices to be implemented during the process of reconstitution of formula (FAO/WHO, 2004, 2006, 2008; CAC 2008).

The WHO strongly emphasizes the need for infants <6 months old to be breast-fed, but when this is not possible, certain standards need to be adhered to for the formulae products that are fed to non-breast fed infants, as well as for the reconstitution of these formulae. Individuals preparing the reconstituted formula for the infants, including mothers, caregivers and medical personnel, should be adequately trained and made aware of possible hazards. WHO recommends that extra care should be taken to maintain aseptic conditions during the preparation as well as to avoid any post-preparation temperature abuse of the formula that can allow for the growth of the organism. The recommended temperature of the reconstitution water is 70^{o}C, which can help to inactivate any bacterial growth. The prepared formula must be consumed within 30 minutes and if stored, must be refrigerated for not more than 24 hours (Iversen and Forsythe 2003; WHO 2001; WHO 2007; Turck 2012).

Despite the FAO/WHO risk assessment meetings starting in 2004, the microbiological criteria of the Codex Alimentarius Commission did not change until 2008. These now apply to PIF for a target age of up to 6 months. However, the criteria were not applied for formula commonly known as ‘follow-up formula’, which are used at the weaning stage (>6 months of age). Although there was microbiological evidence of the frequent isolation of *Cronobacter* spp. from such formula and weaning foods, the epidemiological evidence available was not adequate enough to establish the need for additional microbiological testing by manufacturers (FAO/WHO 2008).

The FAO/WHO (2004) committee had also stated the need for improved research into the organism to gain better understanding into aspects such as taxonomy and virulence. The literature review presented here shows how this has been an ongoing process over the past years, though there is still a fairly long way to go.

1.2 AIMS OF THE STUDY

Research on *Cronobacter* spp. has considerably progressed in recent years, with improved isolation, identification and molecular typing methods. These methods have exhibited the wide diversity of the organism, warranting the need for further detailed investigation. The main aim of this PhD project was to investigate and understand the virulence and diversity of the genus *Cronobacter* and this was proposed to be carried out in two phases, consisting of the following aims:

- **MULTI-LOCUS SEQUENCE TYPING AND ANALYSIS**

- To expand the 7-loci MLST scheme, initially set up for *C. sakazakii* and *C. malonaticus*, to span the entire *Cronobacter* genus.
- To use MLSA as a tool to investigate the population diversity and evolution of the genus *Cronobacter.*
- To analyse the clonality and recombination in the population of the *Cronobacter* genus
- To investigate a possible correlation between the diversity and virulence of the organism

The results obtained in this phase of the study have been presented and discussed in Chapters 3, 4 and 5. These results have also been published in Joseph & Forsythe (2011 & 2012) and Joseph *et al*. (2012a & 2012b).

- **COMPARATIVE GENOMIC ANALYSIS OF THE GENUS**

- To compare the genomes of the representative members of the *Cronobacter* species and study the genomic diversity and phylogeny of the genus *Cronobacter*.
- To establish the *Cronobacter* pan-genome and identify signature regions unique to each species, as well as those shared between the species.
- To identify and characterise the putative genomic regions linked to the physiology and virulence of *Cronobacter* spp. and thereby seek answers for the pathogenicity of the organism.

The results from this phase of the study have been presented and discussed in Chapters 6 and 7. Some of these results have also been published in Joseph *et al*. (2012c).

CHAPTER 2

MATERIALS AND METHODS

2.1 Safety considerations

All experiments were carried out in accordance with the Health and Safety Code of practice for Microbiology Containment Level 2. All materials and protocols were thoroughly assessed and the appropriate COSHH forms were completed. Good laboratory practices were followed for microbes, media and chemicals, as well as for operating laboratory devices. All waste was disposed of as recommended in the material safety data sheets.

2.2 Bacterial strains

All bacterial strains used for this project were either selected from the Nottingham Trent University (NTU) culture collection of *Cronobacter* spp., or recently obtained from culture collections or collaborators following a review of the literature. The strains with isolation and further details have been listed in Table 2.1.

All strains were stored at -80^{o}C in Trypticase Soy Broth (Oxoid, UK) with 10% (v/v) glycerol (Fisher Scientific, UK). They were grown overnight on TSA plates incubated aerobically at 37^{o}C and checked for purity, before performing the experiments.

Table 2.1 List of *Cronobacter* spp. isolates included in this study

Species	NTU ID	Culture collection code	Country	Source	Year	Comments			
						A	**B**	**C**	**D**
C. sakazakii	1	NCTC 11467 ^{T}	USA	Throat	1980		X		
C. sakazakii	2	ATCC 12868	Unk	Unk	Unk		X		
C. sakazakii	4	HPB 2856 SK90	Canada	Clinical	2003		X		
C. sakazakii	5	HPB 2852 LA	Canada	Clinical	2003		X		
C. sakazakii	6	HPB 2853 LB	Canada	Clinical	2003		X		
C. sakazakii	7	-	Switzerland	Infant formula	2003		X		
C. sakazakii	12	-	CR	Faeces	2004		X		
C. sakazakii	14	-	Switzerland	Infant formula	2003		X		
C. sakazakii	20	-	CR	Clinical	2004		X		
C. sakazakii	25	-	Korea	Follow up formula	2004		X		
C. sakazakii	26	-	Korea	Follow up formula	2004		X		
C. sakazakii	27	-	UK	Milk powder	2004		X		
C. sakazakii	33	-	USA	Weaning food	Unk	X			
C. sakazakii	120	-	Korea	Follow up formula	2005		X		
C. sakazakii	121	-	Korea	Follow up formula	2005		X		
C. sakazakii	140	-	Unk	Spice	Unk	X			
C. sakazakii	150	-	Korea	Spice	2005		X		
C. sakazakii	376	-	UK	Unknown	Unk	X			
C. sakazakii	377	NCIMB 8272	UK	Milk powder	1950		X		
C. sakazakii	424	-	France	Infant formula	2006		X		
C. sakazakii	425	-	France	Infant formula	2006		X		
C. sakazakii	426	-	France	Infant formula	2006		X		
C. sakazakii	467	-	France	Infant formula	2006		X		
C. sakazakii	470	ATCC 29004	USA	Unk	1973		X		
C. sakazakii	471	-	Korea	Infant formula	2006		X		
C. sakazakii	511	-	CR	Clinical	1983		X		
C. sakazakii	513	-	CR	Clinical	1983		X		
C. sakazakii	518	-	CR	Clinical	1983	X			
C. sakazakii	520	-	CR	Clinical	1983		X		
C. sakazakii	526	-	CR	Clinical	1983		X		
C. sakazakii	531	-	Denmark	Infant formula	1988	X			
C. sakazakii	532	-	Germany	Infant formula	1988		X		
C. sakazakii	536	-	Russia	Infant formula	1988	X			
C. sakazakii	537	-	Russia	Infant formula	1988		X		
C. sakazakii	538	-	Russia	Infant formula	1988		X		
C. sakazakii	541	-	Netherlands	Infant formula	1988		X		
C. sakazakii	545	-	Netherlands	Infant formula	1988		X		
C. sakazakii	547	-	USA	Infant formula	1988	X			
C. sakazakii	548	-	Germany	Infant formula	1988		X		

C. sakazakii	551	-	Netherlands	Environment	1988		X		
C. sakazakii	552	-	Netherlands	Infant formula	1981		X		
C. sakazakii	553	-	Netherlands	Clinical	1977		X		
C. sakazakii	555	-	Netherlands	Clinical	1979		X		
C. sakazakii	557	-	Netherlands	Clinical	1979		X		
C. sakazakii	558	-	Netherlands	Clinical	1983		X		
C. sakazakii	559	-	Netherlands	Washing brush	1981		X		
C. sakazakii	560	-	Netherlands	Mouse	1988	X			
C. sakazakii	561	-	Netherlands	Infant formula	1988		X		
C. sakazakii	563	CDC 0743-75	USA	Foot wound	1975	X			
C. sakazakii	567	CDC 9363-75	USA	Faeces	1973		X		
C. sakazakii	580	NCTC 9238	UK	Clinical	1953		X		
C. sakazakii	656	BAA-893-1	USA	Unk	2001	X			
C. sakazakii	657	BAA-893-2	USA	Unk	2001	X			
C. sakazakii	658	ATCC BAA-894	USA	Non-infant formula	2001		X		X
C. sakazakii	680	CDC 996-77	USA	CSF	1977		X		X
C. sakazakii	682	-	Unknown	Unknown	Unk	X			
C. sakazakii	683	CDC 407-77	USA	Sputum	1977		X		
C. sakazakii	684	CDC 3128-77	USA	Sputum	1977	X			
C. sakazakii	686	CDC 9369-75	USA	Unk	1975		X		
C. sakazakii	690	-	France	Clinical	1994		X		
C. sakazakii	691	-	France	Clinical	1994	X			
C. sakazakii	692	-	France	Clinical	1994	X			
C. sakazakii	693	-	France	Clinical	1994		X		
C. sakazakii	694	-	France	Clinical	1994	X			
C. sakazakii	695	-	France	Clinical	1994		X		
C. sakazakii	696	-	France	Clinical	1994		X		X
C. sakazakii	698	-	France	Clinical	1994	X			
C. sakazakii	699	-	France	Clinical	1994	X			
C. sakazakii	700	-	France	Clinical	1994	X			
C. sakazakii	701	-	France	Clinical	1994		X		X
C. sakazakii	702	-	France	Clinical	1994	X			
C. sakazakii	703	-	France	Clinical	1994	X			
C. sakazakii	705	-	France	Clinical	1994	X			
C. sakazakii	706	-	France	Clinical	1994	X			
C. sakazakii	707	-	France	Clinical	1994	X			
C. sakazakii	708	-	France	Clinical	1994	X			
C. sakazakii	709	-	France	Clinical	1994		X		
C. sakazakii	711	-	France	Clinical	1994	X			
C. sakazakii	712	-	France	Infant formula	1994	X			
C. sakazakii	713	-	France	Feed	1994	X			
C. sakazakii	714	-	France	Feed	1994	X			
C. sakazakii	715	-	France	Feed	1994	X			

C. sakazakii	716	-	France	Infant formula	1994		X		
C. sakazakii	717	-	France	Infant formula	1994	X			
C. sakazakii	718	-	France	Infant formula	1994	X			
C. sakazakii	721	-	USA	CSF	2007	X			
C. sakazakii	730	-	France	Clinical	1994	X			
C. sakazakii	767	-	France	Clinical	1994		X		
C. sakazakii	890	-	Brazil	Infant formula	2007			X	
C. sakazakii	891	-	Brazil	Infant formula	2007			X	
C. sakazakii	892	-	Brazil	Infant formula	2007			X	
C. sakazakii	894	-	Brazil	Infant formula	2007			X	
C. sakazakii	896	-	Brazil	Infant formula	2007			X	
C. sakazakii	897	-	Brazil	Infant formula	2007			X	
C. sakazakii	978	-	UK	Enteral feeding tube	2007		X		
C. sakazakii	984	-	UK	Enteral feeding tube	2007		X		
C. sakazakii	1019	-	USA	Clinical	2001	X			
C. sakazakii	1105	-	UK	Weaning food	2008		X		
C. sakazakii	1106	-	UK	Weaning food	2008		X		
C. sakazakii	1107	-	UK	Weaning food	2008		X		
C. sakazakii	1108	-	UK	Weaning food	2008		X		
C. sakazakii	1131	-	Korea	Follow up formula	2008	X			
C. sakazakii	1216	CDC 88-035024	USA	Clinical	1988	X			
C. sakazakii	1217	CDC 99-03-381	USA	Clinical	1999	X			
C. sakazakii	1218	CDC 2001-10-01	USA	Clinical	2001	X			
C. sakazakii	1219	CDC 2002-01-09	USA	Clinical	2002	X			
C. sakazakii	1220	CDC 2003-16-01	USA	Clinical	2003	X			
C. sakazakii	1221	CDC 2003-16-02	USA	Clinical	2003	X			
C. sakazakii	1222	CDC 2003-16-03	USA	Clinical	2003	X			
C. sakazakii	1223	CDC 2004-01-01	USA	Clinical	2004	X			
C. sakazakii	1224	CDC 2004-27-01	USA	Clinical	2004	X			
C. sakazakii	1225	CDC 2007-21-01	USA	Clinical	2007	X			
C. sakazakii	1231	NZRM 4297	NZ	Faeces	2005	X			
C. sakazakii	1240	CDC 2009-06-01	USA	Clinical	2009	X			
C. sakazakii	1241	CDC 2009-06-02	USA	Clinical	2009	X			
C. sakazakii	1242	CDC 2009-06-03	USA	Clinical	2009	X			
C. sakazakii	1249	-	UK	Clinical	2010	X			
C. sakazakii	1283	-	UK	Food	2010	X			
C. sakazakii	1331	-	UK	Food	2010	X			
C. sakazakii	1341	-	Unk	Spice	2010	X			
C. sakazakii	1346	-	Unk	Spice	2010	X			
C. sakazakii	1347	-	Unk	Spice	2010	X			
C. sakazakii	1379	-	Unk	Herb	2010	X			
C. sakazakii	1400	-	Australia	Environment	2007	X			
C. sakazakii	1401	-	Australia	Environment	2007	X			

C. sakazakii	1402	-	Australia	Environment	2007	X			
C. sakazakii	1403	-	Australia	Environment	2007	X			
C. sakazakii	1404	-	Australia	Environment	2007	X			
C. sakazakii	1405	-	Australia	Environment	2007	X			
C. sakazakii	1434	-	Turkey	Herb	2010	X			
C. sakazakii	1436	-	Turkey	Flour	2010	X			
C. sakazakii	1437	-	Turkey	Flour	2010	X			
C. sakazakii	1438	-	Turkey	Flour	2010	X			
C. sakazakii	1465	-	Saudi Arabia	Infant formula	2011			X	
C. sakazakii	1466	-	Australia	Environment	2007			X	
C. sakazakii	1467	-	Australia	Environment	2007			X	
C. sakazakii	1470	-	Australia	Environment	2007			X	
C. sakazakii	1471	-	Australia	Environment	2007			X	
C. sakazakii	1472	-	Australia	Environment	2007			X	
C. sakazakii	1473	-	Australia	Environment	2007			X	
C. sakazakii	1474	-	Australia	Environment	2007			X	
C. sakazakii	1475	-	Australia	Environment	2007			X	
C. sakazakii	1476	-	Australia	Environment	2007			X	
C. sakazakii	1477	-	Australia	Environment	2007			X	
C. sakazakii	1479	-	Australia	Environment	2007			X	
C. sakazakii	1480	-	Australia	Environment	2007			X	
C. sakazakii	1481	-	Australia	Environment	2007			X	
C. sakazakii	1482	-	Australia	Environment	2007			X	
C. sakazakii	1483	-	Australia	Environment	2007			X	
C. sakazakii	1484	-	Australia	Environment	2007			X	
C. sakazakii	1485	-	Australia	Environment	2007			X	
C. sakazakii	1486	-	Australia	Environment	2007			X	
C. sakazakii	1487	-	Australia	Environment	2007			X	
C. sakazakii	1488	-	Australia	Environment	2007			X	
C. sakazakii	1489	-	Australia	Environment	2007			X	
C. sakazakii	1490	-	Australia	Environment	2007			X	
C. sakazakii	1491	-	Australia	Environment	2007			X	
C. sakazakii	1492	-	Australia	Environment	2007			X	
C. sakazakii	1493	-	Australia	Environment	2007			X	
C. sakazakii	1494	-	Australia	Environment	2007			X	
C. sakazakii	1495	-	Australia	Environment	2007			X	
C. sakazakii	1496	-	Australia	Environment	2007			X	
C. sakazakii	1497	-	Australia	Environment	2007			X	
C. sakazakii	1498	-	Australia	Environment	2007			X	
C. sakazakii	1499	-	Australia	Environment	2007			X	
C. sakazakii	1500	-	Australia	Environment	2007			X	
C. sakazakii	1501	-	Australia	Environment	2007			X	

C. sakazakii	1502	-	Australia	Environment	2007			X	
C. sakazakii	1503	-	Australia	Environment	2007			X	
C. sakazakii	1504	-	Australia	Environment	2007			X	
C. sakazakii	1505	-	Australia	Environment	2007			X	
C. sakazakii	1506	-	Australia	Environment	2007			X	
C. sakazakii	1507	-	Australia	Environment	2007			X	
C. sakazakii	1508	-	Australia	Environment	2007			X	
C. sakazakii	1529	-	Germany	Environment	2006			X	
C. sakazakii	1530	-	Germany	Environment	2006			X	
C. sakazakii	1531	-	Germany	Environment	2006			X	
C. sakazakii	1532	-	Germany	Environment	2006			X	
C. sakazakii	1533	-	Germany	Environment	2006			X	
C. sakazakii	1534	-	Germany	Environment	2006			X	
C. sakazakii	1535	-	Germany	Environment	2006			X	
C. sakazakii	1536	-	Germany	Environment	2009			X	
C. sakazakii	1537	-	Germany	Environment	2009			X	
C. sakazakii	1538	-	Germany	Environment	2009			X	
C. sakazakii	1539	-	Germany	Unknown	Unk			X	
C. sakazakii	1540	-	Germany	Environment	2009			X	
C. sakazakii	1542	-	Germany	Environment	2009			X	
C. sakazakii	1547	CCM 3460 EB5	CR	Unk	Unk			X	
C. sakazakii	1549	CCM 3479 EB7	CR	Unk	Unk			X	
C. sakazakii	1555	CDC 947-79	USA	Foot wound	1979			X	
C. sakazakii	1557	H Wisglass P-335	Unk	Clinical	1979			X	
C. sakazakii	1559	DBM 3153 148	Portugal	Herb	Unk			X	
C. sakazakii	1565	CDC 2011-12-01	USA	Clinical	2011			X	
C. sakazakii	1566	CDC 2011-12-02	USA	Clinical	2011			X	
C. sakazakii	1567	CDC 2011-12-03	USA	Clinical	2011			X	
C. sakazakii	1568	CDC 2011-12-04	USA	Infant formula	2011			X	
C. sakazakii	1570	CDC 2011-21-01	USA	Clinical	2011			X	
C. sakazakii	1571	CDC 2011-21-03-01	USA	Infant formula	2011			X	
C. sakazakii	1572	CDC 2011-21-03-02	USA	Infant formula	2011			X	
C. sakazakii	1573	CDC 2011-18-05-02	USA	Infant formula	2011			X	
C. sakazakii	1574	CDC 2011-18-01	USA	Clinical	2011			X	
C. sakazakii	1575	CDC 2011-18-07	USA	Clinical	2011			X	
C. sakazakii	1576	CDC 2193-02	USA	Clinical	2011			X	
C. sakazakii	1577	CDC 2193-03	USA	Clinical	2011			X	
C. sakazakii	1578	CDC 2193-08-01	USA	Water	2011			X	
C. sakazakii	1579	CDC 2012-05-05	USA	Clinical	2011			X	
C. sakazakii	1562	-	Slovakia	Chocolate	2010			X	

C. sakazakii	1563	-	Slovakia	Chocolate	2010			X	
C. sakazakii	1564	-	Slovakia	Chocolate	2010			X	
C. sakazakii	BQ1	-	China	Food	2010			X	
C. sakazakii	BQ10	-	China	Food	2010			X	
C. sakazakii	BQ11	-	China	Food	2010			X	
C. sakazakii	BQ12	-	China	Food	2010			X	
C. sakazakii	BQ13	-	China	Food	2010			X	
C. sakazakii	BQ14	-	China	Food	2010			X	
C. sakazakii	BQ15	-	China	Food	2010			X	
C. sakazakii	BQ16	-	China	Food	2010			X	
C. sakazakii	BQ19	-	China	Food	2010			X	
C. sakazakii	BQ2	-	China	Food	2010			X	
C. sakazakii	BQ3	-	China	Food	2010			X	
C. sakazakii	BQ4	-	China	Food	2010			X	
C. sakazakii	BQ6	-	China	Food	2010			X	
C. sakazakii	BQ7	-	China	Food	2010			X	
C. sakazakii	BQ9	-	China	Food	2010			X	
C. sakazakii	DA01	-	Bangladesh	Infant formula	2010			X	
C. sakazakii	E899	-	USA	Clinical	1981			X	X
C. malonaticus	8	-	CR	Weaning food	2004		X		
C. malonaticus	15	-	CR	Faeces	2003	X			
C. malonaticus	18	-	CR	Clinical	2004		X		
C. malonaticus	21	-	CR	Faeces	2003	X			
C. malonaticus	22	-	France	Infant formula	Unk	X			
C. malonaticus	35	-	USA	Weaning food	2005		X		
C. malonaticus	86	-	China	Herb	2004	X			
C. malonaticus	90	-	China	Herb	2005		X		
C. malonaticus	93	-	Unk	Spice	2005	X			
C. malonaticus	101	-	Unk	Spice	2005	X			
C. malonaticus	156	-	China	Herb	2005		X		
C. malonaticus	472	-	Korea	Infant formula	2005	X			
C. malonaticus	507	-	CR	Clinical	1984		X		X
C. malonaticus	510	-	CR	Food	1985		X		
C. malonaticus	512	-	CR	Clinical	1983		X		
C. malonaticus	514	-	CR	Clinical	1983	X			
C. malonaticus	515	-	CR	Clinical	1983		X		
C. malonaticus	521	-	CR	Clinical	1983		X		
C. malonaticus	522	-	CR	Clinical	1983		X		
C. malonaticus	524	-	CR	Clinical	1983		X		
C. malonaticus	527	-	Australia	Infant formula	1988	X			
C. malonaticus	535	-	NZ	Infant formula	1988		X		
C. malonaticus	565	CDC 1895-73	USA	Faeces	1973		X		

C. malonaticus	681	LMG 23826 ^{T}	USA	Clinical	1977		X		X
C. malonaticus	685	CDC 1716-77	USA	Clinical	1977	X			
C. malonaticus	687	-	CR	Sputum	2004	X			
C. malonaticus	688	-	CR	Sputum	2004	X			
C. malonaticus	689	-	CR	Faeces	2005	X			
C. malonaticus	893	-	Brazil	Infant formula	2007			X	
C. malonaticus	1369	-	UK	Herb	2010	X			
C. malonaticus	1432	-	Turkey	Flour	2010	X			
C. malonaticus	1433	-	Turkey	Flour	2010	X			
C. malonaticus	1514	-	Australia	Environment	2007			X	
C. malonaticus	1545	CNCTC 81 6830 EB1	CR	Clinical	Unk			X	
C. malonaticus	1546	CNCTC 83 6831 EB2	CR	Clinical	Unk			X	
C. malonaticus	1558	CNCTC 81 6830 EB2	CR	Clinical	Unk			X	
C. malonaticus	1569	CDC 2193-01	USA	Clinical	2011			X	
C. malonaticus	BQ18	-	China	Food	2010			X	
C. malonaticus	BQ5	-	China	Food	2010			X	
C. turicensis	9	-	UK	Weaning food	2003	X			
C. turicensis	57	-	UK	Milk powder	2004	X			
C. turicensis	92	-	UK	Herb	2004	X			
C. turicensis	109	-	Unk	Herb	2004	X			
C. turicensis	111	-	UK	Herb	2004	X			
C. turicensis	564	CDC 5960-70	USA	Blood	1970	X			X
C. turicensis	566	CDC 3523-75	USA	Bone marrow	1975	X			
C. turicensis	1211	LMG 23827 ^{T}	Switzerland	Blood	2005	X			X
C. turicensis	1303	-	Unk	Unknown	Unk	X			
C. turicensis	1313	-	Unk	Spice	Unk	X			
C. turicensis	1468	-	Australia	Environment	2007			X	
C. turicensis	1478	-	Australia	Environment	2007			X	
C. turicensis	1511	-	Australia	Environment	2007			X	
C. turicensis	1512	-	Australia	Environment	2007			X	
C. turicensis	1513	-	Australia	Environment	2007			X	
C. turicensis	1553	-	Slovakia	Unknown	Unk			X	
C. turicensis	1554	-	Slovakia	Unknown	Unk			X	
C. turicensis	BQ17	-	China	Food	2010			X	
C. turicensis	FN543093	-	Switzerland	Clinical	2005			X	
C. universalis	96	-	UK	Spice	2004	X			
C. universalis	581	NCTC 9529 ^{T}	UK	Freshwater	1956	X			X
C. universalis	731	-	France	Leg infection	2005	X			

C. universalis	1435	-	Turkey	Flour	2010	X			
C. muytjensii	3	-	Unk	Unk	-	X			
C. muytjensii	16	-	Unk	Spice	2005	X			
C. muytjensii	378	-	Unk	Unk	-	X			
C. muytjensii	530	-	Denmark	Infant formula	1988	X			X
C. muytjensii	1025	-	UK	Unk	Unk	X			
C. muytjensii	1026	-	UK	Unk	Unk	X			
C. muytjensii	1129	-	Korea	Follow up formula	2008	X			
C. muytjensii	1229	-	Jordan	Infant formula	2008	X			
C. muytjensii	1371	-	Unk	Spice	2010	X			
C. muytjensii	1522	-	UK	Food	2011	X			
C. muytjensii	1527	-	Slovakia	Food	2010	X			
C. muytjensii	1550	CNCTC EB55	USA	Unk	Unk	X			
C. muytjensii	1552	-	Unk	Unk	Unk	X			
C. muytjensii	1561	ATCC 51329 ^{T}	USA	Unk	Unk	X			
C. muytjensii	CM	-	China	Unk	2010			X	
C. muytjensii	NTUCC825	-	Unk	Unk	Unk	X			
C. muytjensii	NTUCC868	-	Unk	Unk	Unk	X			
C. dublinensis	84	-	China	Herb	2004	X			
C. dublinensis	582	NCTC 9844	UK	Unknown	Unk	X			X
C. dublinensis	583	NCTC 9846	UK	Dairy	1956	X			
C. dublinensis	1130	-	Korea	Follow up formula	2008	X			
C. dublinensis	1132	-	Korea	Follow up formula	2008	X			
C. dublinensis	1133	-	Korea	Follow up formula	2008	X			
C. dublinensis	1134	-	Korea	Follow up formula	2008	X			
C. dublinensis	1210	LMG 23823 ^{T}	Ireland	Environment	2007	X			X
C. dublinensis	1458	-	Korea	Food	2011	X			
C. dublinensis	1460	-	Korea	Food	2011	X			
C. dublinensis	1461	-	Korea	Food	2011	X			
C. dublinensis	1462	-	Korea	Food	2011	X			
C. dublinensis	1463	-	Korea	Food	2011	X			
C. dublinensis	1464	-	Korea	Food	2011	X			
C. dublinensis	1543	-	Unk	Milk powder	2003	X			
C. dublinensis	1548	CCM 3461 EB6	Unk	Unk	Unk	X			
C. dublinensis	1556	CDC 1565-79	USA	Clinical	1979	X			
C. dublinensis	1560	DBM 3154 EB12	CR	Food	Unk	X			
C. dublinensis	BQ8	-	China	Food	2010			X	
C. condimenti	1330	LMG 26250^{T}	Slovakia	Food	2010	X			X

Footnote:

A – MLST laboratory experiments and sequence data analysis performed as part of this project; **B** – MLST laboratory experiments performed by Baldwin *et al*. 2009, sequence data analysed as part of this project; **C** - MLST laboratory experiments performed by collaborators/research group colleagues, sequence data analysed as part of this project; **D** – Genome sequenced strains (Chapters 6 & 7); ^{T} – Species type strain

CR: Czech Republic; NZ: New Zealand; UK: United Kingdom; USA: United States of America;
Unk: Unknown

2.3 Preparation of media

2.3.1 Trypticase Soy Agar (TSA)

Twenty grams of TSA (Oxoid, UK) base were resuspended in 500 ml of distilled water and autoclaved at 121^{o}C for 15 minutes. The agar was then allowed to cool to 50^{o}C and approximately 15-20 ml each was poured into sterile Petri dishes. The plates were stored at 4^{o}C for a maximum of three weeks.

2.3.2 Trypticase Soy Broth (TSB)

Fifteen grams of TSB (Oxoid, UK) base were resuspended in 500 ml of distilled water and autoclaved at 121^{o}C for 15 minutes. The broth was stored at 4^{o}C until required.

2.3.3 Druggan-Forsythe-Iversen (DFI) agar formulation

Twenty one point five grams of DFI agar base (Lab M, UK) were resuspended in 500 ml of distilled water and autoclaved at 121^{o}C for 15 minutes. The agar was then allowed to cool to 50^{o}C and approximately 15-20 ml each was poured into sterile Petri dishes. The plates were stored at 4^{o}C for a maximum of three weeks.

2.4 MLST and analysis

2.4.1 Choice of bacterial strains for MLST

This study included 325 *Cronobacter* spp. strains (Table 2.1), which included 226 *C. sakazakii* strains, 39 *C. malonaticus* strains, 19 *C. turicensis* strains, 19 *C. dublinensis* strains, 17 *C. muytjensii* strains, 4 *C. universalis* strains, and the single isolate of the recently defined species *C. condimenti*. All the strains were either selected from the NTU culture collection of *Cronobacter* spp., or recently obtained from collaborators following a review of the literature. Most of the latter strains had only been identified as *Enterobacter sakazakii* or *Cronobacter* spp. in previous publications. Strains were chosen on the basis of their species, geographic and temporal distribution, source and clinical outcome. This included the type strains for each species, infant formula and clinical isolates from across the world, ranging from 1950 to date, as well as strains from the US which caused high public concern in December 2011 (CDC 2011). As part of the diversity study, the 16S rDNA sequences submitted as *Cronobacter* spp. by third parties to Genbank were also analysed, and any strains of relevant interest were requested from the sequence submitters.

Of these 325 strains, the MLST PCR and sequencing experiments, as well as sequence analysis were performed for 132 of them over the course of the project by myself for establishing the scheme across the genus. Apart from these, there were 75 *C. sakazakii* and *C. malonaticus* strains for which the MLST PCR and sequencing had been performed earlier by Baldwin *et al*. (2009). The sequence data generated from the study has also been included in the further detailed sequence analysis carried out in this project. Once the scheme was successfully working, other members of the research group as well as collaborators from across the world have performed the MLST laboratory experiments on their isolates and submitted the sequence data to the PubMLST database. This sequence data was quality checked and included for the sequence analysis in this project, in order to investigate the diversity of the genus in detail. This breakdown of the strains used has been clearly indicated in Table 2.1. The online *Cronobacter* PubMLST database is now constantly being updated and the analysis and results reported here are based on sequences submitted into the database until March 2012.

The list and details of all the isolates to date, along with their MLST profile details are freely available online at http://www.pubmlst.org/cronobacter, which is a public database of the *Cronobacter* MLST scheme, hosted by the University of Oxford.

The genomes of *Citrobacter koseri* BAA-895 (accession number CP000822) and *Enterobacter* spp. 683 (accession number CP000653) were used as out groups, as previously described (Baldwin *et al*. 2009).

2.4.2 Genomic DNA extraction

Genomic DNA was prepared from 1.5 ml of overnight culture grown in TSB using GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich, UK) as per the manufacturer's instructions. The concentration and purity of the eluted DNA was checked on a Nanodrop 2000 (Thermo Scientific, UK). Only DNA samples with minimum (260/280) nm values of 1.8 and (260/230) nm values of 2 were used for the PCR experiments, else DNA extraction was repeated.

2.4.3 PCR amplification of the MLST loci

Amplification and nested sequencing primers for the MLST loci were as described by Baldwin *et al*. (2009) (Table 2.2). Reaction conditions for all the primers were as follows: initial denaturation at 94°C for 2 min; 30 cycles of denaturation at 94°C for 1 min, primer annealing at 58°C for 1 min, extension at 72°C for 2 min; followed by a final extension step of 72°C for 5 min. Each 25 µl amplification reaction mixture comprised ~10 ng chromosomal DNA, 20 pmol forward and reverse primer (Sigma-Aldrich, UK), 1× PCR buffer (Promega, UK) containing 1.5 mM $MgCl_2$, 0.8 mM deoxynucleotide triphosphates and 1.25 U *Taq* DNA polymerase (Promega, UK).

Primers designed to amplify the *gltB* and *ppsA* loci failed in less than 5% of cases (mainly *C. muytjensii*). Initially, alternative primers were designed using Primer-BLAST (Ye *et al*. 2012), for the amplification of these loci, with an attempt to keep the annealing temperature the same as the other amplification primers. However, when this resulted in a number of non-specific products, then the sequencing primers were used directly for the PCR reactions and sequencing of these problematic strains. A complete set of data was successfully obtained for all strains using this approach.

2.4.4 PCR amplification of the 16S rDNA loci.

PCR amplifications for the partial sequencing of the 16S rDNA loci (528 bp) were performed using primers described by Iversen *et al*. (2004) (5'-TGGAGAGTTTGATCCTGGCTCAG – 3' and 5' – TACCGCGGCTGCTGGCAC – 3'). Reaction conditions for the PCR reactions were as follows: initial denaturation at 95°C for 10 min; 30 cycles each of denaturation at 95°C for 30 s, primer annealing at 60°C for 30 s, and extension at 72°C for 45 s; and a final extension step at 72°C for 10 min.

PCR amplifications for the full length sequencing of the 16S rDNA loci (1300 bp) were performed using the oligonucleotide primers described by Iversen, *et al*. 2007 (5'-AGAGTTTGATCCTGGCTCAG-3' and 5'-GTACGGCTACCTTGTTACGA-3'). Reaction conditions for the PCR reactions were as follows: initial denaturation at 95°C for 3 min; 30 cycles each of denaturation at 95°C for 30 s, primer annealing at 56°C for 30 s, and extension at 72°C for 2 min; and a final extension step at 72°C for 5 min. Each reaction was performed in volumes of 25 µl. The same primers were used for the sequencing as well.

2.4.5 Agarose gel electrophoresis

The amplified PCR products were checked by viewing on a 1% agarose gel. For preparing the gel, 1% (w/v) agarose (Fisher Scientific, UK) was mixed in 1X Tris-acetate-EDTA (TAE) buffer (Geneflow, UK) and heated in a microwave to completely dissolve the agarose. 0.1 µl/ml (v/v) of SYBR® Safe DNA gel stain (Life Technologies – Invitrogen, UK) was then added, dissolved well; after which the gel was poured in a tray and allowed to set. 5 µl aliquots of the PCR samples were loaded into each well along with 5 µl of a 1 kb DNA ladder (Promega, UK) used as a marker, and the gel was run for 40 minutes at 100V in 1X TAE buffer. The gel was then viewed under ultraviolet (UV) light to observe the DNA bands, using the InGenius® gel documentation system (Syngene; UK).

2.4.6 PCR product purification

The amplified products were then purified using MinElute PCR Purification Kits (Qiagen, UK) following the manufacturer's protocol. The final elution step was carried out in 50 µl of molecular biology grade water (Fisher Scientific, UK), after which the concentration and purity of the samples were checked using the Nanodrop 2000 (Thermo Scientific, UK).

2.4.7 DNA Sequencing.

The sequencing was performed using the purified PCR products (10 ng/µl) on ABI 3730XL sequencing machines by Eurofins MWG Operon (London, UK) and Source Bioscience (Nottingham, UK). Using the nested sequencing primers, nucleotide sequences were determined at least once on each DNA strand.

Target	Locus [a]	Putative function		Outer primers (5'-3')	Inner Primers (5'-3')
atpD	ESA_04006	ATP synthase β chain	Forward	CGACATGAAAGGCGACAT	CGAAATGACCGACTCCAA
			Reverse	TTAAAGCCACGGATGGTG	GGATGGCGATGATGTCTT
fusA	ESA_04401	Elongation factor	Forward	GAAACCGTATGGCGTCAG	GCTGGATGCGGTAATTGA
			Reverse	AGAACCGAAGTGCAGACG	CCCATACCAGCGATGATG
glnS	ESA_02658	Glutaminyl tRNA-synthetase	Forward	GCATCTACCCGATGTACG	GGGTGCTGGATAACATCA
			Reverse	TTGGCACGCTGAACAGAC	CTTGTTGGCTTCTTCACG
gltB	ESA_03606	Glutamate synthase large subunit	Forward	CATCTCGACCATCGCTTC	GCGAATACCACGCCTACA
			Reverse	CAGCACTTCCACCAGCTC	GCGTATTTCACGGAGGAG
gyrB	ESA_03973	DNA gyrase β subunit	Forward	TGCACCACATGGTATTCG	CTCGCGGGTCACTGTAAA
			Reverse	CACCGGTCACAAACTCGT	ACGCCGATACCGTCTTTT
infB	ESA_03561	Translation initiation factor IF-2	Forward	GAAGAAGCGGTAATGAGC	TGACCACGGTAAAACCTC
			Reverse	CGATACCACATTCCATGC	GGACCACGACCTTTATCC
ppsA	ESA_02102	Phosphoenol pyruvate synthase	Forward	GTCCAACAATGGCTCGTC	ACCCTGACGAATTCTACG
			Reverse	CAGACTCAGCCAGGTTTG	CAGATCCGGCATGGTATC

Table 2.2 Details of the seven MLST loci and the primers used for their amplification and sequencing

[a] – Genbank locus of the gene on the *C. sakazakii* BAA-894 genome (Kucerova *et al*. 2010)

2.4.8 Allele and Sequence Type designation

The sequence chromatograms were viewed using ChromasLite (Version 2.01; Technelysium Pty Ltd) for quality control and the sequences from both strands of a given locus of the same isolate were aligned and trimmed to the desired allele length using Jalview (Version 2+; University of Dundee) (Waterhouse *et al*. 2009) in ClustalW (http://www.ebi.ac.uk/clustalw) (Larkin *et al*. 2007). Newly identified alleles for each locus were assigned numbers arbitrarily according to the order in the PubMLST database. All alleles within the MLST scheme were in frame, to aid with analysis.

2.4.9 Analysis of DNA sequences

Phylogenetic analyses of the concatenated sequences of the 7 loci (3036 nucleotides) were performed using the Maximum-Likelihood algorithm (Felsenstein. 1981) in Molecular Evolutionary Genetics Analysis (MEGA) Version 5 software (Tamura *et al*. 2011). The various alleles identified for each of the 7 loci were analysed using the Sequence Type Analysis & Recombinational Tests (START) Version 2 software (Jolley *et al*. 2001).

2.4.10 Evolutionary timescale estimations

Mean synonymous substitution values (D_s) were calculated for the individual *Cronobacter* species using the 3036 bp concatenated sequences in MEGA5 (Tamura *et al*. 2011) as well as START2 (Jolley *et al*. 2001). These values were used to calculate the ages of the genus and species based on the formula:

Age = D_s / rate, where rate is the synonymous molecular clock rate.

Since this is based on an estimate, two sets of dates were calculated to obtain an age range, by using the substitution rate of 3×10^{-9} for *E. coli* and *Salmonella* Typhimurium published by Berg and Martelius (1995) and a slower rate of 4.5×10^{-9} calculated for *E. coli* by Ochman *et al*. (1999).

2.4.11 Recombination analysis.

START (Version 2; University of Oxford) was also used for calculating the d_N/d_S ratios (ratio of non-synonymous substitutions to synonymous substitutions) as well as recombination testing (Jolley *et al*. 2001). Further study of recombinational events in the populations was carried out using the Neighbour-N*et al*gorithm in Splitstree (Version 4; Universitat Tubingen)

(Huson & Bryant. 2006) and the Recombination Analysis Tool (RAT) (Etherington *et al*. 2004). Phi test was performed to estimate the recombination using the Splitstree software.

2.4.12 PCR screening for serotype identification

PCR primers designed for the identification of *C. sakazakii* serotypes by Mullane *et al*. (2008) and Jarvis *et al*. (2011) were used for screening a selection of clinical *C. sakazakii* strains to investigate the correlation of serotypes with sequence types. These primers targeted three different genes:

(1) *wehC* (5' – CACGTTCGCCCTGCAAAAAT – 3' & 5' – GCAAGCGGCCAGACTGGATA – 3')

(2) *wehI* (5' – TCCTGCATTTGTGGATTTTGC – 3' & 5' – AACGCATTGCGCTTGAGAAA – 3')

(3) *wzx* (5' – TGGCTGTCATGGTTTTCTTGC – 3' & 5' – TAGTTGGCACCATCAACGCC – 3').

The PCR cycling conditions were: denaturation at 95^{o}C for 2 min, followed by 25 cycles of 95^{o}C for 30 sec, 55^{o}C for 30 sec, 72^{o}C for 1 min, and a final extension step at 72^{o}C for 5 min.

2.4.13 Inter-species nucleotide divergence analysis

The nucleotide divergence values between the sequence types of the type strains of the individual *Cronobacter* species, as well those of the individual members of the *C. muytjensii* and *C. dublinensis* species, were calculated using DnaSP (Librado & Rozas. 2009).

2.4.14 Linkage and clonality

START2 (Jolley *et al*. 2001) was used for calculating the Index of Association (I_A) of the various *Cronobacter* species for linkage analysis of the population. The clonality of the population was analysed using the goeBURST algorithm (Francisco *et al*. 2009) in Phyloviz (Francisco *et al*. 2012). Statistical chi-square analysis for testing associations was carried out using the CHITEST function in Microsoft Excel.

2.5 Whole genome sequencing and comparative genomic analysis

2.5.1 Choice of bacterial strains for genome sequencing

The eleven *Cronobacter* strains which were chosen for this whole genome sequencing study were selected from the NTU *Cronobacter* culture collection based on their species identification by MLST (Chapter 3), sequence type and clinical significance. The details of the strains that were sequenced in this study are listed in Table 2.3.

Apart from these, the study also included two other *Cronobacter* genomes, publicly available and downloaded from Genbank – *C. sakazakii* BAA-894 (Accession Nos. NC_009778 - NC_009780), *C. turicensis* z3032 (Accession Nos. NC_013282-NC_013285). Towards the completion of this study, the whole genome shotgun sequencing project of *C. sakazakii* E899 (Accession No. AFMO01000001-AFMO01000385) became publicly available. This genome has also been incorporated into this study wherever appropriate.

2.5.2 Genomic DNA extraction

The genomic DNA for each of the eleven *Cronobacter* strains was prepared from 1.5 ml of overnight bacterial culture grown in TSB using the Qiagen DNeasy Blood and Tissue DNA Isolation Kit (Qiagen, UK), according to manufacturer's instructions. Quality control checks were performed by running the DNA samples on a 1.5% agarose gel to ensure that the DNA had not sheared during the extraction process. The concentration and purity of the eluted DNA was also checked on a Nanodrop 2000 (Thermo Scientific, UK). Only DNA samples with minimum (260/280) nm values of 1.8 and (260/230) nm values of 2 were sent to Life Technologies (Foster City, CA) for the library preparation and genome sequencing.

Species	Strain ID	Other ID	Source	Country	ST
C. sakazakii	680		Clinical	USA	8
	696		Clinical	France	12
	701		Clinical	France	4
	BAA-894[a]		Powdered formula	USA	1
	E899[a]		Clinical	USA	4
C. malonaticus	507		Clinical	Czech Republic	11
	681	LMG 23826^T	Clinical	USA	7
C. turicensis	564		Clinical	USA	5
	z3032[a]	LMG 23827^T	Clinical	Switzerland	19
C. universalis	581	NCTC 9529^T	Environment	UK	54
C. muytjensii	530		Infant formula	Denmark	49
C. dublinensis	582		Unknown	UK	36
	1210	LMG 23823^T	Environment	Ireland	106
C. condimenti	1330	LMG 26250^T	Food	Slovakia	98

Table 2.3 Details of the eleven newly assembled and three publicly available *Cronobacter* spp. genomes analyzed in this study.

a – Genomes publicly available on Genbank; T – Species type strain

2.5.3 Genome sequencing

The library preparation, genome sequencing and assembly were performed by the research group at Life Technologies, Foster City, CA (SOLiD® 4) and Life Technologies, Darmstadt, Germany (Ion torrent PGM™).

The genomes of all the *Cronobacter* strains except 680, 1210^T and 1330^T were sequenced using the SOLiD®4 system (Life Technologies, Carlsbad, CA). Long mate-paired genomic DNA libraries with approximately 1.8 kb inserts were sequenced to generate 23-36 millions of 2 x 50bp reads for each strain, approximating 500-800 fold coverage of the genome. The colorspace reads were error-corrected and then assembled *de novo* into contigs and scaffolds using the Velvet assembly engine (Zerbino & Birney 2008).

Strains 680, 1210^T and 1330^T were sequenced using Ion Torrent PGM™ system (Life Technologies, Darmstadt, Germany). Fragment library preparation was performed with the Ion Fragment Library Kit (Life Technologies, Darmstadt, Germany). Template preparation was carried out with the Ion Xpress™ Template Kit (Life Technologies). The Ion Sequencing Kit (Life Technologies) was used with the Personal Genome Machine™ (PGM™) sequencer. A single sequencing run (65 cycles) was performed on an Ion 316™ chip for each library. Contigs were assembled from fragment reads using the MIRA 3 assembler [http://sourceforge.net/apps/mediawiki/mira-assembler/index.php].

2.5.4 Genome assembly

The assembled genome scaffolds were aligned to the most closely related publicly available genomes using MUMmer (Kurtz *et al*. 2004). The scaffolds of NTU strains *C. sakazakii* 680, 696, 701 and *C. malonaticus* 507, and 681^T were aligned to the *C. sakazakii* BAA-894 complete genome (accession numbers NC_009778 – NC_009780). The scaffolds of NTU strains *C. turicensis* 564, *C. dublinensis* 582, 1210^T, *C. muytjensii* 530, *C. universalis* 581^T and *C. condimenti* 1330^T were aligned to the *C. turicensis* z3032 complete genome (accession numbers NC_013282-NC_013285). Scaffolds were broken at points where non-contiguous regions of the reference genome were juxtaposed and then re-ordered so that they were syntenic with the reference genome. All scaffolds from a given strain were concatenated into a single draft genome, separated by the sequence, NNNNCACACACTTAATTAATTAAGTGTGTGNNNNN, which contains stop codons in all six reading frames. Scaffolds that did not match the reference genomes were concatenated in random orders at the end of the genome. The sequences of the seven MLST genes of each of the strains were also used as a template assembly to further confirm the ID of each strain. The draft

genomes were then submitted and annotated using the RAST automated annotation server (Aziz *et al*. 2008) at http://rast.nmpdr.org.

The draft genome sequences of the eleven newly sequenced strains have now been deposited to the EBI nucleotide database (see Table 6.1 for accession numbers).

2.5.5 Core and pan-genome identification

The core and pan-genome analysis was performed by Dr. Prerak Desai (Vaccine Research Institute, San Diego, CA). Orthologous and paralogous gene families were initially constructed based on the RAST annotations of the genomes and an 'all-against-all' tblatx matrix was constructed using BLAT (Kent 2002). The blat matrix was used as an input to construct orthologous/paralogous gene families using OrthoMCL (Li *et al*. 2003). The phyletic table generated from the tblatx analysis was consolidated with the phyletic table generated from OrthoMCL analysis to compute a comprehensive pan genome matrix. This phyletic matrix was used as an input for the binomial mixture model software of Snipen *et al*. (2009) to determine the *Cronobacter* core genome and accessory genome values.

2.5.6 Whole genome alignments and phylogenetic analysis

The fourteen *Cronobacter* genomes were aligned using the multiple alignment tool Mugsy (Angiuoli & Salzberg 2010). Mugsy aligns the bacterial chromosomes as chains of syntenic blocks, independent of a reference genome. This alignment was then used for a phylogenetic analysis using the software RaxML (Stamatakis 2006) in the graphical interface raxMLgui (Silvestro and Michalak 2010) to generate the best possible evolutionary tree based on the maximum-likelihood method (Felsenstein 1981). The settings used were ML+rapid bootstrap, 1000 bootstrap replicates and model GTR+GAMMA. The tree was viewed and annotated using TreeDyn (Chevenet *et al*. 2006).

A visual circular alignment was also generated for the fourteen genomes using the Blast Ring Image Generator (BRIG) software package (Alikhan *et al*. 2011) using the *C. sakazakii* BAA-894 genomes (the chromosomes and plasmids were concatenated together) as the reference sequence. The alignments and genome comparisons are performed based on the whole genomes Basic Local Alignment Search Tool (BLAST) algorithm against the reference sequence.

2.5.7 Prophage region identification

Identification and characterisation of the bacteriophage regions present in the fourteen genomes was carried out using ACLAME Prophinder, available at http://aclame.ulb.ac.be/Tools/Prophinder/ (Lima-Mendez *et al*. 2008) and the web server PHAST (Phage Search Tool) available at http://phast.wishartlab.com/ (Zhou *et al*. 2011). The results obtained from both the tools were compiled and compared to make final conclusions about the phage regions in the genomes. Any genes of particular interest were further investigated using the tblastx search options.

2.5.8 Plasmid region identification

The plasmid content of the *Cronobacter* genomes sequenced in this study was predicted by mapping the known plasmids from the published genomes of *C. sakazakii* BAA-894 (pESA2 and pESA3) and *C. turicensis* z3032 (pCTU1, pCTU2, pCTU3) to the draft assemblies. Comparative alignments of all the fourteen genomes using the larger publicly available plasmids (pESA3 and pCTU1) as backbones were performed using BRIG as described above. The plasmid regions were also visualised to identify the presence/absence of genes by using the Artemis genome browser (Rutherford *et al*. 2000) and Artemis Comparison Tool (ACT) (Carver *et al*. 2005).

2.5.9 Plasmid profiling

The plasmid profiling of the thirteen *Cronobacter* spp. strains was carried out according to an alkaline extraction procedure published by Birnboim and Doly (1979).

Reagents: Solution I: Lysozyme solution – 2 mg/ml lysozyme (Fluka), 50 mM glucose (Sigma, UK), 10mM CDTA/EDTA (Sigma, UK), 25 mM Tris-HCl pH 8.0 (Sigma, UK). Stock solution was prepared fresh and stored at 0^{o}C; Solution II: Alkaline SDS solution – 0.2 N NaOH, 1% SDS (Sigma, UK). Solution was stored at room temperature and stable for a week; Solution III: High salt solution –3 moles of sodium acetate (Sigma, UK) was dissolved in a minimal volume of water, adjusted to pH 4.8 with glacial acetic acid, and volume adjusted to 1 litre.

Bacterial culture (2.5 ml) was grown overnight with shaking for 18 hours in TSB. An aliquot (0.5 ml) of the culture was transferred to a 1.5 ml eppendorf tube for plasmid extraction, while the remainder was stored at -20^{o}C after addition of glycerol to 40%. The Eppendorf tube was centrifuged for 15 seconds and the supernatant carefully removed with a fine tip aspirator. The cell pellet was then thoroughly suspended in 100 µl of Solution I. After a 30 min period of

incubation at 0°C, 200 µl of solution II was added and the tube was gently vortexed, until the suspension became almost clear and slightly viscous. The tube was then maintained for 5 min at 0°C and then 150 µl of Solution III was added. The tube was gently mixed by inversion for a few seconds to clot the DNA and maintained at 0°C for 60 min to allow protein, high molecular weight RNA and chromosomal DNA to precipitate. The tube was then centrifugated for 5 min to yield an almost clear supernatant, 400 µl of which was removed and transferred to a second tube. Cold ethanol (1 ml) was added and tube was held at -20°C for 30 min, followed by centrifugation for 2 mins to remove and discard supernatant by aspiration. The pellet was then dissolved in 100 µl of 0.1 M sodium acetate/0.05 M Tris-HCl (pH 8) and reprecipitated with 2 volumes of cold ethanol. After 10 min at -20°C, precipitate was again collected by centrifugation as before. The pellet was then dissolved in 40 µl of water and 10 µl of 5X sample buffer [25% sucrose (Fisher Scientific, UK), 5 mM sodium acetate (Sigma, UK), 0.1% SDS (Sigma, UK)] was added to it. 20 µl of the sample was then loaded in an unstained 1% agarose (Fisher Scientific, UK) gel prepared with 1X TAE buffer (Geneflow, UK) and was run for 50 minutes at 90V in 1X TAE buffer. The gel was then stained in 0.0625 µl/ml Ethidium bromide solution - 25 µl EtBr (Sigma, UK) in 400 ml distilled water - for 1 hour and then viewed under UV light to observe the DNA bands, using the InGenius® gel documentation system (Syngene; UK).

2.5.10 CRISPR region identification

The fourteen *Cronobacter* genomes were analysed for the presence of "Clustered Regularly Interspaced Short Palindromic Repeat" (CRISPR) regions using the CRISPRFinder at the CRISPRs web server hosted at http://crispr.u-psud.fr/ (Grissa *et al*. 2007). The *cas* genes were also identified using the tblastx option on the database.

2.5.11 Genome comparisons and presence/absence of key genes

Genome comparisons to investigate specific unique or shared regions of interest were carried out using the comparative tools in the Seed Viewer on the RAST server (Aziz *et al.* 2008), as well as by further confirmation by whole genome pairwise alignments using ACT (Carver *et al*. 2005).

A number of candidate genes of interest were also identified through an extensive literature search on *Cronobacter* and related organisms, as well as by scanning the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Ogata *et al*. 1999). The presence/absence of these genes was then verified using the BLAST search tool in SEED viewer on the RAST server as well as the Artemis genome browser (Rutherford *et al*. 2000).

2.5.12 Characterisation of O-LPS genes

The O-LPS gene cluster for the sequenced genomes was characterised by ACT comparisons (Carver *et al*. 2005) with the gene regions identified in previous molecular studies by Mullane *et al*. (2008), Jarvis *et al*. (2011) and Sun *et al*. (2012a and 2012b). Genes from the newly identified clusters were subjected to BLAST searches to further characterize their functions.

2.5.13 Phylogenetic analysis of sialic acid gene cluster

The amino acid sequences of the genes *nanK, nanT, nanE, nanA, nanC, nagA* and *nagB* from the *Cronobacter* spp. genomes sequenced in this study were obtained from RAST. The gene sequences from the genome of *C. sakazakii* BAA-894 was used to perform tblastx searches to obtain the corresponding sequences of closely related organisms showing >50% amino acid identity. The corresponding protein sequences were then downloaded from the genomes of *E. coli* O7:K1 strain CE10 (Accession No. NC_017646); *E. cloacae* subsp. *cloacae* ATCC 13047 (Accession No. NC_014121); *E. hormaechei* ATCC 49162 (Accession No. AFHR00000000); *Enterobacter* spp. strain 683 (Accession No. CP000653); *Citrobacter freundii* strain 4_7_47CFAA (Accession No. ADLG00000000); *Citrobacter koseri* ATCC BAA-895 (Accession No. NC_009792); *Edwardsiella tarda* strain EIB202 (Accession No. NC_013508); *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain LT2 (Accession No. NC_003197); *Pantoea agglomerans* strain IG1 (Accession No. BAEF00000000). The sequences of each gene were aligned using ClustalW (Thompson *et al*. 1994) and then phylogenetic analysis was performed using the Maximum-Likelihood algorithm (Felsenstein. 1981) in PhyML (Guindon and Gascuel 2003). Stability of the relationships was assessed by the bootstrap method (1000 replicates) and the trees were viewed and annotated using TreeDyn (Chevenet *et al*. 2006) at http://www.phylogeny.fr (Dereeper *et al*. 2008).

CHAPTER 3

MULTILOCUS SEQUENCE TYPING (MLST) REVEALS RECOMBINATION AND EVOLUTIONARY RELATIONSHIPS OF THE GENUS *CRONOBACTER*

3.1 INTRODUCTION

3.1.1 Aims of this chapter

Our research group, in collaboration with the University of Warwick, had established a 7-loci MLST scheme which had enabled effective differentiation between the species *C. sakazakii* and *C. malonaticus* (Baldwin *et al*. 2009). In this chapter is described how the MLST scheme established by Baldwin *et al*. (2009) was extended from only the two species to cover all seven formally recognised species of the *Cronobacter* genus. MLSA was then applied in order to better quantify the intra-species and inter-species diversity as well as evolution of the genus. The *Cronobacter* MLST database is available online, with open access at www.pubmlst.org/Cronobacter.

The results presented in this chapter have been accepted for publication as Joseph *et al*. (2012b).

3.1.2 Multilocus sequence typing

MLST is a molecular typing technique which involves the identification and clustering of bacterial isolates based on the partial sequence analysis of multiple housekeeping genes which are ideally scattered across the genome of the organism. It was basically developed as a modification of the earlier typing tool, multilocus enzyme electrophoresis (MLEE). MLEE involved the identification of variations within multiple core metabolic or 'housekeeping' genes on the basis of the differing electrophoretic mobilities of the gene products (Maiden *et al*. 1998), while MLST is a sequence based technique. Sequence analysis ensures that even a single nucleotide change gets incorporated for discriminating between isolates and therefore MLST manages to provide greater resolution than a technique such as MLEE.

The use of multiple genes for typing has a number of advantages. The gene targets for MLST are housekeeping genes, which encode for proteins that are absolutely essential for the survival of the organism. As a result, they would be expected to be free of the influence of selective pressures. The combination of the multiple sites on the genome would also provide increased informative characters and better discrimination than a single gene could provide. This would also help to overcome the effect of any recombination events that could have occurred at a single locus (Maiden *et al*. 1998; Enright & Spratt 1999; Spratt 1999).

MLST was first applied and validated on a collection of *Neisseria meningitidis* strains by Maiden *et al*. in 1998. Though this initial scheme was based on six loci, today nearly all

MLST schemes are based on seven housekeeping loci, thus providing an increased level of discrimination and resolution. The procedure for MLST is fairly uncomplicated. The gene targets are first amplified and then sequenced; using either the same set of primers or nested primers (outer primers for PCR amplification and inner primers for nucleotide sequencing). Generally, a 400-500 bp long target region is then trimmed and analysed for each locus. The multiple alleles at each gene locus are aligned and compared with each other and each variant allele is denoted by a unique allele number for the locus. These numbers are assigned arbitrarily and do not reflect any sort of relationship. Thus, in a 7-loci typing scheme, every isolate obtains a combination of seven allele numbers. Each unique combination of these seven numbers forms an allelic profile denoted by a sequence type (ST) for the particular isolate. The isolates with similar STs are considered to be isogenic strains or clones of each other as they are indistinguishable at all seven loci. All this sequence information can then be submitted and stored on a central database, therefore making the technique electronically portable and reproducible in a laboratory in any part of the world (Maiden *et al*. 1998).

The technique is now being applied on a wide range of prokaryotic populations, with the sequence data being hosted on a number of centralised databases, the major ones being:

(1) http://www.pubmlst.org hosted by the Department of Zoology at The University of Oxford

(2) http://www.mlst.net hosted by the Imperial College, London.

(3) http://www.pasteur.fr/mlst hosted by the Institut Pasteur, France.

(4) http://mlst.ucc.ie hosted by the University College, Cork, Ireland.

These databases are supported by central analytical facilities and host a plethora of information on the isolates that have been typed, including sequence data and background information of the strains, such as isolation history. They also provide links to publications that the strains have been used in, thus providing maximum information to the user in one place. All this data is stored with complete open access, enabling convenient downloading without any registrations and enables comparative analysis for researchers working with the said organism in any part of the world. The fact that every one of these databases is strictly curated also ensures the reliability and quality of the data that is stored by them. These portals also have provided links for analysing the sequences with tools to compare alleles and sequence types, detect the presence of linkage and recombination in the population dataset, establish relationships between related STs or isolates by constructing phylogenetic trees or Minimum spanning trees, estimating the %GC and codon usage of the alleles. There are embedded BLAST tools for similarity searches of the loci as well as download options for the profiles of

all the identified sequence types in the scheme. Some of the databases such as PubMLST, have a "Locus Explorer" tool which enables the visualisation of the variable positions within each locus, and also predicts the translated protein product. Since these are housekeeping genes, the majority of the nucleotide differences are in the third 'wobble' base position and therefore do not affect the resulting amino acid sequence, and protein structure (termed 'synonymous substitution'). Although non-synonymous substitutions may also occur in housekeeping genes, their numbers are significantly lower than the synonymous substitutions, thereby not influencing the selective pressure of the gene.

The *Cronobacter* MLST database is based at http://www.pubmlst.org/cronobacter, at the PubMLST group of databases, maintained by Prof. Keith Jolley. PubMLST is home to >40 other major MLST schemes of organisms such as *Aeromonas* spp., *Campylobacter* spp., *Helicobacter pylori, Clostridium difficile, Neisseria* spp., *Pseudomonas aeruginosa, Yersinia* spp. Each organism's database consists of two sections: An information section that introduces the MLST scheme along with the primers and protocols to be used for the amplification and sequencing of the loci; and the main database section that contains all the sequencing data including allele sequences, MLST profiles and isolate information.

The sequencing data obtained by the MLST technique has a number of applications apart from just typing for epidemiological purposes, such as speciation studies and population genetics studies. As a result, the technique is now being used for not just pathogens but also a number of environmental isolates (Brady *et al*. 2008). The considerable amount of sequence data generated by performing MLST provides an excellent starting base for population genetics and evolutionary studies of bacterial species.

3.1.3 Recombination studies using multilocus sequence analysis

Homologous recombination events of partial or whole gene sequences are common occurrences in most bacterial populations. These are especially important in closely related species, as these events could have played a key role in the adaptation and evolution of the genus based on a number of factors such as speciation or virulence (Arnold *et al*. 2008; Talarico *et al*. 2012). The data from MLST experiments of a number of organisms has been previously used for studying the recombination patterns in the population (Didelot *et al*. 2011; Kotetishvili *et al*. 2005).

In this study, we have first considered the allele sharing tendencies between the species, and then gone on to estimate the possibilities of recombination events and breakpoints having

occurred within the entire *Cronobacter* spp. population dataset. The dataset of the whole genus was checked for possible recombination events by constructing a reticulate phylogenetic network using the Neighbour Net algorithm of the software Splitstree 4 (Huson & Bryant. 2006). The individual species were tested for recombination events using the RAT tool, which uses a sliding window mechanism along the length of the concatenated sequence to compare the individual members of the species, with a reference sequence, in this case that of the type strain of the species (Etherington *et al*. 2004).

3.1.4 Evolutionary calibrations using multilocus sequence analysis

The molecular clock rate of an organism is an estimate of the rate of evolution of the species, based on the nucleotide or protein sequences of selected genes. The molecular clock rates for a number of bacterial species have been published in recent times, based on calculations using archaeological or temporal evidence from outbreak studies, or even based on evolutionary rates of host species (Kuo & Ochman. 2009; Ochman *et al*. 1999).

Related studies to calibrate the evolution of populations have been carried out for different bacterial species using a combination of synonymous substitutions in the multilocus sequence data and these clock rates (Achtman *et al*. 1999; Falush *et al*. 2001; McQuiston *et al*. 2008; Waine *et al*. 2009). This is based on the assumption that closely related bacterial species will have similar or comparable rates of evolution. Synonymous substitutions are the silent changes in the nucleotide sequence that do not modify the resulting amino acid sequence, and therefore the expression of the protein. They remain unaffected by purifying selection and are therefore ideal tools for use in evolutionary studies of an organism.

In this study, a similar approach has been used to estimate the age of the *Cronobacter* genus and the divergence dates of the individual species, using the MLST dataset and a range of two different molecular clock rates calculated for two other closely related members of the *Enterobacteriaceae* family – *E. coli* and *Salmonella* species (Berg & Martelius. 1995; Ochman *et al*. 1999).

3.2 RESULTS

3.2.1 MLST of the genus *Cronobacter*.

The *Cronobacter* MLST scheme is based on the seven housekeeping genes: *atpD* (ATP synthase b chain), *fusA* (elongation factor G), *glnS* (glutaminyl tRNA synthetase), *gltB* (glutamate synthase large subunit), *gyrB* (DNA gyrase subunit B), *infB* (translation initiation factor IF-2) and *ppsA* (phosphoenolpyruvate synthase A). The distribution of these seven genes across the genome of *C. sakazakii* BAA-894 has been depicted in Fig. 3.1.

In this study, the alleles of the seven MLST genes were successfully amplified and sequenced in all the 325 strains using the primers described in Chapter 2, Table 2.2; and 115 STs were identified spanning across the genus *Cronobacter* (Fig 3.2; Table 3.1).

The % GC content of the seven sequenced genes in all the STs across the *Cronobacter* genus showed an average of 58%, which is consistent with the overall 57% GC content for the published genome *C. sakazakii* BAA-894 (Kucerova *et al*. 2010).

All sequences are available for free download at http://www.pubmlst.org/cronobacter. The seven MLST genes were successfully sequenced from all the strains used in this study.

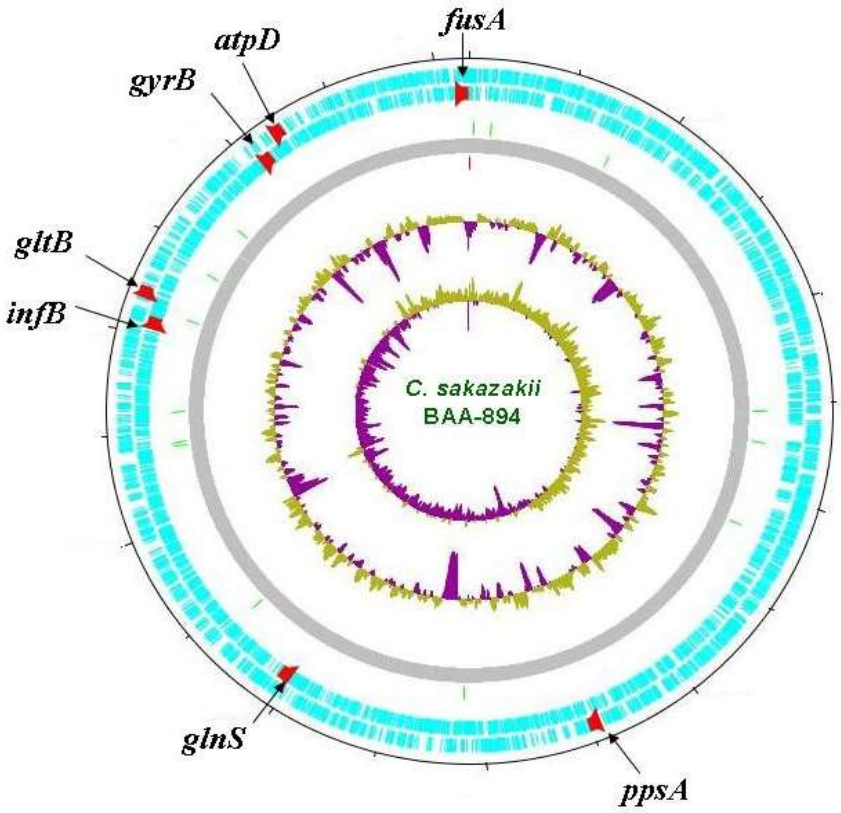


Fig. 3.1 Positions of the seven loci of the *Cronobacter* MLST scheme denoted on the genome of the *C. sakazakii* BAA-894 chromosome of size 4.4 Mb (Kucerova *et al*. 2010). The output was generated using DNAPlotter (Carver *et al*. 2009).

3.2.1.1 *Cronobacter sakazakii*

For *C. sakazakii*, which is also the type species of the genus, 53 STs were identified across the 226 strains of the species and this was found to be the most predominant species. The *C. sakazakii* type strain NCTC 11467^{T} (NTU strain 1) was identified to belong to ST8. STs 1, 4 and 8 were found to be the most significant STs with respect to the number of strains isolated (Fig.3.2) as well as their sources, an aspect that has been dealt with in detail in Chapter 5. Among them ST4 was found to be the most prevalent ST of the entire genus, with 78 isolates. The first genome sequenced strain of genus *Cronobacter* – *C. sakazakii* BAA894 (NTU strain 658) – belongs to ST1 (Kucerova *et al* 2010).

3.2.1.2 *Cronobacter malonaticus*

Seventeen STs were identified for *C. malonaticus* from the 39 strains. Among them, ST7 was the most frequently observed one, with 15 strains belonging to it (Fig.3.2). The type strain of the species - LMG 23826 ^{T} (NTU strain 681) also belonged to ST7.

3.3.1.3 *Cronobacter turicensis*

From the 19 strains of *C. turicensis* in this collection, 13 STs were identified, of which ST5 was found to be the most frequent (Fig.3.2). The type strain of the species, a blood isolate – LMG23827 ^{T} (NTU strain 1211) was found to belong to ST19.

3.2.1.4 *Cronobacter universalis*

Four STs were identified for the recently renamed *C. universalis*, from the only four strains isolated for this species till date. This species was previously known as *Cronobacter* genomospecies1, with only two identified strains then. The reclassification, as a result of this MLST study, is further detailed in Chapter 4. The type strains of this species, a fresh water isolate from UK – NCTC 9529^{T} (NTU strain 581) was identified to be ST54.

3.2.1.5 *Cronobacter muytjensii*

Ten STs were identified for *C. muytjensii* from the 17 strains included in this study. There is a slight bias in this number due to an issue experienced with the type strain ATCC

51329^T. As there was a variant version of the strain found in our in-house strain collection, multiple copies of the type strain were collected from different culture collections, in order to verify this anomaly. The type strain – ATCC 51329^T (NTU strain 1561) was then confirmed to be a ST81.

3.2.1.6 *Cronobacter dublinensis*

From the 19 strains of *C. dublinensis* used in this study, 17 different STs were identified. A number of these strains had been obtained after spotting them among the 16S rDNA submissions in Genbank, then classified only as *Cronobacter* species. The type strain of the species, an isolate from a milk powder manufacturing plant in Ireland – LMG 23823^T (NTU strain 1210) belongs to ST106.

3.2.1.7 *Cronobacter condimenti*

C. condimenti is a newly identified species of the genus *Cronobacter*, with a lone isolate from spiced meat in Slovakia (NTU strain 1330), which was identified to be ST98. The recognition of this species based on this MLST data has been discussed in detail in Chapter 4.

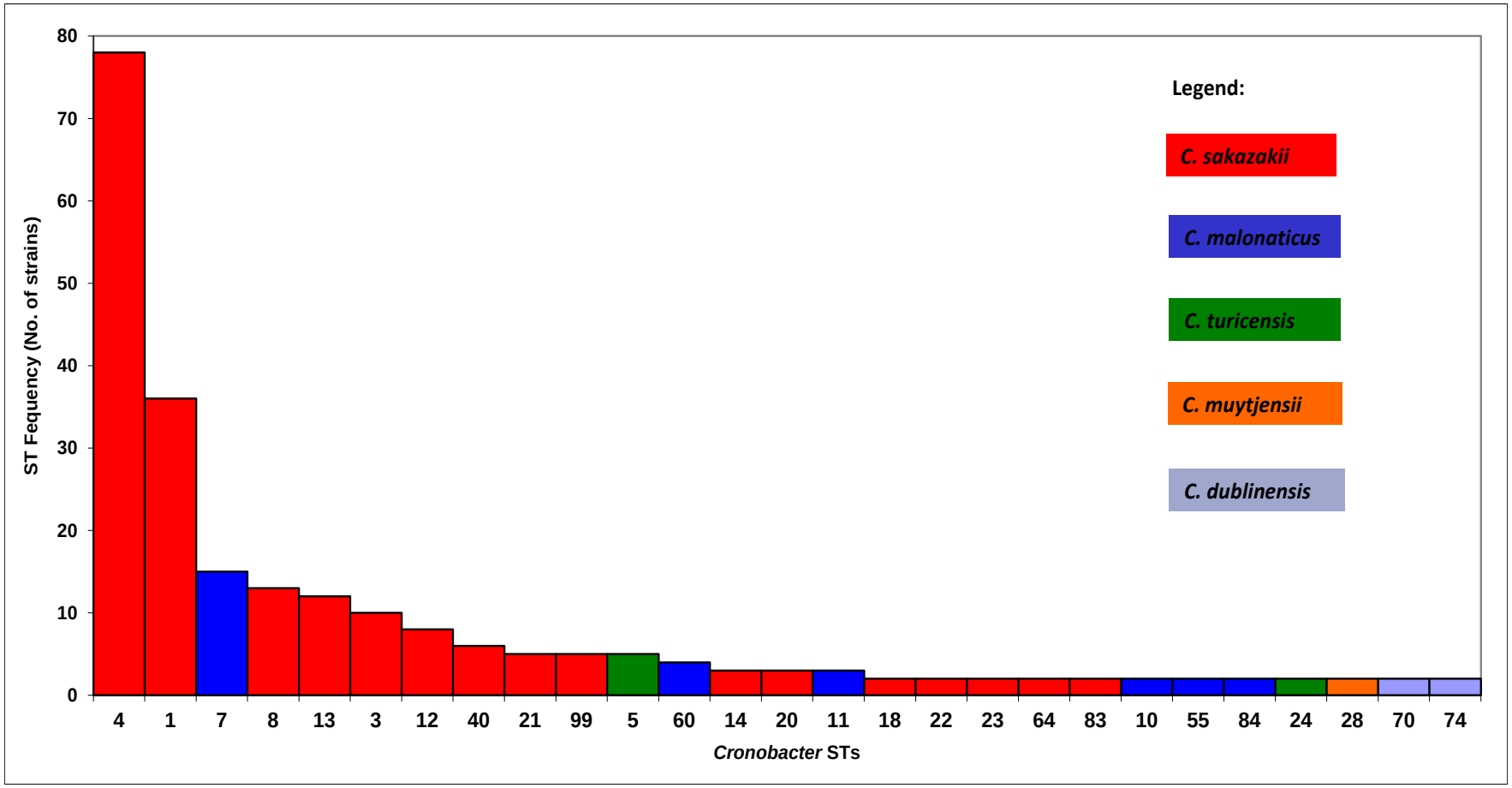


Fig 3.2 Distribution graph of the 27 *Cronobacter* STs in this study with frequencies of 2 or more strains. The STs (on the X-axis) are arranged in decreasing order of frequency of strains (indicated on the Y-axis) and the different *Cronobacter* species indicated according to colour. All the remaining STs (including those of *C. universalis* and *C. condimenti*) are associated with single strains as indicated in Table 3.1, and hence not included in this graph. Further details regarding the individual STs in each species have also been provided in Table 3.1.

Table 3.1 Breakdown of the *Cronobacter* STs according to species, with details of isolate sources.

Species	ST	No. of strains	Isolation source	Country of isolation	Period of isolation
C. sakazakii	1	36	Infant formula, milk powder, clinical, environment, food,	Switzerland, Czech Republic, UK, Russia, Netherlands, China, USA Australia, Brazil, Turkey, Germany	1979-2010
C. sakazakii	3	10	Follow up formula, infant formula, enteral feeding tube, weaning food, environment	Korea, Netherlands, UK, Australia	1988-2008
C. sakazakii	4	78	Clinical, infant formula, milk powder, weaning food, chocolate, washing brush, environment, prepared formula, foot wound	UK, USA, France, China, Canada, Netherlands, Germany, Russia, Czech Republic, Switzerland, Slovakia, New Zealand, Saudi Arabia, Bangladesh	1950-2011
C. sakazakii	8	13	Clinical, infant formula, food	USA, UK, Czech Republic, France, Canada	1977-2011
C. sakazakii	9	1	Weaning food	UK	2008
C. sakazakii	12	8	Clinical, Weaning food	USA, UK, France, Czech Republic	1973-2008
C. sakazakii	13	12	Clinical, infant formula, food, herbs, spice	France, Germany, China, Turkey	1988-2010
C. sakazakii	14	3	Infant formula	France	1994
C. sakazakii	15	1	Clinical	Canada	2003
C. sakazakii	16	1	Spice	Korea	2005
C. sakazakii	17	1	Infant formula	Korea	2006
C. sakazakii	18	2	Clinical, mouse	UK, Netherlands	1953-1988
C. sakazakii	20	3	Food, spice	China, UK	2007-2010
C. sakazakii	21	5	Weaning food, chocolate, food	USA, Slovakia, China	2010
C. sakazakii	22	2	Food	China	2010
C. sakazakii	23	2	Food	Czech Republic, China	2010
C. sakazakii	31	1	Clinical	UK	2010
C. sakazakii	33	1	Environment	Germany	2006
C. sakazakii	37	1	Environment	Australia	2007
C. sakazakii	40	6	Food, spice	UK	2007-2010
C. sakazakii	41	1	Foot wound	USA	1975
C. sakazakii	45	1	Infant formula	Russia	1988
C. sakazakii	46	1	Environment	Australia	2007
C. sakazakii	47	1	Follow up formula	Korea	2008
C. sakazakii	50	1	Sputum	USA	1977
C. sakazakii	52	1	Flour	Turkey	2010
C. sakazakii	56	1	Infant formula	Brazil	2007
C. sakazakii	57	1	Infant formula	Denmark	1988
C. sakazakii	58	1	Flour	Turkey	2010
C. sakazakii	61	1	Clinical	Czech Republic	1983
C. sakazakii	64	2	Environment	UK, Australia	2007
C. sakazakii	65	1	Infant formula	USA	1988
C. sakazakii	67	1	Herbs	UK	2010
C. sakazakii	68	1	Spice	Unknown	2010

C. sakazakii	83	2	Milk powder	Australia	2007
C. sakazakii	86	1	Clinical	France	1994
C. sakazakii	90	1	Herb	Portugal	Unknown
C. sakazakii	93	1	Clinical	Czech Republic	1979
C. sakazakii	96	1	Environment	Australia	2007
C. sakazakii	97	1	Milk powder factory	Australia	2007
C. sakazakii	99	5	Environmental	Germany	2006
C. sakazakii	100	1	Environment	Australia	2007
C. sakazakii	101	1	Environmental	Germany	2006
C. sakazakii	103	1	Environment	Australia	2007
C. sakazakii	104	1	Environment	Australia	2007
C. sakazakii	105	1	Milk powder factory	Australia	2007
C. sakazakii	107	1	Clinical	USA	2011
C. sakazakii	108	1	PIF	USA	2011
C. sakazakii	110	1	Clinical	USA	2011
C. sakazakii	111	1	Water	USA	2011
C. sakazakii	113	1	Infant formula	Brazil	2007
C. sakazakii	115	1	Environment	Australia	2007
C. sakazakii	117	1	Environment	Australia	2007
C. malonaticus	7	15	Food, clinical, PIF, weaning food	Czech Republic, Brazil, France, New Zealand, USA	1973-2007
C. malonaticus	10	2	Herb	China	2005
C. malonaticus	11	3	Clinical	Czech Republic	1983-1984
C. malonaticus	25	1	Food	China	2010
C. malonaticus	26	1	Food	China	2010
C. malonaticus	29	1	Spice	Unknown	2005
C. malonaticus	53	1	Clinical	USA	1977
C. malonaticus	55	2	Flour	Turkey	2010
C. malonaticus	60	4	Faeces, sputum, spice	Czech Republic	2003-2005
C. malonaticus	62	1	Infant formula	Australia	1988
C. malonaticus	63	1	Herb	China	2004
C. malonaticus	66	1	Infant formula	Korea	2005
C. malonaticus	69	1	Herb	UK	2010
C. malonaticus	84	2	Clinical	Czech Republic	Unknown
C. malonaticus	89	1	Clinical	Czech Republic	Unknown
C. malonaticus	102	1	Environment	Australia	2007
C. malonaticus	112	1	Clinical	USA	2011
C. turicensis	5	5	Milk powder, herbs, bone marrow, blood	UK, USA	1970-2004
C. turicensis	19	1	Clinical	Switzerland	2005
C. turicensis	24	2	Herb, food	UK, China	2004-2010
C. turicensis	32	1	Weaning food	UK	2003
C. turicensis	35	1	Herb	UK	2004
C. turicensis	72	1	Spice	Unknown	Unknown

C. turicensis	85	1	Unknown	Slovakia	Unknown
C. turicensis	87	1	Unknown	Slovakia	Unknown
C. turicensis	114	1	Environment	Australia	2007
C. turicensis	116	1	Environment	Australia	2007
C. turicensis	118	1	Environment	Australia	2007
C. turicensis	119	1	Milk powder factory	Australia	2007
C. turicensis	120	1	Milk powder	Australia	2007
C. muytjensii	28	2	Unknown	China	2010
C. muytjensii	30	1	Unknown	Unknown	Unknown
C. muytjensii	34	1	Spice	Unknown	2005
C. muytjensii	44	1	Infant formula	Jordan	2008
C. muytjensii	49	1	Infant formula	Denmark	1988
C. muytjensii	71	1	Spice	Unknown	2010
C. muytjensii	75	1	Food	Slovakia	2010
C. muytjensii	81	1[a]	Unknown	UK, USA	Unknown
C. muytjensii	82	1	Food	UK	2011
C. muytjensii	94	1	Follow up formula	Korea	2008
C. dublinensis	27	1	Food	China	2010
C. dublinensis	36	1	Unknown	Unknown	Unknown
C. dublinensis	38	1	Follow up formula	Korea	2008
C. dublinensis	39	1	Follow up formula	Korea	2008
C. dublinensis	43	1	Herb	China	2004
C. dublinensis	70	2	Follow up formula, food	Korea	2008-2011
C. dublinensis	74	2	Follow up formula, food	Korea	2008-2011
C. dublinensis	76	1	Food	Korea	2011
C. dublinensis	77	1	Food	Korea	2011
C. dublinensis	78	1	Food	Korea	2011
C. dublinensis	79	1	Milk powder	Unknown	2003
C. dublinensis	80	1	Unknown	Czech Republic	Unknown
C. dublinensis	88	1	Clinical	USA	1979
C. dublinensis	91	1	Dairy	UK	1956
C. dublinensis	92	1	Food	Czech Republic	Unknown
C. dublinensis	95	1	Food	Korea	2011
C. dublinensis	106	1	Environment	Ireland	2007
C. universalis	48	1	Spice	UK	2004
C. universalis	51	1	Flour	Turkey	2010
C. universalis	54	1	Fresh water	UK	1956
C. universalis	59	1	Leg infection	France	2005
C. condimenti	98	1	Food	Slovakia	2010

a – 7 isolates of the same strain from various culture collections were tested for profile confirmation

3.2.2 Phylogenetic relationships of the *Cronobacter* species.

A phylogenetic tree based on the concatenated sequences (3036 nucleotides) of the 115 *Cronobacter* STs, constructed using the Maximum-Likelihood algorithm in MEGA 5 was used to study the phylogeny of the genus *Cronobacter* (Fig. 3.3). In comparison to the use of 16S rDNA sequences (Fig 3.4), the sequences of the concatenated seven MLST loci showed better resolved and robust clustering of the individual *Cronobacter* species.

These clusters corresponded to previously defined species, with some exceptions. When investigated, it was found that a few strains had been previously mis-identified due to biotyping errors or discrepancies in 16S rDNA sequences. These mis-identifications have mostly been between the closely related species *C. sakazakii* and *C. malonaticus* (NTU strains 33 and 101), and a few *C. turicensis* isolates which had been mis-identified as *C. dublinensis* (NTU strains 57 and 564). One of the isolates (NTU strain 96) of the reclassified *C. universalis* had been previously incorrectly identified as *C. turicensis* by biotyping.

Overall, the phylogeny of the genus using MLSA showed the species *C. sakazakii* and *C. malonaticus* to be tightly clustered and genetically closely related to each other, yet clearly distinguishable which is not possible using 16S rDNA sequence analysis (Baldwin *et al*. 2009). The recently reclassified *C. universalis* was found to be the closest neighbour to *C. turicensis.* The species *C. muytjensii* and *C. dublinensis* showed independent branches, with diverse branching within each species as well. The newly defined *C. condimenti* was an independent lineage most closely related to *C. dublinensis.*

The tree was rooted using the sequences of the 7 MLST loci from the publicly available genomes of *Citrobacter koseri* BAA-895 [CP000822] and *Enterobacter* spp. 683 [CP000653], two organisms closely related to *Cronobacter* spp. None of their alleles were found to be shared with any of the *Cronobacter* spp. isolates and they were found to be genetically significantly distant from the rest of the *Cronobacter* genus (Fig. 3.3).

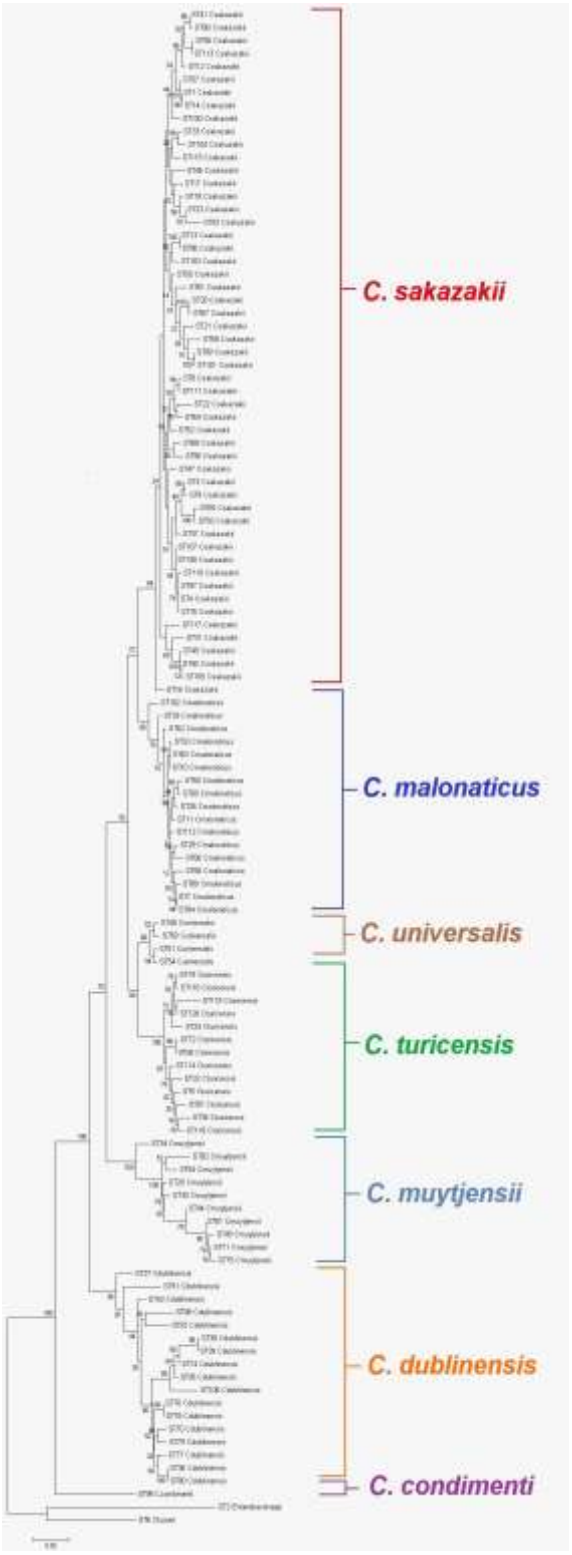


Fig. 3.3 Maximum-likelihood tree based on the concatenated sequences (3036 bp) of the 7 MLST loci for the genus *Cronobacter*. The STs and the corresponding species are indicated at the tip of each branch. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage. doi:10.1128/JCM.00905-12

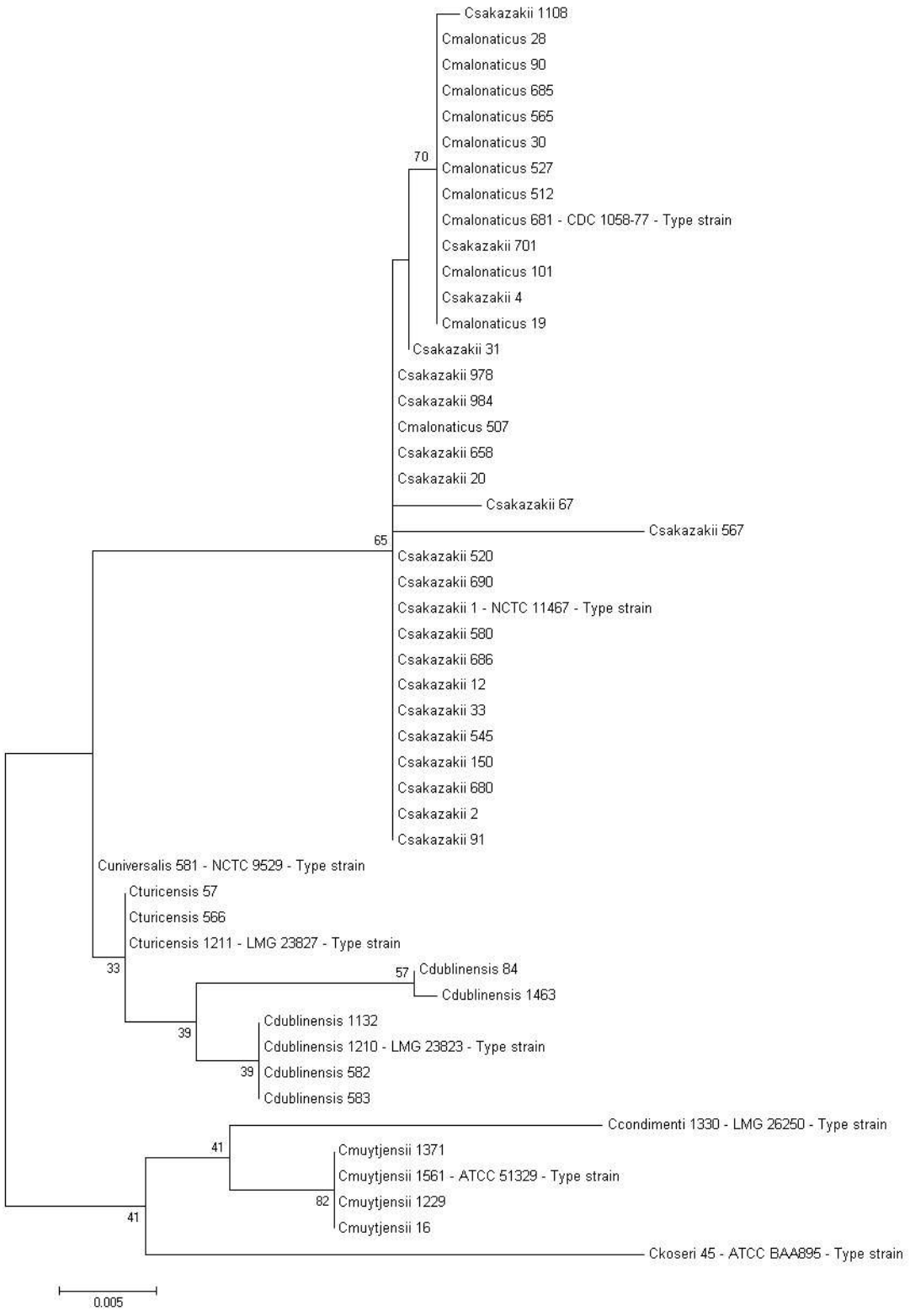


Fig. 3.4 Maximum-likelihood tree based on the partial length 16S rDNA sequences (528 bp) of the members of the genus *Cronobacter*. The NTU strain IDs and the type strains are indicated at the tip of each branch. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage.

3.2.3 Allelic analysis and variations

All the STs and allele numbers were assigned in continuation to the scheme initially established for *C. sakazakii* and *C. malonaticus* (Baldwin *et al*. 2009). The lengths of the alleles for each locus were found to be identical, with no insertions or deletions observed. A pseudo-deletion was detected at the 7th nucleotide position in the sequencing of only the reverse strand of the *glnS* allele for some strains. Hence, extra care was taken to confirm this by aligning the forward and reverse strands, to avoid any errors.

A wide range of alleles have been identified for each of the seven loci, with *ppsA* showing maximum sequence diversity at 78 different profiles. The proportion of the allele lengths that exhibited nucleotide polymorphisms varied from 22.05 % (*atpD*) to 31.71 % (*ppsA*). The ratio of the non-synonymous to synonymous substitutions (d_n/d_s) of all seven alleles was found to be significantly less than one, ranging from 0.009 (*atpD*) to 0.049 (*glnS*). (Table 3.2)

When the individual alleles were compared according to the *Cronobacter* species, there were found some instances of allele numbers being shared between two species. The majority of these shared alleles were observed between *C. sakazakii* - *C. malonaticus* and *C. sakazakii* - *C. muytjensii* species (Appendix Fig. 1 to 6). Among the seven genes, *gltB* had the most (seven) allele numbers shared between species. In order to eliminate the possibility of this allele sharing having influenced the overall genus phylogeny, a phylogenetic tree (Fig 3.5) was constructed with the concatenated sequences of 6 MLST loci, excluding *gltB.* This tree confirmed the robustness of the phylogeny and taxonomic divisions within the genus using this MLST data.

Among the seven loci, *fusA* was observed to be the most stable with none of the alleles being shared between two or more *Cronobacter* species (Fig. 3.6). Our research group has now adopted this approach instead of 16S rDNA gene sequencing for identifying *Cronobacter* spp. isolates.

Gene	Size of fragment analysed (bp)	Number of alleles identified	Number of polymorphic sites	Proportion of fragment as polymorphic sites (%)	% GC content	d_N/d_S
atpD	390	52	86	22.05	59.15	0.009
fusA	438	51	97	22.14	54.03	0.023
glnS	363	56	106	29.20	56.75	0.049
gltB	507	68	151	29.78	61.31	0.023
gyrB	402	67	126	31.34	56.67	0.031
infB	441	67	132	29.93	58.58	0.048
ppsA	495	78	157	31.71	62.52	0.022

d_n/d_s - ratio of the non-synonymous to synonymous substitutions

Table 3.2 START analysis of the 7 MLST loci of the *Cronobacter* spp. isolates. doi:10.1128/JCM.00905-12

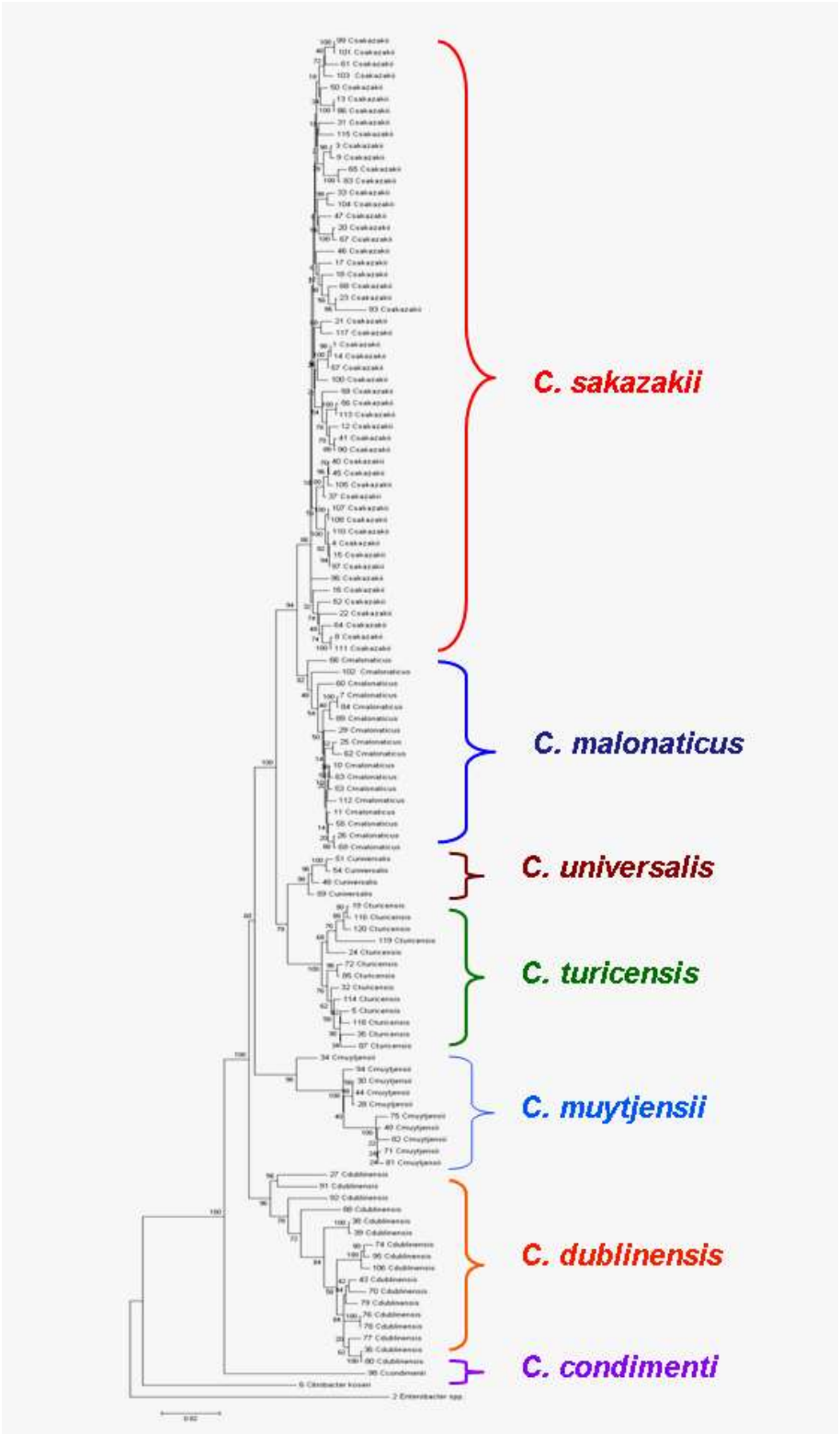


Fig. 3.5 Maximum-likelihood tree based on the concatenated sequences (3036 bp) of 6 MLST loci (*atpD*, *fusA*, *glnS*, *gyrB*, *infB* and *ppsA*) for the genus *Cronobacter*. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage.

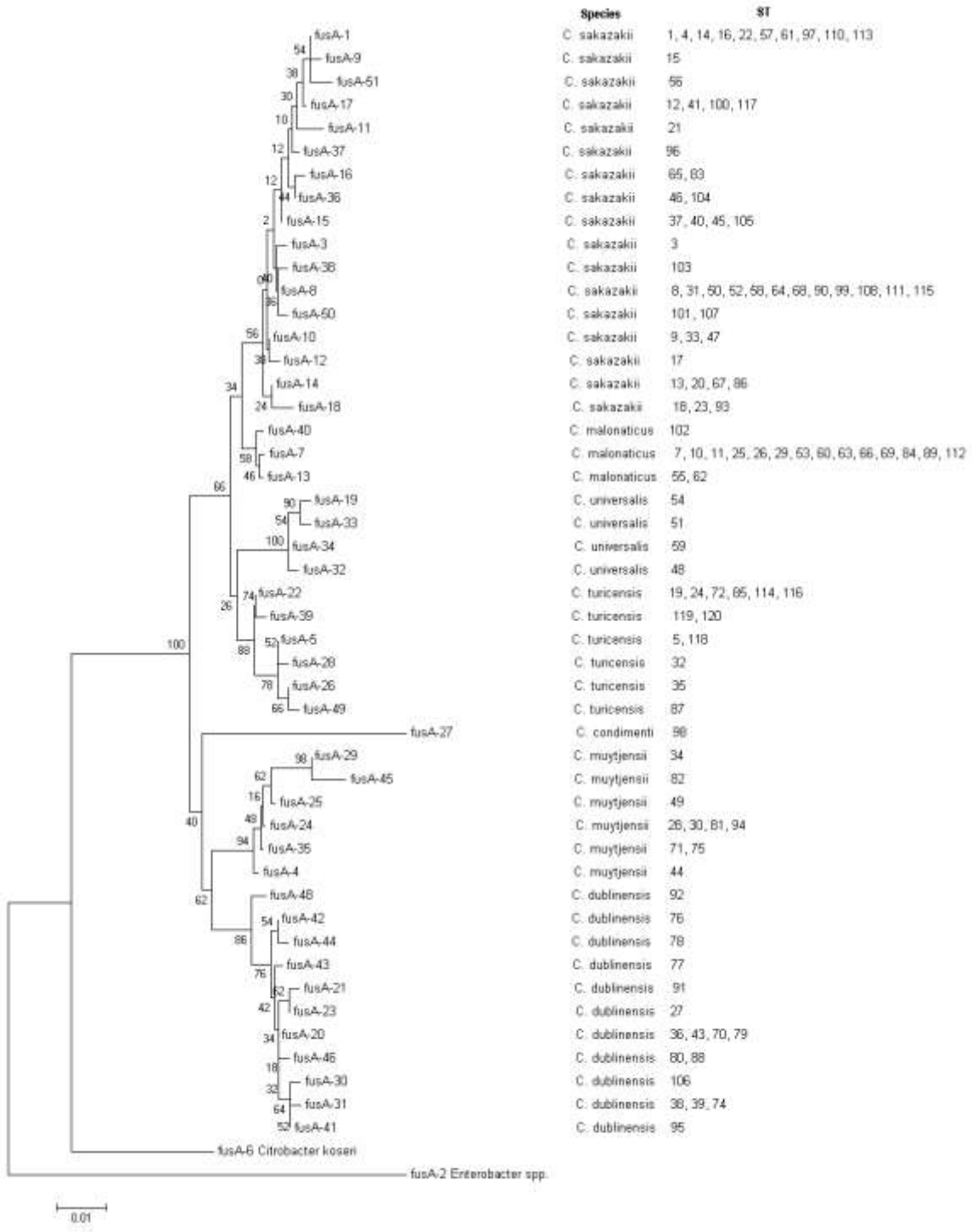


Fig 3.6 Maximum-likelihood tree based on the *fusA* alleles (438 bp) of the *Cronobacter* MLST dataset. The numbers at the end of each branch indicate the allele numbers. The species and corresponding STs have been indicated. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage. doi:10.1128/JCM.00905-12.

3.2.4 Evolutionary timescale estimations

The *Cronobacter* MLST sequence data was used to estimate the relative age of the genus *Cronobacter* and divergence time of the branches for the individual species. The D_s values for the *Cronobacter* genus and the individual species were calculated using the 7 loci sequence data and based on these values, two sets of age values were obtained for the genus and the individual species, providing an estimated range (Table 3.3).

	D_s	**[a] Age based on rate 4.5 x 10^{-9} (in millions of years)**	**[b] Age based on rate 3 x 10^{-9} (in millions of years)**
Genus *Cronobacter*	0.2042	45.37	68.06
Individual species:			
C. sakazakii	0.0703	15.62	23.43
C. malonaticus	0.0531	11.8	17.7
C. turicensis	0.0545	12.11	18.16
C. universalis	0.0377	8.37	12.56
C. muytjensii	0.1139	25.31	37.96
C. dublinensis	0.1222	27.15	40.73

[a] – Ochman *et al*. (1999); [b] – Berg & Martelius (1995).

Table 3.3 Estimated ages of the genus *Cronobacter* and its species, calculated using the MLST dataset based on the formula Age = D_s/Rate, where D_s is the synonymous substitution value and rate is the synonymous molecular clock rate.

Thus, this evolutionary analysis predicts that the *Cronobacter* genus split from its closest ancestor in the *Enterobacteriaceae* family approximately 45-68 million years ago (MYA) which was the Paleogene period of the Cenozoic era. The earliest branches of the genus led to *C. dublinensis* and *C. muytjensii* (Fig. 3.7). *C. condimenti* had to be excluded from this dataset because only one isolate has been identified for this species to date, and hence it was not possible to calculate the D_s values for the species.

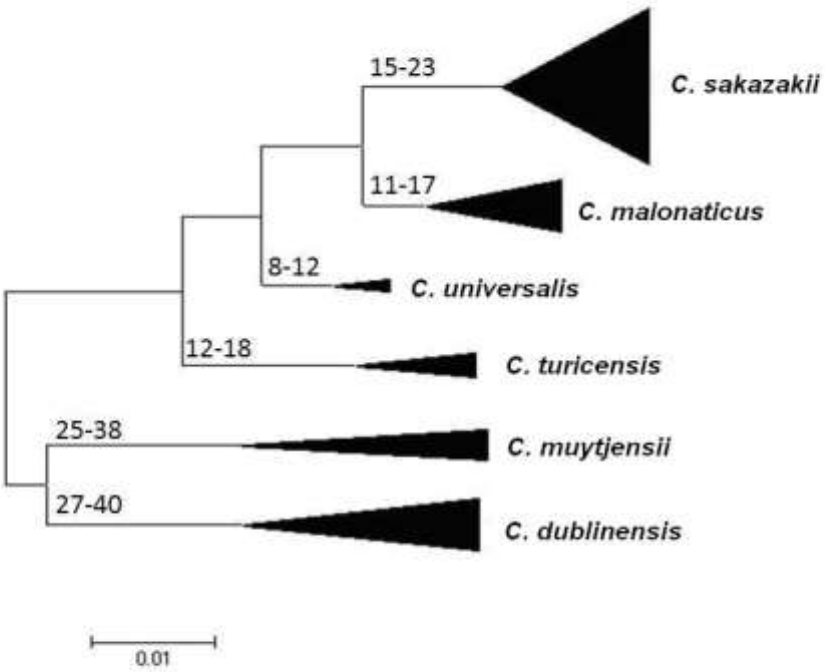


Fig. 3.7 Maximum-likelihood tree of the *Cronobacter* MLST dataset indicating ranges of the hypothetical divergence dates of each species node measured in millions of years before the present. The tree has been drawn to scale using MEGA5. The bases of the triangles indicate the number of isolates used for the analysis, while the heights indicate the diversity of each branch. *C. condimenti* with the single isolate has been excluded from this analysis. doi:10.1128/JCM.00905-12.

3.2.5 Inter-species recombination events.

In addition to the close similarities among many isolates of some species, especially *C. sakazakii*, there was also evidence of homologous recombination (gene conversion) events having occurred in the evolution of the genus. Splitstree4 was used to visualize recombination events and evolutionary relationships in the dataset.

Fig. 3.8 represents a Neighbour-Net generated for the *Cronobacter* genus using the concatenated 7-loci MLST data. This shows the tight clustering of the *C. sakazakii* and *C. malonaticus* species. It also reveals a higher diversity in the *C. muytjensii* and *C. dublinensis* species, with the formation of parallelograms denoting the possibility of recombination events. Phi test showed significant evidence for recombination within the genus ($p<0.001$).

The recombination within the individual *Cronobacter* species was further investigated using the RAT software. It is a distance based tool that analyses the recombination within a population based on pairwise sequence comparisons. As is evident from the graphs, the species *C. sakazakii, C. malonaticus, C. turicensis* and *C. universalis* show minimal diversity and recombination events (Fig. 3.9). However, *C. muytjensii* and *C. dublinensis* show a number of recombination events that could have taken place as well as increased genetic distances between these STs (Fig. 3.9). When compared with the X-axis positions, the major variations appear to be along the *gltB* sequence positions (1192 to 1698), which is also the locus where the most number of allele sharing was observed. This genetic diversity within the *C. muytjensii* and *C. dublinensis* species was further analysed in a speciation study, detailed in Chapter 4.

Splitstree4 was also used to visualise the recombination events at the individual loci, as shown in Fig. 3.10 (a) to (g). The high intensity parallelogram formation in the *gltB and gyrB* loci indicates a greater influence of recombination events. This was confirmed by the phi test which showed significant evidence for recombination at these loci ($p<0.001$). The *gltB* loci also showed the maximum number of inter-species allele sharing events as indicated by the red dots.

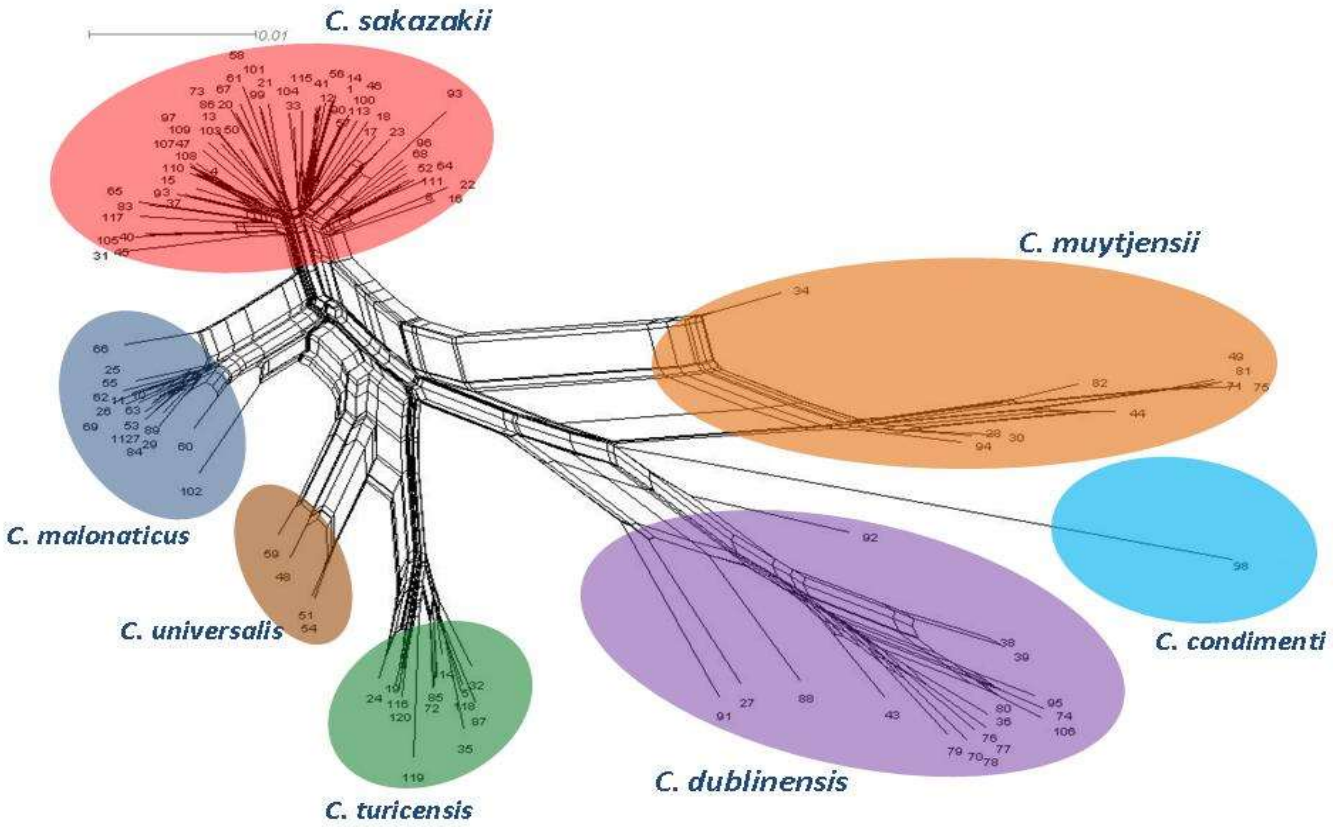


Fig. 3.8 Neighbour-Net of the concatenated 7-loci sequence alignment generated for the *Cronobacter* MLST dataset indicating diversity and recombination events. The figure has been drawn to scale using Splitstree4. The numbers at the tips of the branches indicate STs. The formation of parallelograms indicates possible recombination events. doi:10.1128/JCM.00905-12

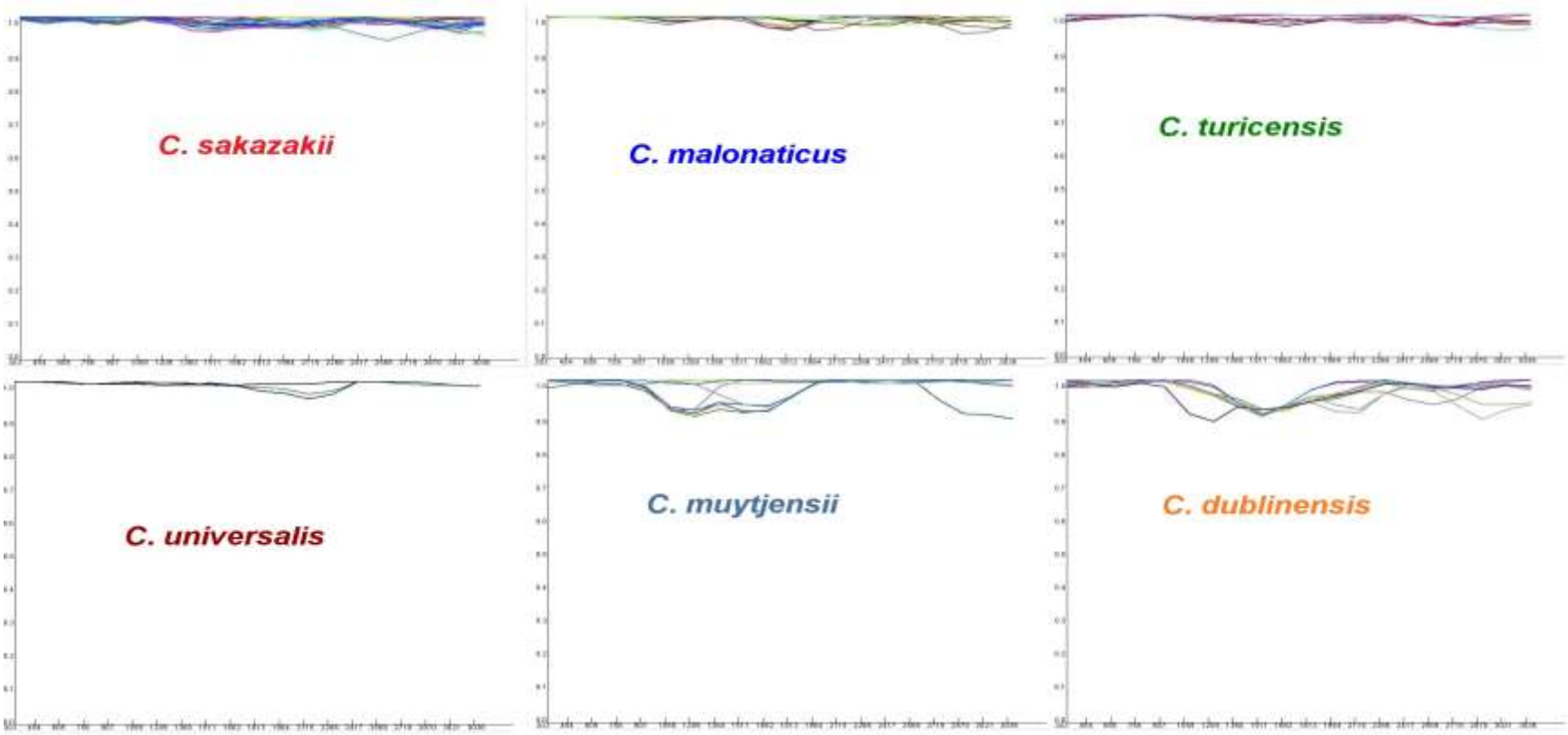


Fig. 3.9 Graphical outputs of the recombination analysis performed in RAT using the concatenated sequences of the 7 MLST loci of the individual *Cronobacter* species. Each coloured line indicates an ST of the species. The X-axis of each graph denotes the nucleotide position along the concatenated 3036 bp of the 7 MLST loci, while the Y axis denotes the genetic distance from the ST of the type strain of the species.

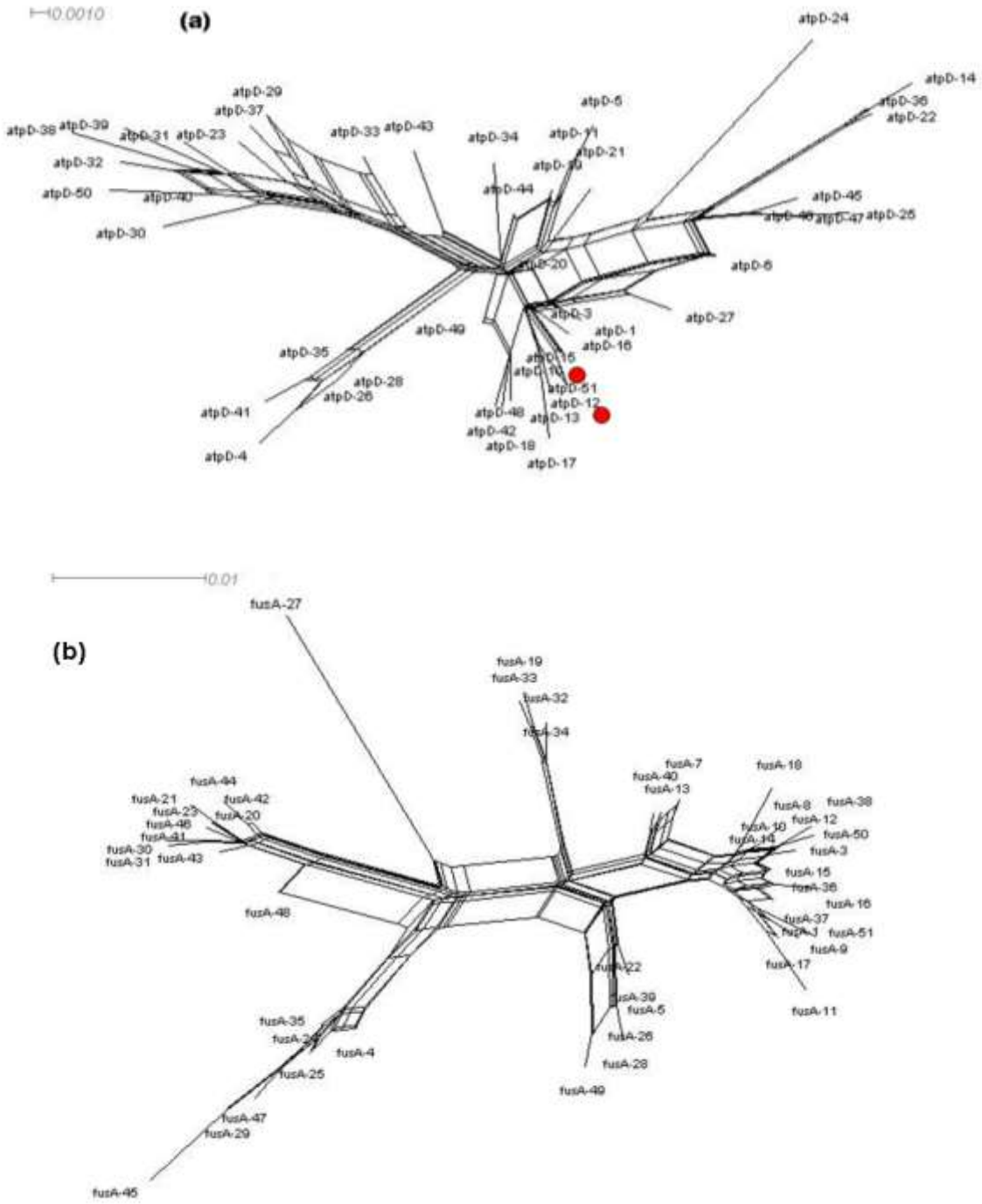


Fig. 3.10 Neighbour-nets of the individual alleles **(a)** *atpD* **(b)** *fusA* of the *Cronobacter* MLST scheme indicating possible recombination events. The numbers at the ends of the branches indicate the allele numbers. The figures have been drawn to scale using Splitstree4. The red dots indicate the alleles shared across species. Fig (b) as originally published in Joseph S & Forsythe SJ (2012) Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Frontiers in Food Microbiology*, 3, 397. doi: 10.3389/fmicb.2012.00397

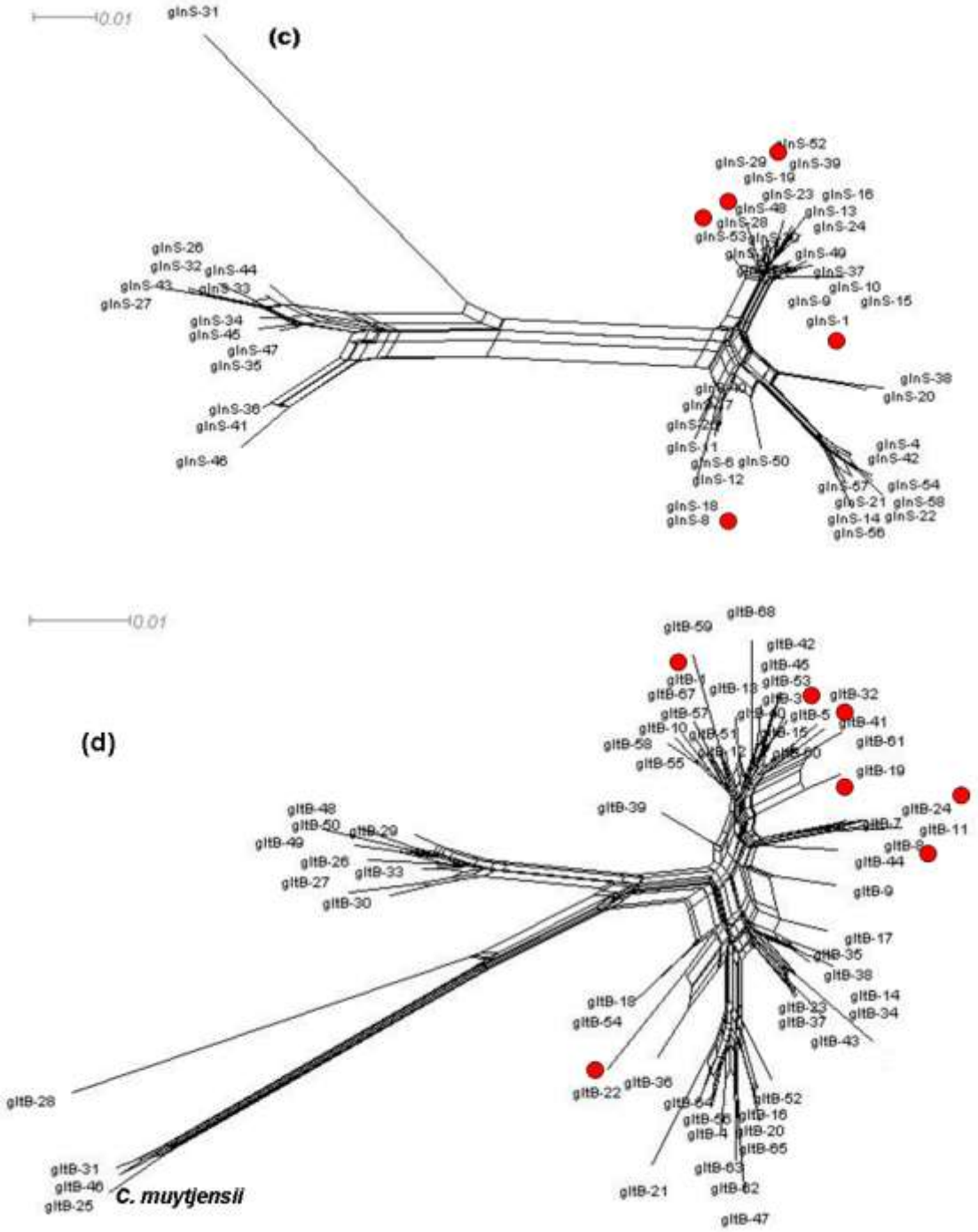


Fig. 3.10 Neighbour-nets of the individual alleles **(c)** *glnS* **(d)** *gltB* of the *Cronobacter* MLST scheme indicating possible recombination events. The numbers at the ends of the branches indicate the allele numbers. The figures have been drawn to scale using Splitstree4. The red dots indicate the alleles shared across species. Fig (d) as originally published in Joseph S & Forsythe SJ (2012) Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Frontiers in Food Microbiology*, 3, 397. doi: 10.3389/fmicb.2012.00397

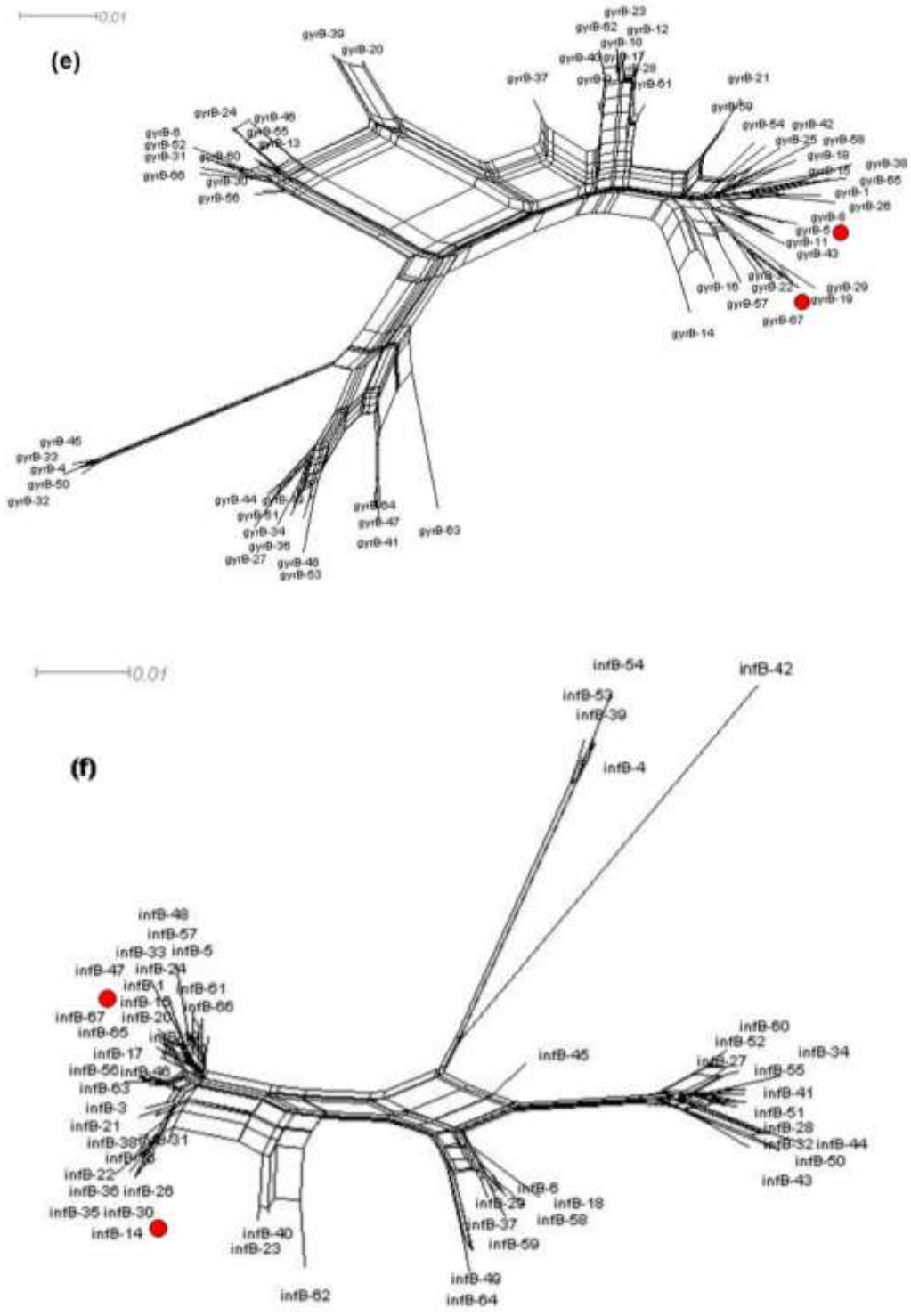


Fig. 3.10 Neighbour-nets of the individual alleles **(e)** *gyrB* **(f)** *infB* of the *Cronobacter* MLST scheme indicating possible recombination events. The numbers at the ends of the branches indicate the allele numbers. The figures have been drawn to scale using Splitstree4. The red dots indicate the alleles shared across species.

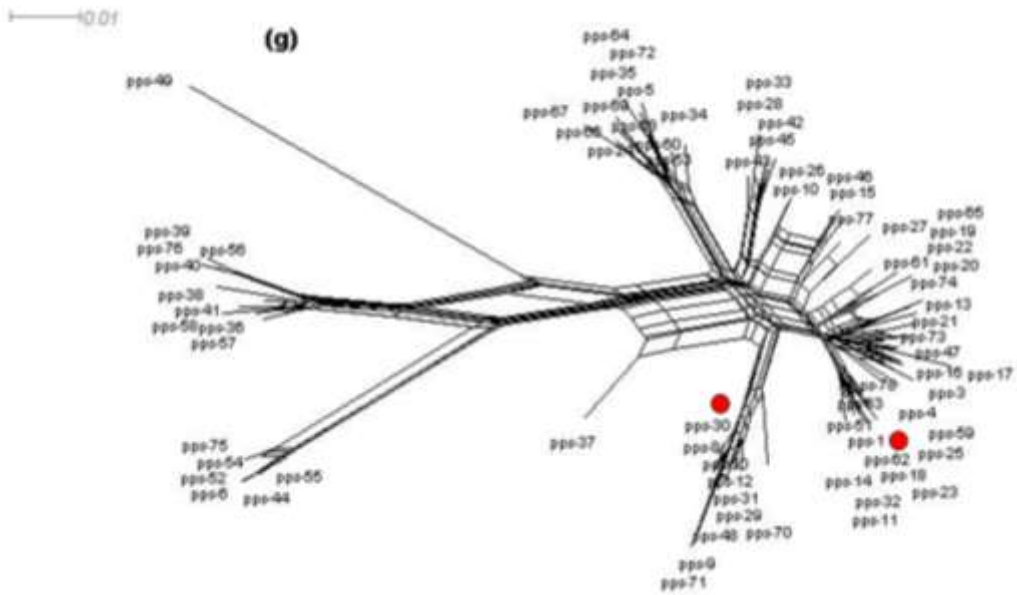


Fig. 3.10 Neighbour-nets of the individual allele **(g)** *ppsA* of the *Cronobacter* MLST scheme indicating possible recombination events. The numbers at the ends of the branches indicate the allele numbers. The figures have been drawn to scale using Splitstree4. The red dots indicate the alleles shared across species.

3.3 DISCUSSION

MLST is a powerful molecular typing tool that has been used in the population genetics and epidemiological studies of a number of bacteria (Wirth *et al*. 2006; Martino *et al*. 2011; Merga *et al*. 2011; Urwin & Maiden 2010). This study involved the use of this technique to undertake a novel analysis of the diversity and evolution of the genus *Cronobacter*.

The study was started as an expansion across the genus of the MLST scheme which was initially established by Baldwin *et al*. (2009) only for the species *C. sakazakii* and *C. malonaticus*, since these species could not be easily distinguished using 16S rDNA sequence analysis. The main aim of this current study was to extend this MLST scheme for the entire *Cronobacter* genus, with accuracy and ease of experimentation.

3.3.1 The *Cronobacter* MLST scheme

The *Cronobacter* MLST scheme comprises of a set of seven housekeeping genes: *atpD, fusA, glnS, gltB, gyrB, infB,* and *ppsA*. A number of these genes have been previously used as targets for typing and phylogeny studies in other members of *Enterobacteriaceae* (Dauga 2002; Brady *et al*. 2008; Hedegaard *et al*. 1999; Paradis *et al*. 2005; Young and Park 2007). Being located at seven different loci around the genome (Fig. 3.1), the combination of these genes provides greater sequence diversity, compared to other traditionally used methods such as 16S rDNA sequencing. The genes had initially been selected by Baldwin *et al*. (2009), who had reported the d_n/d_s ratio for these seven genes to be significantly less than 1 for the species *C. sakazakii* and *C. malonaticus*. Similar results were obtained even after expanding the scheme across the genus (Table 3.2), thus reducing the possibility of selective pressure having influenced these loci. This also validated the use of the same seven loci for the entire genus.

The primers used for the amplification of the seven genes were designed to have similar melting temperatures. Therefore, all seven PCR reactions can be performed in a single thermal cycler run, thus minimizing the duration of the protocol as well. For a few of the strains (<5%), amplification issues were experienced at the *gltB* and *ppsA* loci, as a result of diversity in nucleotide sequence. These were mainly found to be *C. muytjensii* strains, a species which has exhibited considerable diversity in this study (Fig. 3.8), and has been further dealt with in Chapter 4. To overcome this problem, initially new primers had been designed which despite repeated optimisations, exhibited the formation of non-specific products causing errors in the sequencing process. Hence, it was decided to use the nested inner sequencing primers for

amplification purposes as well. This approach was found to be effective for these problematic strains.

The 325 *Cronobacter* spp. strains included in this study generated 115 STs to span the genus (Table 3.1; Fig. 3.2). None of the STs were found to be shared across the species, even though there were a few alleles shared (Fig. 3.10). Thus, the 7-loci MLST scheme enabled successful delineation between the *Cronobacter* species. This is an especially significant achievement, since most traditional MLST schemes are designed only at a single species level, while the *Cronobacter* MLST scheme is successfully able to cover the entire genus, despite the complexity associated with the diversity of this genus. Of the 115 STs, some were observed to be particularly dominant, such as STs 1, 4 and 8 in *C. sakazakii*; STs 7 and 11 in *C. malonaticus* and so on. The associations between the frequencies of the STs and the sources of the isolates have been studied and discussed in detail in Chapter 5.

3.3.2 Phylogenetic relationships of the *Cronobacter* species

The sequences of the seven loci of the 115 identified STs were used for the purpose of a phylogenetic evaluation of the genus. This was carried out by concatenating each set of seven gene sequences to obtain a 3036 bp long sequence. This phylogenetic analysis revealed a clear and robust clustering of the seven *Cronobacter* species (Fig. 3.3).

The phylogeny of the *Cronobacter* genus as well as closely related members of the *Enterobacteriaceae* family has been much investigated in a number of studies in recent years. The polyphyletic nature of the phylogeny of these groups of organisms observed using 16S rDNA sequence analysis has reflected their complex diversity (Iversen *et al*. 2004; Dauga and Breuwer 2008). Even though the use of 16S rDNA sequences has been effective to study distantly related organisms, it has had its share of discrepancies between the species of a genus such as *Cronobacter*, as can be seen in Fig. 3.4. Iversen *et al*. (2004) in their study on the phylogeny of the organism (then *E. sakazakii*) using 16S rDNA sequences, had reported the formation of four clusters within the members of the species, thereby setting the scene for the taxonomic revision of the species *E. sakazakii* into the genus *Cronobacter*. However, it was seen that Cluster 1 comprised of strains that were later identified as *C. sakazakii* and *C. malonaticus*, thus making these two species indistinguishable by this form of analysis. Baldwin *et al*. (2009) had made this differentiation possible through the use of the 7-loci MLST scheme, which is also clearly seen in the genus phylogeny in Fig. 3.3. These two species appear to be genetically very closely related to each other, with the *C. turicensis* and *C. universalis* clusters the nearest neighbours. The *C. dublinensis*, *C. muytjensii* and *C. condimenti* lineages appear to

be fairly distant from the rest of the genus. The former two also exhibited a significant amount of branching within the individual species clusters, providing more information about the genetic relationships between the isolates of a species, discussed in Chapter 4. Some of the clustering and branching in the *C. sakazakii* and *C. malonaticus* lineages could also be correlated to the epidemiology of the strains and has been analysed in further detail from the point of view of clonality and presented in Chapter 5.

The phylogenetic construction of the genus was also important from the point of view of accurate species-level identification of isolates corresponding to each of the STs. This is a very important requirement in *Cronobacter* spp. study, because of the existing confusions in strain identification, especially because of the taxonomic revisions from *E. sakazakii* to *Cronobacter* spp. The direct replacement of *E. sakazakii* with *C. sakazakii* has been an often observed error in some studies. An accurate species level identification is also significant since most of the clinical isolates of *Cronobacter* spp. have predominantly belonged to two species, *C. sakazakii* and *C. malonaticus.* Mis-identification between *Cronobacter* spp. and related organisms has been rampant especially among the 16S sequences submitted as *Cronobacter* spp. in the publicly available Genbank database. As part of the investigation of the diversity of the genus in this study, these sequences were often downloaded and reanalysed by myself to find that some were actually members of related genera such as *Citrobacter* spp. and *Erwinia* spp. (Results not presented here). In most cases, the misidentifications were found to be either due to poor quality sequences or due to not having used appropriate outliers for identification, an essential requirement for 16S rDNA sequence analysis of *Enterobacteriaceae* members because of the fewer variable characters. The MLST database is strictly curated and the sequences thoroughly checked for reliability and quality before being uploaded in the public domain, which ensure that these problems of discrepancies are overcome.

3.3.3 An evolutionary timescale for the genus *Cronobacter*

The diversity and branching observed in the *Cronobacter* MLST data analysis has led to an evolutionary study to estimate the age of the genus *Cronobacter* and divergence time of the branches for the individual species. This has been done based on the assumption that the molecular clock rate of *Cronobacter* spp. would be approximately similar to that of the other *Enterobacteriaceae* members such as *Salmonella* spp. and *E. coli.* Similar age-related studies using MLST data have previously been undertaken for organisms such as *Salmonella* spp. (McQuiston *et al*. 2008) and *Pseudomonas* spp. (Waine *et al*. 2009), among others, though an evolutionary study of this nature has not been carried out for *Cronobacter* spp. to date.

This analysis estimates the *Cronobacter* genus to have split from its closest ancestor in the *Enterobacteriaceae* family approximately 45 to 68 million years ago (Table 3.3). According to the International Commission on Stratigraphy (http://www.stratigraphy.org), this time period was the Palaeogene period of the Cenozoic era, immediately succeeding the Cretaceous period of the Mesozoic era. Flowering plants have been reported to have evolved towards the end of this Cretaceous period, and further widespread evolution of the plants is reported to have occurred during the Palaeogene period **(**Gradstein *et al*. 2008**)**. This coincides well with the suspected natural plant habitat of the organism, as has been proposed by Iversen and Forsythe (2003). It is notable that members of the *Cronobacter* spp. have also been isolated from flies (Mramba *et al*. 2006; Pava-Ripoll *et al*. 2012). It is also possible that the feeding of insect larvae on plants could have led to a host adaptation of the organism.

Within the genus, the earliest branches would have been those of the species *C. muytjensii* and *C. dublinensis* at 25-40 MYA, hinting in particular at their plant association. Recently, in a study characterizing the iron acquisition systems in the genomes of *Cronobacter* spp., Grim *et al*. (2012) reported the presence of plant pathogen associated siderophore receptors unique to *C. muytjensii* and *C. dublinensis*, and therefore they proposed an association of these species with plant-related origins. Also, based on the estimated ages of the individual species of the *Cronobacter* genus, the species that have been most associated with human clinical cases – *C. sakazakii*, *C. malonaticus* and *C. turicensis* – appear to have branched and evolved most recently. This suggests the possibility that the virulence of the organism could have been a more recently acquired trait during the evolution of the organism. Further investigation needs to be carried out for the accurate characterization of this hypothesis.

When compared with the ages of other members of the *Enterobacteriaceae* family estimated by similar MLST studies, the evolution of distinguishable *Cronobacter* species may have occurred over the same period as the divergence of the *Salmonella* species and subspecies, after its split from *E. coli.* McQuiston *et al*. (2008) have reported this split to have occurred 40 to 63 MYA, which correlates with the estimated age of the *Cronobacter* genus (Table 3.3).

3.3.4 Influence of recombination on the evolution of the genus

The species *C. muytjensii* and *C. dublinensis* showed diverse branching within each species in the phylogenetic analysis of the genus (Fig 3.3). A recombination analysis of the *Cronobacter* genus carried out using both Splitstree4 as well as RAT, suggested the possibility of a number of recombination events to have occurred in these two species. The Neighbour-Net graph generated by Splitstree (Fig. 3.8) clearly shows the high level of diversity in these

species, compared to the tight clustering of the *C. sakazakii* and *C. malonaticus* species. The large number of parallelogram formations suggests the strong possibilities of recombination events. According to the estimated divergence times (Table 3.3; Fig 3.7), it appears that the earliest branches of the genus would have been *C. dublinensis* and *C. muytjensii.* Hence, it can also be proposed that high levels of recombination could have led to the splits in the *Cronobacter* genus, which later stabilized and became clonal in the more recently evolved species such as *C. sakazakii* and *C. malonaticus.* There is a certain extent of limitation to this analysis since it assumes a uniform clock rate for the organism, which has been criticised in recent years by some researchers (Kuo & Ochman. 2009), and hence can only be termed as an estimated age factor. The other possible bias in this analysis is that *C. sakazakii* is the most dominant species of the dataset (and also the most frequently isolated species of the genus) and so influences the numbers of the population set. Nevertheless, this analysis is the first attempt to understand the evolution of the *Cronobacter* genus and should throw light on further studies into the diversity and virulence of the organism.

Analysis of the alleles of the individual loci also indicated the presence of recombination events, with 6 of the 7 alleles being affected by possible gene transfer events, mainly indicated by the sharing of alleles between the *Cronobacter* species as well as branching variations in the individual phylogenetic trees of the alleles. The *fusA* locus was found to be the most stable among the seven loci. The effect of recombination was found to be strongest at the *gltB* locus [Fig. 3.10 (d)]. However, a concatenated MLSA of the genus phylogeny after excluding this locus maintained the overall phylogenetic structure of the genus. This proves the necessity of multiple loci for studying the diversity and phylogeny of complex bacterial populations. A number of the allele sharing events occurred between *C. sakazakii* and *C. muytjensii*. This could be described as "replacing horizontal gene transfer" events, which involves the replacement of homologues between lineages. This phenomenon has been recently studied in the *Streptococcus* genus (Choi *et al*. 2012).

3.3.5 The use of the *fusA* locus for typing purposes

Among the seven loci used for the *Cronobacter* MLST scheme, the *fusA* gene was found to be the most stable across the genus, with none of the alleles being shared between species (Fig. 3.6). Currently, our research group has adopted this locus for initial single locus typing of the organism. One of the popular molecular targets for *Cronobacter* spp. identification has been the *rpoB* scheme developed by Stoop *et al*. (2009). However, the protocol includes specific primers for each species, therefore requiring seven different PCR reactions to be set up for identification. On the other hand, for the *fusA* locus, with the use of only two sets of primers

the nucleotide sequence for the strain can be obtained for the purpose of identification up to the species level. Also, a number of the molecular typing tools that have been proposed over the years have included validation with a very small population set. In the current study, the *fusA* locus has given a 100% success rate for all 325 *Cronobacter* spp. isolates, encompassing the wide diversity of the genus, and now continues to be used effectively by our research group.

3.3.6 Conclusions

Thus, the *Cronobacter* MLST scheme has proven to be an effective and robust typing tool for the genus and has exhibited a high level of discrimination between the isolates. PFGE continues to be used for outbreak analysis of clinical *Cronobacter* spp. strains; however this MLST scheme provides an independent form of analysis from PFGE. None of the restriction sites of the enzymes used for PFGE such as *XbaI* or *SpeI* lies within the seven genes of this MLST scheme. Recently, a serotyping scheme has also been in development for the genus *Cronobacter* (Jarvis *et al*. 2011). The current problems associated with the scheme have been discussed in Chapter 1, Section 1.1.7. A comparison of this scheme was carried out with the STs of select strains from the dataset in this study, to reveal a higher discrimination in the MLST results (Appendix Table 1).

With the rapidly falling prices of sequencing and ease of experimentation, the MLST technique does provide an added advantage over the more laborious methods such as PFGE. Also, the large amount of curated sequencing data obtained has additionally helped to reveal the considerable diversity present in the genus as well as the inter-species evolutionary relationships, an aspect that has warranted investigation in the *Cronobacter* genus.

The current set of seven genes used in this MLST study for *Cronobacter* was selected using the only genome sequence available for the genus at that time – *C. sakazakii* BAA-894. With the advances in sequencing technologies and increasing number of genome sequencing projects, there will be ample opportunities to improve upon the choice of loci, if needed, using *in silico* testing.

Apart from the results presented in this chapter, there were two major significant outcomes over the course of this MLST study that have been detailed in the following chapters – the classification of the two latest *Cronobacter* species, *C. universalis* and *C. condimenti* (Chapter 4) and the identification of a genetic signature for *C. sakazakii* neonatal meningitis – ST4 (Chapter 5).

CHAPTER 4

TAXONOMIC RE-EVALUATION OF *CRONOBACTER* SPECIES BASED ON MULTILOCUS SEQUENCE ANALYSIS

4.1 Introduction

4.1.1 Aims of the chapter

This chapter presents how MLSA using the 7 loci (*atpD, fusA, glnS, gltB, gyrB, infB, ppsA*) of the *Cronobacter* MLST scheme, was applied in the taxonomic re-evaluation studies of the genus *Cronobacter.* The results and analysis presented here were obtained in a 2-part study carried out over the course of the year 2011. The first part of the study, conducted in early 2011, resulted in the official classification of two novel *Cronobacter* species, *C. universalis* and *C. condimenti,* and this has now been published as Joseph *et al*. (2012a). The second part of the study was a similar investigation into the diversity of the two species, *C. muytjensii* and *C. dublinensis,* to identify potential new species.

Both the above mentioned studies were co-operative projects, with different members of the co-operating groups contributing to specific sections of the overall study. For the *C. universalis* - *C. condimenti* study, the phenotypic characterization of the strains was performed by Esin Cetinkaya (Ankara University, Turkey) as part of her MSc placement at the NTU laboratory, while for the *C. muytjensii – C. dublinensis* study, the same was performed by Arturo Levican (Rovira i Virgili University, Spain) as part of his PhD placement at the NTU laboratory. The DDH experiments for both studies were performed in the laboratories of Prof. M.J. Figueras in Spain. The 16S rDNA and MLST experiments and sequence analysis for both studies were performed by myself and hence only these details have been described in this chapter.

4.1.2 Concept of bacterial species

In the light of the following chapter, it is pertinent to first consider current approaches and complications in defining bacterial species. Attempts to define the taxonomic classification of bacteria were first proposed as early as the latter half of the 19th century and since then researchers have been working towards accurate identification and classification of various bacterial groups (Cohn 1875). However, with the advances in molecular techniques, better characterisation of bacterial strains has become possible and taxonomists have now proposed certain criteria that need to be fulfilled for defining the bacterial species for a particular isolate. This currently comprises a polyphasic taxonomic approach, with a combination of phylogenetic (DDH and 16S rDNA sequencing) and phenotypic characterization. The approved criteria for a bacterial strain to belong to a novel species are: it must show <70% DNA-DNA reassociation values and <97% 16S rDNA gene sequence similarity with the type strains of the other closely related and well established species in the said genus. Also, phenotypically, it must exhibit the

variation of at least a single phenotypic trait (Wayne *et al*. 1987; Murray *et al*. 1990; Tindall *et al*. 2010). The 16S rDNA gene sequence similarity values are used more as an indicator than definitive value, which always needs to be confirmed by the DDH values. Also, keeping in mind that the regions of variation in these genes are located in certain hotspots, it is important to incorporate as much of the sequence into the analysis to obtain accurate results. Therefore, 16S rDNA sequences of 1300-1400 bp are ideal candidates for use in speciation and taxonomic studies. For phenotypic characterization, cell morphology, physiology and biochemical traits are considered for taxonomic purposes.

In recent times, MLSA of housekeeping genes in the genome has also been incorporated as a part of the speciation process, as they indicate more robust discrimination than 16S rDNA gene sequences (Alperi *et al*. 2010; Brady *et al*. 2012). This has been especially facilitated by the establishment of MLST databases for a number of organisms. With the bacterial genome sequencing boom in recent years, this is now also being extended at a whole genome level. These multilocus sequence and whole genome sequence similarity values can be very useful to complement the DDH values, especially since the latter can be subject to a number of experimental limitations.

Nucleotide sequence analysis for taxonomy and speciation studies is also very important and useful in the phylogenetic, evolutionary and recombination analysis of the said group of organisms – factors that may also play their role in the formation of species but cannot be accounted for using methods such as DDH and phenotypic characterization. When a novel species or taxon has been successfully identified based on these criteria, its nomenclature has to be followed according to the Bacteriological Code (Lapage *et al*. 1992). The code also requires that the type strain for the species be identified and submitted to at least two publicly maintained bacterial culture collections. The novel taxonomic study based on these criteria is published in the *International Journal of Systematic and Evolutionary Microbiology*, which is considered to be "the official journal of record for novel prokaryotic taxa" (http://ijs.sgmjournals.org/site/misc/about.xhtml).

4.2 Results

4.2.1 Identification and classification of *C. universalis* and *C. condimenti* – two novel *Cronobacter* species

In January 2011, during the course of the expansion of the 7 loci MLST scheme across the *Cronobacter* genus (Chapter 3), phylogenetic sequence analysis of the concatenated sequences of the seven loci of the *Cronobacter* dataset had revealed five strains of particular interest based on their clustering pattern within the genus - one was a group of four strains (NTU strains 96, 581, 731, 1435) and the other a single strain (NTU strain 1330).

As these represented potential novel species, a polyphasic study was carried out to conduct a taxonomic re-evaluation of the genus. This study included a combination of MLSA of the seven housekeeping genes, 16S rDNA gene sequencing, DNA-DNA hybridization experiments and phenotypic characterization, in order to establish the taxonomic position of these strains in the genus *Cronobacter*.

Among the cluster of four strains, NTU strain 581 had been previously characterized and assigned as *Cronobacter* genomospecies 1 (Iversen *et al*. 2008). Strain 581 was a freshwater isolate from UK. Strain 731 was isolated in 2005 from a post-operative mixed infection with *S. aureus* of a 9-year-old boy with humeral fracture treated by osteosynthesis using intra-medullar nailing. NTU strains 96 and 1435, isolated from onion powder and rye flour, respectively, had been previously assigned as *C. turicensis.* NTU strain 1330, isolated from spiced sausage meat in Slovakia, had been identified in a separate study as a member of the genus *Cronobacter* and was assigned to biogroup 1 by Turcovsky *et al*. (2011).

Phylogenetic analysis using the concatenated sequences of the seven MLST loci of these strains along with the type strains of all recognized *Cronobacter* species, revealed the four *Cronobacter* strains - 96, 581, 731 and 1435 - to form an independent clade within the genus, with the closest neighbour being *C. turicensis.* Strain 1330 was found to form an independent branch, with the closest neighbour being *C. dublinensis* (Fig 4.1). *Cit. koseri* was used as an outliner to root the tree.

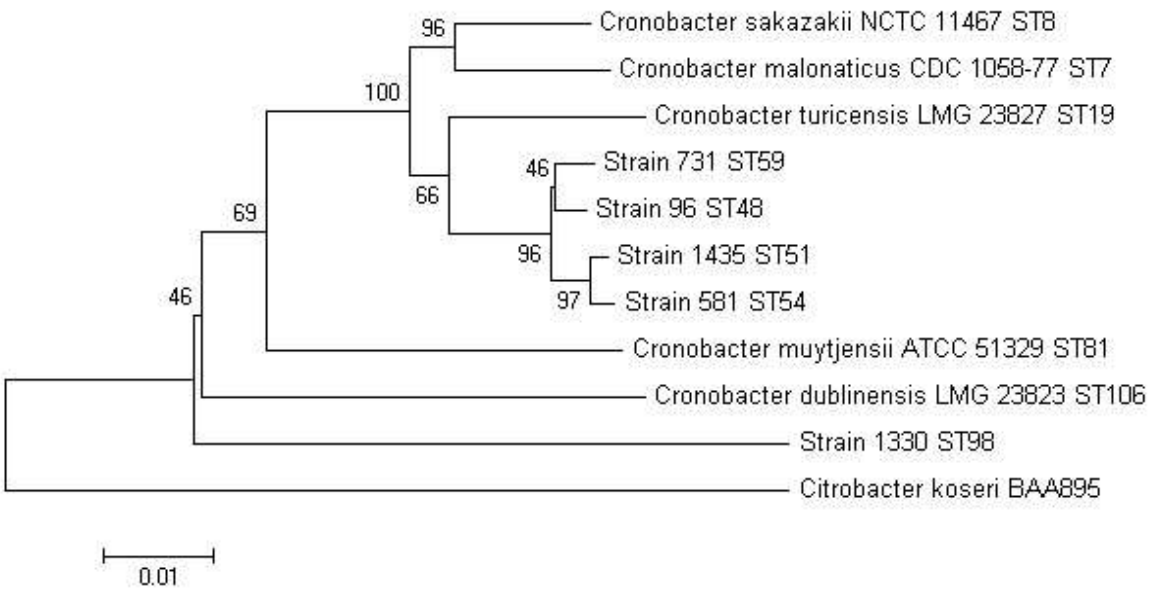


Fig. 4.1 Maximum – likelihood tree constructed based on the concatenated *atpD, fusA, glnS, gltB, gyrB, infB,* and *ppsA* gene sequences (3036 bp), showing the phylogenetic position of the strains 96, 581, 731, 1435 and 1330 within the *Cronobacter* genus. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. *Cit. koseri* has been used as an outliner to root the tree. The numbers at the nodes indicate the bootstrap values expressed in percentage.

The nucleotide divergence values within the genus of these strains was analysed by constructing a pairwise sequence distance matrix (Table 4.1), using the concatenated sequences (3036 bp) of the seven MLST loci. The strains 96, 581, 731 and 1435 showed ≥99% similarity among themselves, while their nucleotide divergence from the other recognized *Cronobacter* species varied from 2.9 to 7.7 %. The minimum nucleotide divergence observed between the other established *Cronobacter* species, was 2.8 % between *C. sakazakii* and *C. malonaticus.* Since the divergence values for the cluster of four strains with the rest of the genus was considerably higher than this value, this indicated a strong possibility of this cluster being a novel *Cronobacter* species.

	C. malonaticus **CDC 1058-77^T** **ST7**	***C. turicensis*** **LMG 23827^T** **ST19**	***C. dublinensis*** **LMG 23823^T** **ST106**	***C. muytjensii*** **ATCC 15329^T** **ST81**	**Strain 581** **ST54**	**Strain 1435** **ST51**	**Strain 96** **ST48**	**Strain 731** **ST59**	**Strain 1330** **ST 98**	***Cit. koseri*** **BAA-895** **ST6**
C. sakazakii **NCTC 11467^T** **ST8**	**2.8**	**4.2**	**5.8**	**7.4**	**3.8**	**3.7**	**3.4**	**3.3**	**8.7**	**12.3**
C. *malonaticus* **CDC 1058-77^T** **ST7**		**4.5**	**6.5**	**7.6**	**3.5**	**3.7**	**3.2**	**2.9**	**8.7**	**12.4**
C. turicensis **LMG 23827^T ST19**			**6.8**	**7.6**	**3.1**	**2.9**	**3.3**	**3.4**	**8.6**	**12.2**
C. dublinensis **LMG 23823^T** **ST106**				**6.8**	**6.2**	**6.3**	**5.8**	**5.8**	**8.5**	**13.2**
C. muytjensii **ATCC 51329^T** **ST81**					**7.7**	**7.6**	**7.5**	**7.5**	**8.5**	**12.8**
Strain 581 **ST54**						**0.4**	**1.0**	**1.0**	**8.8**	**12.6**
Strain 1435 **ST51**							**0.8**	**1.0**	**8.6**	**12.6**
Strain 96 **ST48**								**0.7**	**8.4**	**12.5**
Strain 731 **ST59**									**8.5**	**12.6**
Strain 1330 **ST 98**										**13.6**

Table 4.1 Nucleotide divergence values (expressed in %, n=3036) of the strains 1330, 96, 581, 731 and 1435 (highlighted in yellow) with the types strains of the *Cronobacter* species, calculated using the concatenated sequences of the seven MLST loci. *Cit. koseri* has been used as an outliner species.

Similarly, the 16S rDNA sequence phylogenetic tree of these strains and the type strains of the individual *Cronobacter* species, also showed the four strains 581, 96, 731 and 1435 to be placed together within the *Cronobacter* genus as an independent cluster. Strain 1330 formed an independent lineage, with its closest neighbour on this tree being *C. muytjensii*. (Fig. 4.2)

As reported before in Chapter 3, the 16S rDNA sequences showed considerably less sequence diversity within the *Cronobacter* genus, with the sequence similarity values ranging from 98.2% to 99.9% across the species. The cluster of the strains 96, 581, 731 and 1435 was found to be closest to *C. turicensis* LMG 23827 ^{T} at a mean sequence similarity of 99.7%. Strain 1330 showed highest sequence similarity with *C. turicensis* LMG 23827 ^{T} and *C. dublinensis* LMG 23823 ^{T} at 98.7%. Even the outliner species *Cit. koseri* exhibited a fairly high mean similarity of 97.3% with the *Cronobacter* genus.

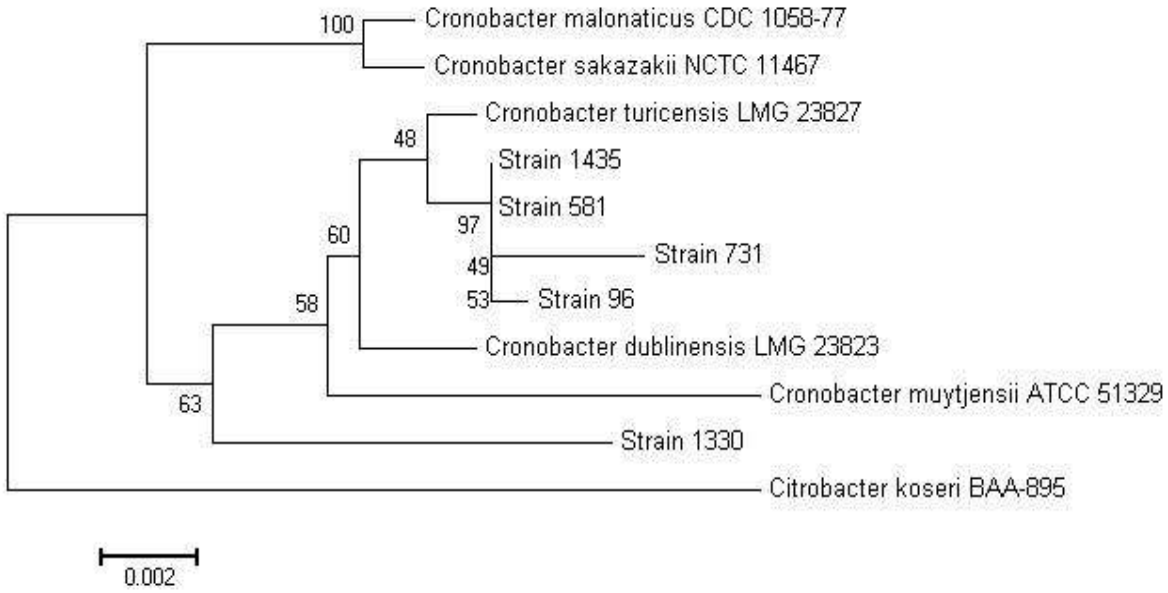


Fig. 4.2 Maximum – likelihood tree constructed based on the full length 16S rDNA gene sequences (1361 bp), showing the phylogenetic position of the strains 96, 581, 731, 1435 and 1330 within the *Cronobacter* genus. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. *Cit. koseri* has been used as an outliner to root the tree. The numbers at the nodes indicate the bootstrap values expressed in percentage.

DDH (direct and reciprocal) experiments were performed between strain 1330 and strain 581 and between the type strains of the currently accepted *Cronobacter* species. The experiments were performed by our collaborators Arturo Levican and Dr. Maria-Jose Figueras in their laboratories in Spain, using the method described by Urdiain *et al*. (2008). All results were below the 70% limit required for the definition of a new species within a genus (Wayne *et*

al. 1987; Stackebrandt & Goebel. 1994). The DDH values have been presented in Appendix Table 2.

The phenotypic characterisation of these strains was carried out in the NTU laboratory by Esin Cetinkaya, a placement student from Turkey as part of her MSc project plan. The phenotypic tests evaluated on strains 581, in the present study were selected from Iversen *et al*. (2006a, 2006b, 2007, 2008) and were the following: catalase and oxidase activity, nitrate reduction, acid production from sugars, malonate utilisation, production of indole from tryptophan, motility, gas from D-glucose, Voges-Proskauer (VP), methyl red, α-glucosidase activity, pigment production on TSA (21 and 37°C), aerobic and anaerobic growth on TSA (37°C), growth on MacConkey agar, and hydrolysis of DNA. The role of these tests in the taxonomic re-evaluation from the species *E. sakazakii* to the genus *Cronobacter* has been presented in Chapter 1, Table. 1.2.

Strain 1330 was found to be biochemically different from all other species of the genus *Cronobacter* by at least 6 different tests (Appendix Table 3). As mentioned earlier, using the criteria of Farmer *et al.* (1980), this strain had been classified as *C. sakazakii* (Biogroup 1), but on the basis of the present results it can be differentiated from this species biogroup because it is non-motile, by the ability to produce indole from tryptophan, to utilize malonate and the lack of production of acid from turanose, inositol, lactulose, putrescine, *cis*-aconitate, 4-aminobutyrate, maltitol and palatinose. It was therefore reclassified as a new biogroup, 10a. Strain 1330 was relatively similar to strains 581, 731, 96, and 1435 but could also be differentiated from them by several tests i.e. indole production, and non acid production from dulcitol, melezitose, inositol, lactulose or maltitol.

Strains 581, 731, 96, and 1435 were found to be biochemically similar to each other and could be differentiated from the other species of the genus *Cronobacter* by at least 3 tests (Appendix Table 3). They can be differentiated from the species *C. turicencis* (biogroups 16, 16a and 16b) because they do not produce acid from turanose, putrescine, or 4-aminobutyrate.

The polyphasic approach using the 16S rRNA gene sequencing, MLSA, DDH results and phenotypic characterisation all clearly differentiated strains 96, 581, 731, 1435 and 1330 from the existing *Cronobacter* species and therefore *Cronobacter condimenti* (type strain 1330^T, =LMG 26250^T, =CECT 7863^T) and *Cronobacter universalis* (type strain NCTC 9529^T, =CECT 7864^T, =LMG 26250^T) were proposed as new species and have been officially published as Joseph *et al*. (2012a).

Description of *Cronobacter condimenti* sp. nov. (Joseph *et al.* 2012a)

Cronobacter condimenti (con.di.men'ti. L. gen. n. condimenti, of spice, seasoning)

Cells of strain 1330 are straight, Gram-stain-negative, non-spore-forming rods, non-motile. Colonies on TSA incubated at 37°C for 24h are 2-3 mm in diameter, opaque, circular and pigmented yellow in colour at 37°C. Strain 1330 grows on MacConkey agar. Optimal growth occurs at 37°C after 24 h in TSB and also at 45°C but no growth is observed at 5°C. No haemolysis is observed on sheep blood agar at 37°C. The strain produces catalase, α-glucosidase, β-galactosidase, and DNase activities, indole from tryptophan, acetoin (Voges-Proskauer positive), hydrolyses gelatine and reduces nitrate. Strain 1330 does not produce oxidase activity, does not produce hydrogen sulphide, does not hydrolyse urea, and does not produce gas from glucose. Strain 1330 is able to utilize ornithine, citrate, lysine, and malonate produce acid from 1-0-methyl α– D-glucopyranoside, glycerol, L-arabinose, ribose, D-xylose, galactose, D-glucose, D-fructose, D-manose, L-rhamnose, D-mannitol, N-acetyl glucosamine, arbutin, esculin, salicin, cellobiose, maltose, lactose, melibiose, sucrose, trehalose, myo-inositol, D-raffinose, β-gentiobiose, D-fucose, L-fucose or galacturonic acid. The strain does not ferment dulcitol, inositol, melezitose, turanose, lactulose, putrescine, *cis*-aconitate, *trans*-aconitate, 4-amino-butyrate, maltitol, palatinose, D-erythritol, D-arabinose, adonitol, β-methyl-D-xyloside, L-sorbose, dulcitol, sorbitol, α-methyl-D-glucoside, inulin, glycogen, xylitol, D-lyxose, D-tagatose, D-arabitol, L-arabitol, gluconate, or (2&) 5-ketogluconate. The API 20E and ID32E profiles obtained for strain 1330 were 3367373 and 342137610030 respectively.

The type strain is 1330 (=CECT 7863^T, = LMG 26250^T), isolated from spiced meat purchased in Slovakia.

Description of *Cronobacter universalis* sp. nov. (Joseph *et al.* 2012a)

Cronobacter universalis (u.ni.ver.sa'lis. L. masc. adj. universalis, of or belonging to all, universal)

Cells of strains NCTC 9529^T, 731, 96 and 1435 are straight, Gram-stain-negative, non-spore-forming rods, variable motility (strain NCTC 9529^T is non-motile). Colonies on TSA incubated at 37°C for 24h are 2-3 mm in diameter, opaque, circular and pigmented yellow in colour at 37°C. All strains grow on MacConkey agar. Optimal growth occurs at 37°C after 24h in TSB and also at 45°C, but no growth is observed at 5°C. No haemolysis is observed on sheep blood agar at 37°C. Strains produce acetoin (Voges-Prokauer positive), catalase, α-glucosidase, β-galactosidase and DNase activities, and reduce nitrate. Strains do not produce oxidase activity, do not produce indole from tryptophan, do not produce hydrogen sulphide, hydrolyse gelatine or urea. Strains are able to use malonate, ornithine or citrate and to produce acid from

glucose, 1-0-methyl α-D-glucopyranoside, dulcitol, inositol, melezitose, lactulose, sucrose, L-arabinose, cellobiose, lactose, myo-inositol, L-rhamnose, D-mannitol, N-acetyl glucosamine, salicin, maltitol, D-fucose, amygdaline or galacturonic acid, but not able to produce acid from turanose, D-sorbitol, putrescine, *trans*-aconitate, L-fucose, adonitol, 5-ketogluconate or 4-aminobutyrate. Variable results are obtained for the acid production from *cis*-aconitate (NCTC 9529^T is negative), palatinose (NCTC 9529^T is negative), production of gas from glucose (NCTC 9529^T is negative). The API 20E and ID32E profiles obtained for strains NCTC 9529^T were 3205373 and 24276777051 respectively.

The type strain is NCTC 9529^T (= CECT 7864^T, =LMG 26250^T), isolated from freshwater and deposited at the NCTC (London) in 1954.

4.2.2 Evaluating the classification and diversity of the species, *C. muytjensii* and *C. dublinensis*

Phylogenetic analysis of the *Cronobacter* MLST dataset carried out using the concatenated sequences of the seven MLST loci as well as recombination analysis using Splitstree4 had shown that the species clusters *C. muytjensii* and *C. dublinensis* had considerably diverse branching patterns (Chapter 3, Fig 3.2), wherein some of the branches appeared to be more genetically distant from the main cluster. This diversity was further investigated to consider if further novel *Cronobacter* species could be defined.

Table 4.2 shows the nucleotide divergence percentage values for the type strains of the established *Cronobacter* species using the concatenated sequences of the seven MLST loci (3036 bp). The least value as reported before was 2.8% between *C. sakazakii* (NCTC 11467^{T}; ST 8) and *C. malonaticus* (CDC 1058-77^{T}; ST 7) and therefore this was used as a cut-off value to identify potential candidates for new species within *C. muytjensii* and *C. dublinensis.* Tables 4.3 and 4.4 show the DNA divergence values calculated for the members of these two species using the concatenated 7-loci sequences, indicated by their sequence types (ST). The DNA divergence values within these species, which were found to be above the cut-off value of 2.8, have been highlighted in Tables 4.3 and 4.4.

Based on these divergence values, the *C. muytjensii* species indicated two main clusters: Cluster I: STs 28, 30 & 94; Cluster II: STs 44, 49, 71, 75, 81 & 82 – plus a lone branch for ST34.

The *C. dublinensis* species divergence, on the other hand, exhibited a far more complicated diversity. The values did indicate two main clusters.
Cluster I: STs 38, 39, 74, 95 & 106; Cluster II: STs 36, 43, 70, 76, 77, 78, 79 & 80.

However in addition, there were four branches that showed conflicting clustering: STs 27, 88, 91 and 92. Of these, ST27 was strain BQ8, a Chinese isolate for which only the MLST sequences had been sent to us for analysis. Repeated requests for a viable culture of the strain to the sequence submitters were in vain, and hence it had to be excluded from any laboratory experiments by our group.

This clustering relationship has been depicted phylogenetically in Fig. 4.3, in a maximum-likelihood tree of the *Cronobacter* genus, using the concatenated 7 MLST loci sequences. The strains in each ST of the species *C. muytjensii* and *C. dublinensis* have also been indicated on each branch. The *C. dublinensis* strains that could be identified to the sub-species level have also been indicated.

Based on the indicative results obtained from this *in silico* analysis, biochemical characterization of the members of these two species were carried out in the NTU laboratory by Arturo Levican during his 6 month placement, in continuation of our previous collaborative taxonomic study. However, detailed phenotypic testing of these clusters of isolates, on the same lines of the one conducted in the previous study, failed to produce a single biochemical trait that could distinguish one from the other. Since current taxonomic standards for recognising new species by the *International Journal for Systematic and Evolutionary Microbiology* require a polyphasic analysis including DNA-DNA hybridization and phenotypic testing for biochemical traits, it was not possible to conclusively re-evaluate the taxonomy of these two species.

Organism	***C. malonaticus*** **CDC 1058-77^T** **ST7**	***C. turicensis*** **LMG 23827^T** **ST19**	***C. dublinensis*** **LMG 23823^T** **ST106**	***C. universalis*** **NCTC 9529^T** **ST54**	***C. condimenti*** **LMG 26250^T** **ST98**	***C. muytjensii*** **ATCC 51329^T** **ST81**
C. sakazakii **NCTC 11467^T** **ST8**	**2.8**	4.2	7.5	3.8	8.7	7.4
C. malonaticus **CDC 1058-77^T** **ST7**		4.5	7.6	3.5	8.8	7.7
C. turicensis **LMG 23827^T** **ST19**			7.4	3.1	8.7	7.6
C. dublinensis **LMG 23823^T** **ST106**				7.3	7.9	5.0
C. universalis **NCTC 9529^T** **ST54**					8.9	7.8
C. condimenti **LMG 26250^T** **ST98**						8.5

Table 4.2 Nucleotide divergence values (%) of type strains of the *Cronobacter* species, calculated using the concatenated sequences of the seven *Cronobacter* MLST loci (n=3036 bp). The least divergence value of 2.8 observed between *C. sakazakii* and *C. malonaticus* has been highlighted in yellow, and was used as an indicative cut-off value for the remaining analysis. doi:10.1128/JCM.00905-12.

C. muytjensii	**ST 30**	**ST 34**	**ST 44**	**ST 49**	**ST 71**	**ST 75**	**ST 81**	**ST 82**	**ST 94**
ST 28	0.7	2.5	1.6	**2.9**	**2.9**	**3.2**	2.6	1.9	1.1
ST 30		2.7	1.7	**3.1**	**3.1**	**3.3**	**2.8**	2.2	1.4
ST 34			**3.4**	**4.4**	**4.4**	**4.5**	**4.4**	**3.3**	**3.0**
ST 44				1.7	1.5	1.7	1.1	**3.0**	2.2
ST 49					0.5	0.7	0.5	1.9	**2.9**
ST 71						0.5	0.4	1.8	**2.8**
ST 75							0.6	2.0	**3.2**
ST 81								1.8	**2.8**
ST 82									2.0

Table 4.3 Nucleotide divergence values (%) for the individual sequence types of the *C. muytjensii* species**,** calculated using the concatenated sequences of the seven *Cronobacter* MLST loci (n=3036 bp). The divergence values above the cut-off value of 2.8 have been highlighted.

C. dublinensis	ST 36	ST 38	ST 39	ST 43	ST 70	ST 74	ST 76	ST 77	ST 78	ST 79	ST 80	ST 88	ST 91	ST 92	ST 95	ST 106
ST 27	**3.5**	**4.8**	**4.7**	2.0	**3.1**	**4.0**	**3.4**	**3.3**	**3.5**	**3.1**	**3.6**	**4.1**	**3.4**	**3.9**	**4.1**	**4.7**
ST 36		**3.0**	**3.0**	1.9	1.3	2.3	1.3	1.1	1.4	1.4	0.0	2.1	**3.3**	**2.9**	2.2	**3.2**
ST 38			0.1	**6.0**	**2.8**	1.2	**2.9**	**3.1**	**2.9**	**2.8**	**3.0**	**4.3**	2.6	**4.3**	1.6	**3.2**
ST 39				**3.2**	2.7	1.1	**2.8**	**3.0**	**2.9**	**2.8**	**3.0**	**4.4**	2.6	**4.3**	1.5	**3.1**
ST 43					1.5	2.6	1.9	1.8	2.0	1.7	1.9	2.4	**3.5**	2.2	2.6	**3.2**
ST 70						2.1	1.3	1.3	1.3	1.0	1.3	**2.8**	**3.0**	**2.8**	2.1	**3.2**
ST 74							2.1	2.2	2.1	2.2	2.3	**3.6**	**3.6**	**3.6**	0.5	2.0
ST 76								1.2	0.0	1.2	1.4	**3.0**	**3.2**	2.6	1.9	**3.0**
ST 77									1.2	1.2	1.2	**2.9**	**3.4**	2.7	2.2	**3.2**
ST 78										1.2	1.4	**3.0**	**3.2**	2.6	2.0	**3.1**
ST 79											1.4	**3.0**	**6.5**	**2.8**	2.1	**3.2**
ST 80												2.0	**3.3**	**2.9**	2.2	**3.3**
ST 88													**4.8**	**3.1**	**3.5**	**4.1**
ST 91														**4.5**	**3.6**	**4.9**
ST 92															**3.6**	**4.3**
ST 95																2.0

Table 4.4 Nucleotide divergence values (%) for the individual sequence types of the *C. dublinensis* species, calculated using the concatenated sequences of the seven *Cronobacter* MLST loci (n=3036 bp). The divergence values above the cut-off value of 2.8 have been highlighted. STs 36 and 70 represent *C. dublinensis* subsp. *lausannensis*, and ST79 represents *C. dublinensis* subsp. *lactaridi.* All the remaining STs belong to *C. dublinensis* subsp. *dublinensis.*

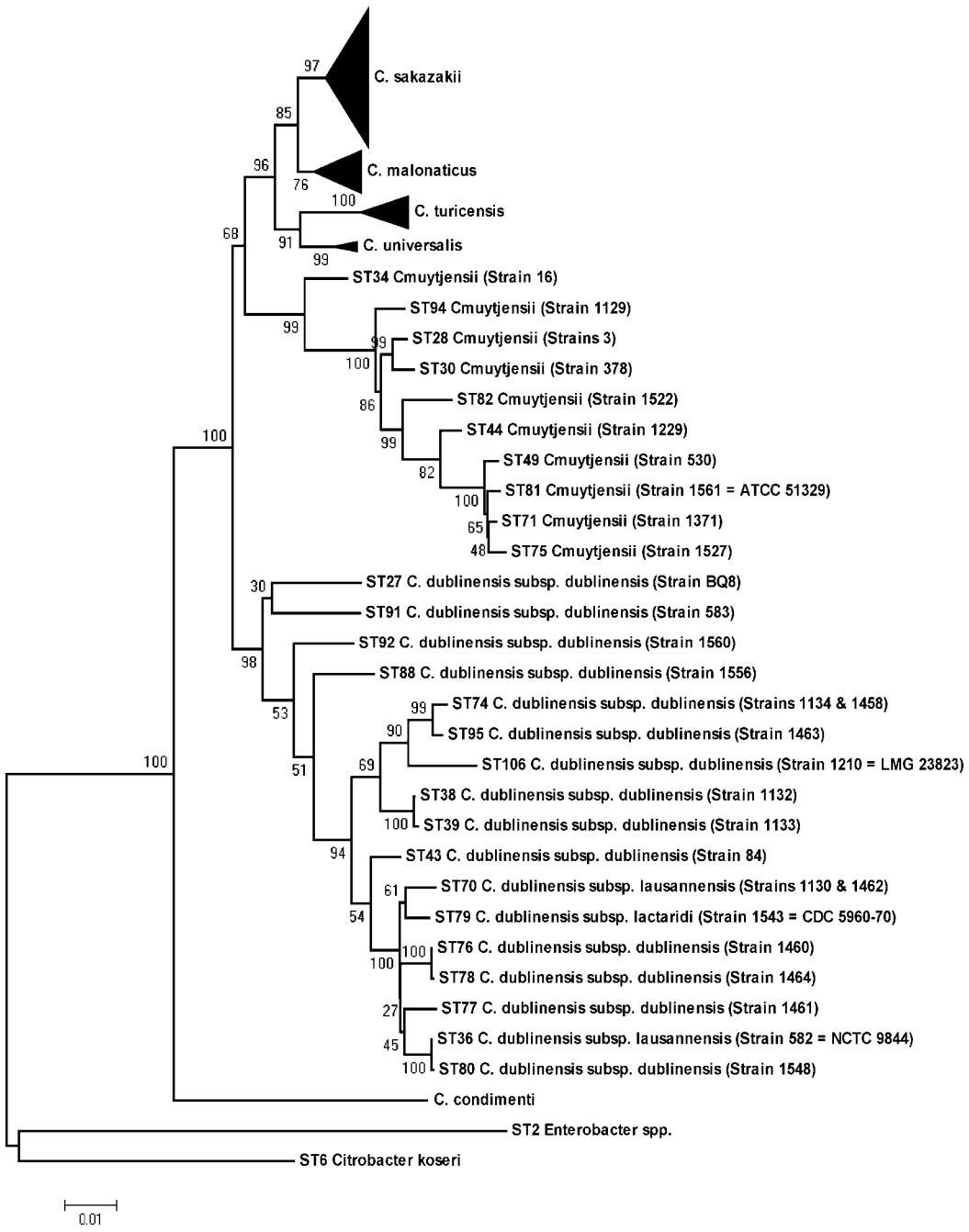


Fig. 4.3 Maximum-likelihood tree constructed using the concatenated sequences (3036 bp) of the 7 *Cronobacter* MLST loci indicating the diversity and clustering within the species *C. muytjensii* and *C. dublinensis*. The tree has been drawn to scale using MEGA5, with 1000 bootstrap replicates. The remaining *Cronobacter* species clusters have been compressed for clarity. The bases of their triangles indicate the number of isolates used in the analysis, while the heights indicate the diversity of each branch. The numbers at the nodes indicate the bootstrap values expressed in percentage.

4.3 Discussion

The diversity of the genus *Cronobacter* has been particularly evident by the number of taxonomic revisions it has undergone from the days of being identified as "yellow pigmented *E. cloacae*" to the present day genus *Cronobacter* as we know it (Farmer *et al*. 1980; Iversen *et al*. 2007 & 2008). At the beginning of this study, there were only five established *Cronobacter* species – *C. sakazakii, C. malonaticus, C. turicensis, C. dublinensis* and *C. muytjensii* – along with a *C.* genomospecies1, which then comprised of only two isolates (NTU strains 581 and 731), and was separately grouped as a Biogroup 16c, because of insufficient evidence for a new species (Iversen *et al*. 2008). Through MLSA, two more strains for this bacterial group was found, one of which (NTU strain 96) had previously been misidentified by biochemical tests, while the other was a fresh isolate from rye flour in Turkey (NTU strain 1435). A thorough investigation according to the criteria for defining novel bacterial species resulted in this cluster of strains to be redefined as a novel *Cronobacter* species, *C. universalis*, the nomenclature being attributed to the wide range of sources the four strains had been isolated from. The study also comprised of the classification of a single *Cronobacter* spp. strain as a hitherto unidentified species, *C. condimenti*, thus expanding the *Cronobacter* genus to seven species.

These results exhibit how the MLSA can and should play a role in the bacterial speciation and taxonomy studies. Current criteria that have been laid down for prokaryotic taxonomy, state the need for a polyphasic approach involving DDH, 16S rDNA gene sequence analysis and phenotypic characterization for the identification of novel species. However, this study has shown how using multiple protein coding gene sequences can be more robust than the use of 16S rDNA sequences which show far less diversity in comparison. A classic example for this is strain 1330, the sole member of the novel *C. condimenti* species. The strain when first isolated from spiced meat was identified solely as *Cronobacter* spp. (Turcovsky *et al*. 2011), based on biochemical testing and 16S rDNA analysis. This indicated the uncertainty of a species level identification of the *Cronobacter* strain using these methods. In contrast, when tested according to the 7 loci *Cronobacter* MLST scheme, this clearly showed the strain to be an independent lineage and potential novel species, which was further confirmed by DDH. This further stresses the importance of the application of the most appropriate typing methods for an organism for its accurate identification, in the absence of which key isolates of a species can get overlooked due to misidentification. Even though the species *C. condimenti* has now been officially established, it still comprises of only a single isolate. Hopefully MLST will become a more established practice for *Cronobacter* researchers and further members of the species will be identified making the taxonomy of this genus more robust.

A similar taxonomic approach was applied in the second part of this study where the diversity observed in the branching of the species *C. muytjensii* and *C. dublinensis* (Fig. 4.3)

was investigated for potential novel species. Considering the very low level of divergence observed with the 16S rDNA gene sequences within the *Cronobacter* genus in the first study, this time only the concatenated sequences (3036 bp) of the seven *Cronobacter* MLST loci were used for the initial screening in the *C. muytjensii – C. dublinensis* study. Even though MLSA indicated the nucleotide divergence between certain members of these species to be considerably higher than that of some of the well established *Cronobacter* species, extensive phenotypic testing of these strains failed to exhibit any traits that could differentiate them, according to the clustering patterns observed according to the MLSA phylogeny. Hence, at present, the results for this investigation does not support the definition of new *Cronobacter* species within the currently defined species *C. muytjensii* and *C. dublinensis,* despite their exhibited diversity.

In spite of this, the presence of the diversity and variability within these species is undeniable. Apart from the main clusters, both species indicated the presence of branches which were genetically distant from the remaining members, and showed conflicting correlation between the nucleotide divergence values and the clustering on the phylogenetic tree (Tables 4.3 & 4.4; Fig. 4.3). These included ST34 in *C. muytjensii* and STs 27, 88, 91 and 92 in *C. dublinensis.* This uncertainty of the species distinction of certain strains or sequence types has also been observed in the MLSA studies in the organism *Neisseria* spp., and this concept was in that case described as "fuzzy species" (Hanage *et al*. 2005).

Another notable trait for both these species is the fact that most of the STs of *C. muytjensii* and *C. dublinensis* comprise of only single strains, the only exceptions being STs 70 and 74 in *C. dublinensis.* This shows the lack of clonality and possible strong role of recombination in the formation of these species. This aspect has also been noted in Chapter 3, where a number of the alleles shared between the species were those of *C. muytjensii* and *C. dublinensis* (Appendix Fig. 1 to 6).

A better understanding of the diversity within these species will be possible only when as many isolates as possible can be obtained to represent each of the two species. Of the various species in the genus *Cronobacter*, *C. muytjensii* and *C. dublinensis* are among lesser isolated and poorly characterized species. This current dataset itself was built over the course of the study through literature reviews and by contacting collaborators. A number of the *C. dublinensis* strains had been obtained for this study after they were spotted as 16S rDNA sequences in Genbank, submitted as *Cronobacter* species, indicating the difficulty involved in their accurate identification. This also brought to light the poor curation and lack of quality control involved in Genbank submissions, and hence their identifications must also be handled warily, until verified.

Previously, during the definition of the genus *Cronobacter*, the species *C. dublinensis* had been further broken down into three subspecies - *Cronobacter dublinensis* subsp.

dublinensis, *Cronobacter dublinensis* subsp. *lausannensis* and *Cronobacter dublinensis* subsp. *lactaridii* (Iversen *et al*. 2008). This sub-speciation was based on the variations in their biochemical profiles for the utilization of lactulose, malonate, maltitol, inositol and melezitose; despite DDH values showing them to be the same species (Table 1.2). However, the variation observed by MLSA seemed to go beyond this classification, as can be seen by the distribution of these sub-species within the branching of the species, in Fig. 4.3. STs 36 and 70 represent *C. dublinensis* subsp. *lausannensis,* and ST79 represents *C. dublinensis* subsp. *lactaridii*. All three of these STs appear to be within the same cluster, while the STs representing *C. dublinensis* subsp. *dublinensis* were found to be more scattered. Hence, appropriate biochemical test results could not be obtained to correlate with the clusters observed in the MLSA. It is also important to point out that certain discrepancies were observed in the classification of these sub-species by Iversen *et al*. (2008). Firstly, their DDH values had been calculated in comparison to the type strain of *C. dublinensis* subsp. *lausannensis* (NCTC 9844^{T}), instead of the type strain of the type species, *Cronobacter dublinensis* subsp. *dublinensis* (LMG 23823^{T}). Also, the DDH % value between the type strains of *C. dublinensis* subsp. *lactaridii* and *C. dublinensis* subsp. *lausannensis* was reported to be 77.4, with a high standard deviation of 7.5. This value being quite close to the species cut-off of 70% does make the accuracy of this sub-speciation questionable.

A key observation made over the course of these studies was the necessity to re-evaluate the present criteria used for the definition of bacterial species. This matter is also being reconsidered by a number of researchers currently. As much as defined rules for speciation are important to maintain uniformity, it is crucial that they keep up with improved analytical techniques. DNA-DNA hybridization, which is considered the gold standard for bacterial taxonomy, is undeniably a key experiment to understand the genomic content of the organism. However, it was proposed during a time period when resources to other sophisticated molecular and genetic methods were limited. The protocol itself has a number of disadvantages – the amount of raw genomic DNA required is considerably high; the entire experimental process lasts 3-4 days and requires expensive resources; the steps involved are quite cumbersome with the possibilities of experimental errors being quite high. Also, it is impossible to standardize the results for use across laboratories, making the reproducibility of results using the same set of isolates a major issue. Similar problems are also seen among phenotypic tests used for bacteria, a major disadvantage in this case being the subjective nature of data collection (Georgiades & Raoult. 2011; Goris *et al*. 2007; Kampfer & Glaeser. 2012; Konstantinidis & Tiedje. 2004).

Presently, with easy accessibility to sophisticated molecular methods and plummeting costs of sequencing, the quantification of the whole genome content can be done far more accurately and efficiently. One such alternative is a concept termed as Average Nucleotide Identity (ANI) calculated by the *in silico* comparisons of whole genome sequences of bacterial

strains. A direct correlation between DDH values and ANI values has been established, where the 70% cutoff criteria for DDH corresponded to a 94% ANI value, which could be used as a convenient speciation tool (Konstantinidis & Tiedje. 2005). A separate study also reported that the DDH value of 70% corresponded to a 69% value when the percentage of DNA conserved between genomes was calculated (Goris *et al*. 2007).

Similar studies also need to be carried out to evaluate the correlation of nucleotide divergence using MLSA of appropriate genetic markers with DDH values. As this was beyond the limitations of this current study, this investigation had to be temporarily terminated at this point, though this diversity study will be continued in our research group for future work.

The boundaries for defining a bacterial species have often been a matter of great debate (Georgiades & Raoult. 2011; Goris *et al*. 2007; Kampfer & Glaeser. 2012; Konstantinidis & Tiedje. 2004). Some might argue about whether there is a real need for further delineation of already identified species. However, for an accurate understanding of an organism, it is very important to constantly evaluate the diversity of the organism at the genus and species level. From an evolutionary standpoint, this aids in an improved understanding of the relationships of the organisms with its possible closest ancestors. In the case of pathogenic organisms, speciation can also play its role in the epidemiology of the organism. A perfect example of this being the genus *Cronobacter* itself, where the transition from the species *E. sakazakii* to the genus *Cronobacter* resulted in the clustering of the pathogenic strains of the organism within two major species, *C. sakazakii* and *C. malonaticus.* This is especially of importance with regards to the *Cronobacter* genus from the standpoint of a number of regulatory issues as well, as has been described in Chapter 1.

CHAPTER 5

CLONALITY AND MOLECULAR EPIDEMIOLOGY OF THE GENUS *CRONOBACTER* AS REVEALED BY MLST AND THE ST4 CLONAL COMPLEX

5.1 Introduction

5.1.1 Aims of the chapter

This chapter presents the relationships that were observed between the various *Cronobacter* STs and how these were used to investigate any possible correlation between the STs and the sources of the strains. Special attention has been given to the epidemiological background of the clinical strains in order to identify the link between the STs and virulence of the organism.

The results presented here are compiled from studies conducted at different points of the PhD study. The clonality and epidemiology with the entire sample dataset of 325 Cronobacter strains was carried out in March 2012 as a final overview presented in Section 5.3.1, and now included in the publications Joseph *et al*. (2012b) and Joseph & Forsythe (2012). The epidemiological study of 40 clinical *C. sakazakii* strains presented in Section 5.3.2 was an initial study conducted in January 2011, published as Joseph & Forsythe (2011). A year later, our laboratory received from CDC all the *Cronobacter* strains from cases and outbreaks in USA in 2011, including those from the highly publicised outbreak cases in December 2011 and these have also been discussed in detail in Section 5.3.3 and linked with the earlier study. This study has now been accepted for publication as Hariri *et al*. (2012).

5.1.2 Linkage and the clonal complex

With the development and increased use of sequencing techniques such as MLST for the studies of bacterial species over the past years; there have been considerable advances in bacterial population genetics research. The data generated has resulted in an improved understanding of the structures of various bacterial populations as well as epidemiological relationships in the case of pathogens. The allele based data of the housekeeping genes not only helps to analyse the impact of evolution and recombination in the population (as discussed in Chapter 3) but also to identify linkage and the presence of clones and their relationships.

Linkage in a population is defined as the possibility of the occurrence of an allele of a certain locus or gene in tandem with the allele of an unrelated locus in the same genome. A population is said to be in linkage equilibrium when the alleles of various loci are arbitrarily present in the genome, such that the presence of one does not indicate or make it possible to predict the presence of another allele. Conversely, linkage disequilibrium in a population is a phenomenon where the presence of a particular allele can aid the prediction of the allele of another locus, and such a population is said to be clonal in nature. A clonal population is said to be under a weak influence of recombination, though recombination may not necessarily be

absent (Feil 2010). The index of association (I_A) is a useful measure to quantify the amount of linkage in a bacterial population. The concept was first applied in the population studies of *Hordeum spontaneum*, a species of wild barley grass, by Brown *et al*. (1980). It is now popularly used in a number of prokaryotic including bacterial population studies (Smith *et al*. 2000). An I_A value of zero is an indicator of a population being in linkage equilibrium. It has been observed from various studies that most bacterial populations often exhibit a certain degree of linkage disequilibrium or clonality (Feil 2010).

Studying multiple genes in a genome is a convenient way of identifying the clonality in a population. Earlier studies into clonality have used multilocus enzyme electrophoresis (MLEE) data for populations of *N. gonorrhoeae*, *N. meningitidis* and *Salmonella* spp. (Maynard Smith *et al*. 1993). Nowadays, this has increasingly progressed into the use of the genotypes identified by the MLST schemes of bacteria for the same. A popular means for this analysis has been an algorithm called eBURST (Feil *et al*. 2004). The algorithm introduces a number of novel terms to characterize the clonality in the MLST dataset of a bacterial population. It uses only the allele numbers and sequence types identified by the MLST scheme and not the nucleotide sequences, thus differentiating it from sequence based phylogenetic analysis (as presented in Chapter 3 and 4). The eBURST algorithm identifies the more frequent and persistent sequence type in the population (the one associated with the maximum number of isolates) which is then denoted as the "founder clone". This founder is then linked to its nearest possible neighbour which in this case is the ST with one among the seven alleles different, therefore known as a single locus variant (SLV) to the founder. These links result in the formation of clusters of closely linked STs within the population, known as "clonal complexes" (CC). The SLV can also then be further linked to its closest neighbours identified as double locus variants (DLVs) which differ at two alleles compared to the founder ST, or triple locus variants (TLVs) which differ at three alleles compared to the founder ST. As a result, the overall population snapshot of the dataset in a graphical output represents a forest composed of a number of unrooted trees as well as isolated STs. Each ST in the eBURST algorithm output is denoted by a circle, where the diameter of the circle is proportional to the frequency of the ST in the said bacterial population (Feil *et al*. 2004). The definition of a CC depends on the threshold value that is decided upon for it. The PubMLST database by default uses only the founder-SLV cluster to define a specific clonal complex, though in the results presented here the DLVs and TLVs are also included for the analysis and discussion of the clonal relationships.

The eBURST algorithm can be used directly via the MLST database websites or by downloading as standalone software (eBURSTv3; http://eburst.mlst.net/). An advanced, globally optimized version of the eBURST algorithm has also now been introduced by Francisco *et al* (2009) and has been used for the presentation of the results in this chapter,

known as goeBURST. It is available through the use of the software Phyloviz at http://goeburst.phyloviz.net/ (Francisco *et al*. 2012).

5.2 Results

5.2.1 Clonality and ST groupings in the *Cronobacter* genus

The I_A is a measure of the linkage of a population. The I_A calculated for the genus *Cronobacter* using the MLST data showed a value of 1.342. Since this was found to be significantly greater than zero ($p<0.001$), this indicated the presence of linkage disequilibrium or clonality in the population.

The goeBURST analysis of the *Cronobacter* MLST data revealed the formation of 13 SLV CCs among the 115 identified STs for the *Cronobacter* genus (Fig. 5.1 (a), (b) and (c); Table 5.1). The minimum requirement for the CC formation was an SLV linkage with the founder ST. These are the STs which shared 6 out of the 7 alleles that comprise a ST and have been indicated by the black lines in the figures. Apart from these, a number of DLVs and TLVs were observed which, though not defined as CCs, have also been indicated.

Nine of these CCs belong to *C. sakazakii,* which is also the type species of the genus and the most dominant in the dataset. Two of them belong to *C. dublinensis* and one each to *C. turicensis* and *C. malonaticus. C. muytjensii, C. condimenti* and *C. universalis* did not show the formation of any such clonal complexes, in this dataset.

Even though *C. muytjensii* species did not show any clonal complexes, there was seen a DLV cluster among them comprising of the STs 28, 30, 44 and 81. Of these, ST81 contains the type strain, ATCC 51329^T, of the species.

Many of these CCs are especially significant with respect to strain clustering according to their isolation sources, and have been discussed below in detail.

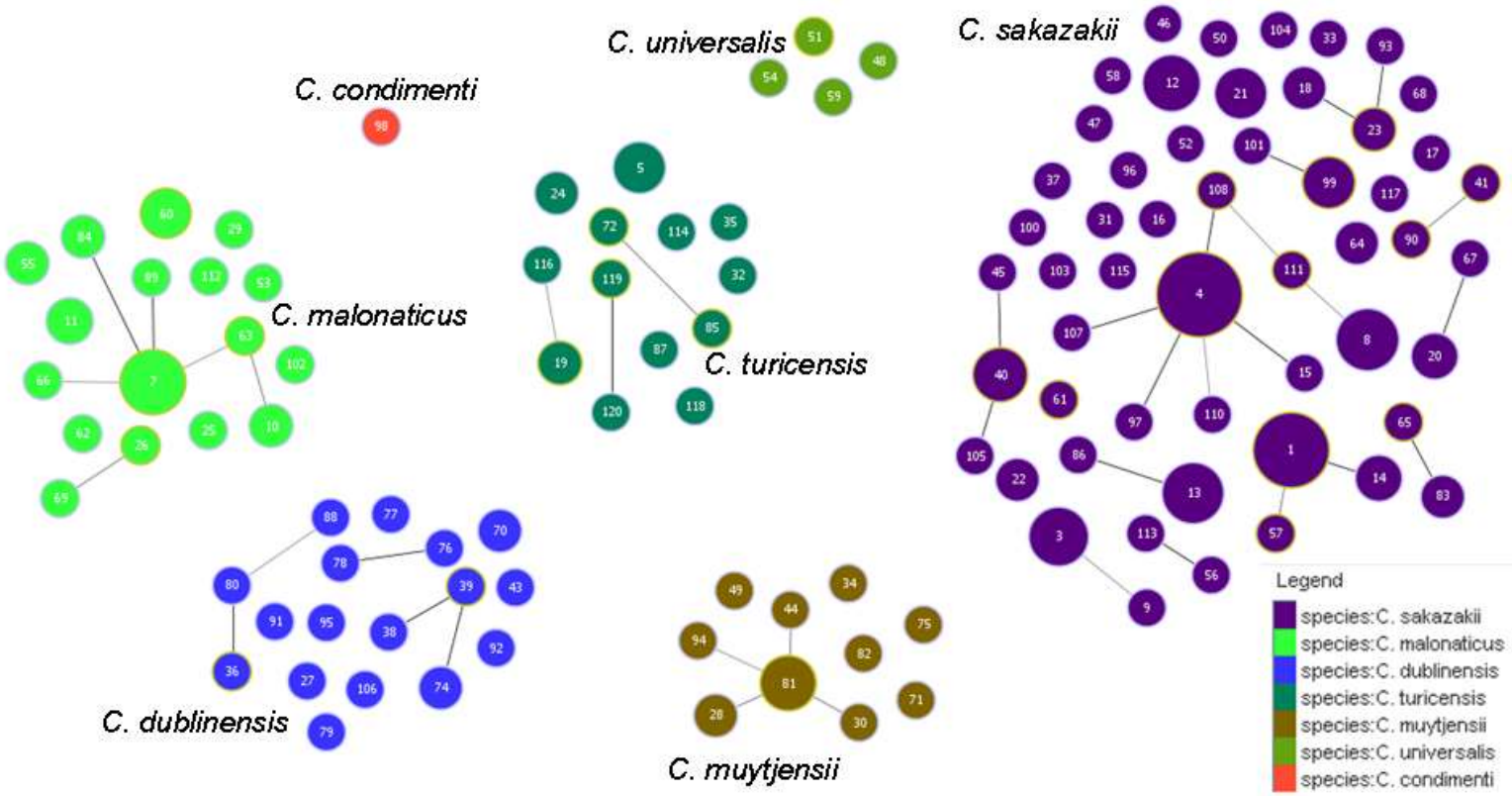


Fig. 5.1 (a) Population snapshot of the *Cronobacter* MLST database generated using goeBURST, indicating the clonal complexes among the different *Cronobacter* species. The threshold for the output was set to triple locus variation. The dominant STs are represented by the circles with larger diameters. Clusters of linked isolates correspond to clonal complexes, listed in Table 5.1. Figure as originally published in Joseph S & Forsythe SJ (2012) Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Frontiers in Food Microbiology*, 3, 397. doi: 10.3389/fmicb.2012.00397

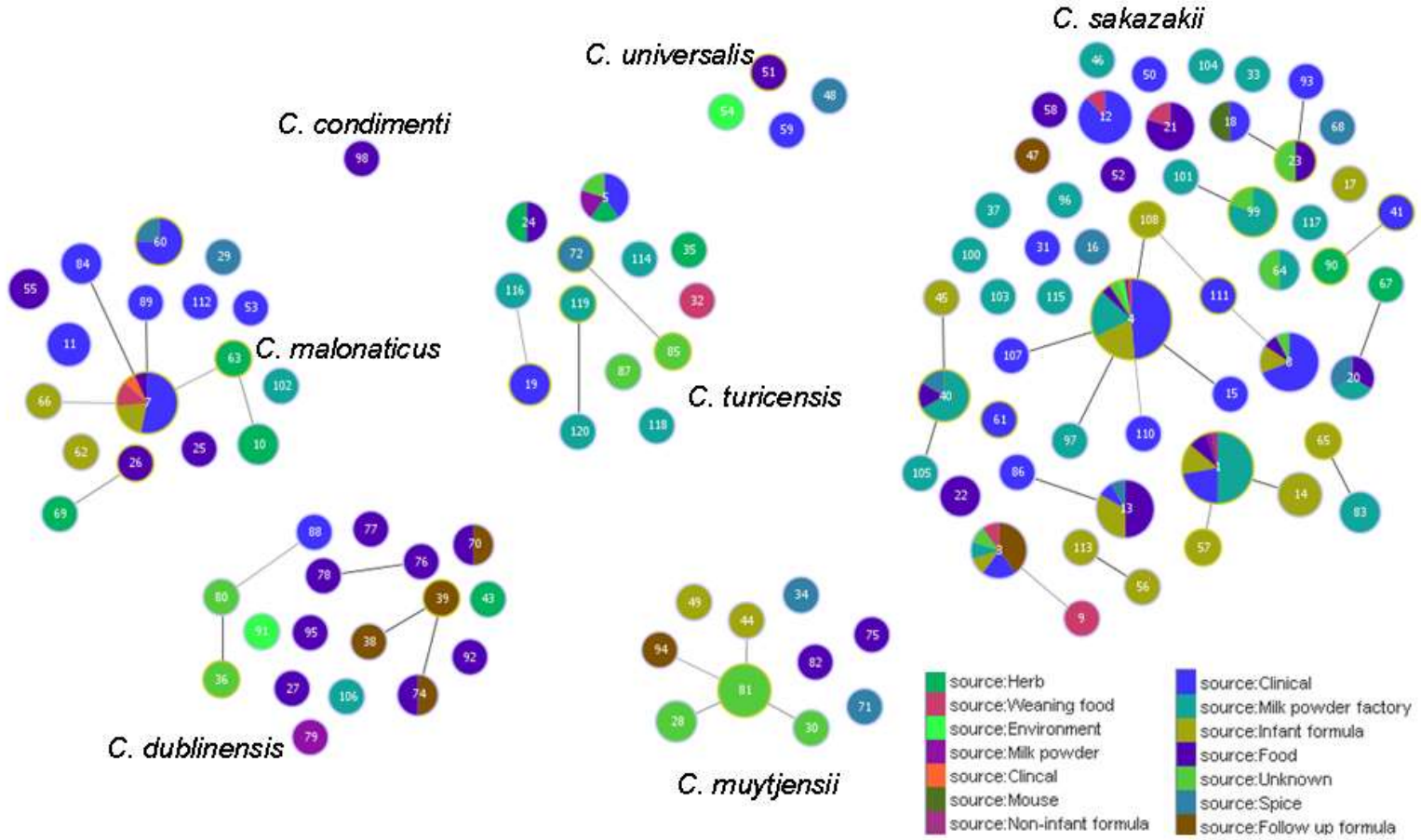


Fig. 5.1 (b) Population snapshot of the *Cronobacter* MLST database generated using goeBURST, indicating the clonal complexes and the diversity of the sources of the strains. The threshold for the output was set to triple locus variation. The dominant STs are represented by the circles with larger diameters. Clusters of linked isolates correspond to clonal complexes, listed in Table 5.1. Figure as originally published in Joseph S & Forsythe SJ (2012) Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Frontiers in Food Microbiology*, 3, 397. doi: 10.3389/fmicb.2012.00397

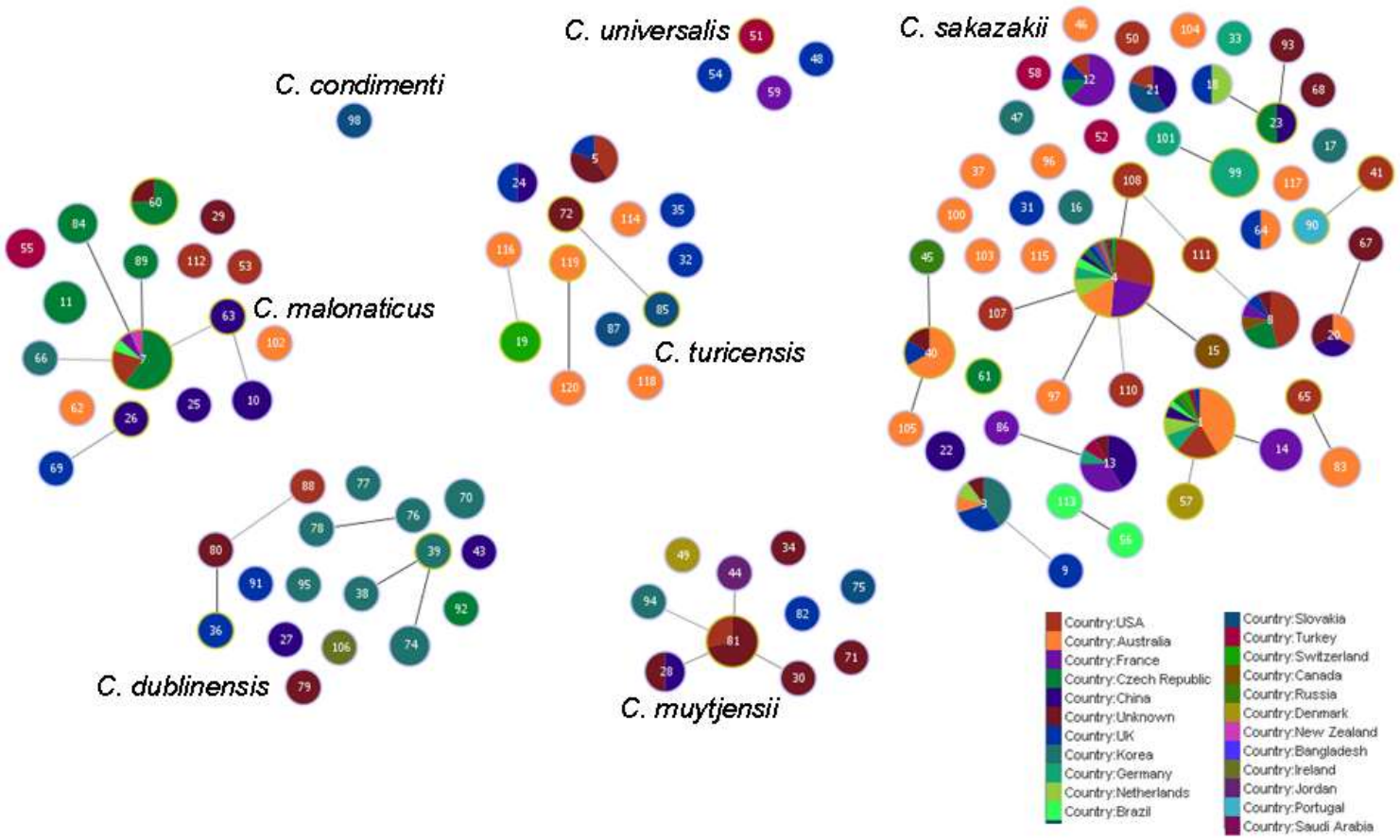


Fig. 5.1 (c) Population snapshot of the *Cronobacter* MLST database generated using goeBURST, indicating the clonal complexes and the diversity of the strains based on country of isolation. The threshold for the output was set to triple locus variation. The dominant STs are represented by the circles with larger diameters. Clusters of linked isolates correspond to clonal complexes, listed in Table 5.1. Figure as originally published in Joseph S & Forsythe SJ (2012) Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Frontiers in Food Microbiology*, 3, 397. doi: 10.3389/fmicb.2012.00397.

CC	Species	ST	Number of Isolates	Isolation Sources	Geographic Distribution
1	*C. sakazakii*	1	36	Clinical, PIF, formula, food, environment	UK, Australia, USA, Germany, China, Brazil, Czech Republic, Switzerland, Turkey, Russia, Netherlands
		14	3	PIF	France
2	*C. malonaticus*	7	15	Food, clinical, PIF, weaning food	USA, New Zealand, France, Czech Republic, Brazil
		84	2	Clinical	Czech Republic
		89	1	Clinical	Czech Republic
3	*C. sakazakii*	18	2	Clinical, mouse	UK, Netherlands
		23	2	Food	Czech Republic, China
		93	1	Clinical	Czech Republic
4	*C. sakazakii*	4	78	Clinical, PIF, milk powder, weaning food, chocolate, washing brush, environment, prepared formula, foot wound	UK, USA, France, China, Canada, Netherlands, Germany, Russia, Czech Republic, Switzerland, Slovakia, New Zealand, Saudi Arabia, Bangladesh
		15	1	Clinical	Canada
		97	1	Milk powder factory	Australia
		107	1	Clinical	USA
		108	1	PIF	USA
5	*C. dublinensis*	38	1	FUF	Korea
		39	1	FUF	Korea
		74	2	FUF, food	Korea
6	*C. sakazakii*	40	6	Food, spices	UK
		45	1	PIF	Russia
		105	1	Milk powder factory	Australia
7	*C. sakazakii*	20	3	Food, spices	China, UK
		67	1	Herbs	UK
8	*C. sakazakii*	13	12	Clinical, PIF, food, Herbs, Spices	France, Germany, China, Turkey
		86	1	Clinical	France
9	*C. sakazakii*	65	1	PIF	USA
		83	2	Milk Powder	Australia
10	*C. sakazakii*	99	5	Environmental	Germany
		101	1	Environmental	Germany
11	*C. sakazakii*	56	1	PIF	Brazil
		113	1	PIF	Brazil
12	*C. turicensis*	119	1	Milk powder factory	Australia
		120	1	Milk powder	Australia
13	*C. dublinensis*	36	1	Unknown	Unknown
		80	1	Unknown	Czech Republic

Table 5.1 Details of the clonal complexes (CCs) identified in the *Cronobacter* MLST dataset revealed using goeBURST.

Key: CC – Clonal complex; PIF – Powdered infant formula; FUF – Follow up formula; UK – United Kingdom; USA – United States of America

CC-1 comprises *C. sakazakii* STs 1 and 14. The two STs differ at the *ppsA* locus by a single nucleotide position (nt 301: A → C). ST1 is one of the dominant STs in the *Cronobacter* MLST database consisting of 36 strains isolated from across the world over a period of more than 30 years. A number of these strains have been isolated from PIF (6/36) and clinical cases (8/36), and also more recently from milk powder processing factories (18/36) in Germany and Australia (Craven *et al*. 2010; Jacobs *et al*. 2011), apart from a few food isolates. The three ST14 strains were isolated in 1994 from infant formula used during an NICU *Cronobacter* spp. outbreak in France. According to PFGE, those strains did not match isolates from the infected neonates (Caubilla-Barron *et al*. 2007). Also linked to this clonal complex is ST57, which is a PIF isolate (NTU strain 531) from Denmark in 1988. This is a DLV to ST1, varying at the *atpD* and *glnS* loci. Thus, ST1 is the founder clone of this cluster of STs.

CC-2 comprises *C. malonaticus* STs 7, 84 and 89. ST7 is the founder clone of this CC, with ST84 varying from it at the *ppsA* locus by a single nucleotide (nt 199: C → T) and ST89 at the *infB* locus (11 nucleotides). Among *C. malonaticus* isolates, ST7 has had the highest frequency (15/39), and is comprised of a mixture of majorly clinical (9/15) and some PIF isolates (3/15) from over 30 years. STs 84 (NTU strains 1545 & 1546) and 89 (NTU strain 1558) comprise clinical isolates from the Czech Republic. These clinical isolates were recovered mainly from faecal, sputum and blood samples. Apart from these, there were also food and weaning food isolates in this cluster of strains.

CC-3 comprises *C. sakazakii* STs 18, 23 and 93 with strains that had been isolated between 1953 and 2010. ST23 is the founder clone of this CC, comprising of a food isolate each from Czech Republic (NTU strain 1549) and China (NTU strain BQ4). ST18 was found to differ from ST23 at the *ppsA* locus (16 nucleotides). It consists of a clinical isolate from the UK (NTU strain 580) and a strain isolated from a mouse in Netherlands (NTU strain 560). ST93 varied from ST23 at the *infB* locus (28 nucleotides). It had a lone strain isolated from the bronchial secretion of a 9 year old girl in the Czech Republic (NTU strain 1557).

CC-4 comprises *C. sakazakii* STs 4, 15, 97, 107 and 108. ST4 is the founder clone of this CC and also the most frequent ST in the *Cronobacter* MLST dataset. ST4 was found to vary at the *fusA* locus from ST15 by one nucleotide (nt 247: G → A), ST107 (6 nucleotides) and ST108 (5 nucleotides). ST4 and ST97 varied at the *gltB* locus by a single nucleotide (nt 321: G → A). This is a key clade with respect to *Cronobacter* spp. epidemiology and clinical significance and details of this CC have been presented and discussed in detail in Section 5.2.2.

CC-5 comprises *C. dublinensis* STs 38, 39 and 74. ST39 is the founder clone of this complex, with ST38 differing from it at the *atpD* locus (3 nucleotides), while ST74 differed from it at the *gyrB* locus (33 nucleotides). All the strains in this CC were isolates from Korea. STs 38 and 39 comprised of single isolates (NTU strains 1132 & 1133 respectively) from

follow-up formula in 2008. ST74 contained one follow up formula isolate from 2008 (NTU strain 1134), while another strain was a Korean food isolate from 2011 (NTU strain 1458).

CC-6 comprises *C. sakazakii* STs 40, 45 and 105. The founder clone of this cluster was ST40, with STs 45 and 105 varying from it at loci *infB* (2 nucleotides) and *atpD* (5 nucleotides) respectively. ST40 consisted of 6 strains isolated between 2005 and 2010. Among these, four had been isolated from various locations in the environment of a milk powder processing factory in Australia (Craven *et al*. 2010), while the other two were spice and food isolates. ST105 also had an isolate (NTU strain 1501) from the roof of the same Australian factory, while ST45 had a single Russian infant formula isolate from 1988 (NTU strain 536).

CC-7 comprises *C. sakazakii* STs 20 and 67. The two STs vary at the *gyrB* locus by 6 nucleotides. ST20 consists of two food isolates from 2010 (NTU strains BQ7 and 1347) and an isolate from the Australian milk powder processing factory from 2010 (NTU strain 1474). ST67 had a lone strain (NTU strain 1379) isolated from herbs (dried dill).

CC-8 comprises of *C. sakazakii* STs 13 and 86. The two STs differ at the *atpD* locus by a single nucleotide (nt 47: T → A). ST13 consists of 12 strains isolated over a period of 1988 to 2010 from four countries. These included six food isolates, 5 Chinese and 1 Turkish. The earliest ST13 strain was isolated in Germany in 1988 from infant formula (NTU strain 532). Three of the ST13 strains were also isolates from feed prepared and used during the NICU outbreak in France in 1994 (NTU strains 713-715). There was also another ST13 strain from the same outbreak isolated from the faeces of an asymptomatic infant (NTU strain 693). ST86 had a single strain isolated during the same French NICU outbreak from the faeces of an infant with digestive problems (NTU strain 700).

CC-9 comprises *C. sakazakii* STs 65 and 83. The two STs vary at their *infB* locus by 10 nucleotides. ST65 consists of a strain isolated from infant formula in the USA in 1988 (NTU strain 547). ST83 had two strains isolated from the environment of the milk powder processing factory in Australia (NTU strains 1498 & 1500).

CC-10 comprises *C. sakazakii* STs 99 and 101, which varied from each other by a single nucleotide at the *fusA* locus (nt 378: G → A). These STs consisted of 6 strains in total, all of them isolated from the towers of a milk powder processing factory in Germany.

CC-11 comprises *C. sakazakii* STs 56 and 113. The two STs differ at the *fusA* locus by two nucleotides. Both the STs contain a strain each isolated from infant formula in Brazil in 2007 (NTU strains 892 & 894 respectively).

CC-12 comprises *C. turicensis* STs 119 and 120, which differed at the *glnS* locus. These STs comprised of a strain each, both isolated from the environment of the Australian milk powder processing factory (NTU strains 1512 & 1513 respectively).

CC-13 comprises *C. dublinensis* STs 36 and 80, differing at the *fusA* locus by a single nucleotide (nt 12: C → T). No further isolation data was available for the strains of these STs.

5.2.2 Predominance of *Cronobacter sakazakii* ST4 with neonatal infections.

In January 2011, a study was conducted to investigate a possible correlation between the sequence types of the *C. sakazakii* strains and the epidemiology of the organism. For this purpose, details of all the clinical *C. sakazakii* strains present in the database at that time were collated. In total, 40 clinical *C. sakazakii* strains were included in this study (Table 5.2).

These 40 strains were found to be from 7 countries, and had been isolated over a period between 1953 and 2008. Of these, patient details were available only for 30 of the isolates, of which only one was found to be an adult isolate (NTU strain 12 isolated from a 74 year old adult). Comparative analysis of these strains was carried out with the online *Cronobacter* MLST database covering isolates from all sources. This showed that these clinical isolates were distributed over 10 of the 25 STs defined for *C. sakazakii* at that time. However, these strains were not evenly distributed across the ten STs. Of particular interest was the fact that half (20/40) of these clinical strains belonged to ST4 (Table 5.3). The remaining strains were in ST8 (7/40), ST1 (3/40), ST12 (3/40), ST3 (2/40) and single strains in ST13, ST15, ST18, ST31, and ST41 (Table 5.2 and 5.3).

Furthermore, of the 20 ST4 strains, 10 were from neonates, 7 from infants, 1 from a child and 2 had no patient details. Similarly, 9/11 of the meningitis isolates were also found to belong to ST4. Of these, 7 had been isolated from the CSF, and the other two from blood and the trachea. The remaining ST4 strains were from cases of bacteraemia (1 isolate), NEC (2 isolates), undefined infection (2 isolates) and 6 from unknown etiologies. ST4 was also the main sequence type associated with neonates, with 10 out of the 17 neonatal isolates belonging to ST4 (Table 5.2). The majority (5/6) of CSF isolates were also *C. sakazakii* ST4, with the remaining strain belonging to *C. sakazakii* ST1 (NTU strain 1218). Similarly 4/5 blood isolates were *C. sakazakii* ST4, and the remaining strain was *C. sakazakii* ST1 (NTU strain 1241).

These twenty ST4 strains were from six countries; Netherlands, France, USA, New Zealand, Czech Republic and Canada. These strains had been isolated over a period of more than 30 years between 1977 and 2008, with the earliest ST4 isolate (NTU strain 553) from a day old neonate in Netherlands (Table 5.2).

A large majority of these strains included those from well cited reports (Himelright *et al*. 2002, Pagotto *et al*. 2003, Caubilla-Barron *et al*. 2008, Jarvis, 2005; Baumbach *et al*. 2009, Hurrell *et al*. 2009) and those which have more historic significance (>20 years) (Farmer *et al*. 1980, Aldova *et al*. 1983, Muytjens *et al*. 1988). Wherever possible, these have been indicated in Table 5.2.

Strain	**Patient type/age (EGA)[a]**	**Clinical presentation**	**Isolation site**	**Year**	**Country**	**ST**	References
553	Neonate/1 d	Unknown	Unknown	1977	Netherlands	4	1
557	Neonate/5 d	Unknown	Unknown	1979	Netherlands	4	1
693	Neonate/13 d (41 weeks)	Asymptomatic	Faeces	1994	France	13	2
695	Neonate/15 d (32 weeks)	Fatal NEC[b] II	Trachea	1994	France	4	2,3
701	Neonate/28 d (28 weeks)	Fatal NEC III	Peritoneal fluid	1994	France	4	2,3
709	Neonate/18 d (29 weeks)	Septicaemia	Trachea	1994	France	4	2,3
767	Neonate/19 d (31 weeks)	Fatal meningitis	Trachea	1994	France	4	2,3
721	Neonate/2 weeks	Meningitis	CSF[c]	2003	USA	4	
978	Neonate/<1 week	Unknown	Enteral feeding tube	2007	UK	3	4
696	Neonate/17 d (32 weeks)	NECII	Faeces	1994	France	12	2,3
984	Neonate/3-4 week	Unknown	Enteral feeding tube	2007	UK	3	4
690	Neonate/27 d (31 week)	Asymptomatic	Faeces	1994	France	12	2
1218	Neonate/ <1 month (30 week)	Fatal meningitis	CSF	2001	USA	1	
1219	Neonate/ <1 month (36 week)	Fatal meningitis	CSF	2002	USA	4	
1221	Neonate/ <1 month	Meningitis adverse neurological outcome	CSF	2003	USA	4	
1225	Neonate/ <1 month (35 week)	Fatal meningitis	Blood	2007	USA	4	
1231	Neonate (33 week)	Fatal neurological damage	Faeces	2004	New Zealand	4	5
1019	Neonate (33 week)	Meningitis	CSF	2001	USA	1	6
1249	Neonate	Fatal infection	Unknown	2009	UK	31	
1220	Infant/6 week (37 week)	Brain abscess, non-fatal	CSF	2003	USA	4	
1223	Infant/6 week (31 week)	Unknown, in ICU[d]	Blood	2004	USA	4	
1240	Infant/ 7 weeks	Fatal meningitis	CSF	2008	USA	4	7
1242	Infant/ 7 weeks	Fatal meningitis	Brain	2008	USA	4	7
1241	Infant/ 7 months	Sudden infant death syndrome	Blood	2008	USA	1	7

1222	Infant/ 8 months	Fever, recovered	Blood	2003	USA	4	
1224	Infant/ 10 months	Fever, severe combined immunodeficiency	Blood	2004	USA	4	
4	Child/ 6 years	Unknown	Unknown	2002	Canada	15	
1	Child	Unknown	Throat	1980	USA	8	8
20	Child/ 6 years	Unknown	Faeces	2004	Czech Republic	4	8
12	Adult/ 74 years	Unknown	Faeces	2004	Czech Republic	1	9
563	Unknown	Foot wound	Wound	1975	USA	41	
683	Unknown	Unknown	Sputum	1977	USA	8	8
680	Unknown	Unknown	Spinal fluid	1977	USA	8	8
580	Unknown	Unknown	Abdomen pus	1953	UK	18	1
6	Unknown	Unknown	Unknown	1990	Canada	4	10
558	Unknown	Unknown	Unknown	1983	Netherlands	4	1
511	Unknown	Unknown	Unknown	1983	Czech Republic	8	11
513	Unknown	Unknown	Unknown	1983	Czech Republic	8	11
526	Unknown	Unknown	Unknown	1983	Czech Republic	8	11
5	Unknown	Unknown	Unknown	1990	Canada	8	10
520	Unknown	Unknown	Unknown	1983	Czech Republic	12	11

a – EGA, Estimated gestation age; values < 37 weeks are considered premature; b - NEC, necrotizing enterocolitis; c - CSF, cerebral spinal fluid; d - ICU, intensive care unit

References : 1. Muytjens *et al*. (1983) 2. Caubilla-Barron *et al*. (2007) 3. Townsend *et al*. (2008) 4. Hurrell *et al*. (2008) 5. Jarvis (2005) 6. Himelright *et al*. (2002) 7. Baumbach *et al*. (2009) 8. Farmer *et al*. (1980) 9. Iversen *et al*. (2004) 10. Pagotto *et al*. (2003) 11. Aldová *et al*. (1983)

Table 5.2 List of *C. sakazakii* strains included in the epidemiology study of January 2011 and their clinical details. The strains have been listed in the ascending order of the ages of the patients, where known. Published in Joseph & Forsythe (2011).

		Patient details						Clinical presentation					
ST	**N=**	**Neonate**	**Infant**	**Child**	**Adult**	**Unknown**		**Meningitis**	**Bacteraemia**	**NEC**	**Infection**	**Asymptomatic**	**Unknown**
1	3	1	1	1				1	1				1
3	2	2											2
4	20	10	7	1		2		9	1	2	2		6
8	7			1		6		1					6
12	3	2				1				1		1	1
13	1	1										1	
15	1			1									1
18	1					1					1		
31	1	1									1		
41	1					1					1		
Total	**40**	**17**	**8**	**4**	**0**	**11**		**11**	**2**	**3**	**4**	**2**	**17**

Table 5.3 Quantitative summary of the *C. sakazakii* STs and source details for the strains included in the epidemiology study in January 2011. Published in Joseph & Forsythe (2011).

5.2.3 The ST4 clonal complex

As described earlier, in the initial study of 40 clinical *C. sakazakii* strains (Section 5.2.2), ST4 was identified as a genetic signature for *C. sakazakii* meningitis, with 75% of the isolates being linked to meningitis cases over a period of 50 years from six different countries. When the statistics were expanded to the current study upto March 2012 consisting of 325 strains, it was found that 73/325 *Cronobacter* spp. strains were related to clinical cases, of which > 50% was found to be *C. sakazakii* ST4 (38/73). ST4 was also still found to be the most dominant ST in this MLST study upto March 2012 with 78/325 isolates (24%), and also the most frequent clinical ST. Also, of the 73 clinical isolates in this study, 18 (~25%) were confirmed to be meningitic isolates belonging to STs 4 (12 isolates), 1 (2 isolates), 8 (1 isolate), 31 (1 isolate), 107 (1 isolate) and 110 (1 isolate). Thus, 12/38 (~32%) clinical ST4 strains were meningitic isolates.

The 78 ST4 strains included twelve strains of *C. sakazakii* isolated from PIF collected in 12 countries. Fifteen strains which had been isolated from milk processing factories in Australia and Germany, including roller dryers, tanker bays, and spray dryers (Craven *et al*. 2010; Jacobs *et al*. 2011) were also revealed as ST4. Other non-clinical sources included isolates from weaning food, chocolate (Turcovsky *et al*. 2011) and a washing brush (van Acker *et al*. 2001). ST15 has a lone isolate from a Canadian clinical case (Pagotto *et al*. 2003). ST97 has one of the strains isolated from a tanker bay at a milk powder processing factory in Australia (Craven *et al*. 2010).

In December 2011, the organism made the headlines after it was implicated in the death of an infant in USA, following which there was a product recall of a major infant formula brand. In February 2012, our research group received from the CDC the *Cronobacter* isolates from all the cases reported in USA in 2011. Applying the MLST scheme to these strains showed a number of the strains to belong to CC-4, while the others showed an interesting correlation with the cluster, as explained below. The epidemiological information of the strains along with their sequence types have been listed in Table 5.4.

NTU ID	CDC ID	ST	Place	Isolation Source	Comment
1579	2012-05-05	4	Lebanon, MO	CSF of <1 month male, term infant; exposed to PIF	Fatal case
1566	2011-12-02	4	Ohio	CSF of 1 month male infant; exposed to PIF	From twin of patient 2011- 12-03
1567	2011-12-03	4	Ohio	Stool of 1 month male infant; exposed to PIF	From twin of patient 2011- 12-02; Asymptomatic
1568	2011-12-04	4	Ohio	Opened PIF	Formula associated with 2011-12-02 and -03
1570	2011-21-01	4	Minnesota	CSF of <1 month male, term infant; exposed to PIF	Brain infarction
1571	2011-21-03-01	4	Minnesota	Opened PIF	Formula associated with 2011-21-01
1576	2193-02	4	Michigan	Tracheal secretion of <1 month male, pre-term infant (30 week EGA); not exposed to PIF	Symptoms were not due to *Cronobacter* infection; fortified breast milk fed only after culture was obtained
1565	2011-12-01	107	Michigan	CSF of <1 month male, term infant; exposed to PIF	Brain abscess; Outcome unknown
1572	2011-21-03-02	108	Minnesota	Opened PIF	-
1577	2193-03	110	Illinois	CSF of 1 month female, term infant; exposed to PIF	-
1578	2193-08-01	111	Illinois	Nursery water	Bottled water associated with 2193-03
1573	2011-18-05-02	8	Ohio	Opened PIF	Formula associated with 2011-18-01 and 2011-18-07
1574	2011-18-01	8	Ohio	Stool of 4 month female, term infant; exposed to PIF	Diarrheal symptoms
1575	2011-18-07	8	Ohio	Stool of ~5 month female, term infant; exposed to PIF	Ongoing diarrhea; same patient as 2011- 18-01

Table 5.4 Details of the *C. sakazakii* isolates received by CDC from the US *Cronobacter* cases in 2011. Published in Hariri *et al.* (2013).

The 14 *C. sakazakii* isolates obtained from CDC exhibited six different STs (Table 5.4). Seven of these strains were ST4. This included the strain from the fatal Lebanon (Missouri, USA) infection case (NTU strain 1579) which received considerable publicity in December 2011 (CDC 2011).

Two other STs that were seen were STs 107 and 108, belonging to CC 4 and DLVs to ST4. ST107 was strain 1565, a CSF isolate of an infant in Michigan. An ST4 strain had also been isolated in Michigan from the tracheal secretion of a <1 month premature infant (NTU strain 1576). ST108 had a strain (NTU strain 1572) isolated from an opened paediatric infant formula can in Minnesota. Two ST4 strains had also been isolated in Minnesota, one from the CSF of a <1 month infant and the other from an opened PIF can that the said infant had been exposed to (NTU strains 1570 & 1571 respectively).

Also seen among these strains was ST110 which was found to be a TLV to ST4, as they share 4 out of the 7 alleles. This was a strain isolated in Illinois from the CSF of a 1 month female infant (NTU strain 1577).

Due to the strong association of the ST4 and loci variants (STs 15, 97, 107, 108 & 110) with severe cases of meningitis, and therefore of high clinical significance, this cluster of STs has been designated as the 'ST4 clonal complex' as an extension of the previous ST4 association (Section 5.3.1). **Out of the 18 meningitic isolates in this 325 isolate study, 14 isolates were found to belong to the ST4 clonal complex (78%).** This association of the ST4 clonal complex with the meningitic *C. sakazakii* isolates was further verified by performing a chi-square (χ^2) analysis on the data. This revealed a significant χ^2 value of 4.53 at a probability of p=0.033. Since a statistically significant probability of $p< 0.05$ was obtained, it can be stated that there is definitely a strong association between the meningitic isolates and the ST4 clonal complex.

The remaining two STs (STs 111 and 8) seen among these US isolates from 2011 also showed a correlation with this ST4 clonal complex. ST111 was found to be a TLV to ST108. ST111 had a strain isolated from bottled nursery water in Illinois (NTU strain 1578). This water sample was also linked to the infant from which strain 1577 (ST110) had been isolated.

ST111 was also, in turn, found to be a TLV to ST8, and therefore the "link" between ST4 and ST8. Among these US isolates, three *C. sakazakii* strains were found to be ST8, all isolates from Ohio. Two of these strains (NTU strains 1574 and 1575) were isolated approximately a month apart from the stool samples of a female infant, who was exhibiting ongoing diarrheal symptoms. The third strain (NTU strain 1573) had been isolated from an opened infant formula can that the infant had been exposed to.

This relationship is also important because of the significance of the *C. sakazakii* ST8 from the point of view of the predominance of clinical isolates. This ST comprises 13 strains,

nine of which were clinical isolates including the US 2011 isolates obtained from the CDC (Table 5.4). Of these nine clinical ST8 isolates, one was from the throat of a child (NTU strain 1), one a sputum isolate (NTU strain 683), one a spinal fluid isolate (NTU strain 680), and two faecal isolates (NTU strains 1574 and 1575) from the CDC US 2011 collection obtained from the same infant suffering diarrhoea, isolated a month apart. The isolation details of the remaining clinical strains were not available. Also, apart from the two diarrhoeal isolates, clinical presentations associated with none of the other ST8 strains could be obtained. The ST8 isolate set also comprised two PIF isolates, one from France (NTU strain 424) and one from an opened PIF tin (NTU strain 1573) from the US 2011 case in Ohio. Formula prepared from this tin had been fed to the infant associated with strains 1574 and 1575, mentioned above (Table 5.4).

This relationship between the ST4 clonal complex and STs 111 and 8 has been graphically depicted in Fig. 5.2, by using the goeBURST algorithm. However, this "connection" could be an observed artefact as discussed in section 5.3.1.

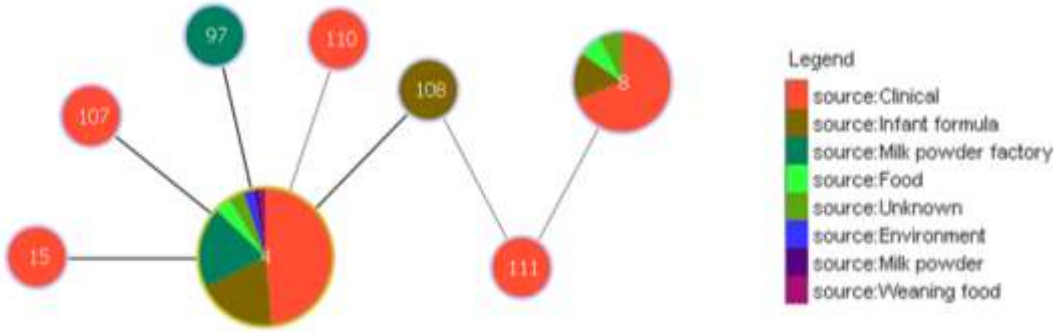


Fig 5.2 Relationship between the clinically significant *C. sakazakii* ST4 clonal complex and ST8 depicted using the goeBURST algorithm in Phyloviz (Francisco *et al*. 2012). The threshold for the output was set to triple locus variation. The black lines denote the SLVs; while the grey lines indicate the TLVs. ST4 is the founder clone of the clonal complex. Figure as originally published in Joseph S & Forsythe SJ (2012) Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Frontiers in Food Microbiology*, 3, 397. doi: 10.3389/fmicb.2012.00397

5.2.4 *C. malonaticus* – clonality and epidemiology

The *Cronobacter* MLST scheme revealed 17 STs among the 39 *C. malonaticus* strains included in this study. Twenty of these strains were clinical isolates from Czech Republic and USA isolates between 1977 and 2011. These 20 strains were found to be concentrated in seven of the *C. malonaticus* STs. Dominant among these STs was ST7 which comprised 9 of the clinical isolates. The list of clinically relevant *C. malonaticus* strains and their available epidemiological data has been collated in Table 5.5. Twelve of the clinical strains were found to belong to clonal complex 2. However, statistical analysis using chi-squared tests did not indicate a strong association with a higher probability value of p=0.18, suggesting that a larger sample set might be required for conclusive evidence of an association between clinical cases and *C. malonaticus* clonal complex 2.

The available isolation data of the strains also revealed a strong predominance of adult *Cronobacter* isolates in the *C. malonaticus* species. Among the strains with epidemiological information available, only a single infant isolate was found (NTU strain 1569). This also happened to be the only strain from the *Cronobacter* cases in USA in 2011 which was not a *C. sakazakii* isolate. This strain belonged to ST112, and had been isolated from the blood of a <1 month old, premature, male infant who had been fed PIF. The child had been diagnosed with clinical meningitis, eventually resulting in fatality.

Apart from these clinically linked cases, there were also five *C. malonaticus* strains that had been isolated from PIF samples from different countries (New Zealand, Australia, France, Brazil and Korea). Three of these belonged to ST7 (NTU strains 535, 22 & 893) while the other two were STs 62 and 66 (NTU strains 527 & 472 respectively).

The remaining *C. malonaticus* strains in the study were isolated from food (5), herbs and spices (6), weaning food (2) and environment of a milk powder processing factory (1).

NTU Strain ID	Country	Year	Isolation data	ST	Clonal Complex
681	USA	1977	Breast abscess isolate	7	2
688	Czech Republic	2004	76 year old adult; sputum isolate	7	2
21	Czech Republic	2003	72 year old adult male; Faecal isolate	7	2
18	Czech Republic	2003	47 year old adult; faecal isolate	7	2
565	USA	1973	Faecal isolate	7	2
515	Czech Republic	1983	No clinical information available	7	2
524	Czech Republic	1983	No clinical information available	7	2
521	Czech Republic	1983	No clinical information available	7	2
522	Czech Republic	1983	No clinical information available	7	2
507	Czech Republic	1984	17 year old adult male; faecal isolate	11	-
512	Czech Republic	1983	No clinical information available	11	-
514	Czech Republic	1983	No clinical information available	11	-
685	USA	1977	Blood isolate	53	-
687	Czech Republic	2004	82 year old adult male; sputum isolate	60	-
689	Czech Republic	2005	26 year old adult; faecal isolate	60	-
15	Czech Republic	2003	66 year old adult; faecal isolate	60	-
1545	Czech Republic	-	Faecal isolate	84	2
1546	Czech Republic	-	Bed swab	84	2
1558	Czech Republic	-	Faecal isolate	89	2
1569	USA	2011	<1 month infant; blood isolate	112	-

Table 5.5 Details of the clinical *C. malonaticus* strains included in this study. The strains have been arranged in ascending order of their STs. The STs 11, 53, 60 and 112 are not associated with any clonal complexes to date.

5.2.5 Correlation of *Cronobacter* STs with biogroups

A study was conducted to investigate the correlation of the *Cronobacter* spp. STs with the various biogroups defined by Farmer *et al*. (1980) and Iversen *et al*. (2006) for the organism. Earlier, the biotyping scheme was one of the popular means of discriminating between or for grouping of bacterial strains. Iversen *et al*. (2006) had also studied the association of 16S rDNA clusters in the *Cronobacter* genus (then *E. sakazakii*) with these biogroups. The biogroups were also, in part, used for the definition of the initial *Cronobacter* genus (Iversen *et al*. 2007). These have been discussed in detail in Chapter 1, Tables 1.1 and 1.2.

For this study, biotyping results available for 163 *Cronobacter* spp. strains spanning all the seven species were obtained by downloading from the *Cronobacter* PubMLST database or by conducting a literature search for strains used in previous publications. A population snapshot of these biotyping results plotted against the STs of the strains has been presented in Fig. 5.3. Eleven of the 31 defined biogroups (35%) were found to be shared between at least two C*ronobacter* species. A number of these were found to be between the closely related species *C. sakazakii* and *C. malonaticus.*

16/31 biogroups were found in *C. sakazakii*, the most frequent ones being Biogroups 1, 2 and 13. Biogroup 1 was found to be exclusive to *C. sakazakii,* and spread across a number of STs, including the key meningitis related ST4. It was also identified to be the "wild type" biogroup in the initial study by Farmer *et al*. (1980). All the *C. sakazakii* ST1 and ST3 strains were found to be Biogroup 2. This biogroup was also observed in one strain of *C. dublinensis* (NTU strain 583; ST91). Both ST4 and ST8 strains were found to comprise of a number of biogroups, though mainly dominated by Biogroups 1 and 13. Among the 49 clinical *C. sakazakii* strains in this sample set, the predominant ones were biogroups 13 (19 isolates) and 1 (10 isolates).

Fourteen biogroups were found to be distributed across the *C. malonaticus* STs, of which ST7 was dominated by Biogroup 9. Biogroups 4a and 13a were found to be shared with *C. sakazakii*. The index strain for Biogroup 4a (NTU strain 515; ST7) as well as that for Biogroup 13a (NTU strain 535; ST7) belonged to *C. malonaticus*, while the rest of the members of these biogroups were *C. sakazakii* strains. The index strain was the representative strain chosen for each biogroup, defined by Iversen *et al*. (2006).

Among the four *C. universalis* strains, the type strain NCTC 9529^{T} (NTU strain 581; ST54) had been reported to be Biogroup 16c by Iversen *et al*. (2006). The remaining three strains of this species, characterised in a later taxonomic study (Chapter 4) were found to belong to Biogroup 16a. This biogroup was shared with the species *C. turicensis*, with the index strain being the *C. turicensis* NTU strain 92 (ST35). The strains of *C. turicensis* comprised 6 different

biogroups among them, four of which were found to be shared between *C. sakazakii, C. malonaticus* or *C. universalis.*

The *C. dublinensis* strains were found to be of five different biogroups, all of which were found to be shared with *C. sakazakii* strains. For the species *C. muytjensii*, apart from one strain, all the rest were found to belong to only Biogroup 15, which was also found to be shared with an isolate each of *C. sakazakii* and *C. turicensis.* The lone exception was *C. muytjensii* strain 1129 which belongs to Biogroup 6.

The lone *C. condimenti* strain, though previously reported to be Biogroup 1 (Turcovsky *et al*. 2011), was found to be incorrect in our laboratory studies. It was identified to be a previously unidentified biogroup and was eventually categorised in this study as Biogroup 10a (described in Chapter 4).

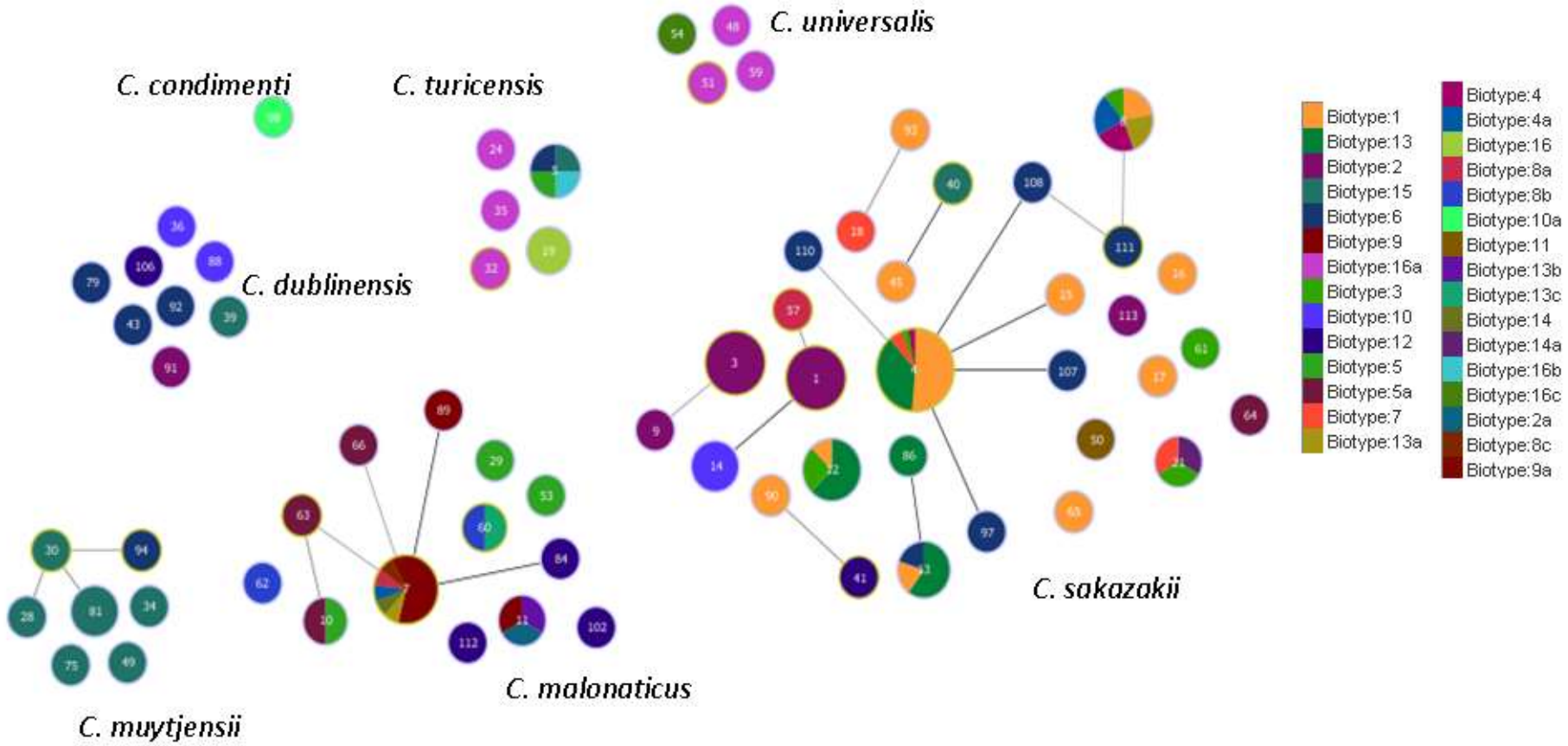


Fig. 5.3 Population snapshot graph created using the goeBURST algorithm to plot the *Cronobacter* biotypes against the STs. The dataset comprised of 163 *Cronobacter* isolates distributed across 68 STs. Figure as published in Joseph *et al*. (2013a). Lack of continuity between *Cronobacter* biotypes and species as determined using multilocus sequence typing. *Molecular and Cellular Probes*, 27, 137-139, http://dx.doi.org/10.1016/j.mcp.2013.02.002

5.3 Discussion

MLST is a very useful and powerful tool for conducting population genetics studies of bacterial species. The results presented here have shown how the *Cronobacter* MLST dataset was applied to investigate the clustering of the sources of the 325 *Cronobacter* spp. strains in the dataset, especially from the point of view of epidemiology.

The previous *C. sakazakii – C. malonaticus* MLST study by Baldwin *et al*. (2009) comprising of 60 clinical and non-clinical strains had indicated that *C. sakazakii* ST4 was the major (22/60) sequence type isolated from a wide range of sources such as clinical, infant formula, food ingredients, and environmental. However, the main focus of that study had been the application of MLST to discriminate between *C. sakazakii* and *C. malonaticus* and hence the study did not consider epidemiologic applications. As a result, the clinical significance of *C. sakazakii* STs was not investigated in detail at that stage.

5.3.1 Clonality of the genus *Cronobacter*

The overall clonality of the genus *Cronobacter* was studied using the goeBURST algorithm and I_A values to identify linkage in the population (Section 5.3.1). In all, 13 clonal complexes were observed across the genus. The fact that the majority (9) of them belonged to *C. sakazakii,* confirmed the results reported in Chapter 3 where much lower levels of recombination events were observed in the *C. sakazakii* species as compared to the rest of the genus. The I_A value of 1.34 observed using the current dataset, was higher than the value of 0.28 reported by Baldwin *et al*. (2009), indicating increased clonality. This is explained by the much larger sample set in this study as well as the greater number of CCs seen in this data as opposed to only two (*C. sakazakii* ST1-ST14 and *C. sakazakii* ST4-ST15) observed in the data presented by Baldwin *et al*. (2009).

The ST4 clonal complex lineage was of great interest revealing an interesting link with *Cronobacter* spp. epidemiology, and has been discussed below in detail in Section 5.3.2. Apart from this lineage, there were also other clusters of interest which were observed in this dataset. Some of these other CCs too were of relevance from the point of view of *Cronobacter* epidemiology (Table 5.1). One of them proved to be CC-1 comprising of ST1 and ST14. ST1 was the second most frequent ST seen in the database, with 36 *C. sakazakii* strains, eight of them clinical strains. All of these clinical strains had been isolated from cases of neonatal infections, though ST4 strains were the ones more associated with neonatal meningitis. Nevertheless, meningitic infections of neonates by *C. sakazakii* ST1 do occur. The most well-known was an outbreak in a neonatal intensive care unit in Tennessee (USA) which was reported by Himelright *et al*. (2001). The isolate (strain ATCC BAA-894) from the associated

formula (non-infant formula) has been the first genome sequenced strain of the genus (Kucerova *et al*. 2010 and 2011). The same strain had also been isolated from the CSF of one of the meningitic infants (NTU strain 679). In a separate case two years later in the USA, an ST1 isolate was obtained from a fatal neonatal meningitis case (NTU strain 1218).

Another ST observed to be of importance was *C. sakazakii* ST8. This ST comprised of 9/13 strains linked to clinical cases, albeit with no confirmed association with meningitis. ST8 has shown itself to be a stable clone isolated over more than 30 years, the most recent ones being faecal and PIF isolates from the *Cronobacter* outbreak in Ohio, USA in December 2011. An interesting aspect of the faecal isolates from the Ohio case was that the two samples were obtained a month apart from the same infant suffering from ongoing diarrhoea, thus indicating the colonisation and persistence of the organism without invasion. This carriage of the organism is an important issue with respect to the transfer of the organism between siblings and carers of the babies. Even though it was not found to be part of a clonal complex, ST8 did reveal a distant link with ST4 and the ST4 clonal complex, as described in Fig. 5.2. Based on the goeBURST analysis, it is difficult to conclude whether ST8 could be a clonal descendant of ST4, since these are only indicative evolutionary relationships. This connection does correlate with the epidemiology of a majority of the clinical strains of these STs, despite the unconfirmed clinical presentations. However, it is highly possible that this link could be an artefact due to the chain formation between the multiple STs. Hence, this result needs to be observed with caution before any further conclusions can be made.

The other *Cronobacter* spp. strains that have been most isolated from clinical cases are members of the *C. malonaticus* species (Section 5.2.4). A review of the clinical strains of this species revealed almost all of them with known patient details to be isolates of adult *Cronobacter* infections. Most of the clinical isolates belonged to CC-2 (STs 7, 84 and 89), dominated by the ST7 strains (Tables 5.1 and 5.5). This is an aspect of the epidemiology of the organism which is quite often overlooked because of the more severe and sensitive neonatal meningitis cases. In fact, *Cronobacter* infections are actually more frequent among the adults, though not as destructive (FAO/WHO 2008). *Cronobacter* spp. have been isolated from adults in cases of bacteraemia, sepsis, pneumonia as well as wound infections (Lai 2001).

These clustering of non-meningitic but clinical strains in STs such as ST1, ST8 and ST7 could also be a reflection of alternative routes of infection of the organism such as other foods, commensal colonization in the throat and the environment. Thus, the *Cronobacter* MLST scheme has been successful in the clustering of the clinically associated *Cronobacter* spp. strains into specific STs/clonal complexes to form stable virulent lineages. A similar phenomenon has also been seen in the MLST of the meningococcus, *Neisseria meningitidis*. Like *Cronobacter* spp., this organism is also a commensal and an opportunistic pathogen, known to cause infections such as meningitis and sepsis, even though the mode of meningitic

infection is quite different than that of *Cronobacter* spp. The MLST analysis of *N. meningitidis* has revealed over the years a large number of clonal complexes. However, only a select few of them have been found to be associated with severe infection and disease. These are classified as the hyper-invasive lineages, as opposed to the STs/clonal complexes which consist of a larger majority of non-disease causing carrier isolates (Caugant 2008; Caugant & Maiden 2009). *N. meningitidis* also happens to be the first bacterial population that the MLST was applied to by Maiden *et al*. (1998).

Clustering patterns were observed among the *Cronobacter* spp. isolates, from the point of view of some other isolation sources as well. Isolation of the organism from infant formula samples has always been a matter of significant interest, especially since contaminated PIF has been implicated in *Cronobacter* infections in the past (Muytjens *et al*. 1983). This sample set of 325 *Cronobacter* strains contained 44 PIF isolates of which more than one third (16/44) belonged to the ST4 clonal complex [Fig. 5.1 (b)]. There were also 9 PIF isolates belonging to CC1 comprising of ST1 and ST14, as well as two ST8 isolates. Majority of these PIF isolates were also found to be *C. sakazakii* (37/44), while the rest were *C. malonaticus* (5/44) and *C. muytjensii* (2/44). Currently, the premise of associating neonatal infections exclusively to powdered infant formula is limited due to the lack of knowledge of *Cronobacter* species and sequence type from sources other than PIF and manufacturing plants.

Another interesting aspect of clustering of strains in STs was observed in a study conducted on 64 *Cronobacter* spp. strains isolated from the processing areas and environment of milk powder manufacturing plants in Australia and Germany (Craven *et al*. 2010; Jacobs *et al*. 2011). These strains had been isolated from six different plants (5 Australian and 1 German) over the period of 2006-2009 [Fig. 5.1 (b) & (c)]. These 64 *Cronobacter* spp. strains were found to be distributed over 30 different STs. However, this was not a balanced distribution. 25% of these strains (16/64) were found to belong to the ST4 clonal complex, while there were also 13 ST1 strains. This further revealed the persistence and dominance of these STs among the *Cronobacter* spp. strains. The results of these studies also demonstrated that *C. sakazakii* ST4, found to be associated with neonatal meningitis, could also be present in the environment of milk powder factories such as tanker bay, shoes, roof, roller-dryer, spray-drying area and milk powder. Also, since both these isolation studies had been carried before the taxonomic revisions of genus *Cronobacter,* this MLST study helped in the accurate species-level identification of the strains, revealing a predominance of species *C. sakazakii* (58/64) among the isolates.

The clustering of infant formula isolates and isolates from milk powder processing environments in specific STs further strengthens the reliability of this MLST scheme as an effective tool for not just species level typing but further subtyping of the strains too. Although a number of previous detailed surveys of *Cronobacter* spp. (then known as the single species *E. sakazakii*) in infant formula and manufacturing environments have been published, their data is

of limited value due to the recent taxonomic revisions, and sequence based genotyping schemes such as MLST. Therefore, the re-analysis of these strain collections is a valuable contribution to the scientific community and public health authorities. These results hold great significance for industries and manufacturing plants, opening up the possibility of using MLST as a typing tool for quality control checks of the final product.

It was also observed that a number of the food isolates were *C. malonaticus* and *C. dublinensis*, apart from *C. sakazakii.* A majority of these were isolates obtained from China and Korea. Also, interestingly all the *Cronobacter* spp. strains in this study that had been isolated from FUF had been obtained from multiple sources in Korea. Of course, the fact that this observation could have been because of a sampling bias cannot be overlooked [Fig. 5.1 (b) & (c)].

5.3.2 A molecular signature for neonatal meningitis

An epidemiological study conducted in January 2011 has been described in Section 5.2.2, wherein a total of 40 clinical *C. sakazakii* strains were analysed by comparing their STs with their clinical presentations and site of isolation (Table 5.2). Nine of these had also been used in the earlier study by Baldwin *et al*. (2009) which did not consider patient details or clinical presentations. As is common in the use of strains collected from a wide range of sources, the source details varied with respect to clinical presentation and site of infection. Nevertheless, this analysis had revealed an interesting pattern of the ST distribution among the 40 clinical *C. sakazakii* strains. ST4 was found to dominate this set of strains with 50% of them being identified as ST4, half of which were also found to be neonatal isolates. When viewed from the point of view of patient details too, it was seen that the majority of the strains from meningitis cases (>80%) were found to belong to ST4. Therefore, it could be clearly proposed that *C. sakazakii* ST4 presented a molecular signature for neonatal meningitis. An epidemiological lineage of this kind has previously not been observed for *Cronobacter* spp. through any other form of technique. In contrast to the correlation of ST4 with meningitis, correlation with ST4 or other STs was not found with other clinical presentations such as necrotizing enterocolitis. This association of meningitis and severity of infection with ST4 was supported by a re-evaluation of the 1994 NICU *Cronobacter* spp. outbreak reported by Caubilla-Barron *et al*. (2007). During this major outbreak in France which included fatalities, *C. sakazakii* had been recovered from 21 neonates. Among these, the strains (NTU strains 695,701, and 767) recovered from the three fatal cases were all ST4 (Table 5.2). The other ST4 strain (NTU strain 709) was isolated from a case of septicaemia. The remaining strains recovered from this outbreak were associated with non-fatal NEC and asymptomatic colonisation. These were in ST12 (NTU strains 696 & 690), and ST13 (NTU strain 693).

Since this initial study, additional strains were obtained from various sources till the endpoint of this *Cronobacter* MLST project was reached with 325 *Cronobacter* spp. strains. At this stage, 53 STs of 226 *C. sakazakii* strains had been determined. Of these, 78 strains were found to belong to ST4 (Section 5.2.3). Thus, more than one third of all the *C. sakazakii* isolates, isolated over a 50 year period, belong to this sequence type, also making it the most frequent ST among the *Cronobacter* genus. Also, 73 strains out of the 226 *C. sakazakii* strains had been obtained from clinical cases and even among them more than 50% of the strains were ST4, thus reinforcing the clinical significance of the sequence type. At this stage, our laboratory had received from CDC, USA fifteen strains from *Cronobacter* cases in five different states in the US in 2011, some of which had resulted in fatalities too (Table 5.4). The MLST of these strains revealed an interesting connection to the ST4 story, by establishing an ST4 clonal complex which revealed a meningitic lineage for *Cronobacter* spp. infections (Section 5.2.3). Out of the 18 meningitic isolates in this 325 isolate study, 14 isolates were found to belong to the ST4 clonal complex (78%). This was also found to be a robust lineage isolated over a 60 year period, with the earliest isolate being an ST4 strain (NCIMB 8272; NTU strain ID 377) by Betty Hobbs from a can of dried milk (Farmer *et al*. 1980).

Thus, after the previous application of MLST to an internationally distributed collection of *Cronobacter* strains from 1953-2008 had revealed a strong association of *C. sakazakii* ST4 with cases of neonatal meningitis, this latter stage MLST analysis of the 15 *Cronobacter* spp. strains from 2011 received from the CDC further reinforced the conclusion of the previous study that CSF isolates are not evenly spread across the seven *Cronobacter* species and are instead predominantly in the *C. sakazakii* ST4 clonal complex. Such infections in neonates are of very high concern due to the associated severe brain destruction of the sensitive target age group of neonates, resulting either in fatalities or permanent brain damage (Bowen & Braden 2008).

5.3.3 Correlation of STs and biotypes

As part of the investigation to observe the diversity of strains across STs, a smaller study was also conducted to observe the correlation of the biogroups of the *Cronobacter* spp. strains and their STs, using the available information for 163 strains (Section 5.2.5). This analysis did reveal considerable discrepancies among the two typing techniques. Among the 31 defined biogroups included in this study, 11 were found to be shared between at least two different species, some of these already reported by Baldwin *et al*. (2009). Interestingly, the species *C. muytjensii* which has earlier been reported to be very genetically diverse (Chapters 3 and 4) was dominated by only a single biotype 15, suggesting that the tests chosen for the biotyping scheme were able to define this species correctly, even though not necessarily

reflecting the diversity present in it. To summarise, this study helped to further prove the lack of reliability in using the biotyping scheme for species level identification of the *Cronobacter* genus. These results correspond to the observations made by Baldwin *et al*. (2009) even though their study had involved only the two species *C. sakazakii* and *C. malonaticus.* It is undeniable that this biotyping scheme consisting of 10 differential tests chosen by Farmer *et al*. (1980) played a crucial role in the defining of the species *E. sakazakii*. However, with more sophisticated and accurate DNA based methods available for typing, the dependency on the scheme has been found to decline over the years, especially also keeping in mind the high degree of subjectivity involved in the reporting of the results of these phenotypic tests.

5.3.4 Conclusions

In conclusion, even though the 7 housekeeping genes used for the MLST scheme are not virulence related, they appeared to demonstrate a clonal distribution of a meningitic trait. The *C. sakazakii* ST4 clonal complex appears to be a very stable lineage as the strains have been isolated from more than six countries, over at least a 60 year period from 1951 to 2011. Whether this association is due to greater neonatal exposure as a result of environmental persistence or particular virulence capabilities due to prolonged host adaptation is as yet uncertain. In addition, it could also be speculated that the predominance of *Cronobacter malonaticus* infections in adults could be due to exposure to different *Cronobacter* STs according to diet and life-style.

This study has showed that the predominant *Cronobacter* species from clinical sources was *C. sakazakii,* and that *C. sakazakii* ST4 was the predominant sequence type of isolates from cases of neonatal meningitis. This coincides with our finding that the majority of isolates from PIF and milk factories were *C. sakazakii*, including ST4. Other STs of significance such as ST1, ST8 and ST7 were also identified. To date, studies on the desiccation resistance and other environmental persistence traits have not focussed on these sequence types, in part due to the recent development of the MLST scheme.

Only by using the partial sequences of seven housekeeping genes, the MLST study has revealed a lot of novel information about the diversity and virulence of the genus *Cronobacter* (Chapters 3, 4 and 5). Therefore, the obvious next step is to expand this study to a whole genome level to further investigate these aspects in details. The following chapters (6 and 7) report a comparative genomic analysis conducted using the draft genome sequences of 13 *Cronobacter* spp. strains spanning the seven species of the genus. These include representative strains of the key *Cronobacter* STs – ST1 (*C. sakazakii* BAA-894; Kucerova *et al*. 2010), ST4 (NTU strain 701), ST8 (NTU strain 680), ST7 (NTU strain 681) along with other species representatives.

CHAPTER 6

THE *CRONOBACTER* PAN-GENOME AS REVEALED BY COMPARATIVE GENOMIC ANALYSIS

6.1 Introduction

6.1.1 Aims of the chapter

This chapter is the first of two presenting the results obtained in a whole genus comparative genomic analysis. This study was conducted on fourteen *Cronobacter* spp. genomes to investigate the diversity and virulence of the organism across the genus. This genomic study is a follow up to the earlier study conducted by our research group (Kucerova *et al*. 2010) which had published the first completely sequenced genome of the *Cronobacter* genus – *C. sakazakii* BAA-894 and had used it in a microarray based comparative genomic hybridisation (CGH) study of five species of the genus – *C. sakazakii*, *C. malonaticus*, *C. turicensis*, *C. muytjensii* and *C. dublinensis*. This genus-level study also aims to expand to a whole genomic level the diversity observed through the MLST based analysis in Chapters 3, 4 and 5.

In this chapter is presented the details regarding the sequencing of the genomes and some of the main features of the *Cronobacter* core and accessory genome, including the mobile genetic elements identified in the genomes. This was a co-operative study conducted with two different research groups - Life Technologies (Foster City, CA) performed the genome sequencing and assemblies of the draft genomes, while Prof. Michael McClelland and Dr. Prerak Desai (Vaccine Research Institute, San Diego) calculated the pan-genome statistics for the fourteen *Cronobacter* spp. genomes. All the remaining results reported here have been obtained by me at NTU following the analysis of the genomes.

As the *C. sakazakii* BAA-894 genome was used as a starting reference point for the analysis of genomic regions, the Genbank loci numbers from the genome have been cited wherever appropriate. When referring to regions from the genomes sequenced in this study, their genes have been indicated as "peg numbers", as provided by the annotations in the RAST server.

6.1.2 Next generation sequencing technology

The concept of DNA sequencing as we understand it today has made considerable process since Sanger and colleagues first devised a means to determine the sequence of DNA. They did this by sequencing the DNA of the bacteriophage ΦX174 by a gel-based technique using dideoxynucleotides (Sanger *et al*. 1977). This original technique itself was adapted in the early 1990s, into a semi-automated, capillary based method, which came to be more popularly known as the Sanger sequencing or "first-generation" sequencing technology (Hunkapiller *et al*. 1991). The following decade saw the boom of more advanced sequencing techniques, called

"next-generation" sequencing (NGS) that involved a combination of template preparation, sequencing, imaging, alignment and assembly, all packaged into automated commercial machines. This brought with it increased competition among companies marketing their NGS machines, with an aim to obtain the highest quality output at much reduced costs. Some of the leading NGS platforms currently in use include 454 Pyrosequencing (Roche), Solexa Genome Analyzer (Illumina), SOLiD® System (Applied Biosystems) and the Ion Torrent PGM™ (Life Technologies), among others. These platforms vary between each other on the basis of various parameters such as sequencing chemistry, library preparation, read lengths and overall cost incurred for the sequencing process (Loman *et al*. 2012a; Mardis 2008; Metzker 2010; Shendure & Ji 2008).

A generic workflow pattern for the genome sequencing that is most commonly used by these machines would be as follows: In shotgun sequencing, the whole genomic DNA is first broken down randomly into a number of overlapping fragments for the purpose of a high-quality library preparation. The fragmented library is then amplified and sequenced to obtain millions of short reads of the original template. The protocols and parameters for the library preparation as well as sequencing vary according to the chemistry and requirements of the sequencing platform being used. The depth of the coverage of the genome depends upon the number of times each region of the DNA is sequenced and read. Longer reads and increased coverage ensures a better quality output. The reads are then aligned to obtain overlapping ends and thereby joined together into many 'contigs', which are further arranged to obtain 'scaffolds'. The scaffolds are then assembled to obtain a final draft version of the genome. An often used parameter for assessing the quality of a draft genome is the N50 value of the contigs or scaffolds, The N50 refers to an identified cut-off value for the lengths of the contigs/scaffolds, such that at least 50% of the contigs/scaffolds of the genome are found to be above this value. A higher N50 value is an indicator of a good quality genome assembly. The genome assembly process can either be *de novo* or by mapping to an available reference genome. There is a number of automated assembly software available today which can be used in conjunction with the NGS platforms, such as MIRA (Chevreux *et al*. 2004) and Velvet (Zerbino & Birney 2008), among others. Once the assembled draft genome sequence is obtained, it is then annotated to identify the genes and their functions. Nowadays, for this purpose too, there are automated annotation pipelines available, such as RAST (Aziz *et al*. 2008) and xBASE (Chaudhuri and Pallen 2005), that are popularly used for bacterial genomes, making the job easier.

A finished or nearly finished, high-quality genome is obtained by a combination of sequencing technologies, as well as manual and automated improvements to resolve the gaps and errors that may be introduced in the sequencing process. The majority of the genomes available in the public domain today, obtained from NGS technologies, are draft genomes, with

≤ 90% of the genome sequenced and the presence of a certain amount of errors incorporated during the sequencing process. Since a number of major studies nowadays are being carried out using these draft genomes, even among them there are certain minimum standards that are being required to be met in order to make them fit for use for the particular study (Shendure & Ji 2008; Chain *et al*. 2009; Loman *et al*. 2012a).

These NGS technologies have considerably advanced the field of microbiology, especially in improving the microbiologists' understanding in areas such as population genetics, epidemiology, metagenomics and so on. The first complete genome of a bacterium was obtained in 1995 for the species *Haemophilus influenzae*, and this was carried out using the traditional Sanger sequencing method (Fleischmann *et al*. 1995). Today, less than two decades later, the NCBI Microbial Genomes database has 13,748 entries listed for bacterial genomes, though only 2,033 of them are complete genomes, while the rest are in various qualities of draft forms (NCBI Microbial Genomes). Of course, these only constitute the number of genomes that have been submitted by researchers into the public domain, and therefore is probably a gross under estimation of the actual number of bacterial genomes that must have been sequenced.

The comparative genomic study reported in this chapter used draft genomes sequenced using the SOLiD® 4 and Ion Torrent PGM™ technologies. Hence, the science behind these two platforms has been described in the following sections.

6.1.3 SOLiD ® 4 sequencing technology

ABI SOLiD ® 4 (Sequencing by oligonucleotide ligation and detection) technology is an NGS platform commercially marketed by Applied Biosystems (Life Technologies, Foster City, CA), first introduced in the market in 2008.

For SOLiD sequencing, the DNA to be sequenced is created into a fragment library and these fragments are then bound to a population of para-magnetic beads (1 μM in size), such that each bead is bound by only one unique fragment of DNA. Universal P1 adapters are then bound to these fragments and the clonal beads are subject to emulsion PCR. In this PCR method, a microreactor environment is created using a water-in-oil emulsion in which independent PCR reactions can be carried out using a clonal bead population. The beads with the amplified products are then enriched by breaking the emulsion and are bound to a solid substrate (usually a glass slide) by covalent linkage to form a dense array of beads. The sequencing reaction is then carried out using this array. In this technology, the sequencing is carried out using a ligase enzyme, as opposed to a polymerase. The ligation is carried out by introducing octamers into the reaction which act as the probes, with fluorescently charged two-base labels attached to them. The fluorescent labels are designed so as to correlate to the 4^{th} and 5^{th} base of the octamer.

The ligation and fluorescent imaging process helps to determine the sequence of the 4th and 5th base, followed by cleavage and removal of the labels by altering the linkage between the 5th and 6th base. Next cycle of ligation then continues until sequences are determined at a distance of every 4th and 5th base along the length of the template (e.g. 4 &5, 9 &10, 14 & 15 and so on). A new cycle is then commenced with a new set of primers, now binding at (n-1) position compared to the previous cycle, and this is repeated until the entire length of each fragment is sequenced. Millions of such sequencing reactions can be carried out in parallel, thus enabling a maximum output in a single machine run.

Read lengths of 30-35 bp can be obtained for SOLiD sequencing technology. Also each sequencing run can generate an output of about 150 Gb of data. The di-base encoding process ensures that each base is sequenced twice, thus acting as an inherent proof reading mechanism by screening for base-calling errors and thus increasing the overall efficiency of the process. This is considered as the main advantage of the SOLiD platform compared to other NGS technologies. However, it does have its share of disadvantages; the main ones being the short read lengths obtained which may introduce complexities in the assembly process as well as the possible difficulties in the emulsion PCR process (Mardis 2008; Shendure & Ji 2008; Loman *et al*. 2012a).

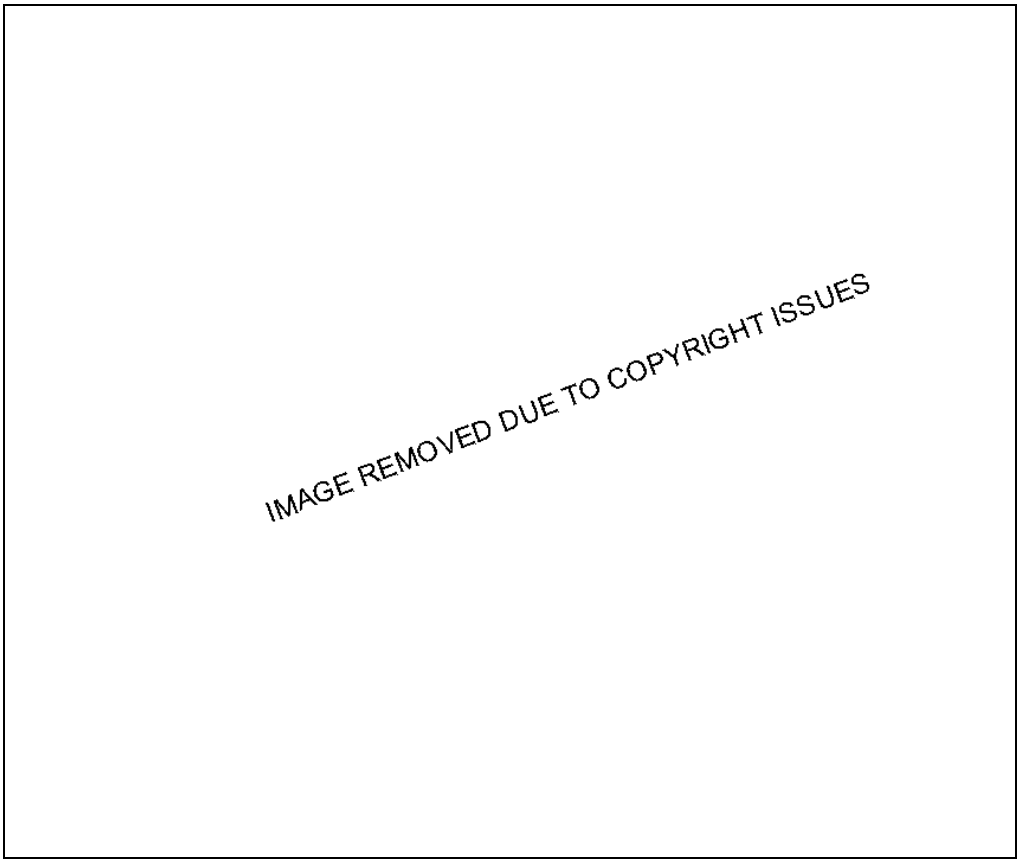


Fig. 6.1 A graphical depiction of the ligase-mediated sequencing approach of the SOLiD® 4 technology (Applied Biosystems) using octamers and di-base encoding process. Source: Mardis (2008).

6.1.4 Ion torrent PGM™

The Ion Torrent Personal Genome Machine (PGM) is a commercial NGS technology owned by Life Technologies, launched in 2010. It differs from other NGS technologies, in that it has a complementary metal-oxide semiconductor (CMOS) technology which uses an ion-sensitive field-effect transistors (ISFET) system based on proton detection instead of optical imaging for the sequencing process.

The library preparation for the Ion Torrent is similar to other platforms such as SOLiD wherein the template DNA is fragmented and amplified on the surface of clonal beads. These beads are then introduced into 'ion chips'. These ion chips have wells with 1.2 million metal oxide sensors. The template DNA-bound beads are introduced into these wells with primers and polymerase for enabling the sequencing reactions. Centrifugation of the chip ensures that each well contains an independent sequencing reaction. The four different nucleotides are added in succession into the reaction. When the appropriate complementary nucleotide is present, the DNA polymerase incorporates it onto the template, resulting in a hydrolysis reaction. The release of the proton causes a change in the pH of the reaction which is eventually detected by the sensors. The signals are then processed by software and thereby used for the base calling and sequence determination of the template DNA (Fig. 6.2).

The Ion Torrent PGM™ can give a final read length of 100-200 bases, with the entire run time being as short as 3 hours. In a recent comparative study with other bench-top NGS machines, the Ion Torrent PGM™ was found to be the best for throughput obtained as well as for the shortest run time (Loman *et al*. 2012b). The lack of optics in the process also reduces the overall cost incurred in the sequencing process. From a microbiological point of view, this technology has a large number of useful applications in the characterization of pathogen genomes, metagenomics as well as transcriptomics studies (Rothberg *et al*. 2011; Loman *et al*. 2012b).

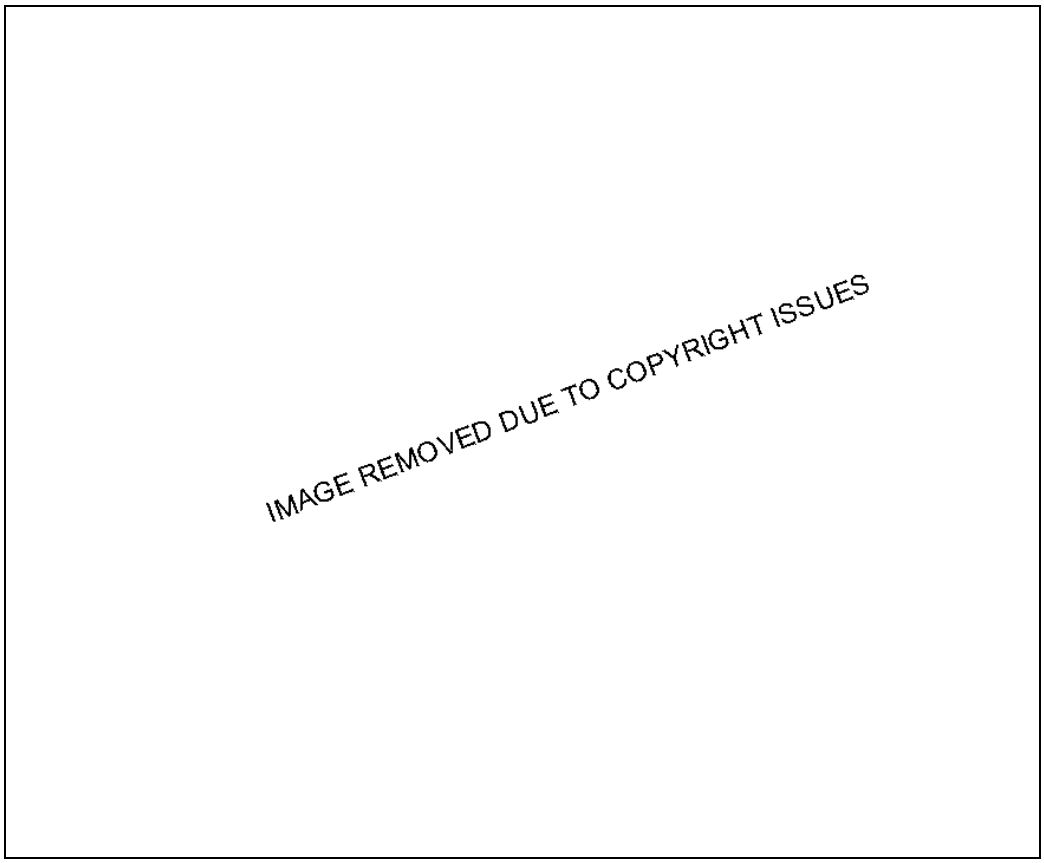


Fig. 6.2 Graphical representation of the sequencing reaction that takes place in an Ion Torrent PGM™ resulting in the release of the proton that alters the pH of the environment. Source: Rothberg *et al*. (2011).

6.1.5 The bacterial pan-genome

The term "pan-genome" refers to the sum total of all the genes present in a bacterial population, either at a species or a genus level. The bacterial pan-genome basically comprises of two sections: the core genome and the accessory genome. The core genome comprises of all the genes that are conserved in all the strains of the population, possibly a representative of the basic backbone of the organism's genome. The accessory genome on the other hand consists of the dispensable genes in the population, which may either be the ones unique to each strain or the ones that are present in two or more strains, but not in the entire population. The accessory genome is especially important in the population studies of pathogenic micro-organisms, as it may contain the key virulence factors of the organism.

A bacterial pan-genome may be classified as either open or closed. Some bacterial species have the tendency to exhibit increasing number of novel genes with each new strain, and such species are said to possess an open pan-genome. These are generally found to be organisms that have been known to be isolated from a wide variety of environments, thus increasing their chances of acquiring or evolving new genes. *E. coli* is a classic example of an open pan-genome, with some researches suggesting a mathematical concept of an infinite pan-genome for the organism. A closed pan-genome, conversely, is generally exhibited by species with a very small gene pool, possibly because of a closed niche or exposure to a unique environment, for example, *Mycobacterium tuberculosis.*

The pan-genome thus enables a quantitative estimation of how dynamic a bacterial population is and can thereby play an important role in understanding the evolution and adaptation of a bacterial species or a genus. With the many advances in bacterial genome sequencing over the last few years, comparative genomic studies of bacterial populations have enabled increasingly accurate estimations of the pan-genome (Medini *et al*. 2005; Tettelin *et al*. 2008; Mira *et al*. 2010).

6.1.6 Prophages

Bacteriophages, the viruses that infect bacterial cells may be either lytic or lysogenic in nature. The lytic phages are the ones that bring about destruction of the infected host cell after replication, and thereby releasing the progeny virus particles by the lysis of the dead cell. On the other hand, the phages that undergo the lysogenic cycle, also known as temperate phages, share a more stable relationship with their host cells. Majority of these temperate phages tend to integrate themselves with the genome of the bacterial host, when they come to be known as prophages. Some also remain in the cell without integration in the form of phage plasmids. Most of the genes in the prophages remain unexpressed in the integrated state. However, in some

situations such as starvation conditions, the lytic genes can get switched on by an induction process. The prophages also possess lysogenic conversion genes, which contribute to alterations in the host cell which may also include conversion from a non-virulent to virulent state.

Prophage genes are known to occupy about 20% of a bacterial host genome. The most frequently characterised Gram-negative bacterial prophages have been the double stranded tailed DNA phages. These prophages may exist either as complete genomes or as incomplete regions in the form of artefacts. Even the complete genomes do not necessarily exist in a functional form. These mobile genetic elements are hugely responsible for variations within genomes of the same bacterial species and also participate in the horizontal gene transfer process; therefore key contributors in shaping the evolutionary history of the organism. Hence, analysis of the prophage regions is a very pertinent aspect of a bacterial comparative genomic study (Casjens 2003; Canchaya *et al*. 2003; Canchaya *et al*. 2004).

6.1.7 CRISPRs

Prokaryotic genomes are often scattered with different types of repetitive nucleotide elements in their genome. One such group of tandem repeat elements are the "clustered regularly interspaced short palindromic repeats", popularly known as CRISPRs. These elements were first identified in *E. coli* by Ishino *et al*. (1987), though detailed studies on these regions have only developed over the last decade. As the name suggests, these elements mainly consist of 23-47 bp long repetitive nucleotide strings, known as "direct repeats" (DRs). Between two DRs is stored a genetic region known as the spacer element. The spacers in each CRISPR are of the same length, and have often been found to be of viral origin. CRISPRs may consist of any number of repeated units consisting of the DRs and spacers forming a genetic array. In the vicinity of the CRISPRs are located a group of *cas* or CRISPR- associated genes (generally present in a cohort of four), coding for conserved proteins. Studies on these genes have shown CRISPRs to be the consequences of horizontal gene transfer events via mega plasmids (Godde & Bickerton 2006; Garneau *et al*. 2010).

The importance of the CRISPR regions lies in their role in the bacterial immune system by protecting the host from phage and plasmid invasions. This involves a three step process of adaptation-expression-interference; wherein the CRISPR acquires a region of the invading DNA and transcribes it into its own mature RNA, which then binds to the complementary target DNA of the invader to form a DNA-RNA complex, which helps to cleave the DNA by endonuclease activity. This whole mechanism is carried out with the help of the *cas* genes (Al-Attar *et al*. 2011; Delihas 2011; Garneau *et al*. 2010).

In addition, CRISPRs have also been reported to play a different type of role by inhibiting motility and biofilm formation in *Pseudomonas aeruginosa* (Zegans *et al*. 2009). In

recent times, these repeat elements have also been used as tools for bacterial diversity studies (Pride *et al*. 2011; McGhee & Sundin 2012).

6.1.8 Plasmids

Plasmids are covalently closed, circular, extra chromosomal DNA elements in the bacterial cell. They are smaller in size than the bacterial chromosome and may be present in multiple copies in the cell, indicated by the copy number of the plasmid. Plasmids have their own self-replicating machinery in place, characterised by the origin of replication or the *ori* gene. They also possess *par* genes that participate in the segregation process to ensure the stability of the plasmid during host cell division. The efficiency of replication of a plasmid is found to increase with a higher copy number (Clowes 1972).

Plasmids have been found to be vehicles of conjugative transfer of DNA between bacterial cells. For this purpose, they possess a special group of genes called *tra* or transfer genes. The entire energy driven, conjugative transfer mechanism is found to be very much like a Type IV secretion system, a common feature in the plasmids of Gram-negative bacteria (also discussed in detail in Chapter 7). These conjugative systems are also found to be associated and working in conjunction with bacterial appendages such as pili and flagella.

Plasmids which undergo the same replication process have been found to be incapable of existing together within the same bacterial cell. Based on this criterion, plasmids are characterized as incompatibility or Inc plasmids. At present, there are 27 Inc groups of plasmids identified in *Enterobacteriaceae*. Another classification and typing method of plasmids in *Enterobacteriaceae* has been a PCR-based Replicon Typing (PBRT) method of the Inc group. In recent times, MLST is also being adapted as a tool for typing plasmids, known as pMLST, and has been applied for the sub-grouping of a number of Inc group plasmids such as IncI1, InchHI2, IncF and IncN (Couturier *et al*. 1988; Carattoli 2011).

Most plasmids possess selectable phenotypic traits, one of the most commonly targeted one being resistance to antibiotics. Plasmid genomes contain clusters of genes called antibiotic cassettes that play a very important role in the defence mechanism of the bacterial cell against the antibiotic activity. Apart from these, most plasmids also carry important virulence related genes related to secretion systems, toxins, siderophores, adhesion factors, metal resistance genes as well as genes related to metabolic functions. A number of these are accessory genes, more beneficial to the host than the plasmid itself. Through the evolutionary process, as a result of the effect of positive selection, plasmids may sometimes lose these accessory regions by conjugative transfer to the host chromosome (Frost *et al*. 2005; Johnson and Nolan 2009; Rankin *et al*. 2011).

Thus, plasmids are very important mobile genetic elements of bacteria that play a crucial role in driving forward the evolution of the organism. Hence, a study of the plasmids is a key aspect of characterizing the genome of a bacterial strain.

6.2 Results

6.2.1 The *Cronobacter* genomes

Using the SOLiD® 4 System and Ion Torrent PGM™ next generation sequencing platforms, the draft genomes of eleven *Cronobacter* spp. strains were sequenced which included three *C. sakazakii*, two *C. malonaticus*, one *C. muytjensii*, one *C. turicensis*, two *C. dublinensis*, one *C. universalis*, and one *C. condimenti* (Table 6.1) strains. The diversity and phylogenetic positions of these selected strains on the *Cronobacter* spp. 7-loci MLSA phylogenetic tree have been represented in Fig. 6.3., justifying the choice of these strains to span the diversity of the genus. The strains were selected based on their ST diversity and their source backgrounds, to represent the population size of each species. Two members of the species *C. dublinensis* were chosen in order to encompass the observed diversity as discussed in Chapters 3 and 4.

The final genome assemblies from the SOLiD® 4 System contain 1,600 to 3,100 contigs with N50 of 3.7 to 5.5kb and 260 to 1,170 scaffolds with N50 of 230 to 600 kb (Table 6.2). The sum of the contig lengths of all the genomes ranged from 4.4 to 4.9 Mb, in length, comparable to the genomes sizes of the previously sequenced *C. sakazakii* BAA-894 and *C. turicensis* z3032 genomes (Kucerova *et al*. 2010, Stephan *et al*. 2011). The publicly available, incomplete genome of *C. sakazakii* E899 (ST4) is only 3.96 Mb and appeared to lack plasmid sequences (Chen *et al*. 2011). Genome comparison revealed that pair-wise DNA sequence identity varies between 89 and 97% across these genomes of the *Cronobacter* genus.

The eleven draft *Cronobacter* spp. genomes sequenced in this study have been deposited to the EBI nucleotide database as whole genome shotgun sequencing projects and their accession numbers are listed in Table 6.1.

Species	Strain ID	Sequencing platform	Total contig length (bp)	N50 of scaffolds (bp)	No. of scaffolds	N50 of contigs (bp)	No. of contigs	Estimated no. of ORFs (bp)	Accession Numbers
C. sakazakii	680	Ion PGM	4,357,873	nd[a]	nd	51,120	194	4,178	CALG01000001-CALG01000201
	696	SOLiD	4,897,138	297,746	920	4,336	2,659	4,661	CALF01000001-CALF01000569
	701	SOLiD	4,752,729	346,235	1,171	3,538	3,148	4,509	CALE01000001-CALE01000768
C. malonaticus	507	SOLiD	4,447,701	373,979	464	3,703	2,361	4,226	CALD01000001-CALD01000249
	681	SOLiD	4,496,745	345,762	263	5,537	1,592	4,291	CALC01000001-CALC01000171
C. turicensis	564	SOLiD	4,500,608	411,105	263	4,796	1,807	4,227	CALB01000001-CALB01000114
C. dublinensis	582	SOLiD	4,677,592	229,230	539	3,822	2,657	4,483	CALA01000001-CALA01000427
	1210	Ion PGM	4,594,228	nd	nd	46,941	210	4,376	CAKZ01000001-CAKZ01000221
C. muytjensii	530	SOLiD	4,533,101	596,924	444	4,925	1,937	4,304	CAKY01000001-CAKY01000365
C. universalis	581	SOLiD	4,450,737	331,248	389	4,506	2,085	4,316	CAKX01000001-CAKX01000231
C. condimenti	1330	Ion PGM	4,469,5362	nd	nd	83,159	137	4,307	CAKW01000001-CAKW01000155

[a] nd indicates that scaffolding was not performed due to lack of mate-paired libraries

Table 6.1 *De novo* assembly statistics for the draft genomes of the eleven *Cronobacter* spp. strains sequenced for this study, using the ABI SOLiD® and Ion Torrent PGM™ systems. Published in Joseph *et al.* (2012c).

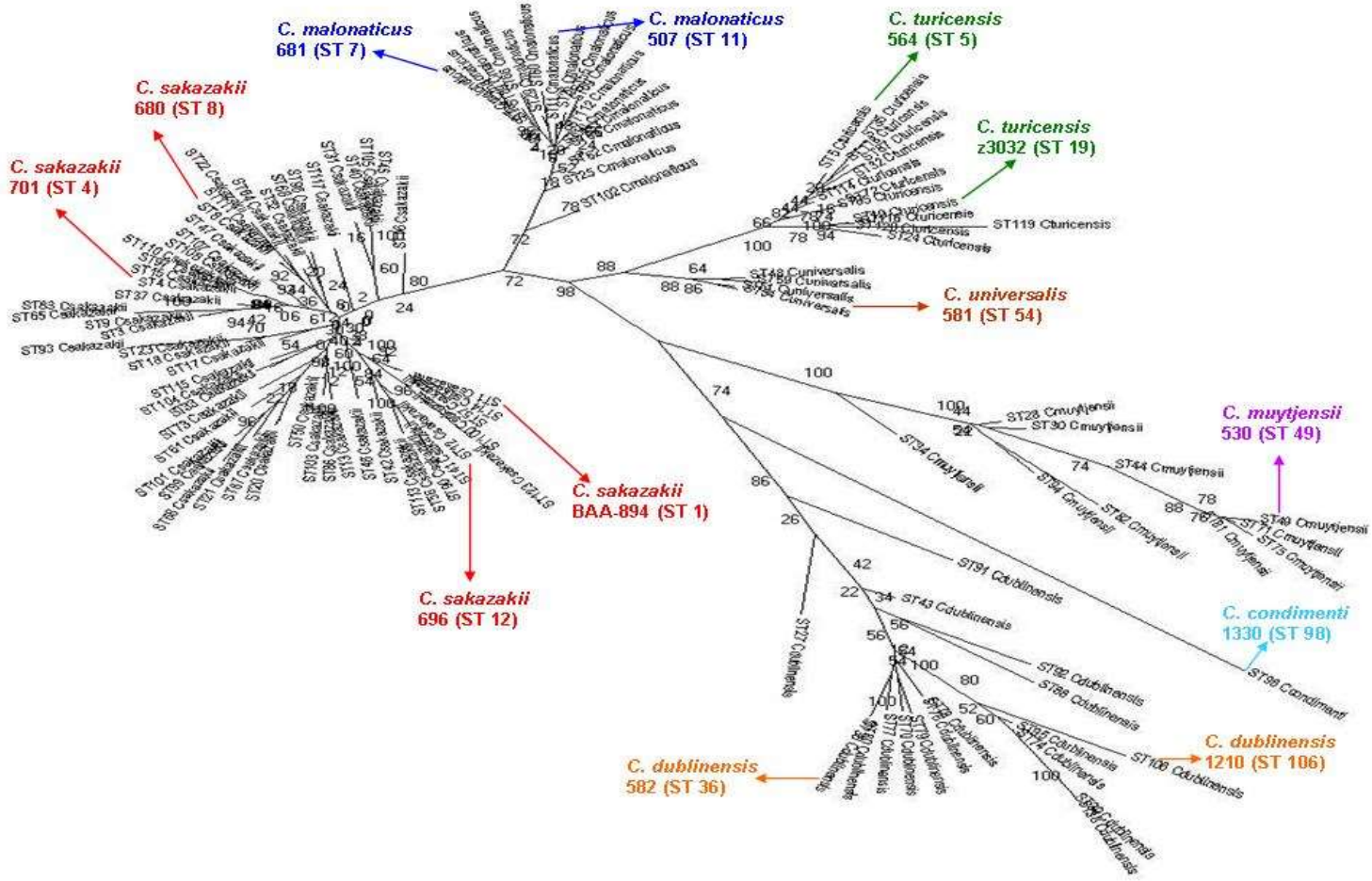


Fig. 6.3 Maximum-likelihood tree based on the concatenated sequences (3036 bp) of the 7 MLST loci for the genus *Cronobacter*, indicating the positions of candidate strains selected for genome sequencing. The STs and the corresponding species are indicated at the tip of each branch. The tree is drawn to scale using MEGA5 (Tamura *et al*. 2011), with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage. Published in Joseph *et al.* (2012c).

6.2.2 The *Cronobacter* pan-genome

All the eleven draft *Cronobacter* spp. genomes sequenced in this study were annotated using the automated RAST annotation server (Aziz *et al*. 2008). The number of annotated genes per genome varied between 3,700 and 4,200. The pan-genome analysis also included the previously published genomes of *C. sakazakii* BAA-894, *C. turicensis* z3032 and *C. sakazakii* E899 comprising 4562, 3718 and 4562 genes respectively.

The core and pan genome identification of the *Cronobacter* spp. genomes was determined with the help of our collaborators Prof. Michael McClelland and Dr. Prerak Desai at the Vaccine Research Institute, San Diego, CA. Using the binomial mixture model approach of Snipen *et al*. (2009), the estimated core genome was found to comprise 2,517 open reading frames (ORFs) while the estimated pan-genome comprises 7,775 ORFs. Thus, 32.37% of the *Cronobacter* pan-genome is estimated to be conserved across the genus. The current dataset represents 6,811 distinct ORFs, suggesting that 964 *Cronobacter* spp. ORFs (about 12% of the pan-genome) remain to be discovered. With additional sequencing, the number of genes in the current core genome is expected to drop from the observed value of 3,154 ORFs to 2,517 ORFs (Fig. 6.4).

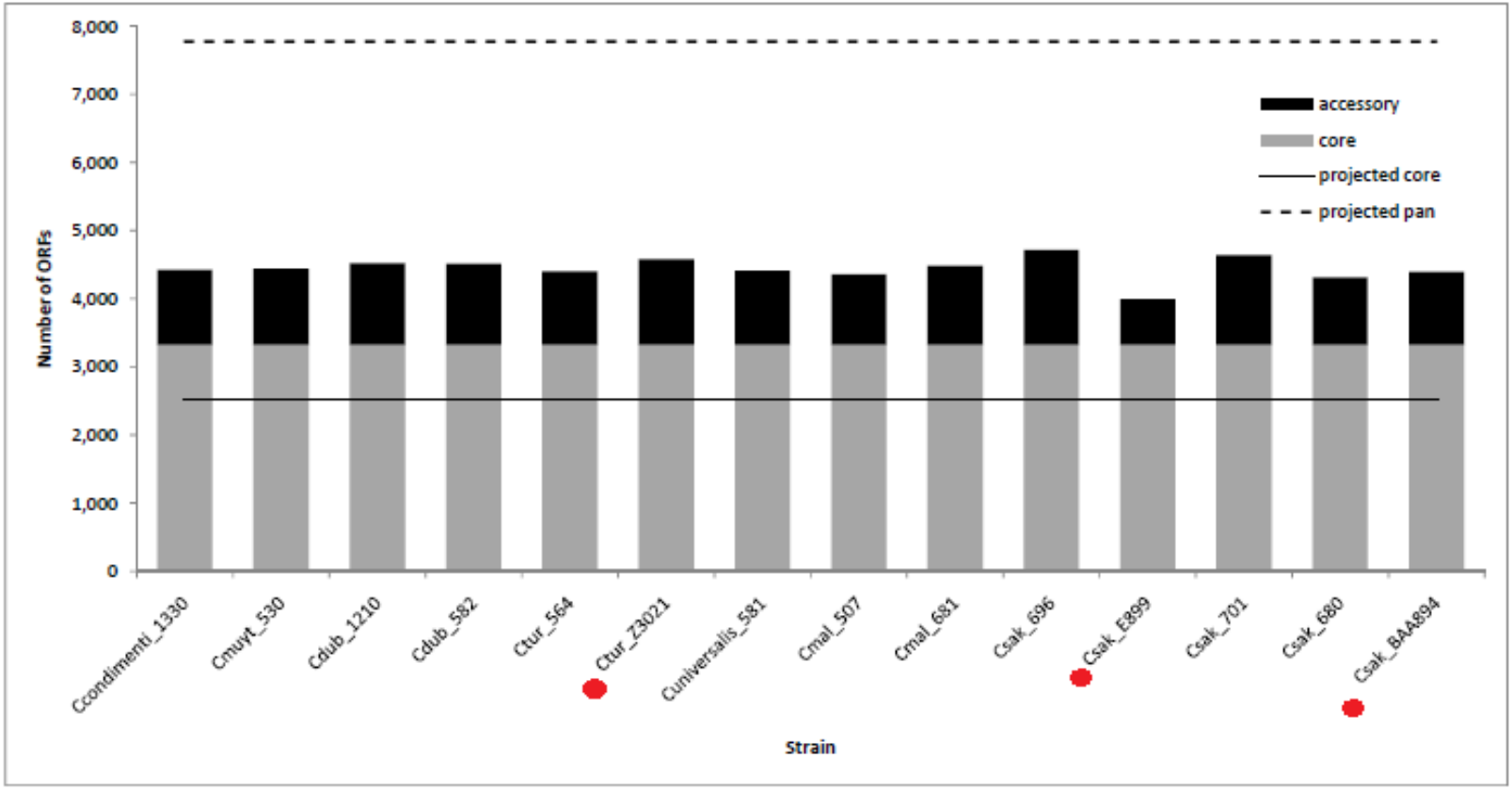


Fig. 6.4 The *Cronobacter* genus pan-genome revealed using the fourteen *Cronobacter* spp. genomes included in this study. The core and accessory genome estimates were determined using the binomial mixture model approach of Snipen *et al*. (2009). The previously published, publicly available genomes have been indicated by the red circles.

6.2.3 Whole genome alignment and phylogeny for the genus *Cronobacter*

The fourteen *Cronobacter* genomes were compared and visually aligned using the BLAST algorithm in BRIG (Alikhan *et al*. 2011), to obtain a circular image of the alignment of the genomes, as represented in Fig. 6.5. The complete and publicly available genome of *C. sakazakii* BAA-894 was used as a reference for the comparison, and therefore the output presented indicates only those conserved regions on the remaining 13 genomes that found positive BLAST hits for the *C. sakazakii* BAA-894 genome. An overall sequence identity > 80% was observed in all the genomes. There were a few gaps/regions with very low identity seen scattered across the genomes of *C. sakazakii* E899 and *C. turicensis* z3032. These regions were all also conspicuous by their high GC-content peaks.

An independent phylogenetic analysis of the *Cronobacter* spp. genomes was also carried out using Mugsy for aligning the syntenic regions of the fourteen genomes and then raxmlGUI for constructing the best maximum-likelihood tree using the same alignment. The evolutionary relationships of the genomes revealed in the maximum-likelihood tree have been indicated in Fig. 6.6 (a). When compared to the phylogenetic tree generated using the concatenated sequences of the seven MLST loci in Fig. 6.6 (b), the topology of the speciation for each *Cronobacter* species was found to be similar, with high levels of confidence at 100% for all branches except an internal branch of the *C. sakazakii* clade. The *C. sakazakii* and *C. malonaticus* genomes formed distinct, neighbouring clades, much more closely related than the rest of the genus. *C. dublinensis* and *C. muytjensii* also clustered in a robust clade, with *C. condimenti* forming a more distant basal branch from it.

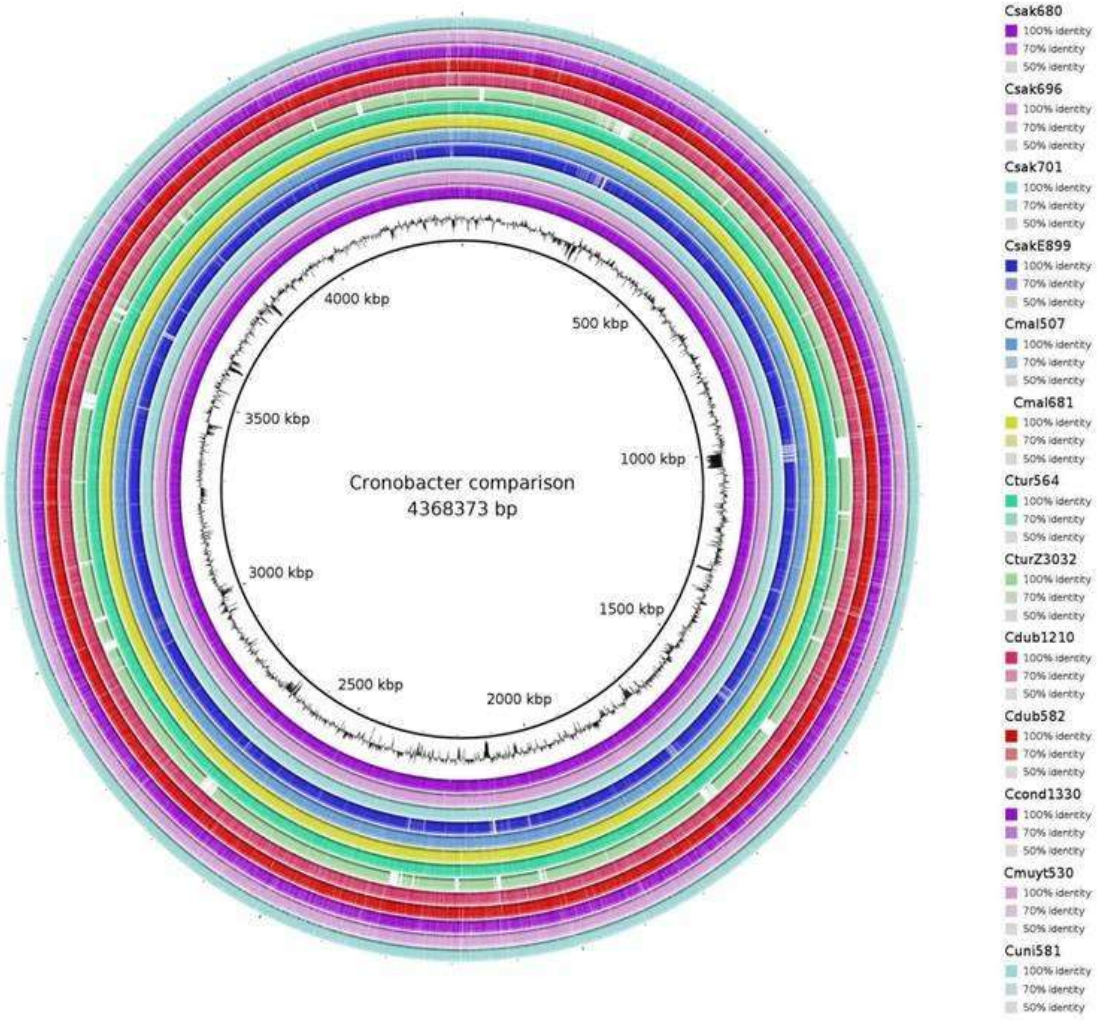


Fig. 6.5 Whole genome alignment of the 14 *Cronobacter* spp. genomes included in this study generated using BRIG (Alikhan *et al*. 2011). The *C. sakazakii* BAA-894 genome was used as a reference backbone for the alignment. The intensity of the colour is indicative of the sequence identity, as represented by the key. Published in Joseph *et al.* (2012c).

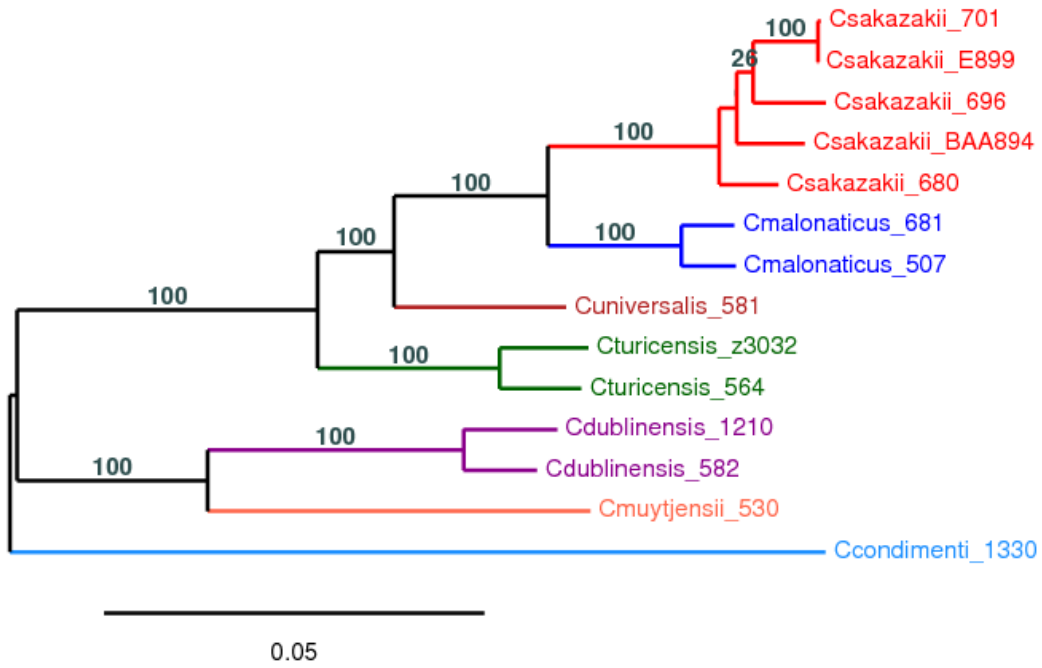


Fig. 6.6 (a) Maximum-likelihood tree indicating the phylogenetic relationships of the fourteen *Cronobacter* spp. genomes used in this study, constructed in RaxML (Stamatakis 2006) using whole genome alignments generated in Mugsy (Angiuoli & Salzberg 2010). The tree has been drawn to scale using 1000 bootstrap replicates, and has been viewed and annotated using TreeDyn (Chevenet *et al*. 2006). The numbers at the nodes indicate the bootstrap values expressed in percentage.

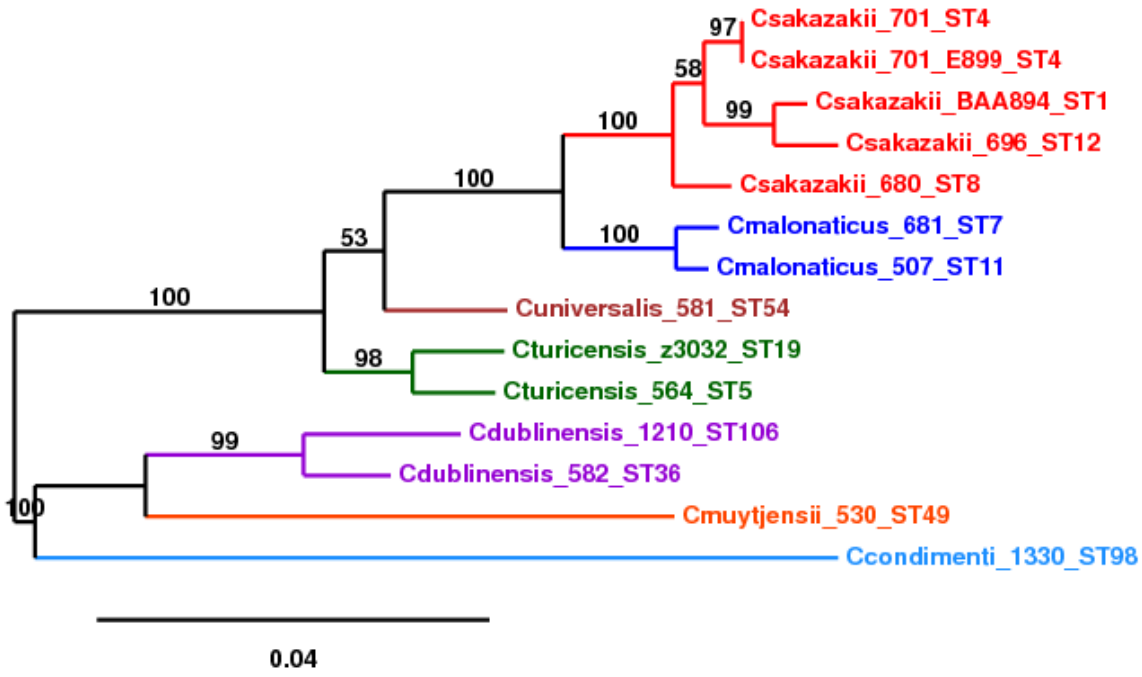


Fig. 6.6 (b) Maximum-likelihood tree based on the concatenated sequences (3036 bp) of the 7 MLST loci for the 14 genome-sequenced strains of the genus *Cronobacter*. The STs and the corresponding species are indicated at the tip of each branch. The tree has been constructed to scale using PhyML with 1000 bootstrap replicates, and has been viewed and annotated using TreeDyn (Chevenet *et al*. 2006). The numbers at the nodes indicate the bootstrap values expressed in percentage.

6.2.4 Individual species specific regions

A pan genome analysis of the genomes of the seven *Cronobacter* species was used to study the diversity of the genus, and to identify genes unique to each species. Some of the unique characteristics of the genomes of each of the species are summarized below.

C. sakazakii:

Comparison of the five *C. sakazakii* genomes (680, 696, 701, E899 and BAA-894) with the genomes of the other species revealed 408 ORFs which were found to be unique to the species. Some of these were unique to certain strains while some others were unique to the species.

A number of these regions were found to be unique to the genomes of strains 696 (ST12) and 701 (ST4), both isolated in a French NICU outbreak, the latter from a fatal case (Caubilla-Barron *et al*. 2007). Notable among these included genes involved with Type I restriction-modification systems (Csak696_peg 1426-29), phosphate-starvation-inducible (Psi) operon genes (Csak696_peg 3058-9), metal resistance genes (Csak696_peg 3056) and a large region with Tra genes belonging to the Inc plasmid family (Csak696_peg 3785-3804).

The *C. sakazakii* 680 (ST8) genome had a unique region comprising of iron uptake (*fecRABCDE*), and transport genes (Csak680_peg 2743-48), metal resistance genes (Csak680_peg 3524), a gene coding for the bacteriocin colicin (Csak680_peg 1434), among others. It also had large regions missing compared with the reference genome *C. sakazakii* BAA-894, including flagella synthesis (ESA_01224-69).

The genome of *C. sakazakii* BAA-894 (ST1) showed regions unique for tellurite resistance (ESA_01796-1801) and some Type I fimbrial assembly proteins (ESA_01972-76). These regions had also been highlighted in the earlier CGH study with this genome by Kucerova *et al*. (2010)

Among the regions conserved across all the genomes of the *C. sakazakii* species and absent from the others, the ones of notable interest were the sialic acid related gene cluster (ESA_03609-13), ABC transport proteins (ESA_01116-19) and a cluster of genes encoding the beta fimbriae major subunit proteins (ESA_03515-20).

C. malonaticus:

The two sequenced genomes of the species *C. malonaticus* revealed 92 ORFs unique to the species. These included genes linked with puromycin biosynthesis (Cmal507_peg 3508), rough lipopolysaccharide formation (Cmal507_peg 1386-88), molybdopterin cofactor

biosynthesis pathway (Cmal681_peg 1456-57) and oxidative stress response (Cmal507_peg 2632-33) among others. Only about 15% of these ORFs were conserved between the two genomes. The *C. malonaticus* 681^T (ST7) unique ORFs included mainly phage related proteins and two genes related to capsular polysaccharide export proteins (Cmal681_peg 881-882).

C. turicensis:

The *C. turicensis* 564 (ST5) genome was analysed along with the publicly available *C. turicensis* z3032 (ST19) genome, to reveal 137 ORFs unique to the species when compared with the rest of the genus. These included genes coding for membrane efflux proteins (Ctur564_peg 2533-36), iron transportation (Ctur564_peg 3267, 3275), DNA sulphur modification operon (Ctu_05230-05270), xanthosine phosphorylase (Ctu_36190) and some transcriptional regulators. *C. turicensis* z3032 had a unique oligogalacturonate protein encoding gene, mis-annotated as *nanC* as it belongs to the same KdgM protein superfamily (Ctu_30960). Less than 10% of these ORFs were shared between the two genomes.

C. universalis:

The *C. universalis* 581 (ST54) genome had 73 unique ORFs when compared to the rest of the *Cronobacter* genus. These included individual genes related to copper resistance (Cuni581_peg 4257), polysaccharide biosynthesis (Cuni581_peg 4275), type I fimbriae (Cuni581_peg 1586), arylsulfatase regulation (Cuni581_peg 3271) and haeme biosynthesis (Cuni581_peg 3272). The genome also has a 36 ORFs long region, which appears to be the remnant of a plasmid (Cuni581_peg 2900-35). This region includes genes encoding for conjugation (*trb* genes), replication initiation and plasmid stabilization.

C. muytjensii:

There were 40 identified ORFs found to be unique to the genome of *C. muytjensii* strain 530 (ST49). A majority of them were found to be hypothetical proteins. There were also genes coding for a range of transcriptional regulators and enzymes such as permease, reductase and methyltransferase. There were some individual genes related to arsenic resistance (Cmuyt530_peg 4251, 4269), haemagglutinin (Cmuyt530_peg 1023) and drug metabolite transport (Cmuyt530_peg 2331).

C. dublinensis:

The genomes of the *C. dublinensis* strains 582 (ST36) and 1210 (ST106) revealed 142 ORFs unique to the species, when compared to the rest of the *Cronobacter* genus. These included a range of interesting genes related to efflux proteins (Cdub582_peg 1570; Cdub1210_peg 3153), capsule synthesis (Cdub582_peg 1423-24), methyl viologen resistance (Cdub582_peg 3154-55), beta glucan synthesis (Cdub582_peg 3867-69), adhesion and heme utilization (Cdub582_peg 4443), among others. Less than 20% of these species-specific ORFs were shared between the two *C. dublinensis* genomes.

C. condimenti:

The genome of the newly defined species, *C. condimenti* strain 1330 (ST98), showed the presence of 79 unique ORFs compared to the rest of the genus. These included genes related to pilin transcription (Ccondi1330_peg 342), cyanate operon (Ccondi1330_peg 3071-73) and colicin receptors (Ccondi1330_peg 2215-16) among others. A cluster of eight genes were also found to be indicating a unique fucose utilization pathway in *C. condimenti* (Ccondi1330_peg 1824-31). Like *C. turicensis,* the genome of *C. condimenti* also showed the presence of a unique gene sequence for an oligogalacturonate protein mis-annotated as *nanC* (Ccondi1330_peg 349). This gene has been discussed further in Chapter 7, Section 7.2.2.3.

6.2.5 Mobile genetic elements and related regions

6.2.5.1 Prophages

In this comparative genomic study of the *Cronobacter* genus, 20 intact prophage genomes (Table 6.2 & 6.3) were identified across the 14 *Cronobacter* genomes. Apart from these, some incomplete or remnant phage regions were also identified (Table 6.2). A number of these regions were found to be shared between two or more *Cronobacter* species. The intact phage regions mainly comprised of genes coding for essential phage assembly proteins such as integrase, terminase, attenuation and lysis related proteins as well as phage head and tail proteins. Compared to the overall GC% of the *Cronobacter* genomes of 56%, these phages regions generally showed GC% values of 51-53%. There were a few that had GC% values of 57-58% (Table 6.3).

C. sakazakii

Eight intact prophage genomes were identified among the four genomes of *C. sakazakii* BAA-894 (ST1), NTU strains 680 (ST8), 696 (ST12) and 701 (ST4). The publicly available genome of *C. sakazakii* E899 (ST4) did not appear to have either completed or remnant prophage regions. Previously, Kucerova *et al*. (2010) had described the prophage regions in *C. sakazakii* BAA-894 and had identified them to be major regions of variation in the CGH study. Majority of the phage regions identified in the *C. sakazakii* genomes were found to be similar to phage genomes previously identified in *E. coli* and *Salmonella* species (Mmolawa *et al*. 2003). Prophage region 3 of *C. sakazakii* BAA-894, representative of the *Salmonella* phage E1, was found to be majorly conserved in a number of other *Cronobacter* species (Table 6.3). A large intact phage genome region shared between the genomes of *C. sakazakii* strains 696 and 701 was found to be representative of the recently characterised *C. sakazakii* temperate bacteriophage phiES15 (Lee *et al*. 2012a). Apart from the phage assembly proteins, these regions included a number of genes related to restriction-modification systems, DNA repair proteins, colicin uptake TolA genes, among others. A tellurite resistance gene (Csak680_peg no. 3524) was found present in one of the remnant phages on the *C. sakazakii* 680 genome. This is different from the tellurite resistance gene cluster found in the genome of *C. sakazakii* BAA-894 discussed in Chapter 7, Section 7.2.1.4. An O-acetyl transferase gene (Csak701_peg no. 1901), which could be responsible for serotype conversion in bacteria, was found among one of the incomplete phage regions in the genome of *C. sakazakii* 701. Except for fragments in *C. sakazakii* 696 (ST12), the gene was not found to be present in the other *Cronobacter* genomes.

C. malonaticus

Between the two *C. malonaticus* genomes, eleven putative prophage regions were identified, only one of which was an intact prophage genome present in the genome of *C. malonaticus* 507 (ST11). This was a region also found conserved in the genome of *C. sakazakii* BAA-894 (Prophage Region 2), as well as *C. dublinensis* 582. The prophage region 1 of *C. sakazakii* BAA-894 was also found conserved in parts as an incomplete phage region in *C. malonaticus* 681 (ST7). Some of the remnant phage regions in the *C. malonaticus* genomes were found to be similar to phages previously isolated in organisms such as *Burkholderia* species and *Pantoea* species. Apart from the assembly regions, these phages also comprised of some genes encoding for stress and membrane-related proteins.

C. turicensis

There were two intact prophage regions and two remnant phage regions identified in the publicly available *C. turicensis* z3032 (ST19) genome which had not been reported previously (Stephan *et al*. 2010). Among these, the intact phage region 1 showed a high degree of similarity to the *C. sakazakii* BAA-894 phage region 3. The *C. turicensis* 564 (ST5) genome showed the presence of 4 putative prophage regions, only one of which is a 35 Kb long intact prophage genome, which was also conserved in the genome of *C. sakazakii* 680.

C. universalis

The *C. universalis* 581 (ST54) genome indicated the presence of a large prophage region and some smaller artefacts. The 39.4 kb long, intact genome was found to be shared with the genomes of *C. sakazakii* BAA-894 and *C. turicensis* z3032, though much smaller in comparison. One of the incomplete regions was unique to *C. universalis*, while the other was shared with the *C. muytjensii* 530 genome.

C. muytjensii

Five putative prophage regions were identified in *C. muytjensii* 530 (ST49), one of which was a 20.1 Kb long intact prophage genome. This prophage region was partially shared with the prophage regions in *C. sakazakii* genomes 696 and 701, as well as the genome of *C. condimenti* 1330. The remnant regions were ones shared with the genomes of *C. sakazakii*, *C. universalis* and *C. dublinensis*.

C. dublinensis

About 15% of the unique ORFs in *C. dublinensis* were putative prophage regions, including a prophage region unique to *C. dublinensis* 1210 (ST108). The largest prophage region found among the *Cronobacter* genomes was a 73 kb long genome present in the genome of *C. dublinensis* 582 (ST36). Apart from genes encoding the phage assembly proteins, this region also comprised of genes involved in arabinose metabolism, copper homeostasis and a number of hypothetical proteins. The two *C. dublinensis* genomes also shared an intact phage genome found to be most similar to an *Aeromonas* phage phiO18p (Beilstein & Dreiseikelmann 2008). This region comprised of a number of hypothetical proteins homologous to genes in organisms like *Pectobacterium carotovorum* and *Yersinia pseudotuberculosis.*

C. condimenti

There were five putative prophage regions in *C. condimenti* strain 1330 (ST98), of which two appeared to be large intact genomes. One was unique to *C. condimenti*, while the other regions were shared with the *C. sakazakii* genomes 696 and 701. Present in the unique phage genome was a gene coding for a ferrichrome outer membrane transporter. One of the incomplete phage regions also had regions with sequence similarity to the very recently characterised *Megavirus chiliensis*, the largest known viral genome till date (Arslan *et al*. 2011).

Species	Strain	Intact Phage Regions	Incomplete Phage Regions	% of total genome
C. sakazakii	BAA-894	3	3	5%
	680	1	6	5%
	696	2	4	4%
	701	2	5	6%
C. malonaticus	507	1	4	3%
	681	0	6	3%
C. turicensis	564	1	3	3%
	z3032	2	2	4%
C. universalis	581	1	2	2%
C. muytjensii	530	1	4	4%
C. dublinensis	582	3	2	6%
	1210	1	5	5%
C. condimenti	1330	2	3	4%

Table 6.2 Number of putative prophage regions identified in the *Cronobacter* spp. genomes included in this study. No prophage regions were identified in the genome of *C. sakazakii* E899. The prophage identification was performed using PHAST (Zhou *et al*. 2011) and ACLAME Prophinder (Lima-Mendez *et al*. 2008). Published in Joseph *et al.* (2012c).

Genome	Intact phage genome characteristics					
	No.	**Size (in Kb)**	**No. of CDS**	**Closest match**	**GC%**	**Position on genome**
C. sakazakii BAA-894	1	46	60	*Enterobacteria* phage SfV	53.41	963828-1009907
	2	35.5	37	*Salmonella* phage ST64B	50.52	1560724-1596249
	3	61.2	86	*Salmonella* phage E1	50.78	2964102-3025401
C. sakazakii 680	1	36.9	53	*Enterobacteria* phage phiV10	57	721333-758307
C. sakazakii 696	1	54	54	*Enterobacteria* phage Fels2	55.6	85322-139409
	2	24.9	66	*Cronobacter* phage phiES15	50.6	3374794-3399783
C. sakazakii 701	1	38.3	49	*Cronobacter* phage phiES15	57.1	821927-860304
	2	25.2	38	*Salmonella* phage E1	54.2	1739809-1765101
C. malonaticus 507	1	30.3	37	*Salmonella* phage ST64B	55.1	676840-707181
C. turicensis 564	1	35	46	*Enterobacteria* phage phiV10	55.5	442416-477444
C. turicensis z3032	1	49.5	80	*Salmonella* phage E1	51.23	1204094-1253674
	2	43.2	51	*Escherichia* phage HK75	51.99	1811622-1854858
C. universalis 581	1	39.4	57	*Salmonella* phage E1	52.7	2695127-2734580
C. muytjensii 530	1	20.1	31	*Cronobacter* phage phiES15	54.5	340633-360819
C. dublinensis 582	1	73	100	*Escherichia* phage HK75	58.7	307477-380490
	2	22.4	36	*Salmonella* phage ST64B	55.3	748192-770616
	3	37.9	36	*Aeromonas* phage phiO18P	56.4	1486588-1524560
C. dublinensis 1210	1	45.4	41	*Aeromonas* phage phiO18P	59	2573599-2619009
C. condimenti 1330	1	43.2	48	*Cronobacter* phage phiES15	56.4	865492-908739
	2	30.9	39	*Enterobacteria* phage P22	57.6	2347505-2378416

Table 6.3 Details of the intact prophage genomes identified in the *Cronobacter* spp. genomes included in this study. This characterisation was performed using the PHAST server (Zhou *et al*. 2011) and BLASTp searches. The conserved phage regions have been shaded in the same colours.

6.2.5.2 Plasmids

The published genome of *C. sakazakii* BAA-894 contains two plasmids pESA2 (31 kb, 51% GC content) and pESA3 (131 kb, 56% GC), encoding for 38 and 127 genes respectively. In contrast, the genome of *C. turciensis* z3032 contains three plasmids pCTU2 (22.5 kb, 49%GC), pCTU3 (53.8 kb, 50% GC) and pCTU1 (138 kb, 56% GC), encoding for 32, 74 and 136 genes respectively. The larger plasmids, pESA3 and pCTU1 were found to share a common backbone and a number of key genes, including virulence related traits (Franco *et al*. 2011). On the other hand, the publicly available genome of *C. sakazakii* E899 had no plasmid sequences uploaded in Genbank, and none were identified in these studies as well. Hence, the genome has been excluded in these results.

All the sequenced genomes appeared to share this common backbone of pESA3/pCTU1. Genes homologous with these plasmids were found in the sequenced strains of all the *Cronobacter* species, though pESA3 was only partial in *C. sakazkaii* 680 and *C. condimenti* 1330^T, and pCTU1 had a partial match in *C. sakazakii* 680, 696, 701, and *C. condimenti* 1330^T (Table 6.4). The other smaller plasmids showed greater degrees of variation. Only partial regions of pCTU3 (53.8 kb) were found in *C. sakazakii* strains 680, 696, 701, *C. malonaticus* 681, *C. muytjensii* 530, and *C. dublinensis* 582. No pCTU3 regions were found in *C. malonaticus* 507, *C. turicensis* 564, *C. dublinensis* 1210, *C. universalis* 581, or *C. condimenti* 1330. No genes of pESA2 (31kb) or pCTU2 (22.5kb) were found in the newly sequenced genomes, except for a small region in *C. sakazakii* 696 (p06563-700) encoding 8 ORFs, one of which was TnpA transposase.

BRIG alignments were also generated using the plasmids pESA3 and pCTU1 as reference backbones and have been shown in Figures 6.7 and 6.8, respectively. It can be seen that the plasmid backbone was conserved in all the strains with varying degrees of homology, indicated by the faded colours of the rings. There were also certain clear regions of variation between the pESA3 and pCTU1 plasmids, some of which have been previously reported by Franco *et al*. (2011). These regions have been annotated (see Tables 6.5 and 6.6). Some of the main regions of variation between the two plasmids included Type VI secretion system related genes, *Cronobacter* plasminogen activator (cpa), outer membrane proteins and efflux pumps, copper sensitivity region on pESA3 (*scsAB*), tellurite resistance protein *tehA* and a two partner secretion system *fhaBC* on the plasmid pCTU1.

However, it should be noted that matches to the plasmid do not necessarily confirm their location. To address this issue, the read coverage of the assembled genomes was also used

to predict the location of the plasmid-borne genes, and most of the results were found to correlate with what has been reported above. The read coverage of the *C. sakazakii* 680 genome indicated that the putative plasmid region also encodes for a *lac* operon and iron-dicitrate transporter (*fecRABCDE*) region. These traits were absent from the other sequenced *Cronobacter* spp. genomes. In contrast, the FHA loci of pCTU1 was found to be located in the chromosomal region of the *C. sakazakii* 680 genome, indicating the mobility of plasmid-related traits within the genome of the bacterium. Large contigs (>2kb) which had not mapped to the reference genomes were also identified in these assemblies and some of these appeared to be small cryptic plasmid regions. These included an arsenic resistance operon region unique to the genome of *C. muytjensii* 530 and a haemagglutinin region on the genome of *C. dublinensis* 1210. Apart from these, a number of hypothetical proteins were found with homology to plasmid regions in other members of the *Enterobacteriaceae* family such as *E. cloacae subsp. cloacae* ATCC 13047 plasmid pECL_A, *Klebsiella pneumoniae* strain KP048 plasmid pKP048 and *E. coli* O55:H7 str. RM12579 plasmid p12579.

Species	Strain	pCTU2 (22.5 kb)	pESA2 (31 kb)	pCTU3 (53.8 kb)	pESA3 (131 kb)	pCTU1 (138 kb)
C. sakazakii	680	no	no	yes (fragmented)	partial	partial
	696	no	partial	Yes	yes	partial
	701	no	no	Partial	yes	partial
C. malonaticus	507	no	no	No	yes	yes
	681	no	no	Yes	yes	yes
C. turicensis	564	no	no	No	yes	yes
C. universalis	581	no	no	No	yes	yes
C. muytjensii	530	no	no	Partial	yes	yes
C. dublinensis	582	no	no	Partial	yes	yes
	1210	no	no	No	yes	yes
C. condimenti	1330	no	no	No	partial	partial

Table 6.4 Presence of the plasmids of the publicly available genomes of *C. sakazakii* BAA-894 (pESA2 and pESA3) and *C. turicensis* z3032 (pCTU1, pCTU2 and pCTU3) on the *Cronobacter* spp. genomes sequenced in this study determined using gene searches as well as the read coverage of the assembled genomes. The plasmids have been listed in increasing order of sizes.

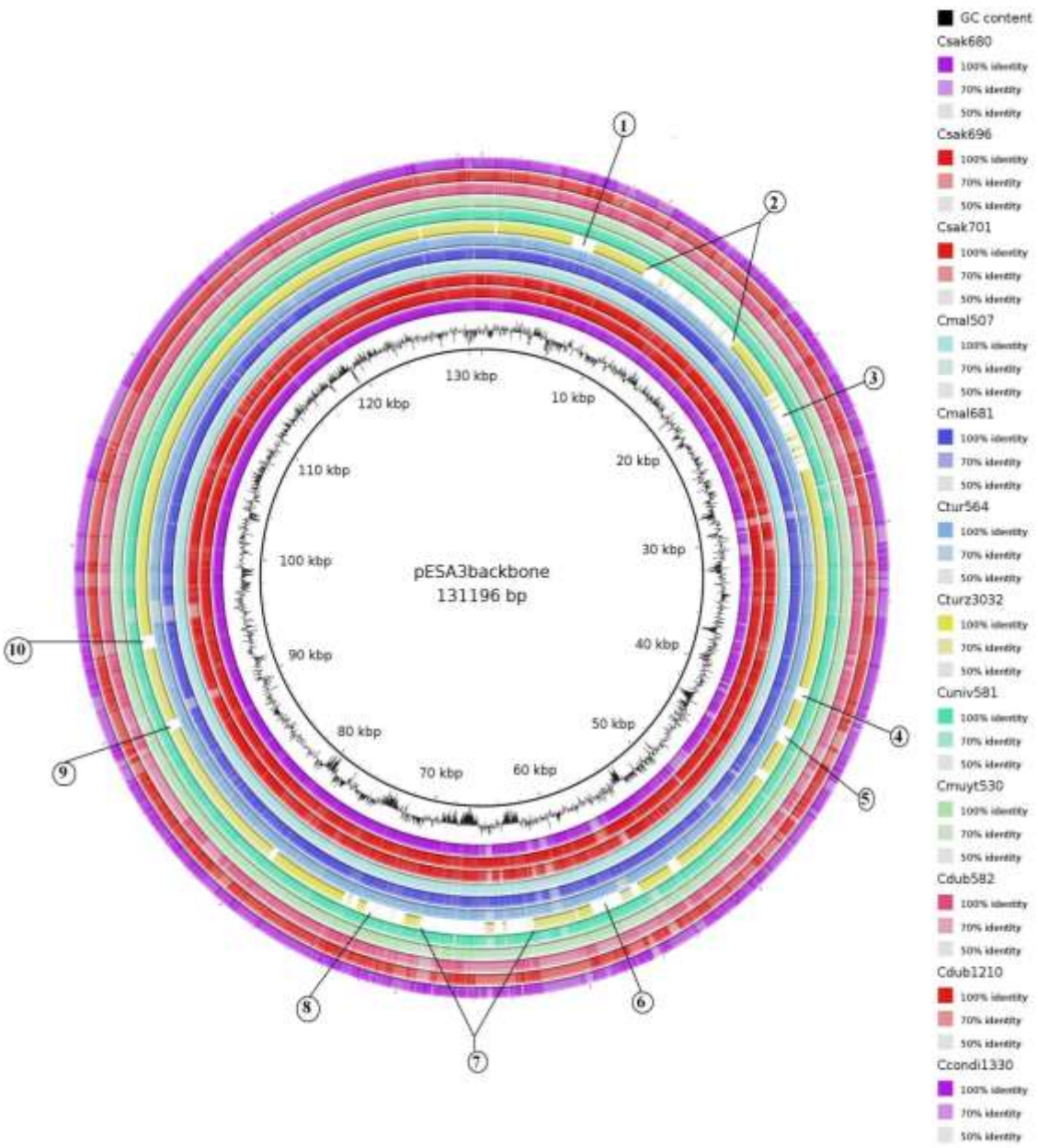


Fig. 6.7 BRIG sequence alignment of putative plasmid-associated regions in the *Cronobacter* spp. genomes using the *C. sakazakii* plasmid pESA3 as a reference backbone. The major regions of variation have been annotated and listed in Table 6.6. Published in Joseph *et al.* (2012c).

Region	Locus	Annotation
1	ESA_pESA3p05434	Plasminogen activator Pla. Aspartic peptidase. MEROPS family A26
	ESA_pESA3p05435	Hypothetical protein - CpmJ protein
2	ESA_pESA3p05440	Outer membrane protein V
	ESA_pESA3p05441	Response regulators with CheY-like receiver domain & winged-helix DNA-binding domain
	ESA_pESA3p05442	Signal transduction histidine kinase
	ESA_pESA3p05443	RND family efflux transporter, MFP subunit
	ESA_pESA3p05444	Cation/multidrug efflux pump
3	ESA_pESA3p05451	ABC-type molybdate transport system, periplasmic component
	ESA_pESA3p05452	Transcriptional regulator, LysR family
4	ESA_pESA3p05478	Diguanylate cyclase (GGDEF) domain
5	ESA_pESA3p05480	Hypothetical protein
	ESA_pESA3p05481	Type VI secretion system effector, Hemolysin coregulated protein1 family
6	ESA_pESA3p05494	Type IV / VI secretion system protein, DotU family
	ESA_pESA3p05495	Outer membrane protein and related peptidoglycan-associated (lipo)proteins
7	ESA_pESA3p05498	Hypothetical protein
	ESA_pESA3p05499	Hypothetical protein
	ESA_pESA3p05500	Type VI secretion system Vgr family protein
	ESA_pESA3p05501	Hypothetical protein
	ESA_pESA3p05502	Hypothetical protein
	ESA_pESA3p05503	Hypothetical protein
	ESA_pESA3p05504	Uncharacterized conserved protein
	ESA_pESA3p05505	Hypothetical protein
	ESA_pESA3p05506	Type VI secretion lipoprotein, VC_A0113 family
8	ESA_pESA3p05508	Thiol:disulfide interchange protein/Uncharacterized protein predicted to be involved in C-type cytochrome biogenesis/ suppressor for copper-sensitivity ScsB
	ESA_pESA3p05509	Suppression of copper sensitivity: putative copper binding protein ScsA
9	ESA_pESA3p05526	Hypothetical protein
10	ESA_pESA3p05530	Glutathione S-transferase (EC:2.5.1.18)

Table 6.5 Plasmids regions of *C. sakazakii* pESA3 found to be absent in the *C. turicensis* plasmid pCTU1, as revealed by the BRIG sequence alignment in Fig. 6.7.

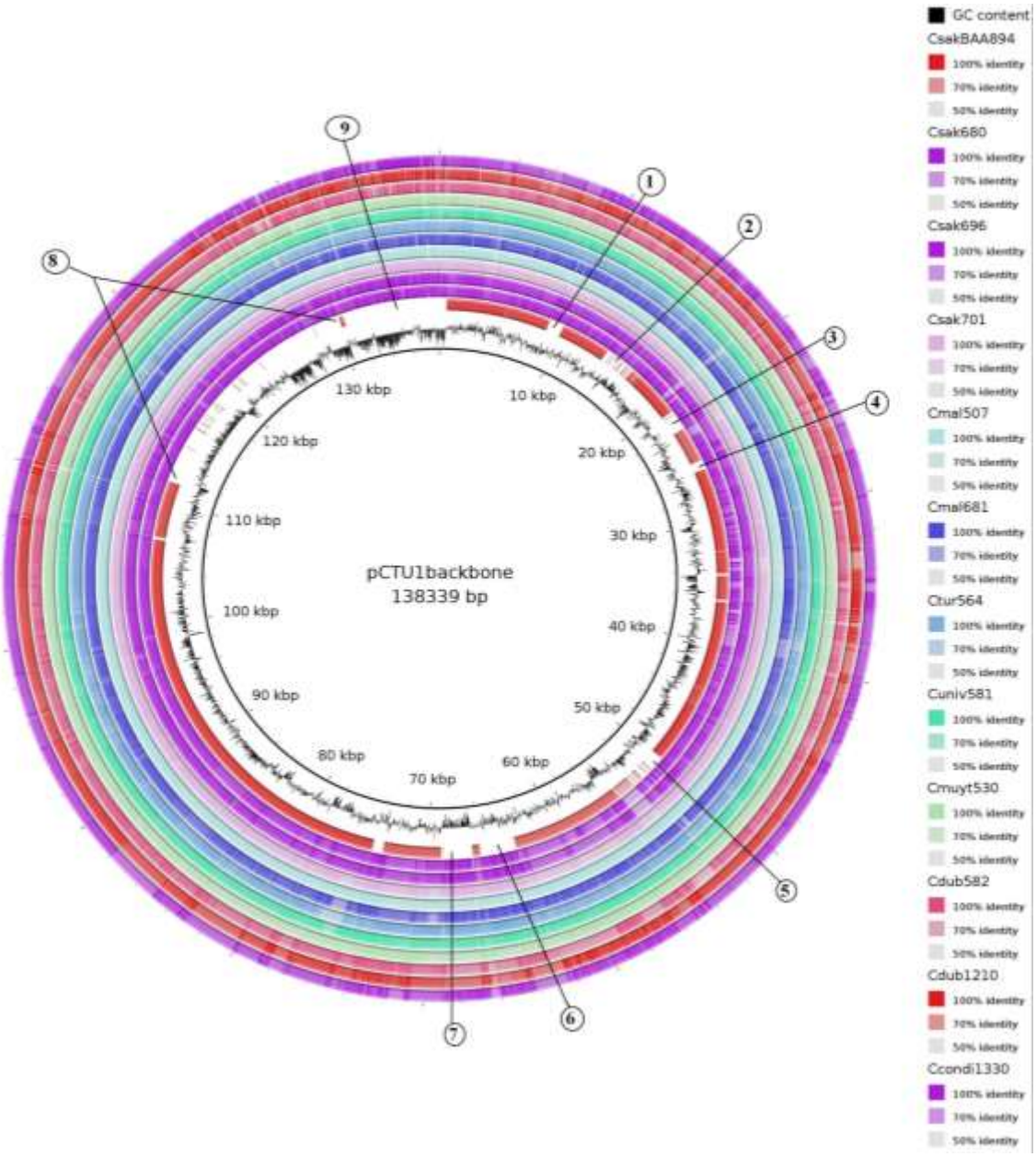


Fig. 6.8 BRIG sequence alignment of putative plasmid-associated regions in the *Cronobacter* spp. genomes using the *C. turicensis* plasmid pCTU1 as a reference backbone. Published in Joseph *et al.* (2012c).

Region	Locus	Annotation
1	Ctu_1p00130	Hypothetical protein, ygjN
	Ctu_1p00140	HTH-type transcriptional regulator, ygjM
2	Ctu_1p00210	Transcriptional regulator, LysR family
	Ctu_1p00220	Permeases of the major facilitator superfamily
3	Ctu_1p00320	Tellurite resistance protein, TehA
4	Ctu_1p00360	Thiol:disulfide interchange protein, DsbC
5	Ctu_1p00640	Beta-glucoside bgl operon antiterminator, BglG family
6	Ctu_1p00740	hypothetical protein
	Ctu_1p00750	hypothetical protein
7	Ctu_1p00780	hypothetical protein
	Ctu_1p00790	hypothetical protein
8	Ctu_1p01150	Two partner secretion protein/Hemolysin activator protein precursor, fhaC
	Ctu_1p01160	fhaB
	Ctu_1p01170 - Ctu_1p01250	hypothetical proteins
9	Ctu_1p01270 - Ctu_1p01360	hypothetical proteins

Table 6.6 Plasmids regions of *C. turicensis* pCTU1 found missing in the *C. sakazakii* plasmid pESA3, as revealed by the BRIG sequence alignment in Fig. 6.8.

Plasmid profiling:

Plasmid profiling experiments were carried out on the *Cronobacter* spp. strains sequenced in this study to verify the results that were observed *in silico* based on the sequences similarities and read coverage analysis. Because of the large sizes of the *Cronobacter* plasmids, appropriate size DNA ladders could not be used and hence the well-characterized plasmid profiles of the genomes of *C. sakazakii* BAA-894 and *C. turicensis* z3032 were used as reference markers.

The profiling of the *C. turicensis* z3032 revealed the three plasmids similar to sizes reported by Stephan *et al*. (2011) - pCTU1 (138 kb), pCTU2 (22.5 kb) and pCTU3 (53.8 kb). *C. sakazakii* BAA-894, on the other hand, though reported only to have two plasmids – pESA2 (31 kb) and pESA3 (131 kb) by Kucerova *et al*. (2010), showed the presence of the three plasmids. A third unreported plasmid similar in size to pCTU3 was observed, suggesting the possibility that this had not been sequenced with the rest of the genome.

An intact plasmid similar to the sizes of pCTU1/pESA3 was observed in the genomes of all the strains except *C. sakazakii* 680 (ST8) and *C. condimenti* 1330 (ST98), as was found by analysing the read coverage of the genomes too. In addition, *C. sakazakii* 701 (ST4) appeared to have two copies of this plasmid, one slightly larger than the other.

An intact plasmid similar to the size of pCTU3 (53.8 kb) was observed in all the strains except *C. sakazakii* 680, 696 and *C. condimenti* 1330. A slightly smaller sized plasmid was also observed in the profiles of *C. malonaticus* 507, *C. dublinensis* 582 and *C. dublinensis* 1210.

An intact plasmid corresponding to the sizes of pCTU2/pESA2 was exhibited by all the strains except *C. sakazakii* 680 & 696, *C. turicensis* 564 and *C. condimenti* 1330. These were not detected by the read coverage analysis of the genomes.

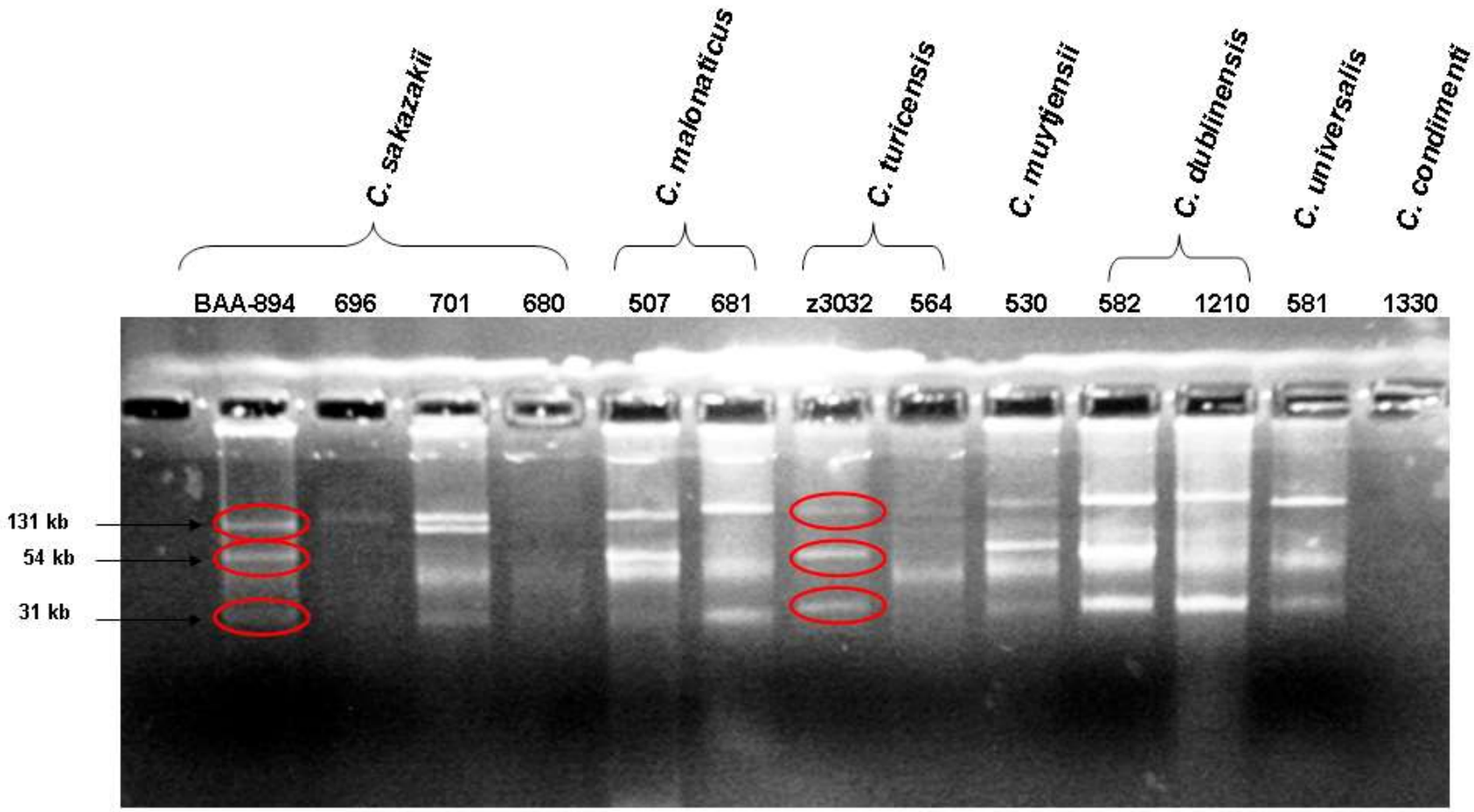


Fig 6.9 Agarose gel electrophoresis image of the plasmid profiles of the *Cronobacter* spp. strains sequenced in this study. The plasmid profiles of the strains *C. sakazakii* BAA-894 and *C. turicensis* z3032 (indicated by the red circles) were used as markers, as their sizes (as denoted) had been determined by sequencing studies (Kucerova *et al*. 2010; Stephan *et al*. 2011).

6.2.5.3 CRISPR regions

The fourteen *Cronobacter* spp. genomes were analyzed for the presence of clustered regularly interspaced short palindromic repeats (CRISPRs). These repeat regions of varying sizes were identified in 10 out of the 14 genomes (Table 6.7). Majority of these were short regions with only two direct repeats and a single spacer separating them.

The *C. dublinensis* 1210 (ST106) genome had a slightly longer CRISPR region with 3 DRs. Only the genomes of *C. sakazakii* BAA-894 (ST1) and *C. turicensis* z3032 (ST19) showed major CRISPR regions, apart from a number of short regions scattered across their genomes. None of these major CRISPR regions were found in the remaining *Cronobacter* genomes.

The three large CRISPR regions of the genome *C. sakazakii* BAA-894 are located at the loci:
(1) ESA_02830-02831 (1.86 kb)
(2) ESA_02853-ESA_02855 (1.62 kb)
(3) ESA_03209-03210 (749 bp)
BLAST searches of these regions revealed them to be most closely related to other CRISPR regions identified in the plant pathogens such as *Erwinia amylovora* and *Erwinia carotovora*.

The two large CRISPR regions of the genome *C. turicensis* z3032 are located at the loci:
(1) CTU_10250-10260 (2.29 kb)
(2) CTU_10520-10530 (2.04 kb)
This CRISPR region 1 of *C. turicensis* z3032 was found to be a region similar to the CRISPR region 2 of *C. sakazakii* BAA-894 (ESA_02853-ESA_02855), mentioned above. Region 2 showed homology to CRISPRs previously identified in strains of *Salmonella enterica.*

No CRISPR regions were found in the genomes of *C. malonaticus* 507, *C. universalis* 581, *C. muytjensii* 530 and *C. dublinensis* 582.

In order to verify whether these identified CRISPR regions were true, functional elements, the genomes were searched for *cas*-like genes, especially in the vicinity of the CRISPRs. This revealed that *cas*-like genes were present only in the genomes of *C. sakazakii* BAA-894 (ESA_02832-02838; adjacent to CRISPR 1) and *C. turicensis* z3032 (CTU_10440-10500; adjacent to CRISPR 2), suggesting that these are the only functional CRISPRs in the genomes.

Genome	No. of candidate CRISPRs	Notes
C. sakazakii BAA-894	15	3 large regions with 31, 27 and 13 DRs respectively; 6 short regions with 2 DRs
C. sakazakii 680	3	Short regions with only 2 DRs
C. sakazakii 696	1	Short region with only 2 DRs
C. sakazakii 701	1	Short region with only 2 DRs
C. sakazakii E899	12	Short regions with only 2 DRs
C. malonaticus 507	0	No CRISPRs identified.
C. malonaticus 681	1	Short region with only 2 DRs
C. turicensis 564	1	Short region with only 2 DRs
C. turicensis z3032	7	2 large regions with 38 and 34 DRs respectively; 5 short regions with 2 DRs
C. universalis 581	0	No CRISPRs identified.
C. muytjensii 530	0	No CRISPRs identified.
C. dublinensis 582	0	No CRISPRs identified.
C. dublinensis 1210	1	Region with 3 DRs
C. condimenti 1330	1	Short region with only 2 DRs

DR = Direct Repeat

Table 6.7 Details of the CRISPR regions identified in the 14 *Cronobacter* spp. genomes included in this study. The CRISPRs were identified using the CRISPRFinder (Grissa *et al*. 2007)

6.3 Discussion

The genomic study presented here is the first whole genus comparative genomic analysis conducted on the seven species of the *Cronobacter* genus. Previously, our research group had, in collaboration with Prof. Michael McClelland, published the first complete genome of *C. sakazakii* BAA-894 (Kucerova *et al*. 2010). The study was also expanded to include the four other species of the *Cronobacter* genus at that point of time in a microarray based CGH study using the *C. sakazakii* BAA-894 genome as a reference, in order to investigate the genetic basis of virulence of the organism. Since then, two more *Cronobacter* genomes have become publicly available – *C. turicensis* z3032 (Stephan *et al*. 2011) and *C. sakazakii* E899 (Chen *et al*. 2011). However, an inter-species comparative study of the genomes was not conducted. Also, the expansion of the *Cronobacter* MLST scheme across the genus had resulted in a taxonomic revision to define two new species – *C. universalis* and *C. condimenti* (Chapter 4). The MLST data analysis had also revealed considerable diversity in the genus, including a genetic signature for neonatal meningitis in ST4 (Chapter 3 and 5). Hence, it was inevitable that the diversity study of the genus was to be expanded to the whole genomic level to investigate key traits of this opportunistic pathogen.

6.3.1 Strain selection

A crucial factor to a successful comparative genomic study is undoubtedly the accurate choice of strains to be sequenced. A strain set of nine *Cronobacter* spp. strains for this genomic study was initially chosen in April 2010, in the early stages of the MLST study, using the STs and source backgrounds of the isolates as the criteria for selection. Later the sample set was expanded to eleven strains in order to include the sole isolate of the newly identified species *C. condimenti* strain 1330, as well as an additional candidate of *C. dublinensis* (NTU strain 582), following the high diversity observed in the species during the MLST study (Chapter 4). Fig 6.3 is a maximum-likelihood phylogenetic tree constructed in MEGA5 using the concatenated sequences of the seven MLST loci for the 325 *Cronobacter* strains included in the MLST study (Chapter 3). The fourteen genome sequenced strains, including the three publicly available *Cronobacter* spp. genomes, have been indicated by the coloured labels on the tree and the figure reflects how the diversity of the entire length of the *Cronobacter* genus is well represented by these genome candidates, thus further validating the choice of strains to represent the genus. The three *C. sakazakii* strains (680, 696 and 701) are well characterised clinical isolates previously reported in a number of publications (Caubilla-Barron *et al*. 2007; Baldwin *et al*. 2009; Townsend *et al*. 2008; Kucerova *et al*. 2010; Kucerova *et al*. 2011). In addition, strain 701 is a *C. sakazakii* ST4 strain that was isolated from a fatality in an NICU outbreak in France and has

been used for *in vitro* tissue culture studies of virulence (Caubilla-Barron *et al*. 2007; Townsend *et al*. 2007), and also represents ST4, the genetic lineage for neonatal meningitis infections in *Cronobacter* (Chapter 5). *C. sakazakii* 680 belongs to ST8, a sequence type associated with clinical infections, with no confirmed association with meningitis (Baldwin *et al*. 2009; Chapter 5). *C. malonaticus* 681^T is ST7, a sequence type that has been primarily associated with adult infections (Chapter 5).

6.3.2 Sequencing and assembly limitations

The eleven genomes were sequenced on two different NGS platforms – eight on the SOLiD ABI system and three on the Ion Torrent PGM. The short sequencing reads generated on the ABI SOLiD platform tends to result in a draft genome assembly consisting of thousands of contigs, as was found in the eight genomes in this study too (Table 6.1). It is only natural that an assembly of this complexity can introduce the issue of fragmentation of some ORFs depending on their positions on the contigs, which in turn can result in mis-annotations, especially in an automated annotation pipeline (Klassen and Currie 2012; Baker 2012). Any possible errors that could have been introduced because of the large number of contigs were overcome to the best possible extent by the high depth read coverage as well as the colour space error correction process in place in the SOLiD system workflow. The same approach has been previously successfully undertaken for the comparative genomics studies of the *Listeria* and *Salmonella* genus (den Bakker *et al*. 2010; den Bakker *et al*. 2011). The three genomes - *C. sakazakii* 680, *C. dublinensis* 1210 and *C. condimenti* 1330 - which were sequenced on the Ion Torrent PGM had significantly lesser number of contigs (Table 6.1), and so this issue was averted.

6.3.3 Phylogenetic analysis and evolution of the genus

Phylogenetic analysis of the core regions of the *Cronobacter* spp. genomes confirmed the species clustering of the genus, with distinct cluster formations within the genus too, a phenomenon which was later found to be reflected in the gene diversity of the genomes (Chapter 7). The overall topology of the maximum-likelihood tree was found to be similar to the tree constructed using the concatenated sequences of the seven MLST loci for the genus (Chapter 3, Fig. 3.2), with a couple of differences in the interior branching. In the whole genome phylogeny, it was found that the species *C. turicensis* and *C. universalis* have evolved in a step wise fashion, while *C. dublinensis* and *C. muytjensii* have evolved from a common ancestor, which was found to be contrary to that observed in the 7-loci MLST tree. It is important to note that this incongruence in the branching occurred only when the outliers were used in the MLST tree, suggesting the role of recombination events within the *Cronobacter*

genus having occurred after the speciation process. It is quite possible for high levels of gene conversion to have occurred between various species of a genus at the whole genome level, as opposed to a more stable phylogeny of isolates belonging to the same species or even very distant species, and this was reflected in the phylogenetic evolution of the genomes of the *Cronobacter* genus. A separate analysis conducted by our collaborators using ClonalFrame (Appendix Fig. 7) to take into account the possible influence of recombination events, overcame these incongruencies, further confirming the role of gene conversion events in the evolution of the genus.

6.3.4 The *Cronobacter* pan-genome

The entire set of fourteen *Cronobacter* spp. genomes involved in this study, including the publicly available genomes, was used for calculating the pan-genome characteristics of the *Cronobacter* genus. This analysis revealed that the core conserved genome consists of approximately 32% of the entire *Cronobacter* spp. pan-genome. Previously, Kucerova *et al*. (2010) in the CGH study comprising five *Cronobacter* species had observed a core genome percentage of 43.3%. The predicted core genome value in this study is expected to further decrease as more genomes of the genus are sequenced in the future, thereby increasing the amount of identified diversity. It is interesting to note that despite this core genome estimation at a whole genus level, the percentage value is much higher than that observed from genomic studies of other *Enterobacteriaceae* species such as *E. coli* (7%), *Salmonella enterica* (9%) and *Yersinia pestis* (14%) (Snipen *et al*. 2009). Recombination and gene transfer events have already been observed for the *Cronobacter* genus (Chapter 3), and therefore this considerably high core genome estimation suggests that a greater amount of the gene exchange could possibly have occurred between the *Cronobacter* species, as compared to exchange with other organisms. Another factor to take into consideration is the fact that half of this dataset is comprised of *C. sakazakii* and *C. malonaticus* genomes, which have been seen to be much more clonal, compared to the rest of the genus, in the MLST studies (Chapter 5). Of course, it is also important to add a word of caution here about this estimation because 12/14 genomes used are assemblies of draft genomes and hence would have a large number of gaps in the genomes. This could have resulted in either incomplete or unannotated fragments, which can further slightly under-estimate the size of the core genome as well as over estimate the size of the pan-genome.

In a whole genomic comparative analysis spanning a genus, an aspect of interest is often the regions unique to each species of the genus, which are major contributors to the genus diversity and speciation. In this study of the 14 genomes of the *Cronobacter* genus, nearly 25%

of the identified accessory genome was found to be regions unique to any one of the seven *Cronobacter* species. This analysis helped identify some interesting aspects of the organism.

When studying the regions unique to *C. sakazakii*, the obvious starting point was to look at regions exclusive to the genome of the ST4 strain 701, as this ST has already been identified to be a molecular signature for neonatal meningitis (Chapter 5). However, this list seemed to be dominated only by genes linked to incomplete phage regions and a few uncharacterized hypothetical proteins. On the other hand, an interesting cluster of shared genes was found in the genomes of *C. sakazakii* 696 (ST12) and 701 (ST4). These two strains were isolated from the same *Cronobacter* spp. outbreak in a French NICU, which had involved three fatalities, one of which was associated with Strain 701 (Caubilla-Barron *et al*. 2007). These genes included two clusters of genes related to Type I and Type III restriction-modification systems, which are defence mechanisms of the bacteria against incoming viruses or phages, by machinery for attacking specific regions in the foreign DNA. Another important region unique to these two *C. sakazakii* genomes comprised of 21 ORFs coding for proteins involved in pilus assembly, linked to the IncF group of plasmids, and therefore potentially linked to the pathogenicity of the organism. These have been further discussed in Chapter 7, Section 7.2.2.1. Interestingly, these regions were not found in the publicly available genome of *C. sakazakii* E899, also belonging to ST4, hinting at the possibility of acquisition events having occurred in these genomes linked to their common isolation environment of the NICU.

Another set of unique genes belonged to the genome of *C. sakazakii* 680 (ST8), an operon (*fecABCDER*) encoding for an iron dicitrate transport system. This region showed a >95% identity to the iron dicitrate transport region in *E. coli* K12, with FecA having been identified to be the main protein involved in the transport process (Wagegg and Braun 1981; Pressler *et al*. 1988). Iron acquisition and transport across membranes could be a potential virulence mechanism in bacteria and have been discussed in Chapter 7; Section 7.2.2.4. Especially considering the history of *Cronobacter* spp. epidemiology, it could play a role in helping the organism to utilize the iron in breast milk and infant formula. The uniqueness of this region in an ST8 strain is intriguing, as this ST has been earlier linked with clinical *C. sakazakii* isolates with no confirmed association with meningitic infections. This suggests a potential unique mode of virulence for this lineage.

The genome of *C. sakazakii* BAA-894 also showed unique regions containing genes for tellurite resistance (ESA_01796-801) and some Type I fimbrial assembly proteins (ESA_01970-76). These regions had previously been highlighted in CGH studies (Kucerova *et al*. 2010; Kucerova *et al*. 2011). These have been discussed in detail in Chapter 7.

There were some important regions that stood out in being unique to the entire *C. sakazakii* species as well. A complete ABC-type multidrug efflux system (ESA_01116–19) was found only in the *C. sakazakii* genomes. It comprised an outer membrane efflux protein from a

family that includes TolC, a permease component (ESA_01118) of the ABC type system, and an ATPase component (ESA_01119) of the efflux system. The *C. sakazakii* region (ESA_03609–13) encodes for the uptake and utilization of exogenous sialic acid, and was also found to be exclusive to the *C. sakazakii* genomes. The acquisition of genes encoding for the utilization of exogenous sialic acid may have a major role in *C. sakazakii* colonisation of the human intestinal tract (via mucins) and the use of sialic acid in breast milk, infant formula, and brain cells as a nutrient source (Almagro-Moreno and Boyd 2009). A detailed section on this genomic region has been included in Chapter 7; Section 7.2.2.3.

This pan-genomic study was also an ideal opportunity to gain a better understanding of the newly identified *Cronobacter* species. The *C. universalis* 581 (ST54) genome was found to share a number of conserved genes with the *C. turicensis* species. However, it also exhibited some interesting regions unique to the species. This included a cluster of genes encoding the Trb group of proteins related to conjugative transfer and the type IV secretion system (Cuni581_peg 2901-2908). This operon has been identified in the conjugal transfer activity of the Ti plasmids in *Agrobacterium tumefaciens*. *A. tumefaciens* also posesses an aternative conjugative transfer system mediated by *vir* genes (Li *et al*. 1999). Interestingly there were a few genes encoding for Vir proteins scattered across the genome of *C. universalis* 581. This suggests the possibility that the organism could have had a partial acquisition of this plasmid-like region over the course of its evolution. Studies have been reported suggesting a natural plant-based habitat for *Cronobacter* spp. including exhibition of traits related to root colonisation and siderophore production (Schmid *et al*. 2009). Also, in the evolutionary estimation of the *Cronobacter* genus based on MLSA, it was found that the early evolution of the *Cronobacter* genus could have occurred over the same period as the evolution of flowering plants (Chapter 3). Hence, the identification of cryptic regions linked to Ti-plasmids correlates well with these findings.

The regions unique to the genome of *C. condimenti* 1330 were mostly found to be uncharacterized hypothetical proteins. One particular region of interest that stood out was a six gene cluster (peg 1824-1829) encoding for enzymes involved in fucose utilization. Fucose ($C_6H_{12}O_5$) is a hexose deoxy sugar, that has been attributed to plant sources as well as dietary sources. It is one of the forms that sugars may exist in the small intestine where it can get utilized by intestinal bacteria. In organisms such as *Campylobacter jejuni*, the utilization of fucose has been linked with the adhesive properties of the organism. Upregulation of a genomic island linked to fucose utilization has been identified in some strains of *C. jejuni* (Stahl *et al*. 2011). To date, only a single strain of *C. condimenti* has been identified, isolated from spiced meat. Therefore, it may be early days to implicate this region with possible pathogenicity of the species. Further investigation and isolation of more samples needs to be carried out in order to clarify this aspect.

6.3.5 Mobile genetic elements

Genomic regions related to the mobile genetic elements such as plasmids and bacteriophages tend to provide interesting insights in the comparative genomic analysis of an organism, and are often found to be major contributors to the accessory genomic regions of the pan-genome of an organism. The study of the variation in plasmid content extends from our group's previous CGH studies (Kucerova *et al*. 2010) and also complements others who have reported that the larger *Cronobacter* spp. plasmids were based on the incompatibility group RepFIB (Franco *et al*. 2011). Kucerova *et al*. (2010) had reported that pESA2 (31 kb) was absent in all except two strains on the micro-array. It was found to be partially present in *C. turicensis* which had 19 (61.3%) genes present, and *C. sakazakii* 696 which had 4 (12.9%) genes present. In this genome study, this was found to be true with the plasmid being partially present on the *C. sakazakii* 696 genome. The plasmid profile experiment however showed a plasmid of a similar size to be present in all genomes except *C. sakazakii* 680 and 696, *C. turicensis* 564 and *C. condimenti* 1330. Therefore, the location of this plasmid region in strain 696 could be chromosomal. The percentage of pESA3 (131 kb) genes present in the strains with the larger plasmid was found to vary from 44 - 89%, across the *Cronobacter* strains in the microarray study (Kucerova *et al*. 2010). This variation was reflected in this genomic study as well. It was seen that the publicly available *C. sakazakii* BAA-894 plasmid pESA3 (131 kb) or *C. turicensis* z3032 pCTU1 (138 kb) formed a conserved backbone, with one copy being present in most of the strains and was confirmed both *in silico* as well by the plasmid profiling. Only remnant regions of the plasmid were found on the sequenced genomes of *C. sakazakii* 680 and *C. condimenti* 1330, while plasmid profiling could not detect the plasmid in these strains. Franco *et al*. (2011) in their laboratory studies of the plasmid regions have also confirmed the presence of this plasmid in 97% of their 229 *Cronobacter* spp. strains tested by PCR. They also identified these plasmids to belong to RepFIB incompatibility group, characterized by the *repA* gene as an origin of replication. This plasmid backbone is especially important for this pathogen as they have been characterised to be virulence plasmids, with genes encoding for potential virulence traits such as iron acquisiton systems (*eitCBAD* and *iucABCD/iutA*). Iron is an essential nutrient for bacteria and the ability of a pathogenic bacterium to acquire iron from its environment is often considered an important virulent determinant. Other regions of variation related to virulence included the presence of plasminogen activator gene (*cpa*) and T6SS in the pESA3 plasmid. The *cpa* gene encodes for an outer membrane protease which has been reported to degrade serum proteins and thereby improve the survival and invasive activity of the bacterium across the blood brain barrier. In the *C. turicensis* plasmid pCTU1, there is a 27kb region with homologies to *fhaB*, *fhaC* and five adhesins (FHA locus), which were found to be absent in the plasmid pESA3 (Franco *et al*. 2011). These are a group of genes that encode for

filamentous agglutinins and their transporters. These differences were also visualised in this study in the BRIG alignments using these plasmids as the reference, where these regions showed varying levels of identity in all the *Cronobacter* genomes (Fig. 6.7 and 6.8). *In silico* analysis found this region to be present in all the genomes except those of the *C. sakazakii* species. These virulence related plasmid-borne traits have been discussed in detail in Chapter 7, Section 7.2.2.

Apart from these plasmid regions conserved with the publicly available regions, a few other cryptic plasmids were also identified in some of the genomes, which mainly comprised of genes encoding important functions in the organism, such as an arsenate transporter in *C. muytjensii* 530 or a group of Inc genes coding for a type four secretion system in *C. sakazakii* 696 and 701. It is interesting to note that even though only some species of the *Cronobacter* genus have been reported to cause infections, all of them appear to carry the same plasmid backbone with genes coding for important virulence functions. Laboratory studies based on plasmid curing and gene expression can help to throw better light on the role of these plasmids in the pathogenicity of this organism. The functionality of these plasmid regions and their role in *Cronobacter* physiology and pathogenicity has been described in further details in Chapter 7.

The prophage regions were found to be major regions of variation between the genomes in this study. There were 20 complete prophage genomes identified in 13 *Cronobacter* spp. genomes, apart from a large number of putative or incomplete prophage regions as well (Tables 6.2 and 6.3). On average, these regions occupied 5% of each genome. Previously, Kucerova *et al*. (2010) had detailed three prophages and 3 partial prophages in *C. sakazakii* BAA-894 contributing to the genome diversity and had identified them to be among the variant regions in the genus-wide micro-array study. None of these prophages were previously reported in the genome of *C. turicensis* z3032 (Stephan *et al*. 2011). Also, no prophage regions were detected in the publicly available genome of *C. sakazakii* E899 (Chen *et al*. 2011). This did come across as an anomaly, and since plasmid sequences were also not sequenced for this strain, the fact that this could have been a sequencing issue cannot be eliminated.

A number of intact prophage regions were found to be shared between the *Cronobacter* species as indicated in Table 6.3. However, apart from the ones indicated, none of the others were found to be completely unique. All the other phage regions exhibited a high level of mosaicism with partial regions being shared between genomes. Also, the genes in these prophage regions found similarity matches with phages that have previously been identified in other bacterial species such as *Enterobacter* spp., *Salmonella* spp., *Pantoea* spp., *E. coli* and *Listeria* spp. This genetic mosaicism has also been observed in the phage genomes of other *Enterobacteriaceae* members, such as the bacteriophage ST64B, isolated from *Salmonella*

enterica (Mmolawa *et al*. 2003). These traits hint at the possibility of these prophage regions having been involved in multiple horizontal gene transfer events.

Temperate bacteriophages have been reported to possess genes that encode O-antigen modification enzymes in organisms such as *Shigella flexneri* and *Pseudomonas aeruginosa* (Allison & Verma 2000; Allison *et al*. 2002; Kropinski 2000). The possibility of this in *Cronobacter* spp. was proposed by Sun *et al*. (2011) when comparing RFLP patterns from PCR amplification of the O-antigen genes of 119 *C. sakazakii* strains. However, they did not identify any sequence for this putative region. In the analysis of the phage regions in this study, a gene coding for the O-acetyl transferase was identified in the genome of *C. sakazakii* 701 in the region of an incomplete phage. This is especially interesting since the strain belongs to ST4, largely responsible for neonatal meningitis in *Cronobacter* (Chapter 5). Whether there is an association between the gene and the virulence of the strain needs to be further investigated. Seroconversion is reported to be an important virulence factor in pathogenic bacteria, especially as an immune defence mechanism. Variations in the antigenic profile of a bacterium may result in the host having to develop a variety of immune responses specific to each of these serotypes. This can in turn play a key role in the improved and enhanced survival properties of the organism. Serotype-converting bacteriophages play an important role in conferring these traits, and recognising their presence in *C. sakazakii* is important due to the recent developments in the application of the serotyping scheme for the genus (Jarvis *et al*. 2011; Sun *et al*. 2011).

In the last couple of years, a number of studies have been carried out to sequence and characterise the genomes of *Cronobacter*-specific phages (Lee *et al*. 2011a & b; Lee *et al*. 2012a, b & c; Shin *et al*. 2012). Of these, an intact phage region shared between *C. sakazakii* 696 and 701 showed high similarity with the genome of *Cronobacter* temperate phage phiES15 (Lee *et al*. 2012a). Parts of this genome were also found conserved in the intact phage regions of *C. muytjensii* 530 and *C. condimenti* 1330. Identification of species-specific phages is very important for the purpose of bacteriophage therapy, especially as it can be useful to overcome the antibiotic resistance in certain pathogens. In the case of *Cronobacter* spp., it has also been applied as a control measure in food-borne environments, including infant formula (Kim *et al*. 2007; Zuber *et al*. 2008). Further studies need to be conducted to investigate any possible association between the prophages and the fitness and virulence of the host species strain.

CRISPR systems provide bacteria with a mechanism of resistance to infection by those phages to which they have previously been exposed. In this study, CRISPR regions were identified in 10 out of the 14 *Cronobacter* spp. genomes analysed (Table 6.7). The majority of them were short regions, with only a single spacer unit, and therefore could be non-functional or artefacts in the genome. The major CRISPR regions were identified in the publicly available finished genomes of *C. sakazakii* BAA-894 and *C. turicensis* z3032. The CRISPRs in the *C.*

sakazakii genome were found to be similar to regions identified in the genomes of the plant pathogens, *Erwinia* spp. (McGhee & Sundin 2012). This again correlates well with the suspected plant habitat of *Cronobacter* spp., and hints at the possible acquisition of these regions through horizontal gene transfer events. Because of the adaptive immune memory mechanism of the CRISPRs, they play a very important role in genome variation and evolution (Horvath and Barrangou 2010). To date, there have been no published studies conducted on the CRISPR regions in *Cronobacter* spp. Further work needs to be carried out to investigate any possible connection between the CRISPRs and the pathogenicity of the organism.

6.3.6 Conclusions

This comparative genomic study of the *Cronobacter* genus has showed how the advances in bacterial genome sequencing coupled with useful tools such as pan-genome analysis can today theoretically tell us a lot more about the diversity of a species or genus than has been possible in laboratory based studies. Analysing the pan-genome also introduces a potentially higher resolution for intra-species characterization between strains. For example, in this study, the strains *C. sakazakii* 701 and E899 are both identified as ST4 by MLST, but revealed a number of variations in the dispensable genes in the pan-genome analysis. The *Cronobacter* genomic study has revealed an open pan-genome with possibilities of increasing diversity being discovered as more genomes are sequenced. Like other enteric bacteria, in *Cronobacter* too, the mobile genetic elements have proven to be major contributors to the diversity and variability, thus driving forward the evolution of the genus.

This genomic study now moves into Chapter 7, which presents results focussing on the analysis of important genomic regions linked to the physiology and virulence of the genus.

CHAPTER 7

DIVERSITY IN THE PHYSIOLOGY AND VIRULENCE OF THE GENUS *CRONOBACTER* AS REVEALED BY COMPARATIVE GENOMIC ANALYSIS

7.1 Introduction

7.1.1 Aims of the chapter

The results reported and discussed in this chapter are a continuation of the comparative genomic study of the *Cronobacter* genus presented in Chapter 6.

This chapter deals with two aspects of the organism. The first set of results is based on the genomic diversity of certain key aspects of the physiology of the organism such as environmental persistence and fitness. The story then continues to the virulence related genomic regions, in order to investigate any possible correlation of the pathogenicity of *Cronobacter* spp. with the species and/or sequence type.

Some of the results presented here have been accepted for publication in Joseph *et al*. (2012c) and Joseph *et al.* (2013b).

7.1.2 Genetics of the O-lipopolysaccharide (O-LPS) in bacteria

The outer membrane (OM) of Gram-negative bacteria is mainly composed of a complex made of lipids and polysaccharides, often referred to as the lipopolysaccharide (LPS). This LPS is comprised of three major components: (i) the lipid A, which is also the endotoxin and is responsible for the pathogenicity of the bacteria (ii) the core polysaccharide which is connected by a lipid molecule to the OM (iii) the O-antigen or the O-LPS that forms the outermost component of the LPS and is attached to the core oligosaccharide.

The O-LPS is basically made up of 2-6 repeating units of oligosaccharides, and can be extremely variable even within species, the variability dependent on the linkage patterns of the sugars. To date, in *E. coli* alone, >100 different types of O-antigens have been identified. This O-antigen structure may be either linear or branched, in the form of either homopolymers or heteropolymers. The synthesis of the O-antigen structure and assembly is encoded by the chromosomal genes referred to as the O-antigen cluster (*rfb* gene cluster). The variability in the O-antigen structures is also reflected in the gene sequences of the O-antigen cluster. On the bacterial chromosome, this gene cluster is located between two genes that are generally found to be conserved within species or related species. For example, in *E. coli*, *Salmonella enterica* and related *Enterobacteriaceae* members (including *Cronobacter* spp.), these genes are *galF* and *gnd* which encode for the enzymes UTP-glucose-1-phosphate uridylyltransferase and 6-phosphogluconate dehydrogenase respectively. The O-antigen clusters in *Enterobacteriaceae* are comprised of a range of 6 to 19 genes depending upon the species.

These genes in the O-antigen cluster encode three groups of proteins (Fig 7.1):

(a) Nucleotide sugar precursor biosynthesis proteins: These proteins are encoded by various gene clusters denoted as *rmlABCD*, *manABC* or *ddhABCD* depending upon the type of sugar being synthesized. This process takes place in the cytoplasm of the cell.

(b) Glycosyl transferase proteins: These transport proteins are responsible for transferring the synthesized sugars to the membrane bound undecaprenol lipid carrier for the formation of the O-antigen structure, a procedure that occurs on the cytoplasmic face of the inner membrane.

(c) O-antigen processing proteins: These proteins function to translocate the O-antigen unit to the periplasmic face of the inner membrane, followed by polymerization by glycosidic linkages to lengthen the structure. In most *Enterobacteriaceae,* they are encoded by the three genes *wzx* (flippase), *wzy* (polymerase) and *wzz* (chain-length determinant).

Finally, this nascent O-antigen unit is transported to the core lipid A by the WaaL protein, and further transported across to the outer surface of the OM by the Lpt group of proteins.

The LPS is an indispensable structure to the Gram-negative bacterial cell, and the O-antigen is especially important for its role in the innate immunity of the host as well as a receptor for bacteriophages. Therefore, detailed analysis of the genomic regions linked to these structures is very important in the study of pathogens. The high levels of variations observed in these regions hint at possibilities of horizontal transfer events, thus also making them interesting targets for evolutionary studies (Samuel and Reeves 2003; Wang and Quinn 2009; Wang *et al*. 2010).

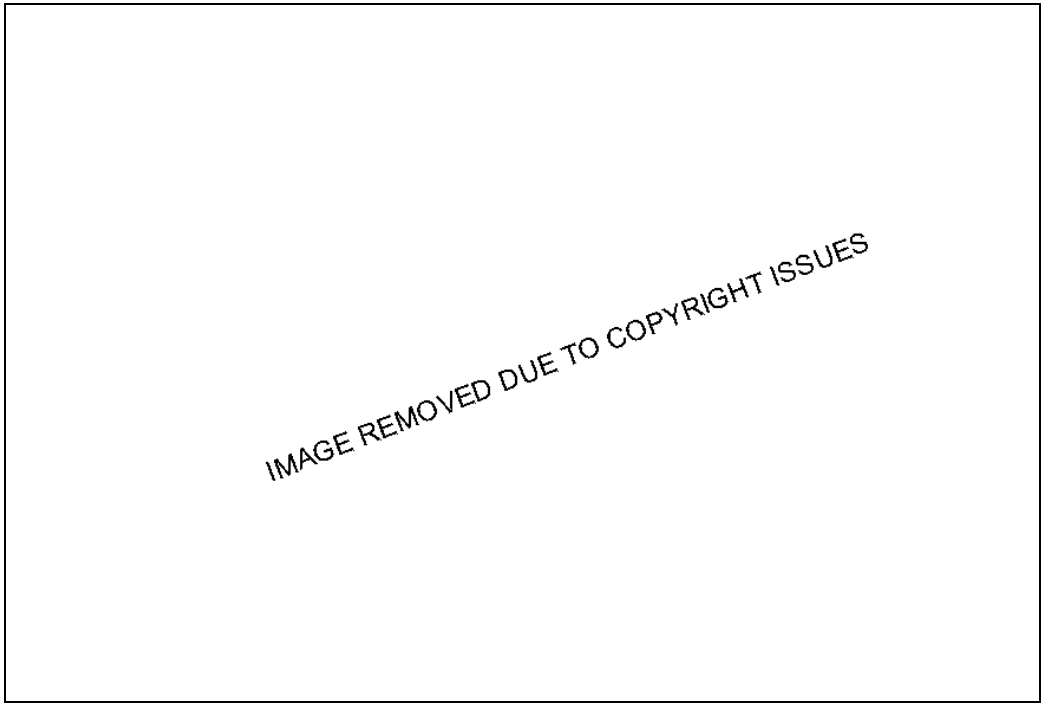


Fig. 7.1 A representation of the O-antigen synthesis in Gram-negative bacteria, indicating the genes involved in the three different stages. Source: Samuel and Reeves (2003).

7.1.3 Bacterial secretion systems: Type IV secretion systems (T4SS) and Type VI secretion systems (T6SS)

Most bacteria possess complex machinery called secretion systems which help them in the transport of molecules across their membranes to be released into the surrounding medium or the eukaryotic host cell. They are common occurrences among most pathogenic bacterial groups, and often the transported proteins are virulence factors. In Gram-negative bacteria, the translocation process may either be a continuous single step procedure across the inner and outer membranes (as seen in types 1, 3, 4 and 6 secretion systems) or a step-by-step process, involving transport into the periplasmic space and thereafter secretion out of the cell across the outer membrane (as observed in types 2 and 5 secretion systems). In both Gram-negative as well as Gram-positive bacteria, this secretion process is enabled by the universal Sec or two-arginine (Tat) pathways. To date, seven different groups of secretion systems have been identified and characterised in bacteria, classified as Type 1 (T1) to Type 7 (T7) secretion systems. Each of these has unique functions and structural properties that set them apart. Of these, types 1 to 6 have been observed both in Gram-negative as well as Gram-positive bacteria, while the T7SS has recently been identified only in Gram-positive bacteria (Tseng *et al*. 2009). Based on the results presented in this chapter, the T4SS and T6SS have been described in further detail here.

The T4SS in bacteria is the translocation machinery that shows a high level of evolutionary similarity with the conjugation systems encoded on bacterial plasmids. The presence of this secretion system in bacteria was initially proposed by Salmond (1994), however the actual structure of the core components were elucidated more recently by Fronzes *et al*. (2009a). The most studied model of the T4SS has been in *A. tumefaciens*, which is composed of a VirB-VirD system, comprising of 11 proteins encoded by the *virB* operon and the protein VirD4, encoded by the *virD* operon (Fig. 7.2). The proteins associated with the T4SS are classified into three major groups. The proteins VirB4, VirB11 and VirD4 are ATPases associated with the inner membrane or the cytoplasm. Of these, VirB11 proteins exhibit chaperone activity while VirD4 have been identified to be coupling proteins. The proteins VirB6-10 may be associated either with the periplasm or the outer membrane, and they form the core structure of the translocation machinery of the T4SS. The proteins VirB2 and VirB5 are the major and minor components respectively of the bacterial pilus, and are therefore most responsible for the virulence and conjugation mechanism (Yeo and Waksman 2004; Fronzes *et al*. 2009b). In *A. tumefaciens,* the T4SS functions in the infection of the plant host cells with the DNA of the Ti plasmid, resulting in the crown gall disease. In recent studies, T4SSs have been implicated in the virulence of human pathogens such as *Helicobacter pylori* (Terradot and Waksman 2011), *E. coli* (Lieberman *et al*. 2012) and *Brucella* spp. (Jong and Tsolis 2012).

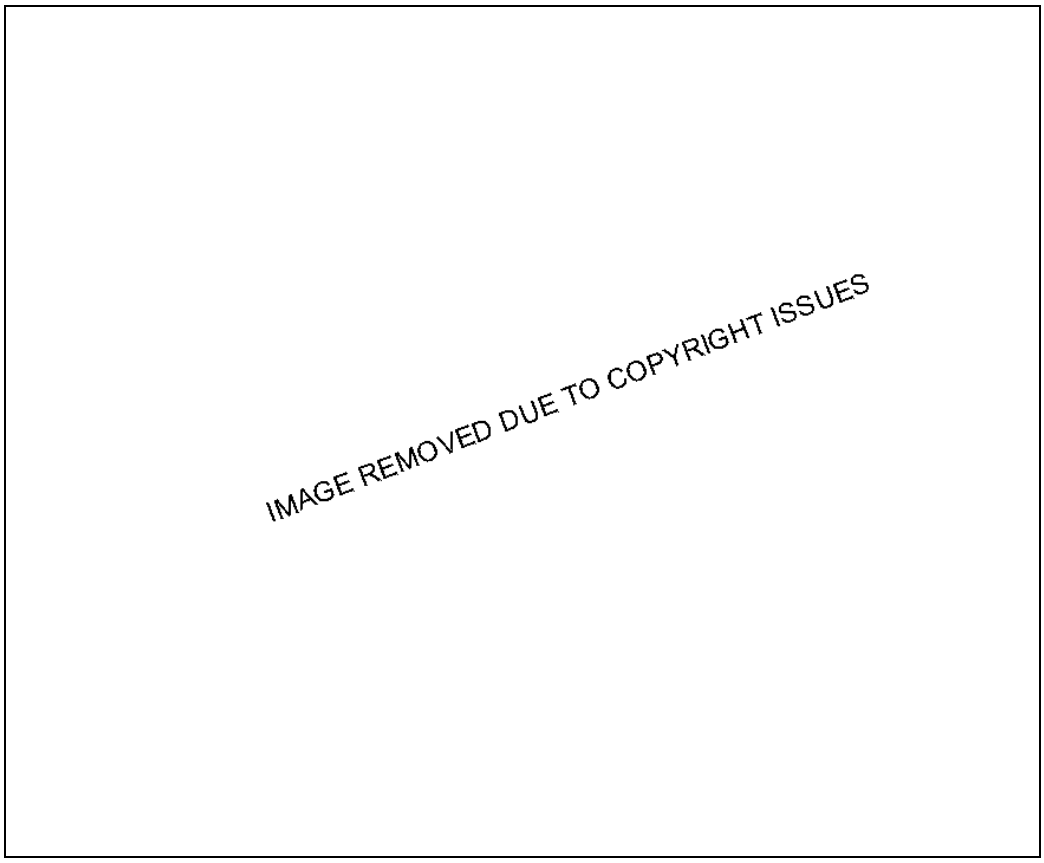


Fig. 7.2 Basic organization of the proteins in a prototype T4SS. Based on putative functions, the ATPases have been denoted in green, the core components in blue, the pilus structures in orange and lytic transglycosylases in yellow. Source: Yeo and Waksman (2004).

The T6SS is one of the more recently characterized bacterial secretory machinery systems, for the transport of proteins across bacterial cell membranes. The composition of T6SSs varies across organisms, with different systems composed of a range of genes from 12 to more than 20. Organisms may possess multiple clusters of T6SS genes, not necessarily identical to each other, and not all of them may be functional. In many organisms such as *P. aeruginosa* and *E. coli*, these regions have been identified as pathogenicity islands.

Indications of the existence of such a protein transport mechanism were first observed as early as 1996, through studies on the secretion of the haemolysin coregulated protein, Hcp, in *Vibrio cholerae* (Williams *et al*. 1996), and subsequently in other groups of micro-organisms. However, it was later in 2003 that further detailed identification of such a region was carried out through an *in silico* study of the *V. cholerae* genome, to be then called IAHP (IcmF-associated homologous protein) clusters, because of the similarity of one of the proteins to IcmF proteins observed in some T4SSs (Das *et al*. 2003). The classification of this protein secretion system as "Type VI" was in 2006 through yet another study on *V. cholerae*, describing the export process of the Hcp and VgrG proteins, and their role in the virulence of the organism (Pukatzki *et al*. 2006).

The core components of a typical T6SS include the IcmF and IcmH-like proteins, CIpV ATPase, a putative lipoprotein and the proteins Hcp and VgrG (valine glycine repeats), the latter two also being the secreted factors (Fig. 7.3). This secretory mechanism may be regulated either by transcriptional activators of the AraC family or σ-54, or by a threonine phosphorylation signalling cascade.

T6SSs have been associated with virulence functions in a number of pathogens such as *E. coli, P. aeruginosa, V. cholerae* among others, performing roles such as host cell adhesion and invasion, macrophage survival and cytotoxicity. However, apart from pathogenicity, T6SS has also been associated with the physiological activities of some organisms such as root colonization by the nitrogen-fixing *Rhizobium* spp., as well as other functions such as quorum sensing and biofilm formation (Bingle *et al*. 2008; Cascales 2008; Leung *et al*. 2011).

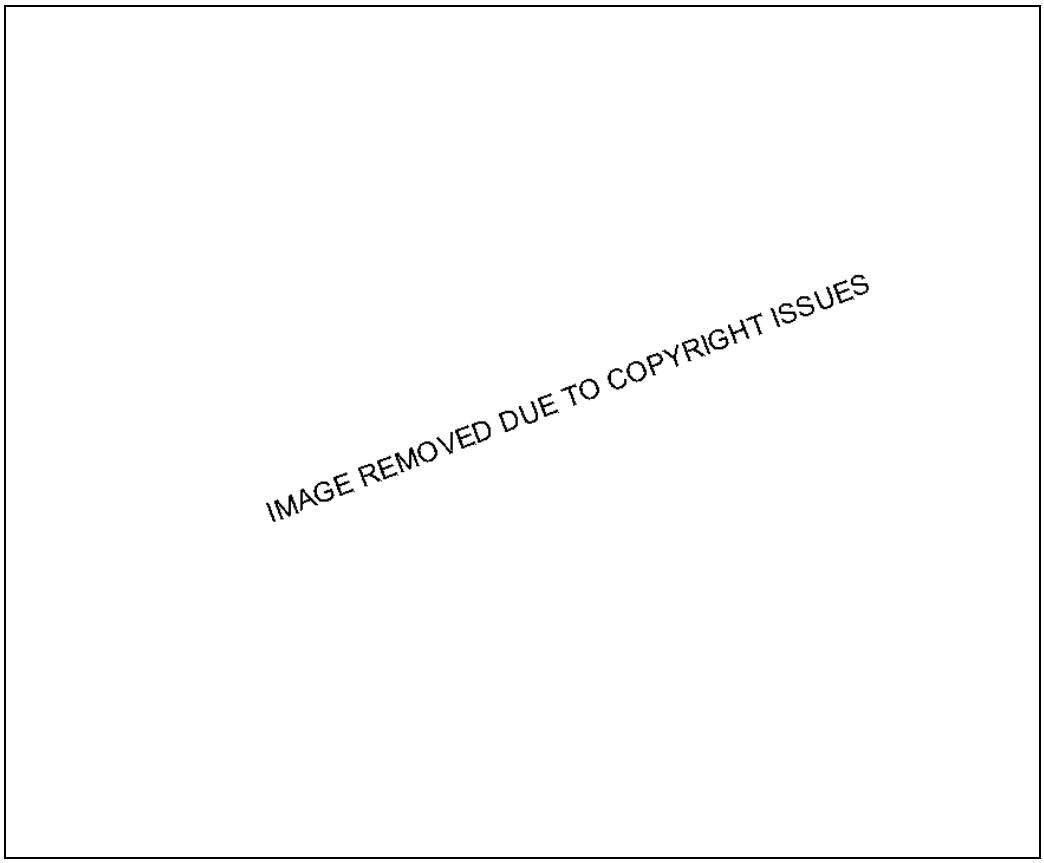


Fig 7.3 A schematic model for the structural assembly and function of a T6SS.
Source: Reproduced from Filloux *et al*. (2008). The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology*, 154, 1570-1583. doi:10.1099/mic.0.2008/016840-0.
Key: C – Bacterial cytoplasm; IM - Bacterial inner membrane; P - Bacterial periplasm; OM – Bacterial outer membrane; ECM - Extracellular milieu; PM - host cell plasma membrane.

7.1.4 Sialic aid utilisation and its role in bacterial pathogenicity

Sialic acid or neuraminic acid refers to a family of nine-carbon keto sugars that may be found on mammalian mucosal surfaces such as the intestinal gut lining, brain, secretions of the mouth and lungs as well as milk. These are sites colonized by a wide range of bacteria for which this sialic acid can act as a source of carbon and nitrogen. With respect to *Cronobacter* spp. epidemiology, it is also noted that sialic acid is often found as an ingredient of PIF.

Sialic acid may exist in nearly 50 different forms, though the most commonly observed and also most studied form is 2-keto-3-deoxy-5-acetamido-D-glycero-D-galacto-nonulosonic acid, often abbreviated as Neu5Ac. This sialic acid is generally found bound to sugars to form polysaccharides, which are then bound to lipids or proteins to form sialo-glycoconjugates. The sialic acids are found to occupy terminal positions on these glycoconjugates exposed on the mucosal surfaces.

Neu5Ac is used by bacteria for various purposes such as nutrient source, colonization and virulence. It has been reported that among bacteria, only human pathogens and commensals have evolved the genetic machinery for the uptake of exogenous sialic acid. For the uptake of sialic acid in bacteria, the Neu5Ac first needs to be released from the glycoconjugate forms, which is brought about by the activity of sialidases. Some bacteria can produce their own sialidase or neuraminidase, encoded by the gene *nanH.* Others may obtain the sialidases exogenously from the environment. These enzymes hydrolyze the sialic acid-terminal sugar linkages in the mucosal glycoconjugates to release free sialic acid. Once freed, the Neu5Ac sugars are available in the solution for bacteria to utilise.

The common mode of sialic acid catabolism in most bacteria is encoded in the basic *nanAKETR* gene cluster. Many variants of this gene cluster have been found in different bacteria, though the basic composition and gene order remains primarily the same for most *Enterobacteriaceae* members. The first step in the metabolism process is the transportation of sialic acid into the bacterial cell, brought about by a transporter protein. There are three different types of transporters in bacteria: NanT, a major facilitator superfamily (MFS) protein; TRAP, a tripartite ATP-independent periplasmic transport system; and an ATP-binding cassette (ABC) transporter. All *Enterobacteriaceae* members studied to date have shown the presence of the single-component NanT transport system.

Once transported into the cell, the Neu5Ac lyase (NanA) converts sialic acid (Neu5Ac) into N-acetylmannosamine (ManNAc) and phosphoenolpyruvate (PEP). NanK is an ATP-dependent kinase specific for ManNAc generating N-acetylmannosamine-6- phosphate (ManNAc-6-P). ManNAc-6-P epimerase (NanE) then converts the ManNAc-6-P into N-acetylglucosamine-6-phosphate (GlcNAc-6-P). GlcNAc-6-P deacetylase (NagA) and glucosamine-6-P deaminase (NagB) converts GlcNAc-6-P into fructose- 6-P (Fru-6-P), which is

a substrate in the glucolytic pathway (Fig. 7.4). NanR is the repressor that regulates the activity of these genes. In most bacteria, the genes for the first three enzymes (*nanA, nanK* and *nanE*) are usually found clustered together forming what is denominated as the Nan cluster. However, there have been a few exceptions such as the genomes of organisms such as *Citrobacter freundii* and *Edwardsiella tarda* where the *nanE* gene has been located separate from the rest of the operon. The genes encoding NagA and NagB are located adjacent to each other, but most often not necessarily in the vicinity of the Nan cluster on the bacterial genomes. Other related genes of interest also controlled by NanR include *nanC,* encoding an outer membrane porin and *nanS,* encoding a sialate O-acetyl esterase (Vimr *et al*. 2004; Severi *et al*. 2007; Vimr 2012).

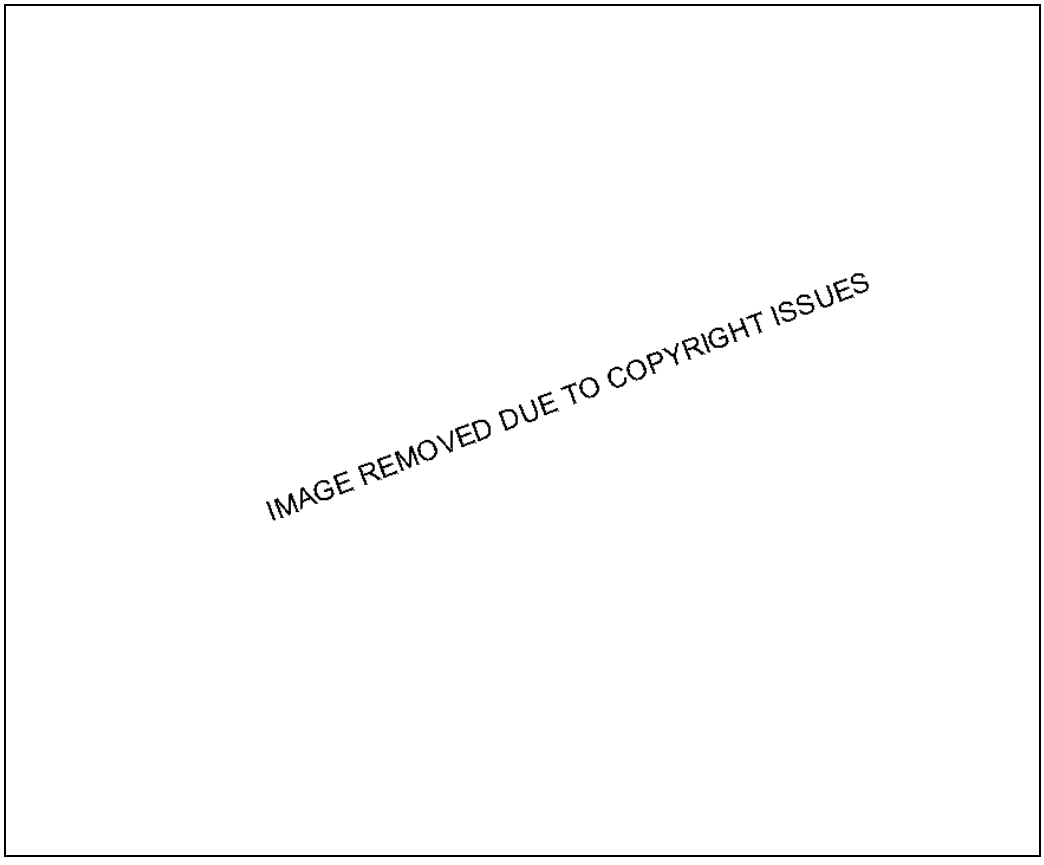


Fig. 7.4 A basic overview of the various pathways that have been characterised for sialic acid utilisation in bacterial pathogens. Source: Reproduced from Severi *et al*. (2007). Sialic acid utilization by bacterial pathogens. *Microbiology,* 153, 2817-2822. doi:10.1099/mic.0.2007/009480-0.

Apart from this transport protein mediated uptake of sialic acid, some bacteria also have alternative means of obtaining the sugar, including *de novo* biosynthesis and scavenging. The *de novo* synthesis is brought about by the action of the proteins NeuB and NeuC, using UDP-GlcNAc as a precursor (Fig. 7.4), and is an activity observed in *E. coli*. Some bacteria that lack the biosynthesis machinery may indulge in host scavenging of either sialic acid itself (*N. gonorrhoea*) or its precursors (*H. influenzae*) (Vimr *et al*. 2004; Severi *et al*. 2007; Vimr 2012).

The uptake of sialic acid into bacterial cells has been associated with a number of virulence factors. In organisms such as *E. coli* K1, sialic acid is used to decorate the cell surface to form glycolipid capsules, and thereby participating in a molecular mimicry of the host. This can help the organism to overcome the immune responses of the host against the pathogen. Studies in organisms such as *Vibrio* spp. and *E. coli* have indicated a possible association of sialic acid catabolism with the adhesion and colonization properties of the organism. The release of free sialic acid from host glycoconjugates reveals masked adherence points on the host ligands for the bacterium (Vimr 2012). Neuraminidase activity has been shown to play an important role in biofilm formation properties of the respiratory pathogen *P. aeruginosa* (Soong *et al*. 2006). Another study conducted by Trappetti *et al*. (2009) has shown the activity of free sialic acid in the biofilm formation, colonization of the nasopharynx and invasion of lungs in pneumococcal bacteria. A more recent study analysed the presence of the sialic acid transport and catabolism gene cluster in Group B *Streptococcus* (GBS) and this revealed the effects of exogenous sialic acid levels on the increased invasive abilities of the organism (Pezzicoli *et al*. 2012).

7.2 Results

7.2.1 Physiological and phenotypic traits in the *Cronobacter* spp. genomes

The presence of known genes conferring physiological and phenotypic traits, categorized according to environmental stress response, cell surface composition, sugar metabolism, and metal resistance was examined.

7.2.1.1 Stress response genes

A number of key stress response genes on the *Cronobacter* genomes were investigated in this study. A gene homologous to the universal stress protein *uspA* (ESA_01955) was detected in all the *Cronobacter* spp. genomes. Similarly, homologues were located for stringent starvation response *sspA* (ESA_03615), carbon starvation sensing protein *rspA* (ESA_01752), and carbon starvation protein (ESA_00801) in all *Cronobacter* spp. strains. However, an additional carbon starvation homologue (ESA_00339) was found in all *Cronobacter* spp. genomes except that of *C. sakazakii* 680.

The organism is known for its ability to survive desiccation for up to two years (Caubilla-Barron & Forsythe 2007; Osaili & Forsythe 2009). Genes responsible for the desiccation resistance and osmotic stress adaption were found in all *Cronobacter* species. This included genes encoding the uptake of the osmoprotectants glycine, betaine (ESA_00587-00589; *proX*, *proW* & *proV*), and trehalose uptake (ESA_01944-46), which were present in all the *Cronobacter* spp. genomes.

Two separate clusters of genes (ESA_01108-01111; ESA_01738-01741) related to osmoprotectant ABC transporters (*OpubA, OpubB* & *OpubC*) found in *Enterobacteriaceae*, were also found to be present in all the *Cronobacter* spp. genomes. Of these the latter cluster was found to be unique to the *Cronobacter* genus, when compared to other closely related *Enterobacteriaceae* members.

A two-component system that regulates the organism's responses to osmotic stress is encoded by the genes ESA_04334 (*ompR*) and ESA_04335 (*envZ*) and was also found to be present across the genus. Another such two-component system involved in the regulation of envelope stress responses is encoded by the genes ESA_04122 (*cpxA*) and ESA_04123 (*cpxR*), was also present in all the *Cronobacter* spp. genomes.

Thermal tolerance in *Cronobacter* spp. has been controversial due to conflicting reports regarding the putative genes involved (Williams *et al*. 2005; Asakura *et al*. 2007; Gajdosova *et al*. 2011). The initiation translation factor (*infB*) proposed by Asakura *et al*. (2007) was present in all the *Cronobacter* spp. genomes. It is a housekeeping gene and also one of the alleles used

in the *Cronobacter* MLST scheme (Chapter 3). In contrast, an 18kb thermoresistance gene cluster (Accession No. FR714908), comprising of 22 genes, identified by Gajdosova *et al*. (2011) showed variations in the *Cronobacter* genus. In the original study, the cluster was identified in the *C. sakazakii* type strain ATCC 29544^T (ST8). In the genomes analysed in this study, the cluster was only partially present in *C. sakazakii* strains 696 (ST12), and 701 (ST4), and completely present in both strains of *C. malonaticus* (ST7 and ST11). The cluster was absent in the remaining *Cronobacter* genomes in this study.

The genes ESA_00341-00344 (*crtZ*) are responsible for β-carotene production, a yellow pigmentation, which is speculated to help protect the organism from light produced by oxygen radicals. This gene cluster was present in all the *Cronobacter* spp. genomes, except *C. sakazakii* 680.

7.2.1.2 Capsule and biofilm formation

A cluster of 21 genes (ESA_01155-01175; *wzABCKM*), encoding proteins involved in the capsular polysaccharide biosynthesis and colanic acid production was identified to be conserved in all the genomes of the sequenced *Cronobacter* spp. strains.

The genes for capsular polysaccharide assembly and export (ESA_03350-59) were identified in the genome of *C. sakazakii* BAA-894. Of this cluster, the genes encoding the export proteins and transporters were present in the genomes of all the *Cronobacter* spp. strains, while the genes involved in the assembly process (ESA_03354-57) were absent from the two clinical isolates, *C. sakazakii* 701 (ST4) and 696 (ST12). These genes were also found missing in the publicly available genome of *C. sakazakii* E899 (ST4).

Also linked to capsule formation is the ability of the organism to form biofilms. Transposon mutant studies by Hartmann *et al*. (2010) have identified the genes ESA_00281 and ESA_00282 in the *C. sakazakii* BAA-894 genome to be linked to the biofilm formation properties of the organism. Both these genes were found conserved in all the sequenced genomes of the genus.

7.2.1.3 Sugar utilisation genes

The uptake and utilisation of sugar sources is a key aspect of microbial metabolism. A number of different sugar pathways are involved in *Cronobacter* metabolism, and the key genes responsible for some of them have been characterised here and tabulated in Table 7.1.

- The gene cluster encoding for maltose utilisation, including α–glucosidase (ESA_02709-14) was found in all the *Cronobacter* spp. genomes.

- A gene cluster for mannose utilisation (ESA_02616-18) was only found in the genomes of *C. sakazakii* and *C. turicensis*, but not in the closely related *C. malonaticus* or the other species of the genus.

- *Cronobacter* spp. are able to ferment β-glucoside sugar substrates from plants, including cellobiose, arbutin, salicin, and esculin. The genes involved in the metabolism of β-glucosides are located at ESA_02544-47. ESA_02544 is a transcriptional antiterminator from the BglG family, while ESA_02545-47 are components of a phosphotransferase (PTS) system involved in the utilization of β-glucosides. This gene cluster was present in all the *Cronobacter* spp. genomes except those of *C. turicensis* z3032 and *C. universalis* 581.

- Genes encoding for a PTS system of trehalose uptake (ESA_00263-265) were found conserved across the genus.

- A cluster of eight genes (peg no. 1824-1831) involved in the utilization of fucose was found unique to the genome of the newly defined species, *C. condimenti* 1330. This cluster was absent in the genomes of other closely related *Enterobacteriaceae* members as well.

- The uptake of galacticol is encoded by a three-gene cluster (CTU_04360-04380; *gatABC*) which was present only in the genomes of the species *C. turicensis, C. universalis* and *C. muytjensii*.

In order to define the genus *Cronobacter*, Iversen *et al*. (2008) had listed fourteen biochemical tests based on carbon utilisation to differentiate the species and subspecies of the genus (Listed in Chapter 1; Table 1.2). A number of the genes involved in these biochemical

pathways were found to be conserved in all the genomes of the *Cronobacter* genus, though some did show variations.

- Two clusters of genes - ESA_02480-02483 (*potFGHI*) & ESA_02222-02224 (*potABCD*) – involved in the metabolism of putrescine were found to be present in all the genomes. A variant of the *potA* gene (CTU_24430) was however found absent in the genomes of *C. sakazakii* and *C. malonaticus*, and present in the rest of the genus.

- The genes required for malonate metabolism - malonyl CoA acyl carrier protein transacylase, phosphoribosyl-dephospho-CoA transferase, and malonate decarboxylase (Ctu_34990-35070) - were found in all *Cronobacter* species except the genomes of *C. sakazakii* and *C. dublinensis* strain 582.

- A number of genes linked to inositol catabolism were found in the *Cronobacter* genomes with a fair degree of variation. The *C. sakazakii* BAA-894 genomes contained a couple of genes (ESA_00718 & ESA_01913) which are also present across the genus. However, a cluster of eleven genes (CTU_08910-09010) showed varying patterns across the genus. A single gene unique to the genome of *C. turicensis* 564 (peg no. 4172) was also found. In a study by Hamby *et al*. (2011), a potential link between inositol fermentation and pathogenicity of *Cronobacter* spp. has been suggested.

Function	Loci	***C. sakazakii* BAA-894**	***C. sakazakii* 696**	***C. sakazakii* 701**	***C. sakazakii* 680**	***C. malonaticus* 507**	***C. malonaticus* 681**	***C. turicensis* 564**	***C. turicensis* z3032**	***C. dublinensis* 582**	***C. dublinensis* 1210**	***C. muytjensii* 530**	***C. universalis* 581**	***C. condimenti* 1330**
Maltose utilisation	ESA_02709-14	+	+	+	+	+	+	+	+	+	+	+	+	+
Mannose utilisation	ESA_02616-18	+	+	+	-	-	-	+	+	-	-	-	-	-
β-glucoside metabolism	ESA_02544-47	+	+	+	+	+	+	+	-	+	+	+	-	+
Trehalose utilization	ESA_00263-265	+	+	+	+	+	+	+	+	+	+	+	+	+
Fucose utilization	Ccondi1330_peg no. 1824-1831	-	-	-	-	-	-	-	-	-	-	-	-	+
Galacticol uptake	CTU_04360-04380	-	-	-	-	-	-	+	+	-	-	+	+	-
Putrescine uptake	ESA_02480-02483	+	+	+	+	+	+	+	+	+	+	+	+	+
	ESA_02222-02224	+	+	+	+	+	+	+	+	+	+	+	+	+
	CTU_24430	-	-	-	-	-	-	+	+	+	+	+	+	+
Malonate metabolism	CTU_34990-35070	-	-	-	-	+	+	+	+	-	+	+	+	+
Inositol catabolism	ESA_00718	+	+	+	+	+	+	+	+	+	+	+	+	+
	ESA_01913	+	+	+	+	+	+	+	+	+	+	+	+	+
	CTU_08910-09010	-	+	+	+	+	-	+	+	-	+	+	+	+
	Ctur564_peg no. 4172	-	-	-	-	-	-	+	-	-	-	-	-	-

Table 7.1 Diversity of the sugar utilisation genes across the sequenced genomes of the genus *Cronobacter.*

7.2.1.4 Resistance to elements

Homologues of tellurite resistance genes (*terACDYZ*) were located in the loci ESA_01775–ESA_01804 of the *C. sakazakii* BAA-894 genome. However, this region was found to be absent in all the other *Cronobacter* spp. genomes (Table 7.2). Homologues of these genes are also found on the IncII plasmid R478 of *Serratia marcescens*, pK29 of *Klebsiella pneumoniae* NK29, pEC-IMPQ of *E. cloacae*, and pAPEC-O1-R of *E. coli* APEC O1 (Gilmour *et al*. 2004), as also reported by Kucerova *et al*. (2010). In the *C. sakazakii* BAA-894 genome, this cluster of genes contains an IS element (ISEhe3) just before ESA_01781 which was also identified on the pK29 plasmid as well as a fragment of an IS4 family transposase within the locus ESA_01803. A lone tellurite resistance gene *tehA* was found on the plasmid pCTU1 of *C. turicensis* z3032 (Ctu_1p00320). Another tellurite resistance gene was found unique to the genome of *C. sakazakii* 680 at Csak680_peg no. 3524.

Two copper and silver resistance gene clusters are located close to each other (ESA_04238-45 and ESA_04248–55) in the annotated chromosome of the *C. sakazakii* BAA-894 genome. A similar cluster was also found on the smallest plasmid (pCTU3) in *C. turicensis* z3032 (pCTU3_3p00600-700 and pCTU3_3p00490-590) (Table 7.2). The first of the two regions (*cusRCFBA/silRECBA*) was present in the genomes of all the *C. sakazakii* strains, in the *C. malonaticus* 681^T genome, and the *C. turicensis* $z3032^T$ genomes. CusA is a membrane protein belonging to the resistance-nodulation-division (RND) protein superfamily. CusF is a periplasmic protein, which in *E. coli* interacts with the cusCBA efflux system. The second plasmid-borne region (*pcoABCDR*) was present in nearly all the sequenced *C. sakazakii* genomes except 701, in the *C. malonaticus* 681^T genome, the *C. turicensis* z3032 genome and in the genome of *C. universalis* 581^T. PcoA encodes a homologue to multicopper oxidase, and PcoB is an outer membrane protein transfering copper into the periplasm where it is oxidised by PcoA. Both these regions are also encoded on the plasmids R478 and APEC-O1-R, referred to above, which carry the tellurite resistance genes (Gilmour *et al*. 2004).

Cluster	Function	Loci	*C.sakazakii* BAA-894	*C. sakazakii* 701	*C. sakazakii* E899	*C. sakazakii* 680	*C. sakazakii* 696	*C. malonaticus* 507	*C. malonaticus* 681	*C. turicensis* 564	*C. turicensis* z3032	*C. universalis* 581	*C. muytjensii* 530	*C. dublinensis* 582	*C. dublinensis* 1210	*C. condimenti* 1330
terACDYZ	Tellurite resistance	ESA_01775–01804	+	-	-	-	-	-	-	-	-	-	-	-	-	-
cusRCFBA/ silRECBA	Copper/silver resistance	ESA_04238-04245	+	+	+	+	+	-	+	-	+	-	-	-	-	-
pcoABCDR	Copper/silver resistance	ESA_04248–04255	+	-	+	+	+	-	+	-	+	+	-	-	-	-

Table 7.2 Diversity of the metal resistance genes in the sequenced genomes of the *Cronobacter* genus.

7.2.1.5 Oligo-lipopolysaccharide (O-LPS) genes

The O-LPS gene cluster was one of the most variable regions in the genomes of the *Cronobacter* genus. In the reference genome of *C. sakazakii* BAA-894, this O-LPS gene cluster was located in the region between ESA_01177 and ESA_01189. The genes ESA_01177 (*galF;* encoding UDP-glucose pyrophosporylase), ESA_01178 (*rfbB;* encoding dTDP-glucose 4, 6-dehydratase) and ESA_01189 (*gnd;* 6-phosphogluconate dehydrogenase) were found to be conserved across the *Cronobacter* genus. The region located between ESA_01179-01188 was the highly variable region, varying across the genus. These variations in the O-antigen region have been applied in the design of serotype-specific PCR-based assays (Mullane *et al*. 2008; Jarvis *et al*. 2011; Sun *et al*. 2011, 2012a & 2012b). The regions between ESA_01177 (*galF*) and ESA_01189 (*gnd*) were amplified to generate distinguishable patterns for the varying O-antigen groups.

Based on the comparison of this molecular analysis, the serogroups of the *Cronobacter* spp. genome sequenced strains were determined.

- The O-antigen gene cluster in the reference genome *C. sakazakii* BAA-894 (ST1) corresponds to the *C. sakazakii* serotype O:1 (Mullane *et al*. 2008) and so did strain 680 (ST8).
- Strains 701 and E899, both belonging to ST4, corresponded to *C. sakazakii* serotype O:2 (Mullane *et al*. 2008).
- Strain 696 (ST12) corresponded to the *C. sakazakii* serotype O:4 (Sun *et al*. 2012).
- The O-antigen gene cluster of *C. malonaticus* 507 (ST11), interestingly, showed a 100% sequence similarity with the gene cluster that was previously defined as *C. sakazakii* O: 5 (Sun *et al*. 2012).
- The *C. malonaticus* type strain LMG 23826 was used by Jarvis *et al*. (2011) to define the *C. malonaticus* serotype O:2, and was reflected in the genome of the same strain NTU 681 (ST7).
- The same was true for the *C. turicensis* serotype O:1 defined using the type strain *C. turicensis* LMG 23827 and was reflected in the publicly available genome of *C. turicensis* z3032 (ST19).
- The O-antigen gene cluster in the genome of *C. muytjensii* 530 (ST49) corresponded to the *C. muytjensii* serotype O:1 defined by Jarvis *et al*. (2011).

All the remaining genomes – *C. turicensis* 564, *C. universalis* 581, *C. dublinensis* 582 & 1210 and *C. condimenti* 1330 – exhibited unique combinations of genes in the O-antigen cluster, that have yet to be fully characterized. These gene clusters have been tabulated with details in Table 7.3.

In addition to the variable O-antigen gene cluster located at ESA_01179-89, a second region, ESA_04101–4109 was also found to encode for genes in O-polysaccharide biogenesis and varies across the genus. ESA_04102 encodes a glycosyltransferase involved in cell wall biogenesis and was present in all *Cronobacter* strains, whereas a putative O-antigen ligase glycosyltransferase (ESA_04103) and lipopolysaccharide heptosyltransferase III (ESA_04105) were found to be exclusive to the *C. sakazakii* genomes.

Another notable O-antigen related gene is located in one of the incomplete phage regions in the genome of *C. sakazakii* 701 (ST4). This gene also encoded an O-antigen acetylase (O-acetyl transferase; Csak701_peg 1901) and was located adjacent to a prophage region and has been discussed in Chapter 6.

(a) *C. turicensis* 564 (peg 3935-3944)		
Gene	**Size (bp)**	**Putative function**
galF	891	UTP-glucose-1-phosphate uridylyltransferase
rmlB	1002	dTDP-glucose 4,6-dehydratase
rmlD	900	dTDP-4-dehydrorhamnose reductase
rmlA	894	Glucose-1-phosphate thymidylyltransferase
rmlC	369	dTDP-4-dehydrorhamnose 3,5-epimerase
wzx	1176	O-antigen flippase
wzy	1059	O-antigen polymerase
wbuV	1083	Glycosyltransferase type 1
gnd	1407	6-phosphogluconate dehydrogenase
(b) *C. universalis* 581 (peg 1044-1053)		
gnd	1407	6-phosphogluconate dehydrogenase
wbuB	1209	glycosyl transferase, group 1
fnlC	1131	UDP-N-acetylglucosamine 2-epimerase
wbjC	1104	glycosyl transferase, group 1
fnlA	1035	UDP-N-acetylglucosamine 4,6-dehydratase
wepR	1176	glycosyl transferase, group 1
wepQ	1122	hypothetical protein
wepQ	1122	hypothetical protein
wzx	1290	hypothetical protein
galF	891	UTP-glucose-1-phosphate uridylyltransferase
(c) *C. dublinensis* 582 (peg 1414-1425)		
gnd	1410	6-phosphogluconate dehydrogenase
kpsT	522	ABC transporter related
uppS	699	Undecaprenyl pyrophosphate synthetase
cdsA	858	Phosphatidate cytidylyltransferase
rseP	1392	Membrane-associated zinc metalloprotease
yaeT	2415	Outer membrane protein assembly factor YaeT precursor
skp	495	Outer membrane chaperone Skp (OmpH) precursor

bcbG	1542	capsular polysaccharide biosynthesis protein
galF	891	UTP-glucose-1-phosphate uridylyltransferase
(d) *C.dublinensis* 1210 (peg 469-483)		
gnd	1338	6-phosphogluconate dehydrogenase
wehL	612	Glycosyl transferase
wehC	135	Glycosyl transferase
wbsE	1185	putative glycosyl transferase
wzxE	1251	O-antigen flippase
wbtC	1050	Aminotransferase
wehA	477	Hypothetical protein
wbtA	222	WxcM-like, C-terminal
fdtA	132	WxcM domain protein isomerase
rfbA	816	Glucose-1-phosphate thymidylyltransferase
galF	1074	dTDP-glucose 4,6-dehydratase
(e) *C. condimenti* 1330 (peg 2746-2754)		
wepQ	1158	Lipopolysaccharide biosynthesis RfbU-related protein
wepR	1164	glycosyl transferase, group 1
fnlA	570	UDP-N-acetylglucosamine 4,6-dehydratase
fnlA	210	UDP-N-acetylglucosamine 4,6-dehydratase
wbjC	1104	glycosyl transferase, group 1
fnlC	867	UDP-N-acetylglucosamine 2-epimerase
wbuB	699	Glycosyl transferase 1
wbuC	405	Glycosyl transferase 1
gnd	1407	6-phosphogluconate dehydrogenase

Table 7.3 The list of ORFs identified in the new O-antigen gene clusters in the genomes of (a) *C. turicensis* 564 (ST5) (b) *C. universalis* 581 (ST54) (c) *C. dublinensis* 582 (ST36) (d) *C. dublinensis* 1210 (ST106) (e) *C. condimenti* 1330 (ST98)

7.2.1.6 ATP-binding cassette (ABC) Transporters

Based on the KEGG Pathway annotations and RAST annotations, 60 clusters of ABC transporters were identified across the *Cronobacter* genomes. Of these, 54 were found to be conserved across the genus. These included transport systems for a wide range of molecules such as sugars, mineral ions, peptides, metallic cations, siderophores and vitamins.

Six of these clusters were found to be unique to one or more *Cronobacter* species and have been listed below:

- Only the genomes of the *C. sakazakii* species contain a complete ABC-type multidrug efflux system (ESA_01116–19).
- An oligopeptide ABC transport system (Cdub1210_peg 3934-3937) was found unique to the genomes of the strains *C. dublinensis* 582 and 1210 and *C. condimenti* 1330. The four gene cluster encoded for the proteins OppABCD.
- A urea carboxylase-related ABC transporter system (Cmuyt530_peg 2292-2294), as part of the urea decomposition pathway, was found only in the genomes of *C. muytjensii* 530 and *C. condimenti* 1330.
- A L-rhamnose ABC transport system (Cmuyt530_peg 2639-2642) was also found unique to the genomes of *C. muytjensii* 530 and *C. condimenti* 1330.
- A cluster of ABC transporter genes (Cmuyt530_peg 4114-4117) putatively involved in putrescine transport was found unique to the genomes of *C. muytjensii* 530 and *C. universalis* 581.
- A cluster of ABC transporter genes (Ctur564_peg 3252-3254) involved in the transport of the iron siderophore hemin was found only in the genomes of *C. turicensis* 564, *C. muytjensii* 530 and *C. dublinensis* 582 (Also discussed in Section 7.2.2.4)

7.2.2 Genomic regions linked to *Cronobacter* virulence

7.2.2.1 Secretion systems

The *Cronobacter* genomes lack Type III and V secretion systems but do contain Type IV and VI secretion systems depending on strain and species.

Type four secretion systems (T4SS)

Three clusters of genes related to the T4SS was identified in the sequenced *Cronobacter* genomes.

- A T4SS, unique to the strains *C. sakazakii* BAA-894 (ST1) and *C. turicensis* z3032 (ST19), was located on the smaller plasmid (pESA2 and pCTU2) of both strains. This is a thirteen gene cluster (ESA_pESA2p06578-90), encoding for the VirB1-11 proteins, involved in an energy driven conjugal transfer system. Genes ESA_pESA2p06579 and ESA_pESA2p06581 encode for the protein components for the pili.

- A second cluster of 21 genes, linked to the T4SS, was found to be present only in the *C. sakazakii* 701 (ST4) and 696 (ST12) genomes (Csak696_peg 3784-3805). This region belongs to the *Tra* group of genes and is annotated as IncF conjugation plasmid related regions. This could also be a plasmid-borne region as these genes have previously been identified in the Inc group of plasmids in *E. coli* and *Serratia* spp.

- The third putative T4SS cluster, was a 9 gene cluster, encoding for the *Trb* group of genes also involved in conjugal transfer and was found unique to the genome of *C. universalis* 581 (ST54) (Cuni581_peg 2901-09). This region was found similar to T4SSs identified in the genomes of *P. aeruginosa.* However, in comparison, it appeared incomplete with some components such as *TrbE*, *TrbG* and *TrbL* missing. Also, in the vicinity of this region, were found genes encoding for Vir proteins, similar to one found on the Ti plasmids of *A. tumefaciens.*

Type six secretion systems (T6SS)

Six putative T6SS clusters have been identified in the fourteen *Cronobacter* spp. genomes, some of which were previously described by Kucerova *et al*. (2010 & 2011). The diversity of the genes associated with T6SS in the *Cronobacter* spp. genomes are given in Table 7.4.

- Cluster 1 (ESA_00140–ESA_00145) encodes proteins such as DotU homologue (ESA_00140), Vgr homologue (ESA_00141), and a putative lipoprotein from the VC_A0113 family (ESA_00145). This cluster was present in all the sequenced genomes in this study, with the exception of *C. malonaticus* 681 genome.

- Cluster 2 (ESA_02035–ESA_02040) encodes proteins such as Vgr-type protein (ESA_02035), a lipoprotein from the VC_A0113 family (ESA_02038), and other genes homologous to proteins encoded in the T6SS clusters. This set of genes was present in all the *C. sakazakii* genomes except 680 (ST8) and 696 (ST12). It was also present in both *C. turicensis* genomes as well as the genomes of *C. muytjensii* 530 and *C. universalis* 581.

- Cluster 3 (ESA_02735–ESA_02740) comprises of genes that encode for SciE-type protein (ESA_02736), Vgr-type protein (ESA_02739), and a protein homologous to phage gp7 protein. This cluster was located adjacent to a phage region. It was present in all the sequenced *C. sakazakii* genomes (except 680), *C. malonaticus* 507, *C. turicensis* z3032 and both the *C. dublinensis* genomes.

- Cluster 4 (ESA_03887–ESA_03946) comprises of genes that encode for Vgr-type proteins (ESA_03905 and ESA_03917), IcmF-type protein (ESA_03945), DotU-type protein (ESA_03946), ClpV ATPase (ESA_03921), SciE-type protein (ESA_03925), Ser/Thr protein phosphatase (ESA_03927), and Ser/Thr protein kinase (ESA_03920). This cluster was found conserved in all the sequenced genomes of the *Cronobacter* genus. It was also found to be the largest T6SS associated region in the studied *Cronobacter* spp. genomes.

- Cluster 5 (ESA_pESA3p05491–ESA_pESA3p05506) was a plasmid borne T6SS gene cluster, observed on the plasmid pESA3 of the *C. sakazakii* BAA-894 genome. ESA_pESA3p05494 encodes DotU-like protein and ESA_pESA3p05495 encodes a protein with a C-terminal extension with similarity to ompA. ESA_pESA3p05497

encodes a ClpV ATPase and ESA_pESA3p05500 encodes a Vgr-like protein. This was found present only in the genomes of *C. sakazakii* BAA-894, 680, 696 and *C. dublinensis* 582.

- Cluster 6 (Ctu_12090-Ctu_12210) was identified on the genome of *C. turicensis* z3032 and showed considerable variations across the genus. It was found only in the genomes of *C. sakazakii* 696 and 701, *C. malonaticus* 507, *C. turicensis* z3032, *C. dublinensis* 582 & 1210. This cluster encoded for the VasABCDEF proteins as well an IcmF related protein.

C. sakazakii 701 (ST4) contains unique remnants of a T6SS with a lone Vgr protein coding gene and Imp protein coding gene. Another T6SS associated gene that was located separately from the main T6SS gene cluster, was *vgrG*, encoding a lipoprotein (ESA_00292-4). It was present in all the *Cronobacter* spp. strains except for the genome of *C. malonaticus* 507 and of *C. muytjensii* 530.

Cluster	Loci	*C.sakazakii* BAA-894	*C. sakazakii* 680	*C. sakazakii* 701	*C. sakazakii* E899	*C. sakazakii* 696	*C. malonaticus* 507	*C. malonaticus* 681	*C. turicensis* 564	*C. turicensis* z3032	*C. universalis* 581	*C. muytjensii* 530	*C. dublinensis* 582	*C. dublinensis* 1210	*C. condimenti* 1330
1	ESA_00140–ESA_00145	+	+	+	+	+	+	-	+	+	+	+	+	+	+
2	ESA_02035–ESA_02040	+	-	+	+	-	-	-	+	+	+	+	-	-	-
3	ESA_02735–ESA_02740	+	-	+	+	+	+	-	-	+	-	-	+	+	-
4	ESA_03887–ESA_03946	+	+	+	+	+	+	+	+	+	+	+	+	+	+
5	ESA_pESA3p05491– ESA_pESA3p05506	+	+	-	-	+	-	-	-	-	-	-	+	-	-
6	CTU_12090-CTU_12210	-	-	+	-	+	+	-	-	+	-	-	+	+	-

Table 7.4 Diversity of the T6SSs in the sequenced genomes of the *Cronobacter* spp. Published in Joseph *et al.* (2012c).

7.2.2.2 Fimbriae

A total of twelve putative fimbriae gene clusters were identified and are summarised in Table 7.5. The functionality of these regions are yet to be investigated.

- Cluster 1 (ESA_01970-76) consists of genes encoding for Type I fimbrial proteins. These genes were found present only in the genome of *C. sakazakii* BAA-894 (ST1). Previously, Kucerova *et al*. (2010) had also identified this cluster in *C. sakazakii* strain 2 (ST3) in the CGH study. However, a representative of this ST has not been sequenced in this study.

- Cluster 2 (ESA_02342-45) was found conserved in all the genomes of the *Cronobacter* genus. These are also genes encoding Type I fimbrial proteins.

- Most of the cluster 3 (ESA_02538-42) was found conserved across the genus. The genes ESA_02541 & ESA_02542 encoding for a fimbrial adhesin precursor and type I fimbriae adaptor subunit (*fimG*) was found missing in the genomes of *C. muytjensii* 530, *C. dublinensis* 582 and 1210, and *C. condimenti* 1330.

- Cluster 4 (ESA_02795-99) encoding for Type I fimbrial proteins was found present only in the genomes of *C. sakazakii* and *C. malonaticus.*

- Cluster 5 (ESA_03512-20) is a set of genes encoding for the beta-fimbrial subunit proteins. These were found to be unique to the genomes of *C. sakazakii*.

- Cluster 6 (ESA_04070-73) was found present in all the sequenced genomes of the genus *Cronobacter,* also found encoding for Type I fimbrial proteins.

- Cluster 7 (ESA_03812-15) was found in all the sequenced *Cronobacter* genomes except *C. muytjensii* 530. These genes were found to encode for the pili assembly proteins.

- Cluster 8 (CTU_36390-450) was found present in all the sequenced genomes except *C. sakazakii* BAA-894 and E899. It was only partially present in the genomes of *C. universalis* 581, *C. dublinensis* 1210 and *C. condimenti* 1330.

- Cluster 9 (ESA_03231-33) are genes encoding for the type IV fimbrial proteins linked to the gliding motility functions of the organism. This cluster was found conserved in all the sequenced *Cronobacter* spp. genomes.

- Cluster 10 (CTU_16160-230) is a group of genes involved in encoding proteins related to curli fimbriae (CsgABCDEFG). These genes were absent in all the *C. sakazakii* genomes, and present only in a select few genomes – *C. malonaticus* 507 and 681, *C. turicensis* z3032, *C. universalis* 581, *C. dublinensis* 1210 and *C. condimenti* 1330.

- Cluster 11 (1210_peg 1323-1330) is a cluster of genes encoding for type I fimbrial proteins. This group was found only in the genomes of *C. dublinensis* 1210 and *C. condimenti* 1330. These genes showed a very high identity to a fimbrial cluster found in *Escherichia hermannii*.

- Cluster 12 (ESA_01343-47) comprises a group of genes encoding sigma fimbrial proteins. These genes were found conserved across the genomes of the *Cronobacter* genus.

Cluster	Loci	*C.sakazakii* BAA-894	*C. sakazakii* 680	*C. sakazakii* 701	*C. sakazakii* E899	*C. sakazakii* 696	*C. malonaticus* 507	*C. malonaticus* 681	*C. turicensis* 564	*C. turicensis* z3032	*C. universalis* 581	*C. muytjensii* 530	*C. dublinensis* 582	*C. dublinensis* 1210	*C. condimenti* 1330
1	ESA_01970-76	+	-	-	-	-	-	-	-	-	-	-	-	-	-
2	ESA_02342-45	+	+	+	+	+	+	+	+	+	+	+	+	+	+
3	ESA_02538-42	+	+	+	+	+	+	+	+	+	+	+	+	+	+
4	ESA_02795-99	+	+	+	+	+	+	+	-	-	-	-	-	-	-
5	ESA_03512-20	+	+	+	+	+	-	-	-	-	-	-	-	-	-
6	ESA_04070-73	+	+	+	+	+	+	+	+	+	+	+	+	+	+
7	ESA_03812-15	+	+	+	+	+	+	+	+	+	+	-	+	+	+
8	CTU_36390-450	-	+	+	-	+	+	+	+	+	+	+	+	+	+
9	ESA_03231-33	+	+	+	+	+	+	+	+	+	+	+	+	+	+
10	CTU_16160-230	-	-	-	-	-	+	+	-	+	+	-	-	+	+
11	1210_peg 1323-30	-	-	-	-	-	-	-	-	-	-	-	-	+	+
12	ESA_01343-47	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 7.5 Fimbrial gene clusters identified across the genomes of the *Cronobacter* genus. Published in Joseph *et al.* (2012c).

7.2.2.3 Sialic Acid utilisation

Distribution of genes

One of the most interesting gene clusters, unique to the genomes of *C. sakazakii* is the gene cluster ESA_03609–13 on the genome of *C. sakazakii* BAA-894. This cluster encodes for the proteins involved in the uptake and utilization of exogenous sialic acid. ESA_03609 encodes a putative sugar isomerase (YhcH). Genes ESA_03610-13 encode the *nanKTAR* genes involved in the N-acetylneuraminate and N-acetylmannosamine degradation pathway. The *nanK* gene (ESA_03610) encodes N-acetylmannosamine kinase; *nanT* (ESA_03611) encodes the sialic acid permease transporter; *nanA* (ESA_03612) encodes N-acetylneuraminate lyase, and *nanR* (ESA_03613) is a transcriptional regulator from the GntR family. The *nanE* locus encoding the enzyme N-acetylmannosamine-6-phosphate-2epimerase was located separate from this cluster at ESA_00529, and unlike the rest was found conserved across the genomes of the *Cronobacter* genus.

Another locus involved in this metabolic cycle was located at ESA_03302, encoding NanC – the N-acetylneuraminic acid outer membrane channel protein. This was found to be unique to the genomes of *C. sakazakii*. Interestingly, there was also another locus, ESA_00448, which was incorrectly annotated as *nanC* by RAST, and was found conserved across the genus. When investigated further, it was identified to be an oligogalacturonate porin, belonging to the same KdgM protein superfamily as NanC. Two other loci also essential to the sialic acid metabolism cycle, were ESA_02661 (*nagB*) and ESA_02662 (*nagA*). They were found conserved across the genomes of the *Cronobacter* genus and are located adjacent to each other in all the genomes.

Some bacteria such as *E. coli* have the ability to synthesize their own sialic acid by the combined action of the genes *neuC* (UDP-N-acetylglucosamine 2-epimerase) and *neuB* (sialic acid synthase). Of these, only the former was found to be present in all the genomes of the *Cronobacter* spp. genomes (ESA_03772). The gene *nanH* that encodes for the enzyme sialidase or neuraminidase is responsible for the cleavage of sialic acid from its bound forms and was found to be absent in all the *Cronobacter* spp. genomes. Other sialic acid related genes include *neuA* (encodes CMP-Neu5Ac synthetases in *E. coli*), *neuS* (polysialyltransferase), *neuO* (Polysialic acid O-acetyltransferase) and *neuD* (sialic acid O-acetyltransferase), which encode proteins responsible for the glycosylation of cell surface structures such as capsular polysaccharide and lipopolysaccharide in organisms such as *E. coli.* These genes were all absent in the genomes of the *Cronobacter* genus.

Interestingly, apart from the sialic acid permease transporter gene *nanT* mentioned earlier, all the *Cronobacter* spp. genomes also encoded the genes for the tripartite ATP-

independent periplasmic (TRAP) transporter, SiaPQM, (ESA_04264-4266) which has mainly been reported in organisms such as *H. influenzae*.

Another gene of possible related interest was a homologue of a stringent starvation protein, *sspA* (ESA_03615), located in the vicinity of the *nanKTAR* gene cluster.

Fimbrial gene cluster 5 (ESA_03512-3520) encoding for beta-fimbrial subunit proteins was also of interest because of their proximity to the sialic acid utilisation gene cluster as well as their uniqueness to *C. sakazakii* (Table 7.5).

The diversity of these genes present across the genomes of the *Cronobacter* genus has been tabulated in Table 7.6.

Gene	Loci	Function	*C.sakazakii* BAA-894	*C. sakazakii* 680	*C. sakazakii* 701	*C. sakazakii* E899	*C. sakazakii* 696	*C. malonaticus* 507	*C. malonaticus* 681	*C. turicensis* 564	*C. turicensis* z3032	*C. universalis* 581	*C. muytjensii* 530	*C. dublinensis* 582	*C. dublinensis* 1210	*C. condimenti* 1330
nanK	ESA_03610	N-acetylmannosamine kinase	+	+	+	+	+	-	-	-	-	-	-	-	-	-
nanT	ESA_03611	Sialic acid transporter (permease) NanT	+	+	+	+	+	-	-	-	-	-	-	-	-	-
nanA	ESA_03612	N-acetylneuraminate lyase	+	+	+	+	+	-	-	-	-	-	-	-	-	-
nanR	ESA_03613	Transcriptional regulator NanR	+	+	+	+	+	-	-	-	-	-	-	-	-	-
nanE	ESA_00529	N-acetylmannosamine-6-phosphate 2-epimerase	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nanC	ESA_03302	N-acetylneuraminic acid outer	+	+	+	+	+	-	-	-	-	-	-	-	-	-

		membrane channel protein													
yhcH	ESA_03609	Putative sugar isomerase	+	+	+	+	+	-	-	-	-	-	-	-	-	-
nagA	ESA_02662	N-acetylglucosamine-6-phosphate deacetylase	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nagB	ESA_02661	Glucosamine-6-phosphate deaminase	+	+	+	+	+	+	+	+	+	+	+	+	+	+
neuC	ESA_03772	UDP-N-acetylglucosamine 2-epimerase	+	+	+	+	+	+	+	+	+	+	+	+	+	+
siaPQM	ESA_04264-4266	tripartite ATP-independent periplasmic (TRAP) transporter	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sspA	ESA_03615	Stringent starvation protein A	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 7.6 Distribution of the sialic acid utilisation and other related genes across the sequenced genomes of the *Cronobacter* genus. The genes *nanH, neuA, neuB, neuD, neuO* and *neuS* were found to be absent in all the *Cronobacter* spp. genomes analysed in this study. Published in Joseph *et al.* (2013b).

%GC Content

An analysis of the %GC content of each of these genes was also carried out to check for influences of horizontal gene transfer events. The *nanA*, *nanR* and *nanT* genes of the *C. sakazakii* genomes showed average GC% values of 57.2%, 56.32% and 57.14% respectively, reflective of the ~56% GC content of the *Cronobacter* genomes. The *nagA* and *nagB* genes too showed GC% values of 56.3% and 54% respectively. On the other hand, the genes *nanE* and *nanK* showed slightly higher values of 63% and 62.2% respectively.

The *nanC* gene interestingly showed a much lower %GC content value of 47.4 %. This was also seen for the nucleotide sequences of the *nanC* gene in the other closely related organisms such as *Cit. koseri* (48%), *E. cloacae* (51%), *Enterobacter* sp. (44.54%), *E. coli* K1 (43%). The genomes of these organims reflected a %GC content of 55-58%.

Phylogenetic analysis

The uniqueness of the core sialic acid-related gene cluster to the *C. sakazakii* genomes hints at a role in the evolution of the virulence of the organism. The predicted amino acid sequences of the proteins encoded by these *nan* cluster genes were individually analysed and their phylogenetic relationships observed with closely related Gram-negative bacteria have been indicated in Fig. 7.5 to 7.12. In the case of each of the genes, the *C. sakazakii* sequences formed an independent cluster of their own, with the other *Enterobacteriaceae* members clustering on the neighbouring branches.

In the NanA (Fig. 7.5) and NanR (Fig. 7.6) phylogenetic trees, the *C. sakazakii* cluster appeared to evolve on the same branch as *E. cloacae* and *Enterobacter* spp., with the others forming a separate clade. In comparison, the NanK (Fig. 7.7) and NanT (Fig. 7.8) *C. sakazakii* clusters appears to have greater phylogenetic distance from the other enteric members, with a clear split of the population into two clades, one of them being that of the *C. sakazakii* cluster.

The *nanE* gene was found across the *Cronobacter* genus, and the phylogenetic analysis of the NanE protein sequences (Fig. 7.9) revealed the *Cronobacter* cluster to have a common and closely related evolutionary clade with *E. cloacae, E. hormaechei, Enterobacter* spp., *C. freundii* and *Pantoea agglomerans.*

The NanC protein in the *C. sakazakii* genomes could be located with >50% homology only in the genomes of *E. cloacae, E. hormaechei, C. koseri* and *E. coli* K1. Of these, the *E. coli* K1 NanC appeared to very distantly related to the rest of the population studied (Fig. 7.10).

The *nagA* and *nagB* genes were both found in the genomes of all the *Cronobacter* species. Phylogenetic analysis of these proteins showed the *Cronobacter* spp. proteins to form a distinct clade, with other *Enterobacteriaceae* members forming a neighbouring clade, both with

a common evolutionary lineage. In both the trees, the *E. tarda* and *P. agglomerans* proteins formed outliers. It is also important to note that newly identified species *C. condimenti* always branched within the *Cronobacter* genus cluster (Fig. 7.11 and 7.12).

Interestingly, within the trees of the Nan protein clusters, *Cit. koseri* and *Cit. freundii* exhibited different patterns in their branching, suggesting independent evolutionary paths for the *nan* genes of these species within the *Citrobacter* genus.

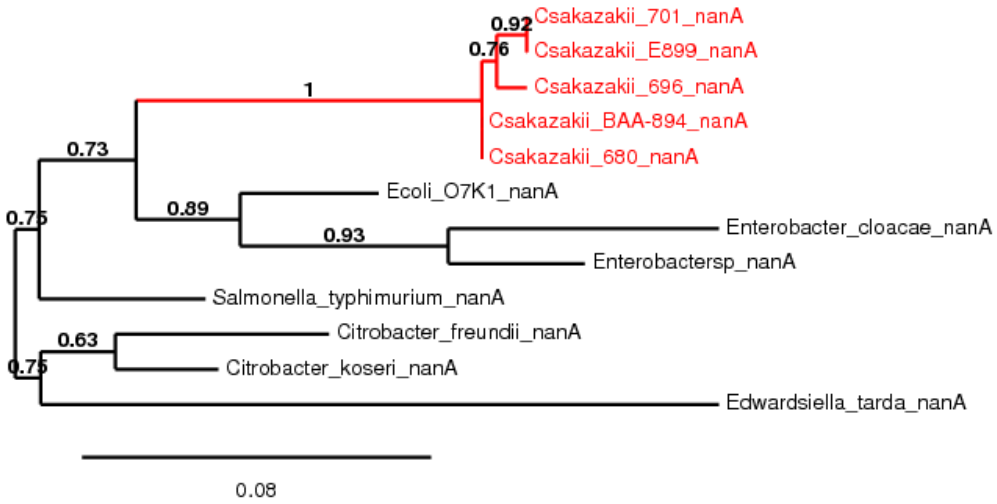


Fig. 7.5 Maximum-likelihood tree of the NanA protein sequences (292 aa) of *C. sakazakii* and related *Enterobacteriaceae* members, constructed using PhyML, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values. Published in Joseph *et al.* (2013b).

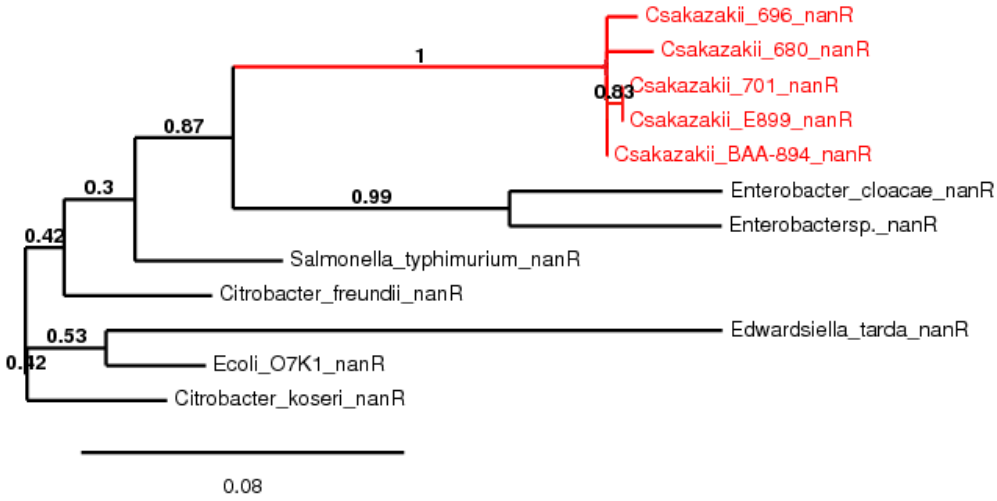


Fig. 7.6 Maximum-likelihood tree of the NanR protein sequences (260 aa) of *C. sakazakii* and related *Enterobacteriaceae* members, constructed using PhyML, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values. Published in Joseph *et al.* (2013b).

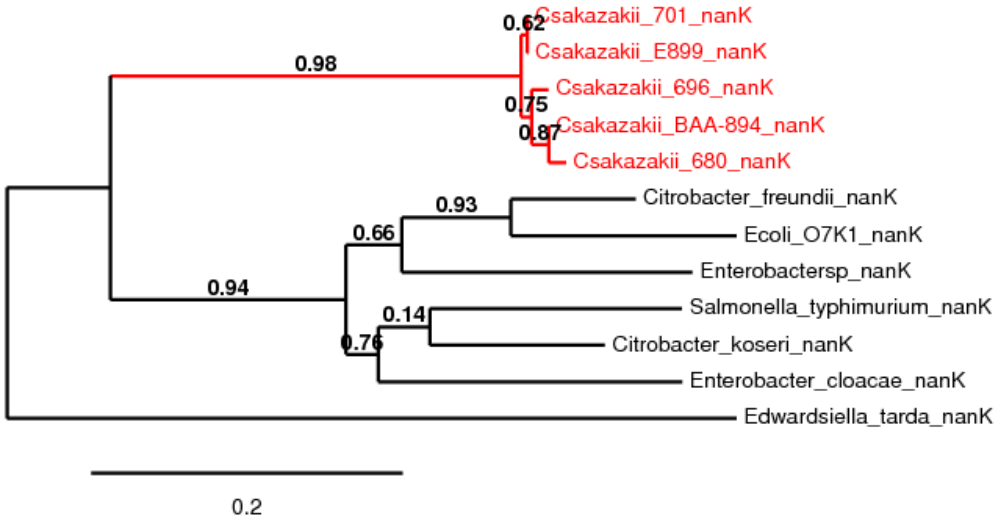


Fig. 7.7 Maximum-likelihood tree of the NanK protein sequences (291 aa) of *C. sakazakii* and related *Enterobacteriaceae* members, constructed using PhyML, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values. Published in Joseph *et al.* (2013b).

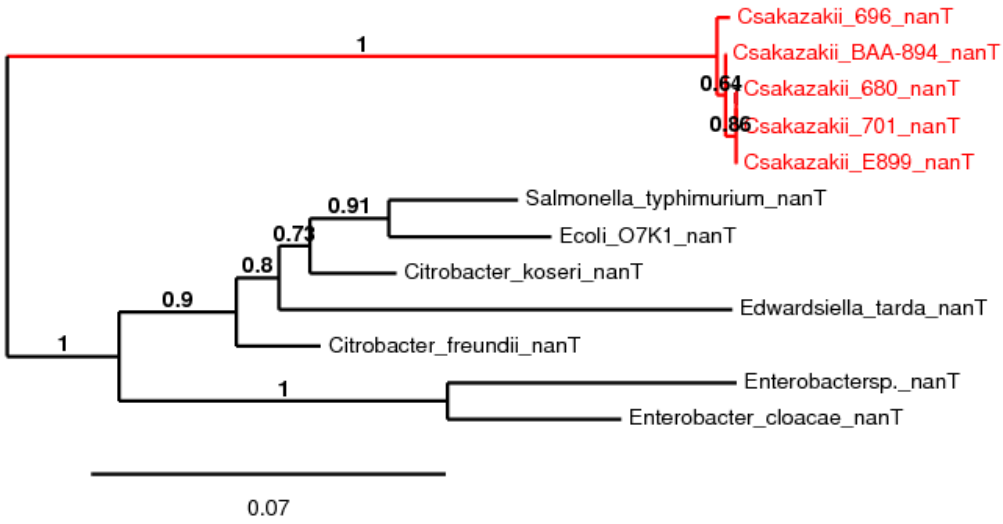


Fig. 7.8 Maximum-likelihood tree of the NanT protein sequences (496 aa) of *C. sakazakii* and related *Enterobacteriaceae* members, constructed using PhyML, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values. Published in Joseph *et al.* (2013b).

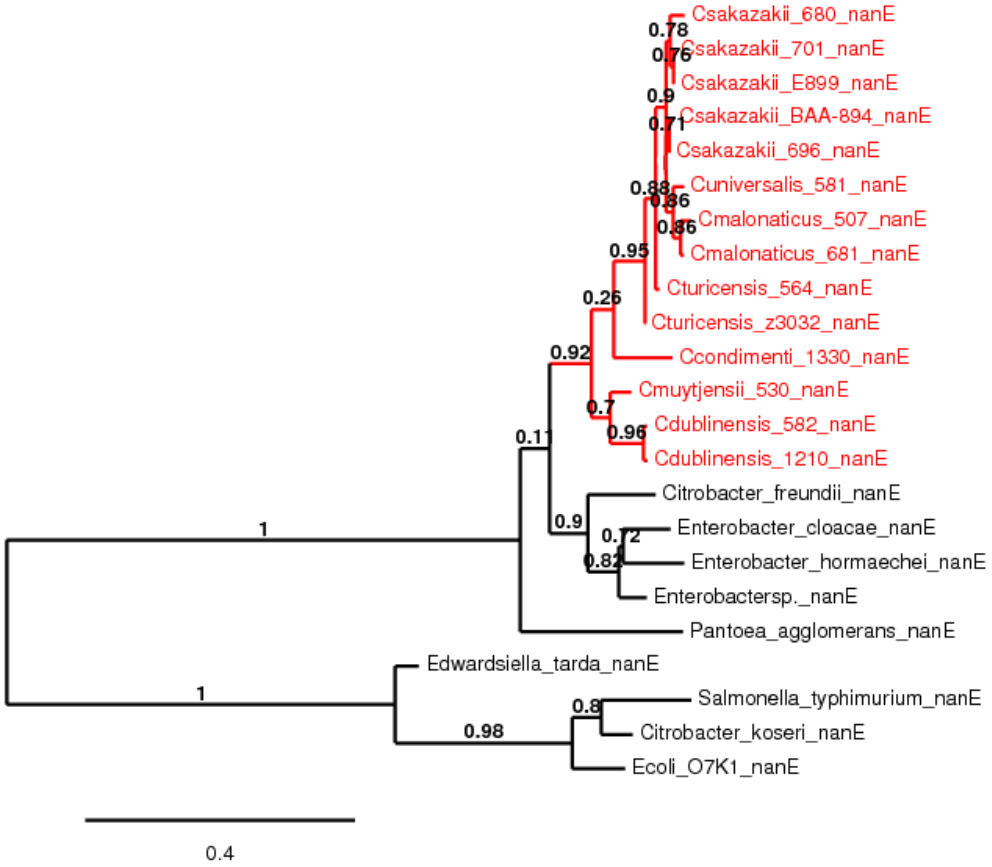


Fig. 7.9 Maximum-likelihood tree of the NanE protein sequences (229 aa) of *Cronobacter* spp. and related *Enterobacteriaceae* members, constructed using PhyML, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values. Published in Joseph *et al.* (2013b).

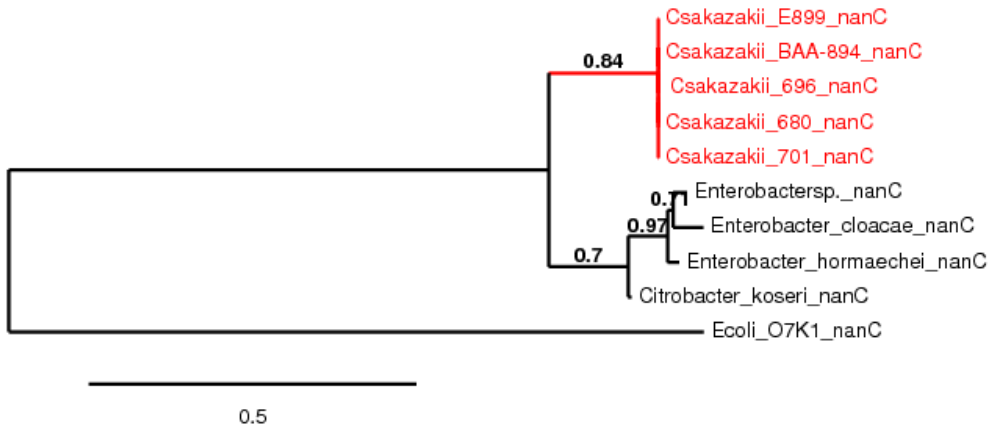


Fig. 7.10 Maximum-likelihood tree of the NanC protein sequences (202 aa) of *C. sakazakii* and related *Enterobacteriaceae* members, constructed using PhyML, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values. Published in Joseph *et al.* (2013b).

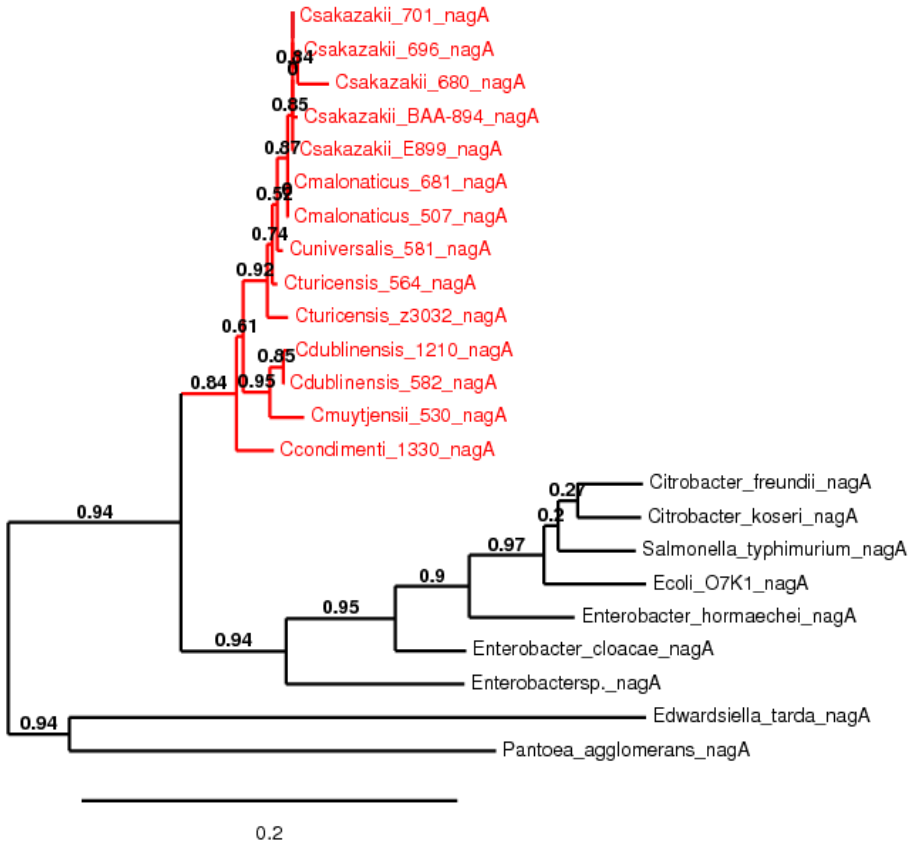


Fig. 7.11 Maximum-likelihood tree of the NagA protein sequences (382 aa) of *Cronobacter* spp. and related *Enterobacteriaceae* members, constructed using PhyML, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values. Published in Joseph *et al.* (2013b).

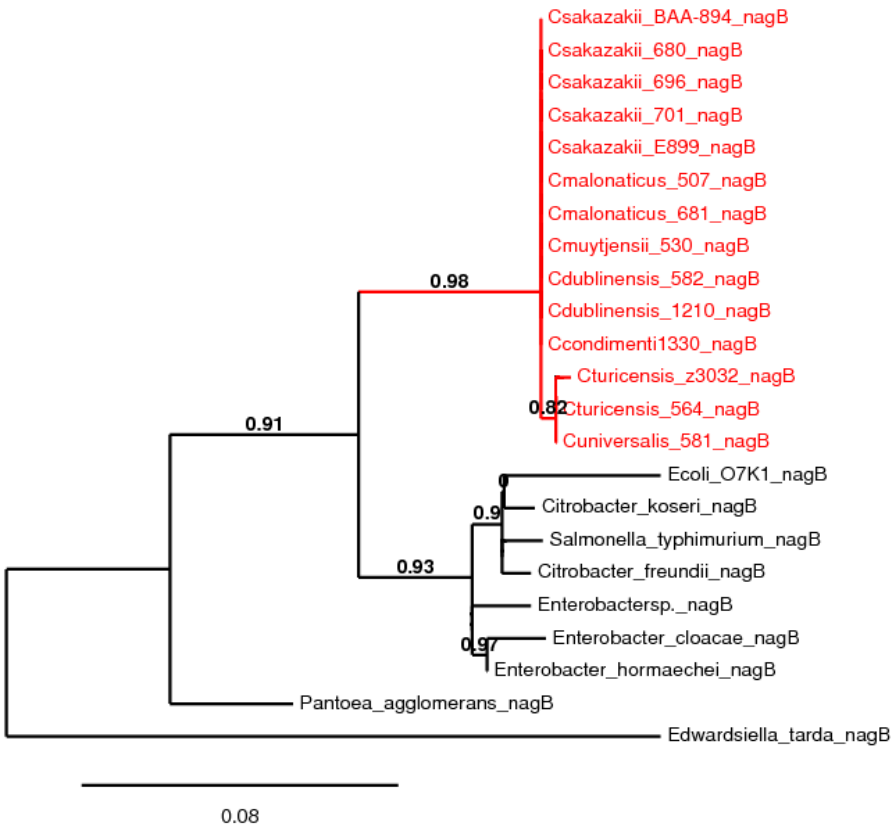


Fig. 7.12 Maximum-likelihood tree of the NagB protein sequences (266 aa) of *Cronobacter* spp. and related *Enterobacteriaceae* members, constructed using PhyML, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values. Published in Joseph *et al.* (2013b).

7.2.2.4 Iron acquisition and transport genes

Like most bacteria, *Cronobacter* spp. too exhibited a number of regions in the genomes linked to the acquisition of iron. Some of these were previously also reported by Kucerova *et al*. (2011) in the genome of *C. sakazakii* BAA-894, and have recently been characterized in detail by Grim *et al*. (2012).

- Nearly all the *Cronobacter* spp. strains examined possessed complete operons for the biosynthesis, receptor and transport of an enterobactin-like siderophore distributed across three locations on the genome: *entABCDEFS* (ESA_00791-800); *fepABCDEG* (ESA_02727-31); *fepE* (ESA_00459). The exceptions were the genomes of *C. sakazakii* strains 701 and 696, which lacked the homologue for EntS, an enterobacterin exporter. The lone *fepE* gene encodes for a ferric enterobactin uptake protein, and was absent in the genomes of *C. dublinensis* and *C. condimenti*.

- All *Cronobacter* species possess a complete plasmid-borne operon for aerobactin synthesis (*iucABCD*) and its receptor *iutA* (ESA_pESA3p05547-51). This region was found absent in the publicly available genome of *C. sakazakii* E899, which appears to lack the plasmids. This aerobactin is now referred to by the term cronobactin (Grim *et al*. 2012).

- Additionally, a gene cluster for hydroxamate-type siderophores (*fhuABCDE*, ESA_03187-90 & ESA_02242) was encoded in all the *Cronobacter* species. Of these, ESA_03187-3189 comprised the genes *fhuBCD* encoding for the ferric hydroxamate ABC transporter proteins, while ESA_03190 encodes the outer membrane receptor protein FhuA. The gene ESA_02242 encoded the receptor precursor protein FhuE.

- As previously referred to in Chapter 6, *C. sakazakii* 680 (ST8) encoded a unique siderophore-related cluster (*fecRABCDE*) which is probably plasmid-borne, at the loci Csak680_peg 2743-2748. The FecA is a Ton-B dependent receptor while the FecBCDE comprise the transport proteins. FecR is a transmembrane sensor protein.

- A unique ABC transport system for the iron B12 siderophore hemin was unique to the genome of *C. turicensis* 564 at the loci Ctur564_peg 3252-3254.

- Ferrous iron transport proteins EfeB and EfeU were encoded at the loci ESA_02350 and ESA_02352. These genes were found to be conserved across the genomes of the *Cronobacter* genus.

- Some other genes linked to the iron acquisition and transport, conserved in the all the sequenced *Cronobacter* spp. genomes were ESA_01885 encoding the siderophore receptor YncD, ESA_03800 encoding the Vitamin B12/cobalamine outer membrane transporter BtuB, and ESA_02242 encoding the ferric-rhodotorulic acid receptor protein FhuE.

- A TonB-dependent ferrichrome receptor FcuA encoding gene (Cdub1210_peg 430) was found unique to the genomes of *C. muytjensii*, *C. dublinensis* and *C. condimenti*. This was also previously identified by Grim *et al*. (2012). Another such TonB-dependent receptor encoding gene (Ccondi1330_peg 2216) for ferrienterocholins was also found unique to the genome of *C. condimenti* 1330.

7.2.2.5 Flagellar gene clusters

Various clusters of genes were found in the *Cronobacter* genomes identified to be involved in the assembly of flagellar proteins of the bacterium.

One cluster of flagellar genes was located at ESA_01248-01261, encoding the proteins FliE-FliR. These proteins are associated with flagellar biosynthesis and motor-switch function. This entire cluster was found missing in the genome of *C. sakazakii* 680.

A large cluster located at ESA_02264-02277 was found in all the genomes, identified as the genes *flgABCDEFGHIJKL*. The proteins encoded by these genes are associated with the hook and basal body of the flagellum.

A third cluster was located at the loci ESA_01337-01340. Of these, the genes ESA_01337 and 01338 encode for flagellar transcriptional activators FlhC and FlhD. The genes ESA_01339 and 01340 encode for the flagellar motor proteins MotA and MotB.

7.2.2.6 Macrophage survival

A gene (ESA_03843) encoding for the superoxide dismutase protein SodA was found conserved in all the *Cronobacter* genomes. One macrophage infectivity potentiator-related protein encoding gene (ESA_01868) was found in all the *C. sakazakii*, *C. malonaticus*, *C.*

turicensis, *C. muytjensii* and *C. condimenti* strains but was absent from *C. universalis* and *C. dublinensis*. This requires further analysis but does correlate with the absence of *C. universalis* and *C. dublinensis* in clinical isolates.

7.2.2.7 Other important virulence associated regions

The genes coding for the outer membrane proteins OmpA and OmpX were found to be conserved across the *Cronobacter* genus. The same was true for the gene (ESA_00752) encoding for a metalloprotease (*zpx*) which was found in all genomes irrespective of source and species.

The *Cronobacter* plasminogen activator gene *cpa* identified on the pESA3 plasmid of *C. sakazakii* BAA-894 (ESA_pESA3p05434) was only found in genomes of *C. sakazakii* strains 696 and 701, and was absent from 680. It was also absent from the genomes of all other *Cronobacter* species with the exception of *C. universalis*.

7.3 Discussion

Previously, Kucerova *et al*. (2010 & 2011) had characterized a number of important regions on the genome of *C. sakazakii* BAA-894, and compared them with four other *Cronobacter* species using a whole-genome microarray study. Through this current study, those results have been further elaborated and expanded to analyse the key genomic regions in all seven species of the *Cronobacter* genus, including the newly characterized *C. universalis* and *C. condimenti*.

7.3.1 Genomic regions related to *Cronobacter* spp. physiology

A number of stress response genes related to osmotic stress adaptation, desiccation survival and carbon starvation were investigated in the *Cronobacter* genus and were found to be conserved across the *Cronobacter* spp. genomes (Section 7.2.1.1). This is not surprising considering the isolation and epidemiological background of *Cronobacter* spp. Isolates of all *Cronobacter* species have been found to be ubiquitous and persistent in a wide range of sources including herbs, spices, dried foods as well as powdered infant formula (Iversen and Forsythe 2004). The desiccation and stress resistant properties can help the organism in its persistence in such extreme conditions. Previously, a study by Caubilla-Barron and Forsythe (2007) had shown how *Cronobacter* spp. strains could be recovered from desiccated conditions even after a period of 2.5 years.

Another aspect related to the thermal resistance of *Cronobacter* spp. was investigated on the basis of a gene cluster identified by Gajdosova *et al*. (2011). When compared to the genomes sequenced in this study, this region was found present only in the genomes of *C. sakazakii* 696 (ST12) and 701 (ST4), and the two *C. malonaticus* genomes. The cluster itself was identified and characterized by Gajdosova *et al*. (2011) in the *C. sakazakii* type strain ATCC 29544 (ST8). Therefore, this region could be unique to a specific lineage of *Cronobacter* spp. strains. The fact that the characterized region was flanked by transposases suggests the possibility of horizontal transfer events having resulted in this region in some *Cronobacter* spp. strains. Thermal resistance is an important feature for *Cronobacter* spp., as it can play an important role in the survival of the organism in factory production environments, and thereby its persistence in substances such as infant formula and milk powder among others. Unfortunately, this region is poorly annotated with a number of hypothetical uncharacterized proteins and the association of these results with thermal resistance is not supported by experimental data in these strains conducted by our research group (Unpublished data; personal communication).

Capsular polysaccharides on the bacterial cell surface can have a critical role in the physiology of an organism, not just from the aspect of interaction with the environment but also from the point of view of pathogenicity. Capsular genes encoding the enzymes responsible for the production and transport of these polysaccharides were found clustered in a large *wca* operon, in all the genomes of the *Cronobacter* genus (Section 7.2.1.2). A similar *wca* operon has previously been characterized in *E. coli* K-12 by mutation studies (Stevenson *et al*. 1996). This gene cluster controls the biosynthesis of the exopolysaccharide colanic acid. Colanic acid is a hexasaccharide containing fucose and glucuronic acid which can form a thick extracellular mucoid matrix contributing to the resistance of bacteria to environmental stresses, such as desiccation, acid and heat treatment, and also contributes to biofilm formation. Highly mucoid colonies on agar plates have often been observed in a number of *Cronobacter* spp. strains, formed as a result of the production of these polysaccharides. *Cronobacter* species have also been reported to form biofilms on inert surfaces including infant feeding equipment and nasogastric feeding tubes (Caubilla-Barron *et al*. 2007; Hurrell *et al*. 2009a, b). Large scale production of *Cronobacter* spp. capsular material has been studied in a strain isolated from Chinese tea and has been patented as a food thickener (Scheepe-Leberkuhne & Wagner, 1986, Harris & Oriel 1989). This strain was found to produce a heteropolysaccharide mainly composed of the sugars D-glucose, D-galactose, D-fucose, D-mannose and glucuronic acid. The colonic acid biosynthesis operon was always found located adjacent to the gene cluster involved in the biosynthesis of the cell wall O-PS (Section 7.2.1.5).

Some of these above mentioned properties of *Cronobacter* spp., such as desiccation resistance and polysaccharide production were also used as a basis of the hypothesis that plants may be a natural habitat for the organism (Iversen and Forsythe 2003; Schmid *et al*. 2009), protecting the bacterium in harsh weather conditions and seasonal variations in their environment. Another trait linked to this hypothesis is the yellow pigmentation of a large majority of *Cronobacter* spp. strains when cultured on TSA agar. It is believed that this pigmentation must serve as a shield for the organism against oxygen radicals generated by the UV radiations of sunlight. The plant-based origin of *Cronobacter* spp. also explains the high frequency of isolation of the organism from sources such as herbs and spices, as well as powdered infant formula (based on the sources of its ingredients, especially starch). Very recently, a similar link based on traits such as biofilm formation and polysaccharide formation was found to play a role in the evolution of plant colonization in certain phylogroups of *E. coli* (Meric *et al*. 2012).

Metabolism of different kinds of sugars is a core physiological function in bacteria. Some of the interesting genomic regions related to this were looked at in the *Cronobacter* spp. genomes (Table 7.1). Genes encoding for α-glucosidase and maltose uptake were found

conserved across the *Cronobacter* spp. genomes. These traits along with osmotic pressure tolerance, have been used as the basis for a number of commercially available enrichment broths and chromogenic agars for the isolation of *Cronobacter* spp. One such example is the Druggan-Forsythe-Iversen (DFI) agar, which uses 5-bromo-4-chloro-3-indolyl-a,D-glucopyranoside (XaGlc) as an indicator for the α-glucosidase reaction, phenotypically observed by the formation of blue-green colonies on the agar (Iversen at al. 2004).

However, the genes required for malonate metabolism - malonyl CoA acyl carrier protein transacylase, phosphoribosyl-dephospho-CoA transferase, and malonate decarboxylase (Ctu_34990-35070) - are found in all *Cronobacter* species except *C. sakazakii* and *C. dublinensis* strain 582. Malonate is also found in plants and *C. malonaticus* was initially a subspecies within *C. sakazakii* that was later differentiated as an independent species by the malonate utilisation test (Iversen *et al*. 2007). The presence of this gene in *Cronobacter* species except *C. sakazakii* suggests possibilities of false positive results based on expression and hence this test cannot be used solely for the purpose of identifying *C. malonaticus*. Apart from malonate, genes encoding for some other species differentiating traits such as putrescine and inositol utilization were found conserved across the genus. However, according to the study by Iversen *et al*. (2007), putrescine utilization was found to be a variable trait in *C. malonaticus, C. universalis* and *C. dublinensis.* They also reported inositol utilization to be variable in *C. sakazakii* and *C. malonaticus*, and absent in *C. dublinensis* subsp. *lausannensis.* This further raises the concern about the high level of subjectivity involved in the use of these phenotypic tests in the bacterial speciation process, also discussed in Chapter 4.

Tellurite (TeO_3^{-2}) is a water soluble anion of tellurium (Te). It is believed that increased tellurite levels can be toxic to bacterial cells, interfering with their enzymatic activities. However, this is hardly a cause for concern as tellurite levels in the environment are very low, with minimal chances of bacteria, especially human pathogens being exposed to it. Different forms of tellurite resistance gene clusters have been identified in pathogenic bacteria, especially on conjugative plasmid regions of some *Enterobacteriaceae* members such as *E. coli* and *K. pneumoniae* (Taylor 1999; Vavrova *et al*. 2006). One such gene cluster (*terACDYZ*) was found unique only to the genome of *C. sakazakii* BAA-894 (ST1) at the loci ESA_01775–ESA_01804 (Table 7.2). As the gene cluster was absent from all other *Cronobacter* strains studied, the reference strain BAA-894 or the ST1 lineage probably acquired the tellurite resistance cluster recently, as has also been suggested by Kucerova *et al*. (2010). The flanking regions included transposases, further hinting at the possibility of a horizontal transfer event.

Copper is essential for bacterial growth, being required for many key enzymes, however like most other elements it is toxic in excess, disrupting cellular functions. Many pathogenic bacteria have exhibited copper resistance operons in their genomes. Some of the *Cronobacter*

spp. genomes too showed the presence of a chromosome based *cusRCFBA* operon, and a plasmid-borne *pcoABCDR* gene cluster (Table 7.2). In *E. coli*, the plasmid-borne *pco* operon has been identified to be the chief determinant for copper resistance regulated by a chromosomal two-component system *cusRS* (Huang *et al*. 1999; Munson *et al*. 2000). The genes *cusRS* were also found to regulate the chromosomal protein CusC in *E. coli.* This protein is of special interest as studies have shown it to be an allelic variant of the *E. coli* K1 gene *ibeB,* implicated in the brain invasive capacity of the organism. This was reported in relation to *Cronobacter* spp. too by Kucerova *et al*. (2010) in the micro-array based study of the genome *C. sakazakii* BAA-894. However, further laboratory experiments need to be performed to confirm this association.

Even though the copper/silver resistance operons as well as tellurite resistance operons appeared to have a common evolutionary association with the R478 plasmids, they were found in distant chromosomal locations from each other. Therefore, these regions do not necessarily appear to be the result of a single plasmid integration event.

The gene clusters related to O-antigen synthesis were found to be one of the most variable regions in the genomes of the *Cronobacter* genus (Section 7.2.1.5). The O-antigen biosynthesis genes were located between the *galF* and *gnd* genes, similar to *E. coli* and *S. enterica* (Samuel and Reeves 2003). These were the two genes of this region that were found conserved across the genus whereas the rest of the genes from the O-antigen locus are highly divergent and were not detected previously by microarray hybridization either (Kucerova *et al*. 2010). Hence, these genes have been used to design a RFLP-based PCR study to elucidate a serotyping scheme for the *Cronobacter* genus (Jarvis *et al*. 2011). A serotyping scheme is not just a typing tool but also provides valuable information about bacterial cell surfaces, which could be linked to the organism's pathogenicity. However, the current PCR-based serotyping assays have found significant limitations because of the gene variations in this region. As a result, some of the defined serotypes have been found in more than one species, leading to false positive results (Sun *et al*. 2011 & 2012; Liu *et al*. 2011; Arbatsky *et al*. 2012). In this genome analysis itself, it has been found that the O-antigen gene cluster of the *C. malonaticus* 507 genome exactly matches the gene cluster characterized previously as *C. sakazakii* serotype O:5 (Sun *et al*. 2012).

The chemical structure of the LPS of a number of *Cronobacter* species have been investigated in detail to reveal a large amount of variation reflective of the diversity and lack of sequence conservation observed in the gene clusters in the genomic comparisons. In the *C. sakazakii* BAA-894 strain (ST1), the LPS was found to be a branched polymer of pentasaccharide units,while *C. sakazakii* strain 767 (ST4) is also a branched polymer but of repeating heptasaccharide units, with variations in the sugar linkages in both the strains

(MacLean *et al*. 2009a; Czerwicka *et al*. 2010). In the *C. malonaticus* strain 3267, the LPS was also a branched pentasaccharide unit (MacLean *et al*. 2009b). On the other hand, the LPS structure elucidated for *C. muytjensii* strain 3270 by MacLean *et al*. (2009c) was found to be a linear unbranched pentasaccharide polymer. Recently, Arbatsky *et al*. (2012) studied the LPS structure of *C. sakazakii* strain G2726 and found it be highly similar to the LPS structure observed in the *C. muytjensii* strain 3270. The specific genes associated with some of these differences in the enzymology of the LPS structure in the *Cronobacter* genus have yet to be elucidated.

7.3.2 Genomic regions linked with *Cronobacter* spp. virulence

Many bacterial species exploit specialized secretion systems to transfer macromolecules across their membranes. These molecules are transported either directly into target cells or into the extracellular environment of the cell. Hence, these can act as important pathogenicity traits in these organisms. The secretory systems are distinguishable by their secreted effector molecules, mechanism of secretion and the conserved components in their structure (Bonemann *et al*. 2010). The *Cronobacter* genomes showed the presence of the T4SS and T6SS detailed in Section 7.2.2.1.

The T4SSs are the secretion systems ancestrally related to the bacterial conjugation machinery. They can transfer both proteins and nucleoprotein complexes and could constitute a conjugal transfer system (Mattick 2002). The three clusters of T4SSs identified in the *Cronobacter* spp. genomes showed very interesting plasmid-borne lineage specificity. The first cluster of 13 plasmid-borne genes (ESA_pESA2p06578-90) was found only in the smaller plasmids, pESA2 and pCTU2 of the genomes of *C. sakazakii* BAA-894 and *C. turicensis* z3032 respectively, encoding for a VirB group of proteins. The second cluster was a 21 gene region of Tra proteins, associated with the IncF plasmid groups, found only in the genomes of *C. sakazakii* 701 (ST4) and 696 (ST12). This suggested the possibility of it being a small cryptic plasmid associated with these strains, both of which were isolated from the same NICU outbreak in France, the former also having resulted in a fatality (Caubilla-Barron *et al*. 2007). The third T4SS cluster was a group of Trb proteins found only in the genome of *C. universalis* 581, and indicate the possibility of being a remnant plasmid. Further studies need to be carried out to evaluate the role of these T4SSs in the *Cronobacter* genus, as well as the possibility of any lineage specific conservation.

The T6SS is a more newly described bacterial secretion system that may be involved in competing with other bacteria in traits such as adherence, host-cell invasion, growth inside macrophages as well as survival within the host. Six clusters of T6SS genes were identified

across the *Cronobacter* spp. genomes with diverse patterns of variations among the species, though not all of them may be functional (Table 7.4). Of them, cluster 4 (ESA_03887–ESA_03946) was found to be the largest cluster and was also the only one conserved across all the genomes of the *Cronobacter* genus. This cluster appeared to possess the complete repertoire of genes involved in the T6SS mechanism and could potentially be a functional region. Zhou *et al*. (2012) have recently reported the role of T6SS in *E. coli* K1 invasion of the human BBB. Therefore, these regions could be crucial from the point of view of *Cronobacter* spp. pathogenicity and its association with neonatal meningitis. Although several T6SSs were found in all the *Cronobacter* species, none were unique to the *C. sakazakii* ST4 strain 701, except for a few remnant genes. Further studies need to be carried out to investigate the role of these genomic regions in the functional virulence of the organism. To date, the only study of the T6SSs in *Cronobacter* spp. has been an *in silico* analysis of the pESA3 plasmid-associated virulence regions by Franco *et al*. (2011). They found a considerable range of size variations among different strains for the T6SSs, though 98% of the plasmid-borne strains were found to possess at least partial T6SS clusters. They have also proposed distinct origins for the chromosomal and plasmid-borne T6SS clusters.

It is also interesting to note that unlike many other enteric pathogens, the *Cronobacter* spp. genomes did not exhibit the presence of any Type III secretion machinery genes.

Appendages such as fimbriae or pili could be very significant in *Cronobacter* spp. infection because of their roles in the attachment and invasion of the intestinal cells by the bacterium. Twelve putative fimbrial clusters (Section 7.2.2.2; Table 7.5) were identified in the *Cronobacter* spp. genomes sequenced for this study, of which only five were found to be conserved across the genus. The genetic content of the majority of the fimbriae clusters was most similar to the type I chaperone/usher-assembled pilus system (Nucci and Baumle 2007). Kucerova *et al*. (2011) had suggested the possible functional role of these fimbrial clusters as some degree of homology was found between the genes in the *C. sakazakii* fimbriae clusters and the remaining components necessary for type-I pilus assembly. Type 1 fimbriae have also been associated with *E. coli* K1 invasion of human brain cells (Teng *et al*. 2005) and are therefore of particular interest in *Cronobacter* pathogenicity studies. Cluster 5 (ESA_03512-20) was of particular interest as it was found unique to the genomes of *C. sakazakii*, and located on the genome near the sialic acid utilisation genes, suggesting a possible association discussed further in detail below.

Cluster 10 (CTU_16160-230) consisted of genes encoding proteins for the synthesis and assembly of curli fimbriae. Interestingly, this cluster was found to be completely absent in all the genomes of *C. sakazakii*, as well as *C. turicensis* 564, *C. muytjensii* 530 and *C. dublinensis* 582 genomes. Curli fimbriae are believed to play an important role in the adhesion of *E. coli* to host cells by interacting with matrix proteins such as fibronectin, laminin, and plasminogen to

initiate adherence and colonization (Sjobring *et al*. 2006). However, since *C. sakazakii* strains dominate clinical isolates, the absence of curli fimbriae genes infers this trait may not be essential for *Cronobacter* pathogenicity. Why the region has been lost in the *C. sakazakii* lineage and retained in the rest of the genus is unclear.

The region ESA_03609–13 on the genome of *C. sakazakii* BAA-894 encodes for the uptake and utilization of exogenous sialic acid, and was found to be unique to the *C. sakazakii* genomes in this study (Section 7.2.2.3). This uniqueness was found to be a very interesting trait with respect to *Cronobacter* spp. epidemiology and virulence. The *C. sakazakii* genomes were found to possess the entire *nanAKE* operon required for the uptake and catabolism of sialic acid. However, the operon was found to be distributed with the *nanE* gene located at a different region of the genome at the locus ESA_00529. This was found true for all the *C. sakazakii* genomes in this study, however since the *C. sakazakii* BAA-894 genome was used as a reference backbone for mapping scaffolds of the *de novo* synthesized genomes, one cannot eliminate the possibility of the assembly of the reference genome having influenced the order of genes in the newly sequenced genomes. However, this split in the operon need not necessarily be seen as an anomaly, as this phenomenon has also been observed in some other Gram-negative pathogens such as *E. tarda* and *C. freundii* (Vimr 2012). This plausibly suggests a separate evolutionary lineage for the gene. All the *C. sakazakii* genomes also had genes encoding for the NanT transporter and the NanC outer membrane porin, the activity of both was reported to be regulated by the repressor protein NanR. Hence, all *C. sakazakii* strains can potentially transport the exogenous sialic acid into the cytoplasm of their cells. Initial laboratory experiments conducted by our research group have confirmed the utilization of sialic acid by *C. sakazakii*, as well as the uniqueness of this trait to only this species in the genus *Cronobacter* (Unpublished results, personal communication). The acquisition of genes encoding for the utilization of exogenous sialic acid may have a major role in *C. sakazakii* colonisation of the human intestinal tract (via mucins) and the use of sialic acid in breast milk, infant formula, and brain cells as a nutrient source (Almagro-Moreno and Boyd 2009).

However, an interesting observation made was that unlike the *nanA*, *nanT* and *nanK* genes, the related gene *nanE* was found conserved across the genomes of the *Cronobacter* genus (Table 7.6), and as already mentioned also located in a different position on the genome. Among the genes, the %GC content of the *nanC* gene was found to be 47%, considerably less than the 56% GC content of the overall *Cronobacter* genome. Slight aberrations were also seen in the %GC content values of the *nanK* and *nanE* genes, an observation noted in the *nan* clusters of other *Enterobacteriaceae* members such as *Salmonella enterica*, *E. coli* and *Yersinia* spp. in an evolutionary study conducted by Almagro-Moreno and Boyd (2009). These observations suggest a strong possibility of horizontal transfer events having influenced separate

acquisitions for the *nanAKT* cluster in *C. sakazakii*, as well as the *nanC* and *nanE* clusters in the whole genus. Hence, it is possible that the *nan* clusters could have evolved in a mosaic pattern in this bacterial genus. A phylogenetic analysis was also conducted using the protein sequences encoded by the sialic acid utilization genes in *C. sakazakii* (Fig. 7.5 to 7.12). These figures reveal that the *nanATKR* genes have evolved as an independent lineage in *C. sakazakii*, closely related to other pathogens of the *Enterobacteriaceae* family. The NanC protein sequences did not find suitable homology with a number of the *Enterobacteriaceae* NanC sequences, suggesting that the acquisition of the associated gene in *C. sakazakii* could have been from a more distantly related organism, currently unidentified.

It is also notable that the *nanATK* cluster in *C. sakazakii* is located adjacent to a stringent starvation gene homologue (*sspA*, ESA_03615; Table 7.6) and therefore suggesting that expression of this gene cluster could be responsive to environmental nutrient levels. The presence and activity of other related genes such as *nagA* and *nagB* suggest the role of sialic acid as a carbon or nitrogen source in *C. sakazakii*. The gene (*nanH*) encoding the sialidase enzyme required to obtain free sialic acid by the bacterium was found to be absent in the *Cronobacter* spp. genomes. It has been proposed that such organisms that lack the *nanH* gene often obtain the enzyme from other sialidase-producing bacteria in the environment or from the host that may produce the enzyme in conditions of inflammation (Shakhnovich *et al*. 2002; Severi *et al*. 2007). It is also important to note the absence of genes such as *neuS*, *neuO* and *neuD* in the genomes of the *Cronobacter* genus. In *E. coli*, these genes are responsible for the formation of polysialic acid (PSA) capsules as well as sialylation of the O-lipopolysaccharide component of the cell wall. The absence of these genes in *Cronobacter* spp. suggests that the organism does not decorate its cell surface with sialic acid, unlike others such as *E. coli* and *N. meningitidis* (Severi *et al*. 2007).

Three rich mammalian sources of sialic acid for pathogenic or commensal bacteria are in the gastrointestinal tract, the brain and in human milk. The human intestinal epithelium shows very high sialic acid concentrations, and especially in the intestinal lining of a newborn, the concentration of sialic acid and N-acetylglucosamine residues have been found to be higher compared to adults. The cell membranes in the brain contain 20 fold higher sialic acid compared to other mucosal membranes, and the Neu5Ac is especially found concentrated on the gangliosides of the brain to form sialylated glycolipids (Wang 2009). Human milk has been found to be an abundant source of sialic acid, with highest levels observed in the colostrum until 3 months after the birth of the child. Therefore, an infant is exposed to sialic acid very early in life, and this in turn is believed to influence the levels of sialic acid in the brain as well (Wang *et al*. 2001). These three locations of sialic acid accumulation exhibit an interesting clinical association with the epidemiology of *Cronobacter* spp., through the NEC and intensive brain damage during neonatal meningitis that the organism causes. This is especially pertinent since

most of the neonatal infections in *C. sakazakii* have been reported at a very early stage of growth with half in the first week and three quarters during the first month of age (Lai 2001). Apart from these mammalian sources of sialic acid, another key link with the organism is also the fact that sialic acid is often added to PIF products, though at much lower concentrations compared to human milk (<25%). This sialic acid in infant formula is mostly glycoprotein-bound as opposed to the oligosaccharide-bound form in human milk (Wang *et al*. 2001).

Another interesting aspect to this topic was the presence of a sialic acid specific S-fimbrial adhesin (sfa) found in *E. coli* K1, a major causative agent of neonatal meningitis (Ott *et al*. 1986). With the strong evidence of the role of fimbriae in bacterial invasion and infections in the host as well as similar epidemiology of *Cronobacter* spp., this makes this a topic worth pursuing for future studies. In the analysis of the fimbrial regions of the *Cronobacter* spp. genomes, cluster 5 (ESA_03512-3520) was found to be unique to *C. sakazakii*. Further investigation needs to be conducted to verify any possible correlation between the adhesive properties of these fimbrial clusters to intestinal and brain cells, the sialic acid utilisation and the pathogenicity of the bacterium.

Iron is an essential growth factor for micro-organisms and bacteria have evolved a number of mechanisms for its acquisition from the environment. Iron uptake has been reported as a virulence mechanism in pathogenic bacteria (Gao *et al*. 2012; Penwell *et al*. 2012). This is of special importance with respect to *Cronobacter* spp. epidemiology because of the plausible utilisation by the organism of iron in breast milk and infant formula. The analysis of the sequenced *Cronobacter* spp. genomes revealed a range of gene clusters related to the acquisition of iron through the synthesis of siderophores such as aerobactin, enterobactin and hemin, as well as ferrous and ferric iron transporters (Section 7.2.2.4). Majority of these were conserved across the genus and therefore correlated with previous findings by Kucerova *et al*. (2011) and Grim *et al*. (2012). One of these plasmid-borne gene clusters (*iucABCD* and *iutA*) encodes for an aerobactin-like siderophore. This region characterized on the pESA3 plasmid of *C. sakazakii* BAA-894, was also previously analysed by Franco *et al*. (2011), and is now identified as a functional siderophore of *Cronobacter* spp (Grim *et al*. 2012). This region was found present in 97% of the plasmid-borne strains tested in their studies. On the other hand, the enterobactin-like chromosomal-borne siderophores were found to be non-functional in the studies conducted by Grim *et al*. (2012).

The genome of *C. sakazakii* 680 (ST8) encoded a six-gene cluster (*fecRABCDE*), not found in any of the other *Cronobacter* genomes. Grim *et al*. (2012) also identified this cluster in only one of their *C. sakazakii* strains, the sequence type of which is not yet known. Further studies need to be carried out to confirm whether this could be a lineage-acquired cluster in *C.*

sakazakii. Homologues of this region have been identified in virulence-associated plasmid-borne locations in other enteric bacteria such as *E. coli* and *Shigella* spp. (Payne *et al*. 2006).

Genes encoding for TonB-dependent receptors for ferrichrome (FcuA) and ferrienterochelins were found only in the *C. muytjensii, C. dublinensis* and *C. condimenti* genomes. Grim *et al*. (2012) also found the ferrichrome receptors Fct and FcuA exclusive to *C. dublinensis* and *C. muytjensii*. The *C. condimenti* species had not been included in their study. They also reported the close association of the *fcuA* gene with homologous regions in plant pathogens such as *Pantoea* spp. and *Erwinia* spp., hinting at a specific plant based association for these species. This correlates well with the evolutionary timescale estimations made earlier in this study (Chapter 3) using the *Cronobacter* MLST dataset, where *C. dublinensis* and *C. muytjensii* were found to be the earliest evolved members of the genus, the lineages evolving during the same period as the evolution of flowering plants.

Since a majority of the iron uptake mechanisms were found in all the *Cronobacter* species, including the ones not yet associated with infections, further studies need to be carried out to evaluate their functionality and association with virulence.

Flagella are important in bacteria not just for functions associated with motility and chemotaxis, but also for virulence related traits such as adhesion, biofilm formation and immune response elicitation. The role of flagella has been implicated in the virulence of many bacterial pathogens (Duan *et al*. 2012). In *C. sakazakii*, flagellar structure genes have been associated with the biofilm formation properties of the organism (Hartmann *et al*. 2011). Analysis of the *Cronobacter* spp. genomes revealed three main gene clusters across the genus related to flagellar biosynthesis as well as motor activities. These gene clusters were found present in all the *Cronobacter* spp. genomes, with one exception. The genome of *C. sakazakii* 680 (ST8) had missing the *fliE-fliR* set of proteins associated with flagellar biosynthesis. Laboratory experiments conducted by our research group have confirmed that the strain *C. sakazakii* 680 is non-motile. Since all the other *Cronobacter* spp. genomes had the flagellar machinery intact, this anomaly in *C. sakazakii* 680 could be the result of a deletional mutation. This strain belongs to ST8, which has been associated with a number of clinical cases, though not confirmed to be linked with meningitis (Chapter 5). It needs to be investigated further whether this deletion could be a lineage specific trait and whether it bestows an environmental disadvantage to the physiology and virulence of the organism.

Superoxide dismutase activity has been proposed as one of the mechanisms for macrophage survival in *Cronobacter* spp. (Townsend *et al*. 2007). Two other key genes linked to *Cronobacter* spp. virulence have been *ompA* and *ompX*, mutant studies on which have implicated their role in the basolateral invasion of human epithelial cells (Kim *et al*. 2010). In

particular, the *ompA* gene has been associated with the invasion of HBMEC both in *C. sakazakii* as well as *E. coli* K1, which makes it a key determinant for neonatal meningitic infections (Wang *et al*. 2002; Nair *et al*. 2009). The *zpx* gene encoding a zinc-containing metalloprotease has been characterized for its possible role in the necrosis and cellular destruction observed during NEC infections (Kothary *et al*. 2007). All these genes were found to be conserved across the genomes of the *Cronobacter* genus. Therefore, no correlation could be made between *sodA, ompA, ompX* or *zpx* and the variance in virulence with *Cronobacter* species or sequence type.

7.4.3 Conclusions

In summary, the comparison of the draft genomes representing the *Cronobacter* genus has revealed interesting patterns of diversity of the regions related to the physiology and virulence of the organism. One of the obvious aspects of interest for this genomic study was to investigate any possible unique virulence trait for the *C. sakazakii* ST4 lineage. However, despite detailed genomic analysis, the reason for the association of *C. sakazakii* ST4 with neonatal meningitis remains unclear. The lack of clearly identifiable virulence genes unique to this lineage may indicate that its prevalence in neonatal meningitis cases is due to environmental persistence and increased host exposure, as has been suggested earlier in Chapter 5.

The presence of a number of key virulence associated genes in all the *Cronobacter* species, including the ones that have not yet been associated with infections, proves that the genotype does not always necessarily predict the phenotype of the organism. However, the sialic acid utilisation *nan* gene cluster did stand out for its uniqueness to *C. sakazakii*, the species that has been most frequently associated with infections in the *Cronobacter* genus, as well as the mosaic evolution of the genes involved in the cluster. The actual significance of these regions in the pathogenicity of the organism needs to be further assessed by laboratory studies such as gene expression analysis, mutational studies as well as *in vitro* tissue culture experiments. Other genes of possible significance in *C. sakazakii* virulence due to their diversity observed in the genomes were the beta-fimbrial gene cluster (ESA_03512-3520), the *cpa* gene (ESA_pESA3p05434) and an ABC-type multidrug efflux system (ESA_01116-19).

Further improvement of the draft genome sequences, bioinformatics searches for novel or known virulence markers present in bacteria with similar modes of infection such as *Cit. koseri* and *E. coli* K1, and *in silico* and *in vitro* ascertainment of the contribution of single nucleotide polymorphisms, genome rearrangements, and other sequence features to pathogenicity may shed light on acquired pathogenicity of ST4 strains.

CHAPTER 8
CONCLUSIONS AND FUTURE DIRECTIONS

Cronobacter spp. are opportunistic pathogens that have been implicated in a number of infections affecting all age groups, in particular neonatal meningitis, NEC and bacteraemia. In this project, a population study of the *Cronobacter* genus was carried out using 325 strains spanning the seven species in order to gain better insights into the diversity and virulence of the organism, using the tools of MLST and comparative whole genomic analysis.

The *Cronobacter* MLST scheme is based on the seven genes *atpD, fusA, glnS, gltB, gyrB, infB* and *ppsA.* The expansion of the MLST scheme across the entire genus and subsequent sequence analysis has contributed to a number of key aspects of the genus:

- Revealed the phylogenetic relationships and diversity between the species.
- Revealed the evolutionary descent of the genus.
- Helped in the recognition of the two new species, *C. universalis* and *C. condimenti*.
- Revealed the strong clonality of the *C. sakazakii* and *C. malonaticus* species.
- Identified a clonal lineage for a majority of the neonatal meningitic cases – the *C. sakazakii* ST4 clonal complex.

Thus, the study of this bacterial population using MLST and analytical tools has revealed a novel structure for the *Cronobacter* genus which is supported by the reliability of DNA sequencing over subjective phenotyping. This study is now being carried forward in a number of directions. The MLST scheme spanning the seven species of the *Cronobacter* genus is available as an open access, curated database at http://www.pubmslt.org/cronobacter. The database continues to be updated beyond this study with the continuing characterization of *Cronobacter* spp. isolates from across the world, by other researchers in our group as well as by contributions from collaborators and *Cronobacter* researchers in different parts of the globe. At the end of November 2012, the total number of *Cronobacter* spp. strains in the database was 406.

The identification of the ST4 clonal complex as the lineage responsible for the majority of neonatal meningitis infections in the *Cronobacter* spp. has revealed a unique molecular signature for this severe and fatal infection caused by the organism. Apart from the ST4 clonal complex, the MLST scheme also revealed some other key lineages linked with the virulence of the organism such as ST1 and ST8. Based on these results, a core set of clinical strains have been chosen for future laboratory experiments on the organism. These results are now being used to expand the clinical study of the *C. sakazakii* species to a whole genome level. For this purpose, 20 representative clinical and 2 non-clinical *C. sakazakii* strains have been chosen from the strain set for whole-genome pair-ended 36 bp Illumina resequencing, in collaboration

with the University of Exeter. These genomes are now being analysed by Naqash Masood, in our research group, as part of his PhD studies. The core strain set of clinical significance is also being used for laboratory studies such as plasmid profiling, PCR probes for virulence genes and *in vitro* tissue culture assays, by other members of our group.

Another important outcome of the MLST study was that it formed the basis for the selection of strains for the whole genus genomic study, in order to effectively represent the diversity of the *Cronobacter* genus. The comparative genomic analysis study that formed the latter part of this project, using the 14 high quality draft genomes of the *Cronobacter* spp., is the first genus level genomic study to be performed on the organism. The analysis revealed a phylogenetic structure of the genus largely resembling the one observed by MLSA, with influence of recombination on the individual speciation events. Apart from identifying the core genome content of the entire genus, the pan-genome analysis also revealed novel regions unique to the individual *Cronobacter* species.

A number of key traits such as type six secretion systems, fimbrial gene clusters, iron acquisition systems and metal resistance were found to be scattered across the genus, with varying levels of diversity. Some of the virulence traits were also identified to be plasmid-borne making the plasmids one of the obvious targets for future studies in our research group. The diversity of the physiological and virulence traits identified in the genomes of the organism are now being individually characterized by laboratory studies as part of ongoing PhD projects, in order to gain a better understanding of the pathogenicity of the organism. Even though the strain set for the comparative genomic study included a representative of the ST4 lineage (*C. sakazakii* 701), the genome did not reveal any unique virulence traits exclusive to ST4, hinting at the possibility of gene expression playing a greater role in the virulence phenotype of the organism, rather than just the presence or absence of the virulence associated genes.

A key finding from the comparative genomic study was the unique cluster of genes in *C. sakazakii* encoding for the utilization of exogenous sialic acid. Since this is also the species most associated with the neonatal meningitic infections, this association could prove to be a crucial link to the pathogenicity of the organism. The functionality of this trait in *C. sakazakii* as well as the possible mechanisms of the pathogenicity in the neonatal brain are being studied further in laboratory studies. Initial results have already confirmed the uniqueness of the sialic acid utilization trait in the *C. sakazakii* species, as suggested by the genomic regions.

Plans are now being undertaken to expand the current *Cronobacter* PubMLST database in order to incorporate the Bacterial Isolate Genome Sequence Database (BIGSdb) for the

organism (Jolley and Maiden 2010), using the fourteen genomes analysed in this study. Further expansion of this centralised genomic database will enable improved comparative genomic analysis studies, identification of targets for typing schemes as well as a possible *in silico* expansion of the MLST scheme to include greater number of loci, if needed.

Thus, this is not the end of this study, but is only the end of the beginning. This whole genus population study has enabled us to better understand this bacterium which can cause irreparable damage to a newborn baby's brain. This study has also opened up many new avenues into the research of this opportunistic food-borne pathogen. It is hoped that continuing research will help find answers about the destructive pathogenic mechanism of *Cronobacter* spp., which can thereafter help to put better control measures in place.

REFERENCES

ACHTMAN, M., ZURTH, K., MORELLI, G., TORREA, G., GUIYOLE, A. & CARNIEL, E. 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis. Proceedings of the National Academy of Sciences of the USA*, 96, 14043-14048.

ACHTMAN, M. & WAGNER, M. 2008. Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology*, 6, 431-440.

ACINAS, S.G., MARCELINO, L.A., KLEPAC-CERAJ, V. & POLZ, M.F. 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *Journal of Bacteriology*, 186, 2629-2635.

ADEGBOLA, R. A. & OLD, D. C. 1983. Fimbrial haemagglutinins in *Enterobacter* species. *Journal of General Microbiology*, 129, 2175-2180.

ADEKUNTE, A., VALDRAMIDIS, V. P., TIWARI, B. K., SLONE, N., CULLEN, P. J., DONNELL, C. P. O. & SCANNELL, A. 2010. Resistance of *Cronobacter sakazakii* in reconstituted powdered infant formula during ultrasound at controlled temperatures: A quantitative approach on microbial responses. *International Journal of Food Microbiology*, 142, 53-59.

AGOGUÉ, H., CASAMAYOR, E. O., BOURRAIN, M., OBERNOSTERER, I., JOUX, F., HERNDL, G. J. & LEBARON, P. 2006. A survey on bacteria inhabiting the sea surface microlayer of coastal ecosystems. *FEMS Microbiology Ecology*, 54, 269-280.

AL-ATTAR, S., WESTRA, E.R., VAN DER OOST, J. & BROUNS, S.J. 2011. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biological Chemistry*, 392, 277-289.

AL‐HOLY, M. A., CASTRO, L. F. & AL‐QADIRI, H. M. 2010. Inactivation of *Cronobacter* spp. (*Enterobacter sakazakii*) in infant formula using lactic acid, copper sulfate and monolaurin. *Letters in Applied Microbiology*, 50, 246-251.

ALDOVÁ, E., HAUSNER, O. & POSTUPA, R. 1983. Tween esterase activity in *Enterobacter sakazakii*. *Zentralblatt fuer Bakteriologie Mikrobiologie und Hygiene Series A*, 256, 103-108.

ALIKHAN, N. F., PETTY, N. K., BEN ZAKOUR, N. L. & BEATSON, S. A. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, 12, 402.

ALLISON, G.E. & VERMA, N.K. 2000. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends in Microbiology*, 18, 17-23.

ALLISON, G.E., ANGELES, D., TRAN-DINH, N. & VERMA, N.K. 2002. Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of *Shigella flexneri*. *Journal of Bacteriology*, 184, 1974-1987.

ALMAGRO-MORENO, S. & BOYD, E.F. 2009. Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evolutionary Biology*, 9, 118.

ALPERI, A., MARTINEZ-MURCIA, A.J., MONERA, A., SAAVEDRA, M.J. & FIGUERAS, M.J. 2010. *Aeromonas fluvialis* sp. nov., isolated from a Spanish river. *International Journal of Systematic and Evolutionary Microbiology*, 60, 72-77.

ANGIUOLI, S.V. & SALZBERG, S.L. 2010. Mugsy: Fast multiple alignment of closely related whole genomes. *Bioinformatics*, 27, 334-342.

ANONYMOUS. 2006. Milk and Milk Products — Detection of *Enterobacter sakazakii*. ISO/TS 22964:2006.

ARBATSKY, N.P., SUN, Y., SHASHKOV, A.S., WANG, M., LIU, B., DAEVA, E.D., WANG, L. & KNIREL, Y.A. 2012. Structure and genetics of the O-antigen of *Cronobacter sakazakii* G2726 (serotype O3) closely related to the O-antigen of *C. muytjensii* 3270. *Carbohydrate Research*, 355, 50-55.

ARNOLD, M.L., SAPIR, Y. & MARTIN, N.H. 2008. Genetic exchange and the origin of adaptations: prokaryotes to primates. *Philosophical Transactions of the Royal Society,* 363, 2813-2820.

ARSLAN, D., LEGENDRE, M., SELTZER, V., ABERGEL, C. & CLAVERIE, J.M. 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proceedings of the National Academy of Sciences USA*, 108, 17486-17491.

ASAKURA, H., MORITA-ISHIHARA, T., YAMAMOTO, S. & IGIMI, S. 2007. Genetic characterization of thermal tolerance in *Enterobacter sakazakii*. *Microbial Immunology*, 51, 671-677.

AZIZ, R.K., BARTELS, D., BEST, A.A., DEJONGH, M., DISZ, T., EDWARDS, R.A., FORMSMA, K., GERDES, S., GLASS, E.M., KUBAL, M., MEYER, F., OLSEN, G.J., OLSON, R., OSTERMAN, A.L., OVERBEEK, R.A., MCNEIL, L.K., PAARMANN, D., PACZIAN, T., PARRELLO, B., PUSCH, G.D., REICH, C., STEVENS, R., VASSIEVA, O., VONSTEIN, V., WILKE, A. & ZAGNITKO, O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9, 75.

BACK, S. Y., JIN, H. H. & LEE, S. Y. 2009. Inhibitory effect of organic acids against *Enterobacter sakazakii* in laboratory media and liquid foods. *Food Control*, 20, 867-872.

BADGER, J. L., STINS, M. F. & KIM, K. S. 1999. *Citrobacter freundii* invades and replicates in human brain microvascular endothelial cells. *Infection and Immunity*, 67, 4208-4215.

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, 223.

BAKER, M. 2012. *De novo* genome assembly: what every biologist should know. *Nature Methods*, 9, 333-337.

BARDOWSKI, J., EHRLICH, S.D. & CHOPIN, A. 1994. BglR protein, which belongs to the BglG family of transcriptional antiterminators, is involved in beta-glucoside utilization in *Lactococcus lactis*. *Journal of Bacteriology,* 176, 5681-5685.

BAUMBACH, J., ROONEY, K., SMELSER, C. & TORRES, P. 2009. *Cronobacter* species isolation in two infants – New Mexico, 2008. *Morbidity and Mortality Weekly Report*, 58, 1179-1183.

BAUMGARTNER, A., GRAND, M., LINIGER, M. & IVERSEN, C. 2009. Detection and frequency of *Cronobacter* spp. (*Enterobacter sakazakii*) in different categories of ready-to-eat foods other than infant formula. *International Journal of Food Microbiology*, 136, 189-192.

BEILSTEIN, F. & DREISEIKELMANN, B. 2008. Temperate bacteriophage PhiO18P from an *Aeromonas media* isolate: characterization and complete genome sequence. *Virology,* 373, 25-29.

BERG, O.G. & MARTELIUS, M. 1995. Synonymous substitution-rate constants in *Escherichia coli* and *Salmonella* Typhimurium and their relationship to gene expression and selection pressure. *Journal of Molecular Evolution,* 41, 449-56.

BINGLE, L.E.H., BAILEY, C.M. & PALLEN, M.J. 2008. Type VI secretion: a beginner's guide. *Current Opinion in Microbiology,* 11, 3-8.

BIRNBOIM, H.C. & DOLY, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research,* 7, 1513-1523.

BLOCK, C., PELEG, O., MINSTER, N., BAR-OZ, B., SIMHON, A., ARAD, I. & SHAPIRO, M. 2002. Cluster of neonatal infections in Jerusalem due to unusual biochemical variant of *Enterobacter sakazakii*. *European Journal of Clinical Microbiology and Infectious Diseases,* 21, 613-616.

BÖNEMANN, G., PIETROSIUK, A. & MOGK, A. 2010. Tubules and donuts: a type VI secretion story. *Molecular Microbiology*, 76, 815–821.

BOWEN, A. B. & BRADEN, C. R. 2006. Invasive *Enterobacter sakazakii* disease in infants. *Emerging Infectious Diseases*, 12, 1185.

BOWEN, A.B. & BRADEN, C.R. 2008. *Enterobacter sakazakii* disease and epidemiology. Edited by Farber, J.M. & Forsythe, S.J. *Enterobacter sakazakii*, Washington, D.C.: ASM Press. pp 101-125.

BOYER, F., FICHANT, G., BERTHOD, J., VANDENBROUCK, Y. & ATTREE, I. 2009. Dissecting the bacterial type VI secretion system by a genome wide *in silico* analysis: what can be learned from available microbial genomic resources? *BMC Genomics,* 10, 104.

BRADY, C., CLEENWERCK, I., VENTER, S., VANCANNEYT, M., SWINGS, J. & COUTINHO, T. 2008. Phylogeny and identification of *Pantoea* species associated with plants,

humans and the natural environment based on multilocus sequence analysis (MLSA). *Systematic and Applied Microbiology,* 31, 447-460.

BRADY, C.L., CLEENWERCK, I., VAN DER WESTHUIZEN, L., VENTER, S.N., COUTINHO, T. & DE VOS, P. 2012. *Pantoea rodasii* sp. nov., *Pantoea rwandensis* sp. nov. and *Pantoea wallisii* sp. nov., isolated from *Eucalyptus*. *International Journal of Systematic and Evolutionary Microbiology*, 62, 1457-1464.

BREEUWER, P., LARDEAU, A., PETERZ, M. & JOOSTEN, H. M. 2003. Desiccation and heat tolerance of *Enterobacter sakazakii*. *Journal of Applied Microbiology*, 95, 967-973.

BRENGI, S.P., O'BRIEN, S.B., PICHEL, M., IVERSEN, C., ARDUINO, M., BINSZTEIN, N., JENSEN, B., PAGOTTO, F., RIBOT, E.M., STEPHAN, R., CERNELA, N., COOPER, K. & FANNING, S. 2012. Development and validation of a PulseNet standardized protocol for subtyping isolates of *Cronobacter* species. *Foodborne Pathogens and Disease*, 9, 861-867.

BROWN, A.H., FELDMAN, M.W. & NEVO, E. 1980. Multilocus structure of natural populations of *Hordeum spontaneum. Genetics*, 96, 523-536.

BUTLER, J. F., GARCIA-MARUNIAK, A., MEEK, F. & MARUNIAK, J. E. 2010. Wild Florida house flies (*Musca domestica*) as carriers of pathogenic bacteria. *Florida Entomologist*, 93, 218-223.

CANCHAYA, C., PROUX, C., FOURNOUS, G., BRUTTIN, A. & BRUSSOW, H. 2003. Prophage genomics. *Microbiology and Molecular Biology Reviews*, 67, 238-276.

CANCHAYA, C., FOURNOUS, G. & BRUSSOW, H. 2004. The impact of prophages on bacterial chromosomes. *Molecular Microbiology*, 53, 9-18.

CARATTOLI, A. 2011. Plasmids in Gram negatives: Molecular typing of resistance plasmids. *International Journal of Medical Microbiology,* 301, 654-658.

CARVER, T.J., RUTHERFORD, K.M., BERRIMAN, M., RAJANDREAM, M.A., BARRELL, B.G. & PARKHILL, J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics*, 21, 3422-3423.

CARVER, T., THOMSON, N., BLEASBY, A., BERRIMAN, M. & PARKHILL, J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics,* 25, 119-120.

CASCALES, E. 2008. The type VI secretion toolkit. *EMBO Reports,* 9, 735-741.

CASJENS, S. 2003. Prophages and bacterial genomics: what have we learned so far? *Molecular Microbiology,* 49, 277-300.

CARTER, L., LINDSEY, L. A., GRIM, C. J., SATHYAMOORTHY, V., JARVIS, K. G., GOPINATH, G., LEE, C., SADOWSKI, J.A., TRACH, L., PAVA-RIPOLL, M., MCCARDELL, B.A., TALL, B.D. & HU, L. 2012. A multiplex PCR assay, targeting a di-guanylate cyclase encoding gene, *cdgc1* to differentiate species within the genus *Cronobacter. Applied and Environmental Microbiology,* doi: 10.1128/AEM.02898-12.

CAUBILLA-BARRON, J., HURRELL, E., TOWNSEND, S., CHEETHAM, P., LOC-CARRILLO, C., FAYET, O., PRÈRE, M.F. & FORSYTHE, S.J. 2007. Genotypic and phenotypic analysis of *Enterobacter sakazakii* strains from an outbreak resulting in fatalities in a neonatal intensive care unit in France. *Journal of Clinical Microbiology*, 45, 3979-3985.

CAUBILLA-BARRON, J. & FORSYTHE, S. 2007. Dry stress and survival time of *Enterobacter sakazakii* and other *Enterobacteriaceae*. *Journal of Food Protection*, 70, 2111-2117.

CAUGANT, D.A. 2008. Genetics and evolution of *Neisseria meningitides*: Importance for the epidemiology of meningococcal disease. *Infection, Genetics and Evolution*, 8, 558-565.

CAUGANT, D.A. & MAIDEN, M.C.J. 2009. Meningococcal carriage and disease – Population biology and evolution. *Vaccine*, 27, B64-B70.

CDC. 2011. CDC Update: investigation of *Cronobacter* infections among infants in the United States. Available at: http://www.cdc.gov/foodsafety/diseases/cronobacter/investigation.html/ (Accessed 5th November 2012).

CHAIN, P.S., GRAFHAM, D.V., FULTON, R.S., FITZGERALD, M.G., HOSTETLER, J., MUZNY, D., ALI, J., BIRREN, B., BRUCE, D.C., BUHAY, C., COLE, JR., DING, Y., DUGAN, S., FIELD, D., GARRITY, G.M., GIBBS, R., GRAVES, T., HAN, C.S., HARRISON, S.H., HIGHLANDER, S., HUGENHOLTZ, P., KHOURI, H.M., KODIRA, C.D., KOLKER, E., KYRPIDES, N.C., LANG, D., LAPIDUS, A., MALFATTI, S.A., MARKOWITZ, V., METHA, T., NELSON, K.E., PARKHILL, J., PITLUCK, S., QIN, X., READ, T.D., SCHMUTZ, J., SOZHAMANNAN, S., STERK, P., STRAUSBERG, R.L., SUTTON, G., THOMSON, N.R., TIEDJE, J.M., WEINSTOCK, G., WOLLAM, A.; GENOMIC STANDARDS CONSORTIUM HUMAN MICROBIOME PROJECT JUMPSTART CONSORTIUM & DETTER, J.C. 2009. Genome project standards in a new era of sequencing. *Science*, 326, 236-237.

CHAP, J., JACKSON, P., SIQUEIRA, R., GASPAR, N., QUINTAS, C., PARK, J., OSAILI, T., SHAKER, R., JARADAT, Z., HARTANTYO, S.H.P., SANI, N.A., ESTUNINGSIH, S., & FORSYTHE, S.J. 2009. International survey of *Cronobacter sakazakii* and other *Cronobacter* spp. in follow up formulas and infant foods. *International Journal of Food Microbiology*, 136, 185-188.

CHAUDHURI, R.R. & PALLEN, M.J. 2005. *x*BASE, a collection of online databases for bacterial comparative genomics. *Nucleic Acids Research*, 34, D335-D337.

CHEN, Y., HAMMACK, T.S., SONG, K. & LAMPEL, K.A. 2004. Evaluation of a revised U.S. Food and Drug Administration method for the detection and isolation of *Enterobacter sakazakii* in Powdered Infant Formula: Precollaborative study. *Journal of AOAC International*, 92, 3.

CHEN, Y., SONG, K.Y., BROWN, E.W. & LAMPEL, K.A. 2010. Development of an improved protocol for the isolation and detection of *Enterobacter sakazakii* (*Cronobacter*) from powdered infant formula. *Journal of Food Protection,* 73, 1016-1022.

CHEN, Y., NOE, K.E., THOMPSON, S., ELEMS, C.A., BROWN, E.A., LAMPEL, K.A. & HAMMACK, T.S. 2011a. Evaluation of a revised FDA method for the detection of *Cronobacter* in powdered infant formula: Collaborative study. *Journal of Food Protection,* 73, 1016-1022.

CHEN, Y., STRAIN, E.A., ALLARD, M. & BROWN, E.W. 2011b. Genome sequence of *Cronobacter sakazakii* E899, a strain associated with human illness. *Journal of Bacteriology,* 193, 581.

CHEN, Y., LAMPEL, K. & HAMMACK, T. 2012. Chapter 29: *Cronobacter*. Bacteriological Analytical Manual. *U.S. Food and Drug Administration.* Available at: http://www.fda.gov/Food/ScienceResearch/LaboratoryMethods/BacteriologicalAnalyticalManualBAM/ucm289378.htm (Accessed 10th November 2012).

CHEVENET, F., BRUN, C., BANULS, A.L., JACQ, B. & CHRISTEN, R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics,* 7, 439.

CHEVREUX, B., PFISTERER, T., DRESCHER, B., DRIESEL, A.J., MÜLLER, W.E., WETTER, T. & SUHAI, S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research,* 14, 1147-1159.

CHOI, S. C., RASMUSSEN, M. D., HUBISZ, M. J., GRONAU, I. & STANHOPE, M. J. 2012. Replacing and additive horizontal gene transfer in *Streptococcus*. *Molecular Biology and Evolution,* 29, 3309-3320.

CLOWES, R.C. 1972. Molecular structure of bacterial plasmids. *Bacteriological Reviews,* 36, 361-405.

CODEX ALIMENTARIUS COMMISSION. 2008. Code of hygienic practice for powdered infant formulae for infants and young children. CAC/RCP 66-2008.

Available at: http://www.codexalimentarius.net/search/advancedsearch.do. (Accessed 10th November 2012).

COHN, F., 1875. Untersuchungen ueber Bakterien. *Beitraege zur Biologie der Planzen,* 1, 127-222. Translated and edited by Brock, T.D. 1998. *Milestones in Microbiology: 1556 to 1940*. ASM Press, pp 210.

COUTURIER, M., BEX, F., BERGQUIST, P.L. & MAAS, W.K. 1988. Identification and classification of bacterial plasmids. *Microbiology and Molecular Biology Reviews,* 52, 375-395.

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, 1044-1052.

CRUZ, A., XICOHTENCATL-CORTES, J., GONZÁLEZ-PEDRAJO, B., BOBADILLA, M., ESLAVA, C. & ROSAS, I. 2011a. Virulence traits in *Cronobacter* species isolated from different sources. *Canadian Journal of Microbiology*, 57, 735-44.

CZERWICKA, M., FORSYTHE, S.J., BYCHOWSKA, A., DZIADZIUSZKO, H., KUNIKOWSKA, D., STEPNOWSKI, P. & KACZYNSKI, Z. 2010. Structure of the O-polysaccharide isolated from *Cronobacter sakazakii* 767. *Carbohydrate Research*, 345, 908-913.

DAS, S. & CHAUDHURI, K. 2003. Identification of a unique IAHP (IcmF associated homologous proteins) cluster in *Vibrio cholerae* and other proteobacteria through in silico analysis. *In Silico Biol*ogy, 3, 287-300.

DAUGA, C. 2002. Evolution of the *gyrB* gene and the molecular phylogeny of *Enterobacteriaceae*: a model molecule for molecular systematic studies. *International Journal of Systematic and Evolutionary Microbiology*, 52, 531-547.

DAUGA, C. & BREUWER, P. 2008. Taxonomy and physiology of *Enterobacter sakazakii*. Edited by Farber, J.M. & Forsythe, S.J. *Enterobacter sakazakii*, Washington, D.C.: ASM Press. pp 1-26.

DELIHAS, N. 2011. Impact of small repeat sequences on bacterial genome evolution. *Genome Biology and Evolution,* 3, 959-973.

DIDELOT, X., BOWDEN, R., STREET, T., GOLUBCHIK, T., SPENCER, C., McVEAN, G., SANGAL, V., ANJUM, M.F., ACHTMAN, M., FALUSH, D., DEN BAKKER, H.C., CUMMINGS, C.A., FERREIRA, V., VATTA, P., ORSI, R.H., DEGORICIJA, L., BARKER, M., PETRAUSKENE, O., FURTADO, M.R. & WIEDMANN, M. 2010. Comparative genomics of the bacterial genus *Listeria*: Genome evolution is characterized by limited gene acquisition and limited gene loss. *BMC Genomics*, 11, 688.

DEN BAKKER, H.C., MORENO SWITT A.I., GOVONI, G., CUMMINGS, C.A., RANIERI, M.L., DEGORICIJA, L., HOELZER, K., RODRIGUEZ-RIVERA, L.D., BROWN, S., BOLCHACOVA, E., FURTADO, M.R. & WIEDMANN, M. 2011. Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of *Salmonella enterica. BMC Genomics,* 12, 425.

DEREEPER, A., GUIGNON, V., BLANC, G., AUDIC, S., BUFFET, S., CHEVENET, F., DUFAYARD, J.F., GUINDON, S., LEFORT, V., LESCOT, M., CLAVERIE, J.M., & GASCUEL, O. 2008. *Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research,* 36, W465-W469.

DONNELLY, P. 2011. Recombination and Population Structure in *Salmonella enterica*. *PLoS Genetics*, 7, e1002191.

DUAN, Q., ZHOU, M., ZHU, L. & ZHU, G. 2012. Flagella and bacterial pathogenicity. *Journal of Basic Microbiology,* doi: 10.1002/jobm.201100335.

EDELSON‐MAMMEL, S., PORTEOUS, M. K. & BUCHANAN, R. L. 2006. Acid resistance of twelve strains of *Enterobacter sakazakii*, and the impact of habituating the cells to an acidic environment. *Journal of Food Science*, 71, M201-M207.

EGGERS, C.T., MURRAY, I.A., DELMAR, V.A., DAY, A.G. & CRAIK, C.S. 2004. The periplasmic serine protease inhibitor ecotin protects bacteria against neutrophil elastase. *The Biochemical Journal,* 379, 107–118.

EMAMI, C.N., MITTAL, R., WANG, L., FORD, H.R. & PRASADARAO, N.V. 2011. Recruitment of dendritic cells is responsible for intestinal epithelial damage in the pathogenesis of necrotizing enterocolitis by *Cronobacter sakazakii*. *Journal of Immunology*, 186, 7067-79.

ENRIGHT, M.C. & SPRATT, B.G. 1999. Multilocus sequence typing. *Trends in Microbiology,* 7, 482-487.

ETHERINGTON, G.J., DICKS, J. & ROBERTS, I.N. 2004. Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. *Bioinformatics*, 21, 278-281.

FALUSH, D., KRAFT, C., TAYLOR, N.S., CORREA, P., FOX, J.G., ACHTMAN, M. & SUERBAUM, S. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: Estimates of clock rates, recombination size, and minimal age. *Proceedings of the National Academy of Sciences of the USA*, 98, 15056-15061.

FANNING, S. & FORSYTHE, S.J. 2008. Isolation and identification of *Enterobacter sakazakii*. Edited by Farber, J.M. & Forsythe, S.J. *Enterobacter sakazakii*, Washington, D.C.: ASM Press. pp 27-59.

FARMER, J.J. III, ASBURY, M.A., HICKMAN, F.W., BRENNER, D.J. & THE ENTEROBACTERIACEAE STUDY GROUP. 1980. *Enterobacter sakazakii*: a new species of "*Enterobacteriaceae*" isolated from clinical specimens. *International Journal of Systematic Bacteriology,* 30, 569-584.

FDA. 2002. Isolation and enumeration of *Enterobacter sakazakii* from dehydrated powdered infant formula. *Washington, DC USA: Food and Drug Administration.*

Available at: http://www.fda.gov/food/scienceresearch/laboratorymethods/ucm114665.htm (Accessed 11th November 2012).

FEIL, E.J., HOLMES, E.C., BESSEN, D.E., CHAN, M-S., DAY, N.P.J., ENRIGHT, M.C., GOLDSTEIN, R., HOOD, D.W., KALIA, A., MOORE, C.E., ZHOU, J. & SPRATT, B.G. 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proceedings of the National Academy of Sciences USA,* 98, 182–187.

FEIL, E.J., LI, B.C., AANENSEN, D.M., HANAGE, W.P. & SPRATT, B.G. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of Bacteriology*, 186, 1518-30.

FEIL, E.J. 2010. Linkage, selection and the clonal complex. Edited by Robinson, D.A., Falush, D., Feil, E.J. *Bacterial Population Genetics in Infectious Disease*, John Wiley & Sons, Inc., pp 19-35.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, 17, 368-376.

FELSENSTEIN, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, 5, 164-166.

FILLOUX, A., HACHANI, A. & BLEVES, S. 2008. The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology*, 154, 1570-1583.

FLEISCHMANN, R.D., ADAMS, M.D., WHITE, O., CLAYTON, R.A., KIRKNESS, E.F., KERLAVAGE, A.R., BULT, C.J., TOMB, J.F., DOUGHERTY, B.A., MERRICK, J.M., MCKENNEY, K., SUTTON, G., FITZHUGH, W., FIELDS, C., GOCYNE, J.D., SCOTT, J., SHIRLEY, R., LIU, L., GLODEK, A., KELLEY, J.M., WEIDMAN, J.F., PHILLIPS, C.A., SPRIGGS, T., HEDBLOM, E., COTTON, M.D., UTTERBACK, T.R., HANNA, M.C., NGUYEN, D.T., SAUDEK, D.M., BRANDON, R.C., FINE, L.D., FRITCHMAN, J.L., FUHRMANN, J.L., GEOGHAGEN, N. S. M., GNEHM, C.L., MCDONALD, L.A., SMALL, K.V., FRASER, C.M., SMITH, H.O. & VENTER, C.J. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269, 496-512.

FOOD AND AGRICULTURE ORGANIZATION-WORLD HEALTH ORGANIZATION (FAO-WHO). 2004. *Enterobacter sakazakii* and other microorganisms in powdered infant formula: meeting report. *Microbiological Risk Assessment Series 6.* WHO-FAO of the United Nations, Geneva and Rome. WHO Press, Geneva, Switzerland.
Available at: http://www.who.int/foodsafety/publications/micro/mra6/en/index.html. (Accessed 11th November 2012).

FOOD AND AGRICULTURE ORGANIZATION-WORLD HEALTH ORGANIZATION (FAO-WHO). 2006. *Enterobacter sakazakii* and *Salmonella* in powdered infant formula: meeting report. *Microbiological Risk Assessment Series 10.* WHO-FAO of the United Nations, Geneva and Rome. WHO Press, Geneva, Switzerland. Available at: http://www.who.int/foodsafety/publications/micro/mra10/en/index.html. (Accessed 11th November 2012).

FOOD AND AGRICULTURE ORGANIZATION-WORLD HEALTH ORGANIZATION (FAO-WHO). 2008. *Enterobacter sakazakii* (*Cronobacter spp.*) in powdered follow-up formulae. *Microbiological Risk Assessment Series 15*, WHO-FAO of the United Nations, Geneva and Rome. WHO Press, Geneva, Switzerland.

Available at: http://www.who.int/foodsafety/publications/micro/mra_followup/en/index.html. (Accessed 11th November 2012).

FOX, G.E., WISOTZKEY, J.D. & JURTSHUK, JR., P. 1992. How close is close: 16s rRNA sequence identity may not be sufficient to guarantee species identity. *International Journal of Systematic Bacteriology*, 42, 166-170.

FRANCISCO, A.P., BUGALHO, M., RAMIREZ, M. & CARRICO, J.A. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics*, 10, 152.

FRANCISCO, A.P., VAZ, C., MONTEIRO, P.T., MELO-CRISTINO, J., RAMIREZ, M. & CARRIÇO, J.A. 2012. PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics*, 13, 87.

FRANCO, A. A., KOTHARY, M. H., GOPINATH, G., JARVIS, K. G., GRIM, C. J., HU, L., DUTTA, A.R.., MCCARDELL, B.A. & TALL, B. D. 2011a. Cpa, the outer membrane protease of *Cronobacter sakazakii*, activates plasminogen and mediates resistance to serum bactericidal activity. *Infection and immunity*, 79, 1578-1587.

FRANCO, A.A., HU, L., GRIM, C.J., GOPINATH, G., SATHYAMOORTHY, V., JARVIS, K.G., LEE, C., SADOWSKI, J., KIM, J., KOTHARY, M.H., MCCARDELL, B.A. & TALL, B.D. 2011b. Characterization of putative virulence genes on the related RepFIB plasmids harbored by *Cronobacter* spp. *Applied and Environmental Microbiology*, 77, 3255-67.

FRIEDEMANN, M. 2007. *Enterobacter sakazakii* in food and beverages (other than infant formula and milk powder). *International Journal of Food Microbiology*, 116, 1-10.

FRIEDEMANN, M. 2009. Epidemiology of invasive neonatal *Cronobacter (Enterobacter sakazakii)* infections. *European Journal of Clinical Microbiology and Infectious Disease*, 28, 1297-1304.

FRONZES, R., REMAUT, H. & WAKSMAN, G. 2008. Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria. *The EMBO Journal*, 27, 2271-2280.

FRONZES, R., SCHÄFER, E., WANG, L., SAIBIL, H.R., ORLOVA, E.V. & WAKSMAN, G. 2009a. Structure of a type IV secretion system core complex. *Science*, 323, 266-268.

FRONZES, R., CHRISTIE, P.J. & WAKSMAN, G. 2009b. The structural biology of type IV secretion systems. *Nature Reviews Microbiology*, 7, 703-714.

FROST, L.S., LEPLAE, R., SUMMERS, A.O. & TOUSSAINT, A. 2005. Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*, 3, 722-732.

GAJDOSOVA, J., BENEDIKOVICOVA, K., KAMODYOVA, N., TOTHOVA, L., KACLIKOVA, E., STUCHLIK, S., TURNA, J. & DRAHOVSKA, H. 2011. Analysis of the DNA region mediating increased thermotolerance at 58°C in *Cronobacter* sp. and other enterobacterial strains. *Antonie Van Leeuwenhoek*, 100, 279-89.

GALLAGHER, P.G. & BALL, W.S. 1991. Cerebral infarctions due to CNS infection with *Enterobacter sakazakii*. *Pediatric Radiology*, 21, 135–136.

GAO, Q., WANG, X., XU, H., XU, Y., LING, J., ZHANG, D., GAO, S. & LIU, X. 2012. Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiology*, 12, 143.

GARNEAU, J.E., DUPUIS, M., VILLION, M., ROMERO, D.A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADAN, A.H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468, 67-72.

GEORGIADES, K. & RAOULT, D. 2011. Defining pathogenic bacterial species in the genomic era. *Frontiers in Microbiology*, 1, 151.

GILMOUR, M.W., THOMPSON, N.R., SANDERS, M., PARKHILL, J. & TAYLOR, D. 2004. The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. *Plasmid*, 52, 182-202.

GODDE, J.S. & BICKERTON, A. 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *Journal of Molecular Evolution*, 62, 718-729.

GORIS, J., KONSTANTINIDIS, K.T., KLAPPENBACH, J.A., COENYE, T., VANDAMME, P. & TIEDJE, J.M. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, 57, 81-91.

GRADSTEIN, F. M., OGG, J. G. & VAN KRANENDONK, M. 2008. On the geologic time scale 2008. *Newsletters on Stratigraphy*, 43, 5-13.

GRIM, C.J., KOTHARY, M.H., GOPINATH, G., JARVIS, K.G., BEAUBRUN, J.J., MCCLELLAND, M., TALL, B.D. & FRANCO, A.A. 2012. Identification and characterization of *Cronobacter* spp. iron acquisition systems. *Applied and Environmental Microbiology*, 78, 6035-6050.

GRISSA, I., VERGNAUD, G. & POURCEL, C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, 35 (suppl 2): W52-W57.

GUINDON, S. & GASCUEL, O. 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52, 696-704.

HAMBY, S. E., JOSEPH, S., FORSYTHE, S. J. & CHUZHANOVA, N. 2011. *In Silico* identification of pathogenic strains of *Cronobacter* from biochemical data reveals association of inositol fermentation with pathogenicity. *BMC Microbiology*, 11, 204.

HANAGE, W.P., FRASER, C. & SPRATT, B.G. 2005. Fuzzy species among recombinogenic bacteria. *BMC Biology*, 7.

HANAGE, W.P., FRASER, C. & SPRATT, B.G. 2006. Sequences, sequence clusters and bacterial species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361, 1917-1927.

HARIRI, S., JOSEPH, S. & FORSYTHE, S.J. 2012. Predominance of *Cronobacter sakazakii* ST4 clonal complex strains in *Cronobacter* neonatal meningitis infections in US 2011. *Emerging Infectious Diseases*, In Press.

HARRIS, L.S. & ORIEL, P.J. 1989. Heteropolysaccharide produced by *Enterobacter sakazakii*. *United States Patent number*, 4, 806–636.

HARTMANN, I., CARRANZA, P., LEHNER, A., STEPHAN, R., EBERL, L. & RIEDEL, K. 2010. Genes Involved in *Cronobacter sakazakii* biofilm formation. *Applied and Environmental Microbiology*, 76, 2251-61.

HEALY, B., COONEY, S., O'BRIEN, S., IVERSEN, C., WHYTE, P., NALLY, J., CALLANAN, J.J. & FANNING, S. 2010. *Cronobacter* (*Enterobacter sakazakii*): an opportunistic foodborne pathogen. *Foodborne Pathogens and Disease*, 7, 339-350.

HEDEGAARD, J., STEFFENSEN, S.A., NORSKOV-LAURITSEN, N., MORTENSEN, K.K. & SPERLING-PETERSEN, H.U. 1999. Identification of *Enterobacteriaceae* by partial sequencing of the gene encoding translation initiation factor 2. *International Journal of Systematic and Evolutionary Microbiology*, 49, 1531-1538.

HEIN, I., GADZOV, B., SCHODER, D., FOISSY, H., MALORNY, B. & WAGNER, M. 2009. Temporal and spatial distribution of *Cronobacter* isolates in a milk powder processing plant determined by pulsed-field gel electrophoresis. *Foodborne Pathogens and Disease*, 6, 225-233.

HIMELRIGHT, I., HARRIS, E., LORCH, V. & ANDERSON, M. 2002. *Enterobacter sakazakii* infections associated with the use of powdered infant formula—Tennessee, 2001. *Journal of American Medical Association*, 287, 2204–2205.

HOLÝ, O., MATOUSKOVÁ, I., HOLÝ, V., KOUKALOVÁ, D. & CHMELAR, D. 2011. Isolation of *Cronobacter* spp. (formerly *Enterobacter sakazakii*) from nostrils of healthy stable horse--short communication. *Epidemiologie, mikrobiologie, imunologie: casopis Spolecnosti pro epidemiologii a mikrobiologii Ceské lékarské spolecnosti JE Purkyne*, 60, 167.

HORLACHER R. & BOOS W. 1997. Characterization of TreR, the major regulator of the *Escherichia coli* trehalose system. *Journal of Biological Chemistry*, 272, 13026–13032.

HORVATH, P. & BARRANGOU, R. 2010. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science*, 327, 167-170.

HUANG, S. H., CHEN, Y.H., FU, Q., STINS, M., WANG, Y., WASS, C. & KIM, K.S. 1999. Identification and characterization of an *Escherichia coli* invasion gene locus, *ibeB*, required for penetration of brain microvascular endothelial cells. *Infection and Immunity*, 67, 2103–2109.

HUANG, S. H., STINS, M.F. & KIM, K.S. 2000. Bacterial penetration across the blood-brain barrier during the development of neonatal meningitis. *Microbes and Infection*, 2, 1237-1244.

HUNKAPILLER, T., KAISER, R.J., KOOP, B.F. & HOOD, L. 1991. Large-scale and automated DNA sequence determination. *Science*, 254, 59-67.

HURRELL, E., KUCEROVA, E., LOUGHLIN, M., CAUBILLA-BARRON, J., HILTON, A., ARMSTRONG, R., SMITH, C., GRANT, J., SHOO, S. & FORSYTHE, S. 2009a. Neonatal enteral feeding tubes as loci for colonisation by members of the *Enterobacteriaceae*. *BMC Infectious Diseases*, 9, 146.

HURRELL, E., KUCEROVA, E., LOUGHLIN, M., CAUBILLA-BARRON, J. & FORSYTHE, S. 2009b. Biofilm formation on enteral feeding tubes by *Cronobacter sakazakii*, *Salmonella* serovars and other *Enterobacteriaceae. International Journal of Food Microbiology*, 136, 227-231.

HUSON, D.H. & BRYANT, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254-67.

INTERNATIONAL COMMISSION ON STRATIGRAPHY. Available at: http://www.stratigraphy.org. (Accessed 3rd October 2012).

ISHINO, Y., SHINAGAWA, H., MAKINO, K., AMEMURA, M. & NAKATA, A. 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, 169, 5429–5433.

IVERSEN, C. & FORSYTHE, S. 2003. Risk profile of *Enterobacter sakazakii*, an emergent pathogen associated with infant milk formula. *Trends in Food Science and Technology*, 14, 443-454.

IVERSEN, C., DRUGGAN, P. & FORSYTHE, S.J. 2004. A selective differential medium for *Enterobacter sakazakii*. *International Journal of Food Microbiology*, 96, 133-139.

IVERSEN, C., WADDINGTON, M., ON, S.L.W. & FORSYTHE, S. 2004a. Identification and phylogeny of *Enterobacter sakazakii* relative to *Enterobacter* and *Citrobacter* species. *Journal of Clinical Microbiology*, 42, 5368-5370.

IVERSEN, C., LANE, M. & FORSYTHE, S. J. 2004b. The growth profile, thermotolerance and biofilm formation of *Enterobacter sakazakii* grown in infant formula milk. *Letters in Applied Microbiology*, 38, 378-382.

IVERSEN, C. & FORSYTHE, S. 2004. Isolation of *Enterobacter sakazakii* and other *Enterobacteriaceae* from powdered infant formula milk and related products. *Food Microbiology*, 21, 771-777.

IVERSEN, C., LANCASHIRE, L., WADDINGTON, M., FORSYTHE, S. & BALL, G. 2006a. Identification of *Enterobacter sakazakii* from closely related species: the use of artificial neural networks in the analysis of biochemical and 16S rDNA data. *BMC Microbiology*, 6, 28.

IVERSEN, C., WADDINGTON, M., FARMER, J.J. III & FORSYTHE, S. 2006b. The biochemical differentiation of *Enterobacter sakazakii* genotypes. *BMC Microbiology*, 6, 94.

IVERSEN, C. & FORSYTHE, S. J. 2007. Comparison of media for the isolation of *Enterobacter sakazakii*. *Applied and Environmental Microbiology*, 73, 48-52.

IVERSEN, C., LEHNER, A., MULLANE, N., BIDLAS, E., CLEENWERCK, I., MARUGG, J., FANNING, S., STEPHAN, R. & JOOSTEN, H. 2007. The taxonomy of *Enterobacter sakazakii*: proposal of a new genus *Cronobacter* gen. nov. and descriptions of *Cronobacter sakazakii* comb. nov. *Cronobacter sakazakii* subsp. *sakazakii*, comb. nov., *Cronobacter sakazakii* subsp. *malonaticus* subsp. nov., *Cronobacter turicensis* sp. nov., *Cronobacter muytjensii* sp. nov., *Cronobacter dublinensis* sp. nov. and *Cronobacter* genomospecies 1. *BMC Evolutionary Biology*, 7, 64.

IVERSEN, C., MULLANE, N., MCCARDELL, B., TALL, B.D., LEHNER, A., FANNING, S., STEPHAN, R. & JOOSTEN, H. 2008. *Cronobacter* gen. nov., a new genus to accommodate the biogroups of *Enterobacter sakazakii*, and proposal of *Cronobacter sakazakii* gen. nov., comb. nov., *Cronobacter malonaticus* sp. nov., *Cronobacter turicensis* sp. nov., *Cronobacter 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 Biology*, 58, 1442-1447.

JACOBS, C., BRAUN, P. & HAMMER, P. 2011. Reservoir and routes of transmission of *Enterobacter sakazakii* (*Cronobacter* spp.) in a milk powder-producing plant. *Journal of Dairy Science*, 94, 3801-3810.

JARRELL, K.F. & MCBRIDE, M.J. 2008. The surprisingly diverse ways that prokaryotes move. *Nature Reviews Microbiology*, 6, 466-476.

JARVIS, C. 2005. Fatal *Enterobacter sakazakii* infection associated with powdered infant formula in a neonatal intensive care unit in New Zealand. *American Journal of Infection Control*, 23, e19.

JARVIS, K.G., GRIM, C.J., FRANCO, A.A., GOPINATH, G., SATHYAMOORTHY, V., HU, L., SADOWSKI, J.A., LEE, C.S. & TALL, B.D. 2011. Molecular characterization of *Cronobacter* lipopolysaccharide O-antigen gene clusters and development of serotype-specific PCR assays. *Applied and Environmental Microbiology*, 77, 4017–4026.

JOHNSON, T.J. & NOLAN, L.K. 2009. Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiology and Molecular Biology Reviews*, 73, 750-774.

JOKER, R. N., NORHOLM, T. & SIBONI, K. E. 1965. A case of neonatal meningitis caused by a yellow enterobacter. *Danish Medical Bulletin*, 12, 128.

JOLLEY, K.A., FEIL, E.J., CHAN, M.S. & MAIDEN, M.C. 2001. Sequence type analysis and recombinational tests (START). *Bioinformatics*, 17, 1230-1231.

JOLLEY, K.A. & MAIDEN, M.C. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level**.** *BMC Bioinformatics*, 11, 595.

JONG, M.F. & TSOLIS, R.M. 2012. Brucellosis and type IV secretion. *Future Microbiology,* 7, 47-58.

JOSEPH, S. & FORSYTHE, S.J. 2011. Predominance of *Cronobacter sakazakii* ST4 in neonatal infections. *Emerging Infectious Diseases,* 17, 1713-1715.

JOSEPH, S., CETINKAYA, E., DRAHOVSKA, H., LEVICAN, A., FIGUERAS, M.J. & FORSYTHE, S.J. 2012a. *Cronobacter condimenti* sp. nov., isolated from spiced meat and *Cronobacter universalis* sp. nov., a novel species designation for *Cronobacter* sp. genomospecies 1, recovered from a leg infection, water, and food ingredients. *International Journal of Systematic and Evolutionary Microbiology*, 62, 1277-1283.

JOSEPH, S., SONBOL, H., HARIRI, S., DESAI, P., McCLELLAND, M. & FORSYTHE, S. 2012b. Diversity of the *Cronobacter* genus as revealed by multi locus sequence typing. *Journal of Clinical Microbiology,* 50, 3031-3039.

JOSEPH, S., DESAI, P., JI, Y., HAMBY, S. E., MASOOD, N., HARIRI, S., SONBOL, H., CUMMING, C.A., RICO, A., SHIH, S. M., DEGORICJA, L., BRZOSKA, P., CHUZHANOVA, N., MCCLELLAND, M.., FURTADO, M. R. & FORSYTHE, S.J. 2012c. Comparative analysis of genome sequences covering the seven *Cronobacter* species. *PLoS ONE*, 7, e49455.

JOSEPH, S. & FORSYTHE, S.J. 2012. Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Frontiers in Food Microbiology,* 3, 397.

JOSEPH, S., HARIRI, S., FORSYTHE, S.J. 2013a. Lack of continuity between *Cronobacter* biotypes and species as determined using multilocus sequence typing. *Molecular and Cellular Probes,* 27, 137-139.

JOSEPH, S., HARIRI, S., MASOOD, N., FORSYTHE, S. 2013b. Sialic acid utilization by *Cronobacter sakazakii*. *Microbial Informatics and Experimentation,* 3, 3.

KAMPFER, P. & GLAESER, S.P. 2012. Prokaryotic taxonomy in the sequencing era – the polyphasic approach revisited. *Environmental Microbiology*, 14, 291-317.

KANDHAI, M. C., REIJ, M. W., GORRIS, L. G., GUILLAUME-GENTIL, O. & VAN SCHOTHORST, M. 2004. Occurrence of *Enterobacter sakazakii* in food production environments and households. *The Lancet*, 363, 39-40.

KENT, W.J. 2002. BLAT-the BLAST-like alignment tool. *Genome Research*, 12, 656-664.

KEYSER, M., WITTHUHN, R.C., RONQUEST, L.C. & BRITZ, T.J. 2003. Treatment of winery effluent with upflow anaerobic sludge blanket (UASB) – granular sludges enriched with *Enterobacter sakazakii*. *Biotechnology Letters*, 25, 1893–1898.

KIM, H. & BEUCHAT, L. R. 2005. Survival and growth of *Enterobacter sakazakii* on fresh-cut fruits and vegetables and in unpasteurized juices as affected by storage temperature. *Journal of Food Protection*, 68, 2541-2552.

KIM, H., RYU, J. H. & BEUCHAT, L. R. 2006. Attachment of and biofilm formation by *Enterobacter sakazakii* on stainless steel and enteral feeding tubes. *Applied and Environmental Microbiology*, 72, 5846-5856.

KIM, K.P., KLUMPP, J. & LOESSNER, M.J. 2007. *Enterobacter sakazakii* bacteriophages can prevent bacterial growth in reconstituted infant formula. *International Journal of Food Microbiology*, 115, 195–203.

KIM, K. S. 2008. Mechanisms of microbial traversal of the blood–brain barrier. *Nature Reviews Microbiology*, 6, 625-634.

KIM, K. P. & LOESSNER, M. J. 2008. *Enterobacter sakazakii* invasion in human intestinal Caco-2 cells requires the host cell cytoskeleton and is enhanced by disruption of tight junction. *Infection and Immunity*, 76, 562-570.

KIM, T. J., WENG, W. L., SILVA, J. L., JUNG, Y. S. & MARSHALL, D. 2010a. Identification of natural antimicrobial substances in red muscadine juice against *Cronobacter sakazakii*. *Journal of Food Science*, 75, M150-M154.

KIM, K., KIM, K. P., CHOI, J., LIM, J. A., LEE, J., HWANG, S. & RYU, S. 2010b. Outer membrane proteins A (OmpA) and X (OmpX) are essential for basolateral invasion of *Cronobacter sakazakii*. *Applied and Environmental Microbiology*, 76, 5188-5198.

KLASSEN, J.L. & CURRIE, C.R. 2012. Gene fragmentation in bacterial draft genomes: extent, consequences and mitigation. *BMC Genomics*, 13, 14.

KLINE, M.W. 1988. *Citrobacter* meningitis and brain abscess in infancy: epidemiology, pathogenesis and treatment. *The Journal of Pediatrics*, 113, 430-434.

KONSTANTINIDIS, K.T. & TIEDJE, J.M. 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences USA*, 102, 2567-2572.

KONSTANTINIDIS, K.T., RAMETTE, A. & TIEDJE, J.M. 2006. Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Applied and Environmental Microbiology*, 72, 7286-7293.

KOTETISHVILI, M., KREGER, A., WAUTERS, G., MORRIS, Jr., J.G., SULAKVELIDZE, A. & STINE, O.C. 2005. Multilocus sequence typing for studying genetic relationships among *Yersinia* species. *Journal of Clinical Microbiology*, 43, 2674-2684.

KOTHARY, M. H., MCCARDELL, B. A., FRAZAR, C. D., DEER, D. & TALL, B. D. 2007. Characterization of the zinc-containing metalloprotease encoded by *zpx* and development of a species-specific detection method for *Enterobacter sakazakii*. *Applied and Environmental Microbiology*, 73, 4142-4151.

KROPINSKI, A.M. 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *Journal of Bacteriology*, 182, 6066-6074.

KUCEROVA, E., CLIFTON, S.W., XIA, X.Q., LONG, F., PORWOLLIK, S., FULTON, L., FRONICK, C., MINX, P., KYUNG, K., WARREN, W., FULTON, R., FENG, D., WOLLAM, A., SHAH, N., BHONAGIRI, V., NASH, W.E., HALLSWORTH-PEPIN, K., WILSON, R.K., MCCLELLAND, M. & FORSYTHE, S.J. 2010. Genome sequence of *Cronobacter sakazakii* BAA-894 and comparative genomic hybridization analysis with other *Cronobacter* species. *PLoS ONE,* 5, e9556.

KUCEROVA, E., JOSEPH, S. & FORSYTHE, S. 2011. The *Cronobacter* genus: ubiquity and diversity. *Quality Assurance and Safety of Crops & Foods*. 3, 104-122.

KUHNERT, P., KORCZAK, B.M., STEPHAN, R., JOOSTEN, H. & IVERSEN, C. 2009. Phylogeny and prediction of genetic similarity of *Cronobacter* and related taxa by multilocus sequence analysis (MLSA). *International Journal of Food Microbiology,* 136, 152-158.

KUO, C. & OCHMAN, H. 2009. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biology Direct,* 4, 35.

KURTZ, S., PHILLIPPY, A., DELCHER, A.L., SMOOT, M., SHUMWAY, M., ANTONESCU, C. & SALZBERG, S.L. 2004. Versatile and open software for comparing large genomes. *Genome Biology,* 5, R12.

KUZINA, L. V., PELOQUIN, J. J., VACEK, D. C. & MILLER, T. A. 2001. Isolation and identification of bacteria associated with adult laboratory Mexican fruit flies, *Anastrepha ludens* (*Diptera: Tephritidae*). *Current Microbiology,* 42, 290-294.

LACHER, D.W., STEINSLAND, H., BLACK, T.E., DONNENBERG, M.S. & WHITTAM, T.S. 2007. Molecular evolution of typical enteropathogenic *Escherichia coli*: clonal analysis by Multilocus sequence typing and virulence gene allelic profiling. *Journal of Bacteriology*, 189**,** 342-350.

LAI, K. K. 2001. *Enterobacter sakazakii* infections among neonates, infants, children and adults: case reports and a review of the literature. *Medicine (Baltimore)*, 80, 113-122.

LAPAGE, S. P., SNEATH, P. H. A., LESSEL, E. F., SKERMAN, V. B. D., SEELIGER, H. P. R. & CLARK, W. A. 1992. *International Code of Nomenclature of Bacteria (1990 Revision). Bacteriological Code.* Washington, DC: American Society for Microbiology.

LARKIN, M.A., BLACKSHIELDS, G., BROWN, N.P., CHENNA, R., MCGETTIGAN, P.A., MCWILLIAM, H., VALENTIN, F., WALLACE, I.M., WILM, A., LOPEZ, R., THOMPSON, J.D., GIBSON, T.J. & HIGGINS, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

LEE, Y.D., PARK, J.H. & CHANG, H.I. 2011a. Genomic sequence analysis of virulent *Cronobacter sakazakii* bacteriophage ES2. *Archives of Virology,* 156, 2105-2108.

LEE, Y.D., PARK, J.H. & CHANG, H.I. 2011b. Complete genomic sequence of virulent *Cronobacter sakazakii* phage ESSI-2 isolated from swine faeces. *Archives of Virology,* 156, 721-724.

LEE, J.H., CHOI, Y., SHIN, H., LEE, J. & RYU, S. 2012a. Complete genome sequence of *Cronobacter sakazakii* temperate bacteriophage phiES15. *Journal of Virology,* 86, 7713-7714.

LEE, Y.D. & PARK, J.H. 2012b. Complete genome of temperate phage ENT39118 from *Cronobacter sakazakii*. *Journal of Virology,* 86, 5400-5401.

LEE, Y.D., KIM, J.Y., PARK, J.H. & CHANG, H. 2012c. Genomic analysis of bacteriophage ESP2949-1, which is virulent for *Cronobacter sakazakii*. *Archives of Virology,* 157, 199-202.

LEHNER, A., TASARA, T. & STEPHAN, R. 2004. 16S rRNA gene based analysis of *Enterobacter sakazakii* strains from different sources and development of a PCR assay for identification. *BMC Microbiology*, 4, 43.

LEHNER, A., FRICKER-FEER, C. & STEPHAN, R. 2012. Identification of the recently described *Cronobacter condimenti* by an *rpoB*-gene-based PCR system. *Journal of Medical Microbiology*, 61, 1034-1035.

LEUNG, K.Y., SIAME, B.A., SNOWBALL, H. & MOK, Y-K. 2011. Type VI secretion regulation: crosstalk and intracellular communication. *Current Opinion in Microbiology,* 14, 9-15.

LI, P., HWANG, I., MIYAGI, H., TRUE, H. & FARRAND, S.K. 1999. Essential components of the Ti plasmid *trb* system, a Type IV macromolecular transporter. *Journal of Bacteriology*, 181, 5033-5041.

LI, L., STOECKERT, C.J. Jr. & ROOS, D.S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research,* 13, 2178-2189.

LIBRADO, P. & ROZAS, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25, 1451-1452.

LIEBERMAN, J.A., FROST, N.A., HOPPERT, M., FERNANDES, P.J., VOGT, S.L., RAIVIO, T.L., BLANPIED, T.A. & DONNENBERG, M.S. 2012. Outer membrane targeting, ultrastructure, and single molecule localization of the enteropathogenic *Escherichia coli* type IV pilus secretin BfpB. *Journal of Bacteriology,* 194, 1646-1658.

LIMA-MENDEZ, G., VAN HELDEN, J., TOUSSAINT, A. & LEPLAE, R. 2008. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics*, 24, 863–865.

LIN, P. W. & STOLL, B. J. 2006. Necrotising enterocolitis. *The Lancet,* 368, 1271-1283.

LIU, Y., GAO, Q., ZHANG, X., HOU, Y., YANG, J. & HUANG, X. 2006a. PCR and oligonucleotide array for detection of *Enterobacter sakazakii* in infant formula. *Molecular and Cellular Probes*, 20, 11-17.

LIU, Y., CAI, X., ZHANG, X., GAO, Q., YANG, X., ZHENG, Z., LUO, M. & HUANG, X. 2006b. Real time PCR using TaqMan and SYBR Green for detection of *Enterobacter sakazakii* in infant formula. *Journal of Microbiological Methods*, 65, 21-31.

LOMAN, N.J., CONSTANTINIDOU, C., CHAN, J.Z.M., HALACHEV, M., SERGEANT, M., PENN, C.W., ROBINSON, E.R. & PALLEN, M.J. 2012a. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*, 10, 599-606.

LOMAN, N.J., MISRA, R.V., DALLMAN, T.J., CONSTANTINIDOU, C., GHARBIA, S.E., WAIN, J. & PALLEN, M.J. 2012b. Performance comparison of bench-top high-throughput sequencing platforms. *Nature Biotechnology*, 30, 434-439.

MACLEAN, L.L., PAGOTTO, F., FARBER, J.M. & PERRY, M.B. 2009a. Structure of the antigenic repeating pentasaccharide unit of the LPS O-polysaccharide of *Cronobacter sakazakii* implicated in the Tennessee outbreak. *Biochemistry and Cell Biology (Biochimie et Biologie Cellulaire)*, 87, 459–465.

MACLEAN, L.L., VINOGRADOV, E., PAGOTTO, F., FARBER, J.M. & PERRY, M.B. 2009b. Characterization of the O-antigen in the lipopolysaccharide of *Cronobacter (Enterobacter) malonaticus* 3267. *Biochemistry and Cell Biology*, 87, 927-32.

MACLEAN, L.L., PAGOTTO, F., FARBER, J.M. & PERRY, M.B. 2009c.The structure of the O-antigen in the endotoxin of the emerging food pathogen *Cronobacter* (*Enterobacter*) *muytjensii* strain 3270. *Carbohydrate Research*, 344, 667–671.

MAIDEN, M.C., BYGRAVES, J.A., FEIL, E., MORELLI, G., RUSSELL, J.E., URWIN, R., ZHANG, Q., ZHOU, J., ZURTH, K., CAUGANT, D.A., FEAVERS, I.M., ACHTMAN, M. & SPRATT, B.G. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the USA*, 95, 3140-3145.

MALORNY, B. & WAGNER, M. 2005. Detection of *Enterobacter sakazakii* strains by real-time PCR. *Journal of Food Protection*, 68, 1623-1627.

MANGE, J. P., STEPHAN, R., BOREL, N., WILD, P., KIM, K. S., POSPISCHIL, A. & LEHNER, A. 2006. Adhesive properties of *Enterobacter sakazakii* to human epithelial and brain microvascular endothelial cells. *BMC Microbiology*, 6, 58.

MARDIS, E.R. 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, 9, 387-402.

MARTINO, M.E., FASOLATO, L., MONTEMURRO, F., ROSTEGHIN, M., MANFRIN, A., PATARNELLO, T., NOVELLI, E. & CARDAZZO, B. 2011. Determination of microbial diversity of *Aeromonas* strains on the basis of multilocus sequence typing, phenotype, and presence of putative virulence genes. *Applied and Environmental Microbiology*, 77, 4986-5000.

MATTICK, J.S. 2002. Type IV pili and twitching motility. *Annual Reviews of Microbiology,* 56, 289–314.

MAYNARD SMITH, J., SMITH, N.H., O'ROURKE, M. & SPRATT, B.G. 1993. How clonal are bacteria? *Proceedings of the National Academy of Sciences USA*, 90, 4384-4388.

MCGHEE, G.C. & SUNDIN, G.W. 2012. *Erwinia amylovora* CRISPR elements provide new tools for evaluating strain diversity and for microbial source tracking. *PLoS ONE,* 7, e41706.

MCQUISTON, J.R., HERRERA-LEON, S., WERTHEIM, B.C., DOYLE, J., FIELDS, P.I., TAUXE, R.V. & LOGSDON, J.M. JR. 2008. Molecular phylogeny of the *Salmonellae*: Relationships among *Salmonella* species and subspecies determined from four housekeeping genes and evidence of lateral gene transfer events. *Journal of Bacteriology,* 190, 7060-7067.

MEDINI, D., DONATI, C., TETTELIN, H., MASIGNANI, V. & RAPPUOLI, R. 2005. The microbial pan-genome. *Current Opinion in Genetics and Development,* 15, 589-594.

MERGA, J.Y., LEATHERBARROW, A.J., WINSTANLEY, C., BENNETT, M., HART, C.A., MILLER, W.G. & WILLIAMS, N.J. 2011. Comparison of *Arcobacter* isolation methods, and diversity of *Arcobacter* spp. in Cheshire, United Kingdom. *Applied and Environmental Microbiology,* 77, 1646-1650.

MÉRIC, G., KEMSLEY, E. K., FALUSH, D., SAGGERS, E. J. & LUCCHINI, S. 2012. Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*. *Environmental Microbiology*. doi: 10.1111/j.1462-2920.2012.02852.x.

METZKER, M.., L. 2010. Sequencing technologies – the next generation. *Nature Reviews Genetics,* 11, 31-46.

MIRA, A., MARTIN-CUADRADO, A.B., D'AURIA, G. & RODRIGUEZ-VALERA, F. 2010. The bacterial pan-genome: a new paradigm in microbiology. *International Microbiology,* 13, 45-57.

MITTAL, R., WANG, Y., HUNTER, C. J., GONZALEZ-GOMEZ, I. & PRASADARAO, N. V. 2009. Brain damage in newborn rat model of meningitis by *Enterobacter sakazakii*: a role for outer membrane protein A. *Laboratory Investigation*, 89, 263-277.

MMOLAWA, P.T., SCHMEIGER, H. & HEUZENROEDER, M.W. 2003. Bacteriophage ST64B, a genetic mosaic of genes from diverse sources isolated from *Salmonella enterica* serovar Typhimurium DT 64. *Journal of Bacteriology*, 185, 6481-6485.

MOHAN NAIR, M. K., VENKITANARAYANAN, K., SILBART, L. K. & KIM, K. S. 2009. Outer membrane protein A (OmpA) of *Cronobacter sakazakii* binds fibronectin and contributes to invasion of human brain microvascular endothelial cells. *Foodborne Pathogens and Disease*, 6, 495-501.

MOSSO, M. D. L. A., ROSA, M. D. C., VIVAR, C. & MEDINA, M. D. R. 1994. Heterotrophic bacterial populations in the mineral waters of thermal springs in Spain. *Journal of Applied Microbiology*, 77, 370-381.

MRAMBA, F., BROCE, A. & ZUREK, L. 2006. Isolation of *Enterobacter sakazakii* from stable flies, *Stomoxys calcitrans* L.(*Diptera: Muscidae*). *Journal of Food Protection*, 69, 671-673.

MULLANE, N. R., RYAN, M., IVERSEN, C., MURPHY, M., O'GAORA, P., QUINN, T., WHYTE, P., WALL, P.G. & FANNING, S. 2008a. Development of multiple-locus variable-number tandem-repeat analysis for the molecular subtyping of *Enterobacter sakazakii. Applied and Environmental Microbiology*, 74, 1223-1231.

MULLANE, N.P., O'GAORA, P., NALLY, J.E., IVERSEN, C., WHYTE, P., WALL, P.G. & FANNING, S. 2008b. Molecular analysis of the *Enterobacter sakazakii* O-antigen gene locus. *Applied and Environmental Microbiology*, 74, 3783-3794.

MUNSON, G.P., LAM, D.L., OUTTEN, F.W. & O'HALLORAN, T.V. 2000. Identification of a copper-responsive two-component system on the chromosome of *Escherichia coli* K-12. *Journal of Bacteriology*, 182, 5864-5871.

MURRAY, R.G.E., BRENNER, D.J., COLWELL, R.R., DE VOS, P., GOODFELLOW, M., GRIMONT, P.A.D., PFENNIG, N., STACKEBRANDT, E. & ZAVARZIN, G.A. 1990. Report of the ad hoc committee on approaches to taxonomy within the *Proteobacteria. International Journal of Systematic Bacteriology*, 40, 213-215.

MUYTJENS, H.L., ZANEN, H.C., SONDERKAMP, H.J., KOLLEE, L.A., WACHSMUTH, K. & FARMER, J.J. 1983. Analysis of eight cases of neonatal meningitis and sepsis due to *Enterobacter sakazakii. Journal of Clinical Microbiology*, 18, 115-120.

MUYTJENS, H. L. & VAN DRUTEN, H. A. 1984. Enzymatic profiles of *Enterobacter sakazakii* and related species with special reference to the alpha-glucosidase reaction and reproducibility of the test system. *Journal of Clinical Microbiology*, 20, 684-686.

MUYTJENS, H.L., ROELOFS-WILLEMSE, H. & JASPER, G.H.J. 1988. Quality of powdered substitutes for breast milk with regard to members of the family *Enterobacteriaceae*. *Journal of Clinical Microbiology*, 26, 743–746.

NAIR, M.K.M. & VENKITANARAYANAN, K.S. 2006. Cloning and sequencing of the *ompA* gene of *Enterobacter sakazakii* and development of an *ompA*- targeted PCR for rapid detection of *Enterobacter sakazakii* in infant formula. *Applied and Environmental Microbiology*, 72, 2539-2546.

NAIR, M. K. M. & VENKITANARAYANAN, K. 2007. Role of bacterial OmpA and host cytoskeleton in the invasion of human intestinal epithelial cells by *Enterobacter sakazakii*. *Pediatric Research*, 62, 664-669.

NAIR, M.K.M., VENKITANARAYANAN, K., SILBART, L.K. & KIM, K.S. 2009. Outer membrane protein A (OmpA) of *Cronobacter sakazakii* binds fibronectin and contributes to invasion of human brain microvascular endothelial cells. *Foodborne Pathogens and Disease*, 6, 495-501.

NAZAROWEC-WHITE, M. & FARBER, J.M. 1997. Incidence, survival, and growth of *Enterobacter sakazakii* in infant formula. *Journal of Food Protection*, 60, 226-230.

NAZAROWEC-WHITE, M. & FARBER, J. M. 1999. Phenotypic and genotypic typing of food and clinical isolates of *Enterobacter sakazakii*. *Journal of Medical Microbiology*, 48, 559-567.

NAZAROWEC-WHITE, M., MCKELLAR, R. C. & PIYASENA, P. 1999. Predictive modelling of *Enterobacter sakazakii* inactivation in bovine milk during high-temperature short-time pasteurization. *Food Research International*, 32, 375-379.

NCBI MICROBIAL GENOMES.
Available at: http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html
(Accessed 7th September 2012).

NEELAM, M., NAWAZ, Z. & RIAZUDDIN, S. 1987. Hydrocarbon biodegradation: Biochemical characterization of bacteria isolated from local soils. *Pakistan Journal of Scientific and Industrial Research*, 30, 382-385.

NUCCI, S. & BÄUMLE, A.J. 2007. Evolution of the chaperone/usher assembly pathway. Fimbrial classification goes Greek. *Microbiology and Molecular Biology Reviews*, 71, 551.

OCHMAN, H., ELWYN, S. & MORAN, N.A. 1999. Calibrating bacterial evolution. *Proceedings of the National Academy of Sciences of the USA*, 96, 12638-12643.

OGATA, H., GOTO, S., SATO, K., FUJIBUCHI, W., BONO, H. & KANEHISA, M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 27, 29-34.

OSAILI, T. & FORSYTHE, S. 2009. Desiccation resistance and persistence of *Cronobacter* species in infant formula. *International Journal of Food Microbiology*, 136, 214-220.

OTT, M., HACKER, J., SCHMOLL, T., JARCHAU, T., KORHONEN, T.K. & GOEBEL, W. 1986. Analysis of the genetic determinants coding for the S-fimbrial adhesin (sfa) in different *Escherichia coli* strains causing meningitis or urinary tract infections. *Infection and Immunity*, 54, 646-653.

PAGOTTO, F. J., NAZAROWEC-WHITE, M., BIDAWID, S. & FARBER, J. M. 2003. *Enterobacter sakazakii*: infectivity and enterotoxin production in vitro and in vivo. *Journal of Food Protection*, 66, 370-375.

PARADIS, S., BOISSINOT, M., PAQUETTE, N., BELANGER, S. D., MARTEL, E. A., BOUDREAU, D. K., PICARD, F.J., OUELLETTE, M., ROY, P.H. & BERGERON, M. G. 2005. Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase β-subunit. *International Journal of Systematic and Evolutionary Microbiology*, 55, 2013-2025.

PAVA-RIPOLL, M., PEARSON, R. E. G., MILLER, A. K. & ZIOBRO, G. C. 2012. Prevalence and relative risk of *Cronobacter* spp., *Salmonella* spp. and *Listeria monocytogenes* associated with the body surface and the guts of individual filth flies. *Applied and Environmental Microbiology*, 78, 7891-7902.

PAVON, A.B.I. & MAIDEN, M.C.J. 2009. Multilocus sequence typing. Edited by D.A. Caugant. *Molecular Epidemiology of Microorganisms*, Methods in Molecular Biology, Vol. 551: Humana Press. pp. 129-140.

PAYNE, S. M., WYCKOFF, E. E., MURPHY, E. R., OGLESBY, A. G., BOULETTE, M. L., & DAVIES, N. M. 2006. Iron and pathogenesis of *Shigella*: iron acquisition in the intracellular environment. *Biometals*, 19, 173-180.

PENWELL, W.F., ARIVETT, B.A. & ACTIS, L.A. 2012. The *Acinetobacter baumannii entA* gene located outside the acinetobactin cluster is critical for siderophore production, iron acquisition and virulence. *PLoS ONE*, 7, e36493.

PEZZICOLI, A, RUGGIERO, P., AMERIGHI, F., TELFORD, J.L. & SORIANI, M. 2012. Exogenous sialic acid transport contributes to Group B Streptococcus infection of mucosal surfaces. *The Journal of Infectious Diseases*, 206, 924-931.

PRESSLER, U., STAUDENMAIER, H., ZIMMERMANN, L. & BRAUN, V. 1988. Genetics of the iron dicitrate transport system of *Escherichia coli*. *Journal of Bacteriology*, 170, 2716-2724.

PRIDE, D.T., SUN, C.L., SALZMAN, J., RAO, N., LOOMER, P., ARMITAGE, G.C., BANFIELD, J.F. & RELMAN, D.A. 2011. Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome Research*, 21, 126-136.

PROUDY, I., BOUGLE, D., COTON, E., COTON, M., LECLERCQ, R. & VERGNAUD, M. 2008. Genotypic characterization of *Enterobacter sakazakii* isolates by PFGE, BOX-PCR and sequencing of the *fliC* gene. *Journal of Applied Microbiology*, 104, 26-34.

PUKATZKI, S., MA, A.T., STURTEVANT, D., KRASTINS, B., SARRACINO, D., NELSON, W.C., HEIDELBERG, J.F. & MEKALANOS, J.J. 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proceedings of the National Academy of Sciences USA*, 103, 1528-1533.

RANKIN, D.J., ROCHA, E.P.C. & BROWN, S.P. 2011. What traits are carried on mobile genetic elements, and why? *Heredity*, 106, 1-10.

RIEDEL, K. & LEHNER, A. 2007. Identification of proteins involved in osmotic stress response in *Enterobacter sakazakii* by proteomics. *Proteomics*, 7, 1217-1231.

ROTHBERG, J.M., HINZ, W., REARICK, T.M., SCHULTZ, J., MILESKI, W., DAVEY, M., LEAMON, J.H., JOHNSON, K., MILGREW, M.J., EDWARDS, M., HOON, J., SIMONS, J.F., MARRAN, D., MYERS, J.W., DAVIDSON, J.F., BRANTING, A., NOBILE, JR., PUC, B.P., LIGHT, D., CLARK, T.A., HUBER, M., BRANCIFORTE, J.T., STONER, I.B., CAWLEY, S.E., LYONS, M., FU, Y., HOMER, N., SEDOVA, M., MIAO, X., REED, B., SABINA, J., FEIERSTEIN, E., SCHORN, M., ALANJARY, M., DIMALANTA, E., DRESSMAN, D., KASINSKAS, R., SOKOLSKY, T., FIDANZA, J.A., NAMSARAEV, E., MCKERNAN, K.J.,

WILLIAMS, A., ROTH, G.T. & BUSTILLO, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475, 348-352.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M.A. & BARRELL, B. 2000. Artemis: Sequence visualization and annotation. *Bioinformatics*, 16, 944-945.

SALMOND, G.P.C. 1994. Secretion of extracellular virulence factors by plant pathogenic bacteria. *Annual Review of Phytopathology*, 32, 181-200.

SAMUEL, G. & REEVES, P. 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate Research*, 338, 2503-2519.

SANGER, F., NICKLEN, S. & COULSON, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA*, 74, 5463-5467.

SCHEEPE-LEBERKUHNE, M. & WAGNER F. 1986. Optimization and preliminary characterization of an exopolysaccharide synthesised by *Enterobacter sakazakii*. *Biotechnology Letter*, 8, 695–700.

SCHMID, M., IVERSEN, C., GONTIA, I., STEPHAN, R., HOFMANN, A., HARTMANN, A., JHA, B., EBERL, L., RIEDEL, K. & LEHNER, A. 2009. Evidence for a plant-associated natural habitat for *Cronobacter* spp. *Research in Microbiology*, 160, 608-614.

SEO, K.H. & BRACKETT, R.E. 2005. Rapid, specific detection of *Enterobacter sakazakii* in infant formula using a real-time PCR assay. *Journal of Food Protection*, 68, 59-63.

SEVERI, E., HOOD, D.W. & THOMAS, G.H. 2007. Sialic acid utilization by bacterial pathogens. *Microbiology*, 153, 2817-2822.

SHAKHNOVICH, E. A., KING, S. J. & WEISER, J. N. 2002. Neuraminidase expressed by *Streptococcus pneumoniae* desialylates the lipopolysaccharide of *Neisseria meningitidis* and *Haemophilus influenzae*: a paradigm for interbacterial competition among pathogens of the human respiratory tract. *Infection and Immunity*, 70, 7161–7164.

SHENDURE, J. & JI, H. 2008. Next-generation DNA sequencing. *Nature Biotechnology*, 26, 1135-1145.

SHIN, H., LEE, J., KIM, Y. & RYU, S. 2012. Complete genome sequence of *Cronobacter sakazakii* bacteriophage CR3. *Journal of Virology*, 86, 6367-6368.

SILVESTRO, D. & MICHALAK, I. 2011. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution*, 12, 335-337.

SINGAMSETTY, V. K., WANG, Y., SHIMADA, H. & PRASADARAO, N. V. 2008. Outer membrane protein A expression in *Enterobacter sakazakii* is required to induce microtubule condensation in human brain microvascular endothelial cells for invasion. *Microbial Pathogenesis*, 45, 181-191.

SJÖBRING, U., POHL, G. & OLSÉN, A. 2006. Plasminogen, absorbed by *Escherichia coli* expressing curli or by *Salmonella enteritidis* expressing thin aggregative fimbriae, can be activated by simultaneously captured tissue‐type plasminogen activator (t‐PA). *Molecular Microbiology,* 14, 443-452.

SMITH, J. M., FEIL, E. J. & SMITH, N. H. 2000. Population structure and evolutionary dynamics of pathogenic bacteria. *Bioessays,* 22, 1115-1122.

SNIPEN, L., ALMOY, T. & USSERY, D.W. 2009. Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics,* 10, 385.

SOONG, G., MUIR, A., GOMEZ, M.I., WAKS, J., REDDY, B., PLANET, P., SINGH, P.K., KANETKO, Y., WOLFGANG, M.C., HSIAO, Y-S., TONG, L. & PRINCE, A. 2006. Bacterial neuraminidase facilitates mucosal infection by participating in biofilm production. *The Journal of Clinical Investigation,* 116, 2297-2305.

SPRATT, B.G. 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Current Opinion in Microbiology,* 2, 312-316.

STACKEBRANDT, E. & GOEBEL, B.M. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology,* 44, 846-849.

STAHL, M, FRIIS, L.M., NOTHAFT, H., LIU, X., LI, J., SZYMANSKI, C.M. & STINTZI, A. 2011. L-Fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proceedings of the National Academy of Sciences of the United States of America,* 108, 7194-7199.

STAMATAKIS, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* 22, 2688-2690.

STEPHAN, R., LEHNER, A., TISCHLER, P. & RATTEI, T. 2010. Complete genome sequence of *Cronobacter turicensis* LMG 23827, a food-borne pathogen causing deaths in neonates. *Journal of Bacteriology,* 193, 309-310.

STEVENSON, G., ANDRIANOPOULOS, K., HOBBS, M. & REEVES, P.R. 1996. Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *Journal of Bacteriology,* 178, 4885-4893.

STOOP, B., LEHNER, A., IVERSEN, C., FANNING, S. & STEPHAN, R. 2009. Development and evaluation of *rpoB* based PCR systems to differentiate the six proposed species within the genus *Cronobacter*. *International Journal of Food Microbiology,* 136, 165-168.

SUN, Y., WANG, M., LIU, H., WANG, J., HE, X., ZENG, J., GUO, X., LI, K., CAO, B. & WANG, L. 2011. Development of an O-Antigen Serotyping Scheme for *Cronobacter sakazakii*. *Applied and Environmental Microbiology,* 77, 2209-2214.

SUN, Y., WANG, M., WANG, Q., CAO, B., HE, X., LI, K., FENG, L. & WANG, L. 2012a. Genetic Analysis of the *Cronobacter sakazakii* O4 to O7 O-antigen gene clusters and development of a PCR Assay for identification of all *C. sakazakii* O serotypes. *Applied and Environmental Microbiology*, 78, 3966-3974.

SUN, Y., ARBATSKY, N. P., WANG, M., SHASHKOV, A. S., LIU, B., WANG, L. & KNIREL, Y. A. 2012b. Structure and genetics of the O−antigen of *Cronobacter turicensis* G3882 from a new serotype, *C. turicensis* O2, and identification of a serotype−specific gene. *FEMS Immunology & Medical Microbiology*, 66, 323-333.

SWAMINATHAN, B., BARRETT, T.J., HUNTER, S.B., TAUXE, R.V. & THE CENTERS FOR DISEASE CONTROL AND PREVENTION PULSENET TASK FORCE. 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerging Infectious Diseases*, 7, 382-389.

TALARICO, S., WHITEFIELD, S.E., FERO, J., HAAS, R. & SALAMA, N.R. 2012. Regulation of *Helicobacter pylori* adherence by gene conversion. *Molecular Microbiology*, 84, 1050-1061.

TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M. & KUMAR, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28, 2731-2739.

TAYLOR, D.E. 1999. Bacterial tellurite resistance. *Trends in Microbiology*, 7, 111-115.

TENG, C-H., CAI, M., SHIN, S., ZIE, Y., KIM, K-J., KHAN, N.A., DI CELLO, F. & KIM, K.S. 2005. *Escherichia coli* K1 RS218 interacts with human brain microvascular endothelial cells via type 1 fimbria bacteria in the fimbriated state. *Infection and Immunity*, 73, 2923-31.

TERRADOT, L. & WAKSMAN, G. 2011. Architecture of the *Helicobacter pylori* Cag-type IV secretion system. *The FEBS Journal*, 278, 1213-1222.

TETTELIN, H., RILEY, D., CATTUTO, C. & MEDINI, D. 2008. Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, 12, 472-477.

THOMPSON, J.D., HIGGINS, D.G. & GIBSON, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673-80.

TINDALL, B.J., ROSELLO-MORA, R., BUSSE, H.J., LUDWIG, W. & KAMPFER, P. 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. *International Journal of Systematic and Evolutionary Microbiology*, 60, 249-266.

TOWNSEND, S.M., HURRELL, E., GONZALEZ-GOMEZ, I., LOWE, L., FRYE, J.G., FORSYTHE, S. & BADGER, J.L. 2007. *Enterobacter sakazakii* invades brain capillary

endothelial cells, persists in human macrophages influencing cytokine secretion and induces severe brain pathology in the neonatal rat. *Microbiology*, 153, 3538-3547.

TOWNSEND, S.M., HURRELL, E., CAUBILLA-BARRON, J., LOC-CARRILLO, C. & FORSYTHE, S.J. 2008a. Characterization of an extended-spectrum beta-lactamase *Enterobacter hormaechei* nosocomial outbreak, and other *Enterobacter hormaechei* misidentified as *Cronobacter (Enterobacter) sakazakii*. *Microbiology*, 154, 3659-3667.

TOWNSEND, S., HURRELL, E. & FORSYTHE, S. 2008b. Virulence studies of *Enterobacter sakazakii* isolates associated with a neonatal intensive care unit outbreak. *BMC Microbiology*, 8, 64.

TRAPPETTI, C., KADIOGLU, A., CARTER, M., HAYRE, J., IANNELLI, F., POZZI, G., ANDREW, P.W. & OGGIONI, M.R. 2009. Sialic acid: A preventable signal for pneumococcal biofilm formation, colonization, and invasion of the host. *The Journal of Infectious Diseases*, 199, 1497-1505.

TSENG, T, TYLER, B.M. & SETUBAL, J.C. 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiology*, 9, S2.

TURCK, D. 2012. Safety aspects in preparation and handling of infant food. *Annals of Nutrition and Metabolism*, 60, 211-214.

TURCOVSKÝ, I., KUNIKOVÁ, K., DRAHOVSKÁ, H. & KACLÍKOVÁ, E. 2011. Biochemical and molecular characterization of *Cronobacter* spp. (formerly *Enterobacter sakazakii*) isolated from foods. *Antonie van Leeuwenhoek*, 99, 257-269.

URDIAIN, M., LÓPEZ-LÓPEZ, A., GONZALO, C., BUSSE, H. J., LANGER, S., KÄMPFER, P. & ROSSELLÓ-MORA, R. 2008. Reclassification of *Rhodobium marinum* and *Rhodobium pfennigii* as *Afifella marina* gen. nov. comb. nov. and *Afifella pfennigii* comb. nov., a new genus of photoheterotrophic Alphaproteobacteria and emended descriptions of *Rhodobium*, *Rhodobium orientis*and *Rhodobium gokarnense*. *Systematic and Applied Microbiology*, 31, 339–351.

URMENYI, A. M. & FRANKLIN, A. W. 1961. Neonatal death from pigmented coliform infection. *Lancet*, 1, 313.

URWIN, R. & MAIDEN, C.J. 2010. Multi-locus sequence typing: a tool for global epidemiology. *Trends in Microbiology*, 11, 479-487.

VAN ACKER, J., DE SMET, F., MUYLDERMANS, G., BOUGATEF, A., NAESSENS, A., & LAUWERS, S. 2001. Outbreak of necrotizing enterocolitis associated with *Enterobacter sakazakii* in powdered milk formula. *Journal of Clinical Microbiology*, 39, 293–297.

VAN OS, M., VAN WIKSELAAR, P. G. & SPOELSTRA, S. F. 1996. Formation of biogenic amines in well fermented grass silages. *Journal of Agricultural Science - Cambridge*, 127, 97-108.

VAVROVA, S., VALKOVA, D., DRAHOVSKA, H., KOKAVEC, J., MRAVEC, J. & TURNA, J. 2006. Analysis of the tellurite resistance determinant on the pNT3B derivative of the pTE53 plasmid from uropathogenic *Escherichia coli*. *Biometals*, 19, 453-460.

VIMR, E. R., KALIVODA, K.A., DESZO, E.L. & STEENBERGEN, S.M. 2004. Diversity of microbial sialic acid metabolism. *Microbiology and Molecular Biology Reviews,* 68, 132-153.

VIMR, E. 2012. Unified theory of bacterial sialometabolism: how and why bacteria metabolize host sialic acids. *ISRN Microbiology,* In Press.

WAGEGG, W. & BRAUN, V. 1981. Ferric citrate transport in *Escherichia coli* requires outer membrane receptor protein FecA. *Journal of Bacteriology,* 145, 156-163.

WAINE, D.J., HONEYBOURNE, D., GRACE SMITH, E., WHITEHOUSE, J.L. & DOWSON, C.G. 2009. Cross-sectional and longitudinal multilocus sequence typing of Pseudomonas aeruginosa in cystic fibrosis sputum samples. *Journal of Clinical Microbiology*, 47, 3444-3448.

WANG, B., BRAND-MILLER, J., MCVEAGH, P. & PETOCZ, P. 2001. Concentration and distribution of sialic acid in human milk and infant formulas. *The American Journal of Clinical Nutrition,* 74, 510-515.

WANG, Y. & KIM, K. S. 2002. Role of OmpA and IbeB in *Escherichia coli* K1 invasion of brain microvascular endothelial cells *in vitro* and *in vivo*. *Pediatric Research*, 51, 559-563.

WANG, B. 2009. Sialic acid is an essential nutrient for brain development and cognition. *Annual Review of Nutrition.* 29, 177-222.

WANG, X. & QUINN, P.J. 2009. Lipopolysaccharide: Biosynthetic pathway and structure modification. *Progress in Lipid Research,* 49, 97-107.

WANG, L., WANG, Q. & REEVES, P. R. 2010. The variations of O-antigens in gram-negative bacteria. *Sub-cellular Biochemistry,* 53, 123-152.

WATERHOUSE, A.M., PROCTER, J.B., MARTIN, D.M.A, CLAMP, M. & BARTON, G. J. 2009. Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189-1191.

WAYNE, L.G., BRENNER, D.J., COLWELL, R.R., GRIMONT, P.A.D., KANDLER, O., KRICHEVSKY, M.I., MOORE, L.H., MOORE, W.E.C., MURRAY, R.G.E., STACKEBRANDT, E., STARR, M.P. & TRUPER, H.G. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic Bacteriology,* 37, 463-464.

WHO (WORLD HEALTH ORGANIZATION). 2001. 54th World Health Assembly. Infant and young child nutrition.

Available at: http://apps.who.int/gb/archive/pdf_files/WHA54/ea54id4.pdf (Accessed 10th November 2012).

WHO (WORLD HEALTH ORGANIZATION) AND FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. 2007. Safe preparation, storage and handling of powdered infant formula. Guidelines.
Available at: http://www.who.int/foodsafety/publications/micro/pif_guidelines.pdf (Accessed 10th November 2012).

WILLIAMS, S.G., VARCOE, L.T., ATTRIDGE, S.R. & MANNING, P.A. 1996. *Vibrio cholerae* Hcp, a secreted protein coregulated with HlyA. *Infection and Immunity*, 64, 283-289.

WILLIAMS, T.L., MONDAY, S.R., EDELSON-MAMMEL, S., BUCHANAN, R. & MUSSER, S.M. 2005. A top-down proteomics approach for differentiating thermal resistant strains of *Enterobacter sakazakii*. *Proteomics*, 5, 4161-4169.

WIRTH, T., FALUSH, D., LAN, R., COLLES, F., MENSA, P., WIELER, L.H., KARCH, H., REEVES, P.R., MAIDEN, M.C., OCHMAN, H. & ACHTMAN, M. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Molecular Microbiology*, 60, 1136-1151.

YE, J., COULOURIS, G., ZARETSKAYA, I., CUTCUTACHE, I., ROZEN, S. & MADDEN, T. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134.

YEO, H-J. & WAKSMAN, G. 2004. Unveiling molecular scaffolds of the Type IV secretion system. *Journal of Bacteriology*, 186, 1919-1926.

YOUNG, J. M. & PARK, D. C. 2007. Relationships of plant pathogenic *Enterobacteria* based on partial *atpD*, *carA*, and *recA* as individual and concatenated nucleotide and peptide sequences. *Systematic and Applied Microbiology*, 30, 343-354.

ZEGANS, M.E., WAGNER, J.C., CADY, K.C., MURPHY, D.M., HAMMOND, J.H. & O'TOOLE, G.A. 2009. Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviours of *Pseudomonas aeruginosa. Journal of Bacteriology*, 191, 210-219.

ZERBINO, D.R. & BIRNEY, E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research*, 18, 821-829.

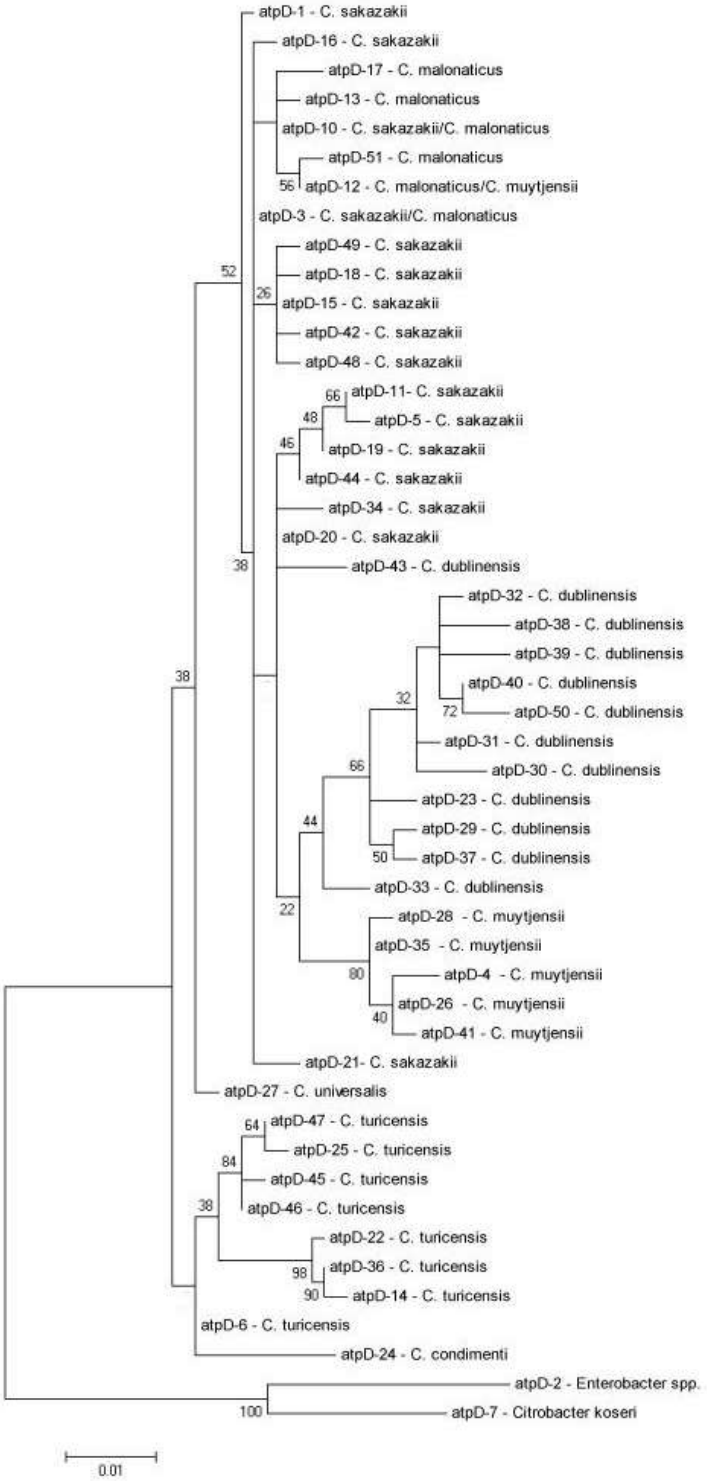
ZHOU, Y., LIANG, Y., LYNCH, K., DENNIS, J.J. & WISHART, D. S. 2011. PHAST: A fast phage search tool. *Nucleic Acids Research*, 39 (suppl 2), W347-W352.

ZHOU, Y., TAO, J., YU, H., NI, J., ZENG, L., TENG, Q., KIM, K.S., ZHAO, G.P., GUO, X., & YAO, Y. 2012. Hcp Family proteins secreted via the Type VI secretion system coordinately regulate *Escherichia coli* K1 interaction with human brain microvascular endothelial cells. *Infection and Immunity*, 80, 1243-1251.

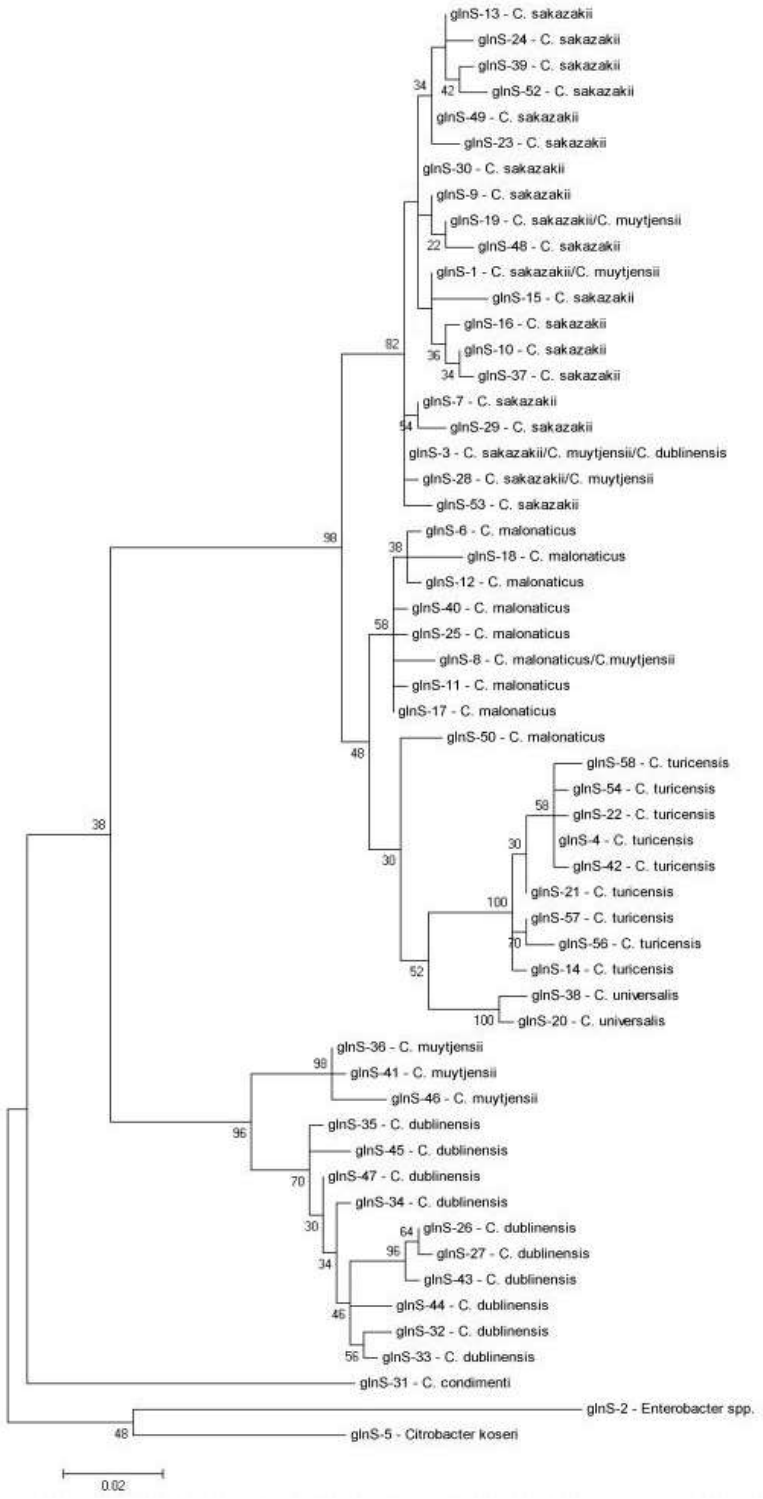
ZUBER, S., BOISSIN-DELAPORTE, C., MICHOT, L., IVERSEN, C., DIEP, B., BRUSSOW, H. & BREEUWER, P. 2008. Decreasing *Enterobacter sakazakii* (*Cronobacter* spp.) food contamination level with bacteriophages: prospects and problems. *Microbial Biotechnology*, 1, 532-543.

APPENDIX

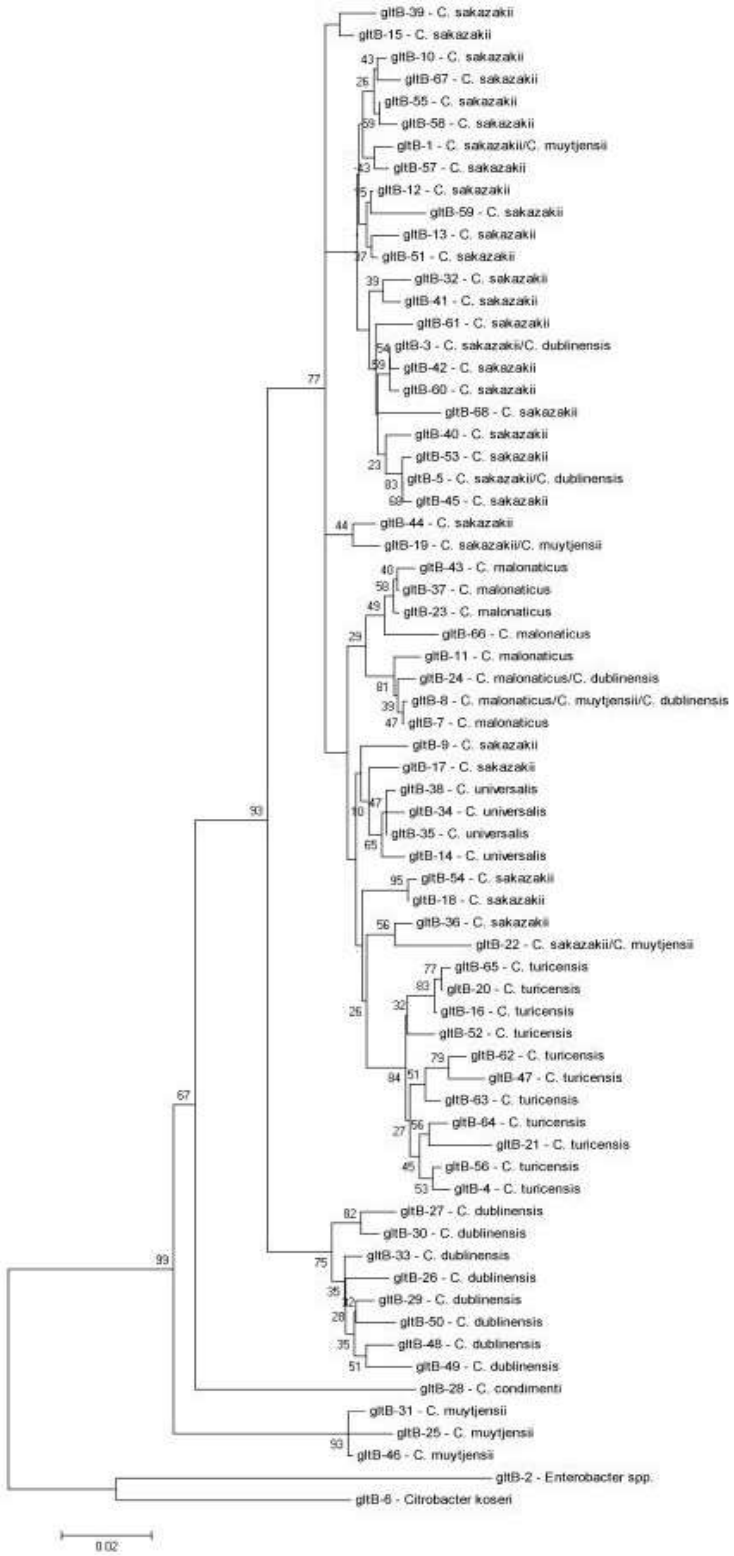
Appendix Fig. 1 Maximum-likelihood tree of the *atpD* alleles (390 bp) of the *Cronobacter* MLST dataset. The numbers at the end of each branch indicate the allele numbers. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage.



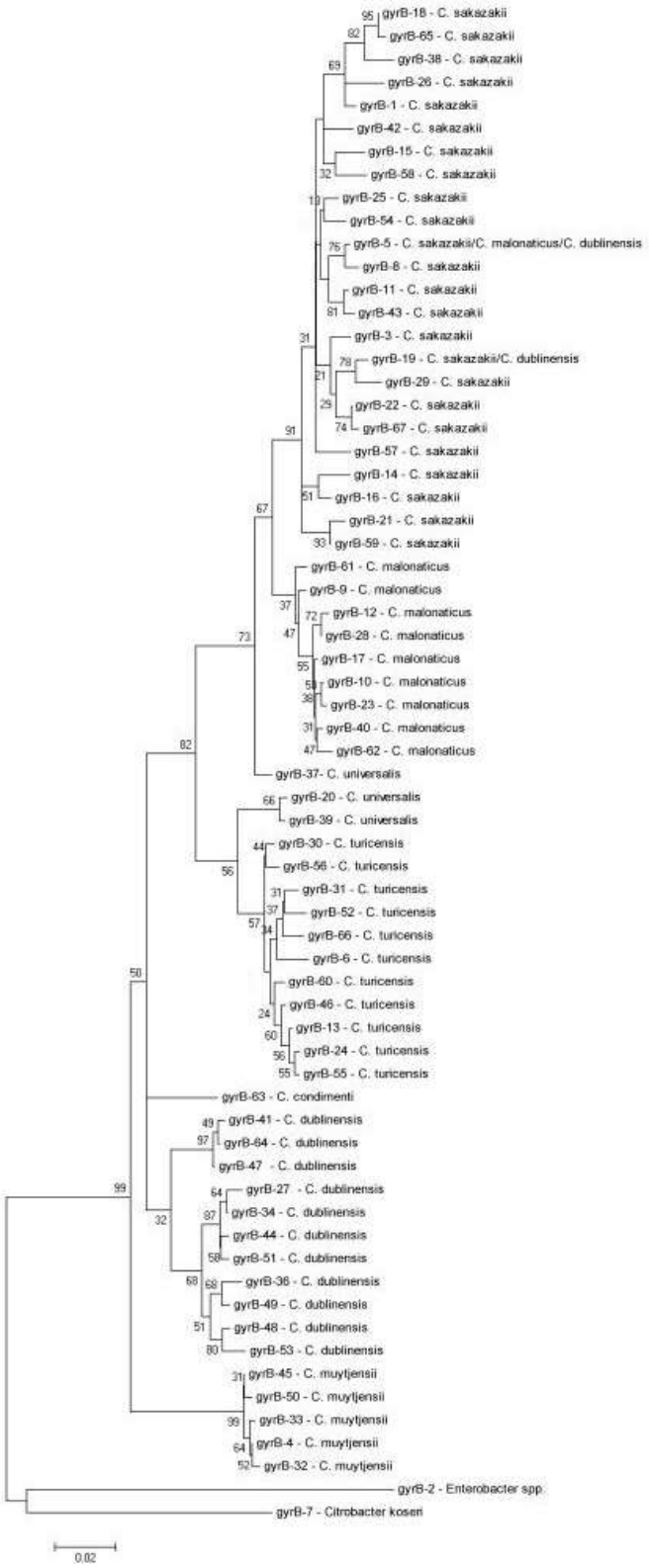
Appendix Fig. 2 Maximum-likelihood tree of the *glnS* alleles (363 bp) of the *Cronobacter* MLST dataset. The numbers at the end of each branch indicate the allele numbers. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage.



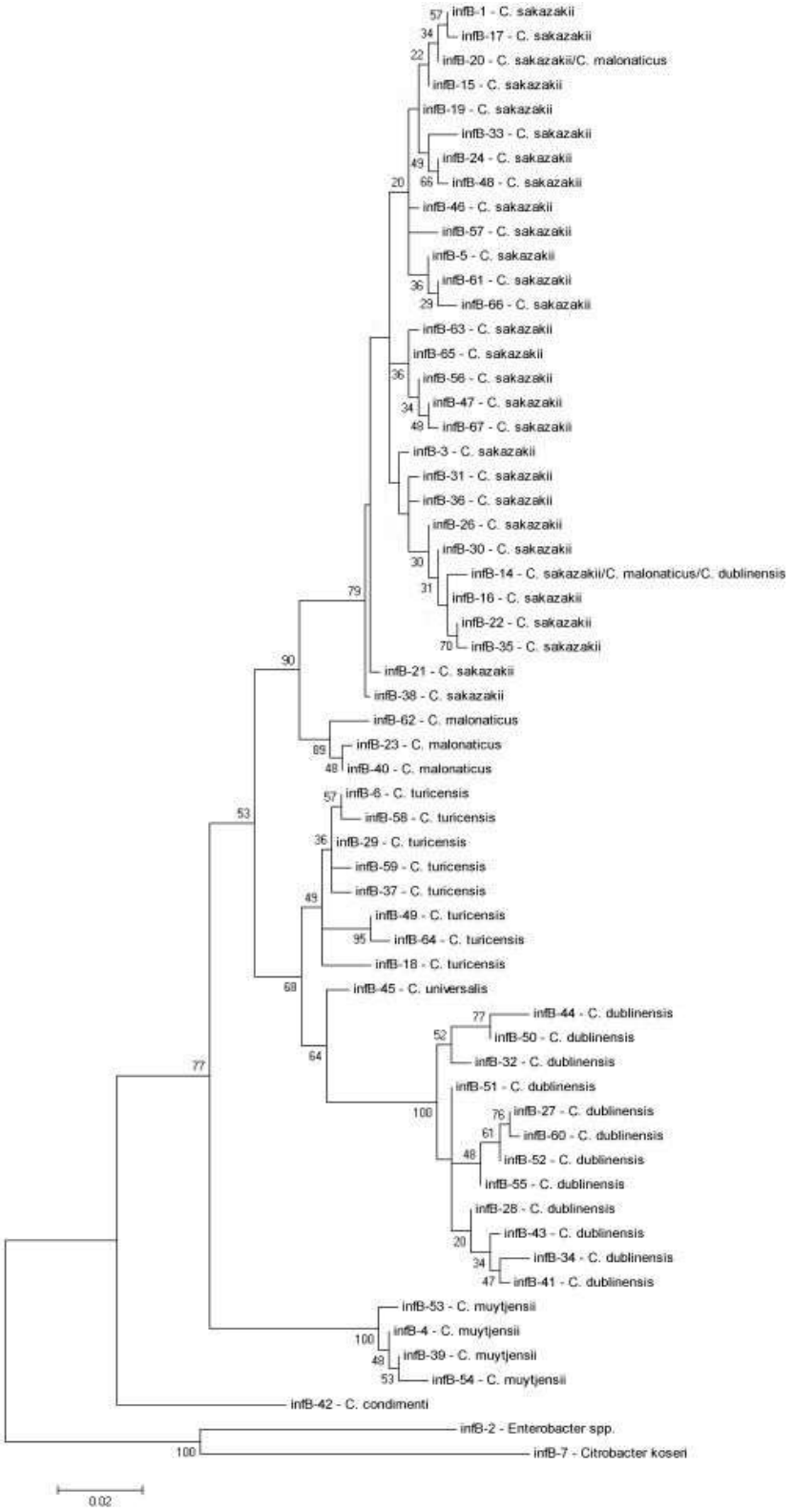
Appendix Fig. 3 Maximum-likelihood tree of the *gltB* alleles (507 bp) of the *Cronobacter* MLST dataset. The numbers at the end of each branch indicate the allele numbers. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage.



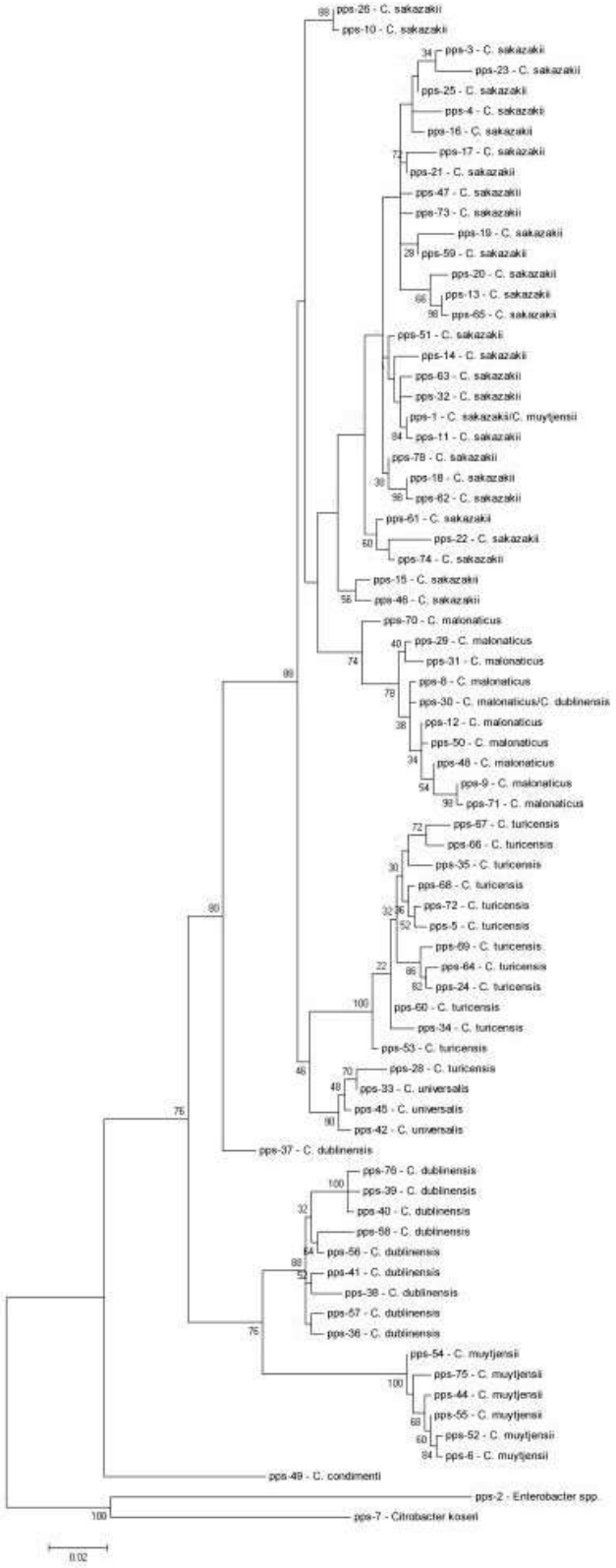
Appendix Fig. 4 Maximum-likelihood tree of the *gyrB* alleles (402 bp) of the *Cronobacter* MLST dataset. The numbers at the end of each branch indicate the allele numbers. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage.



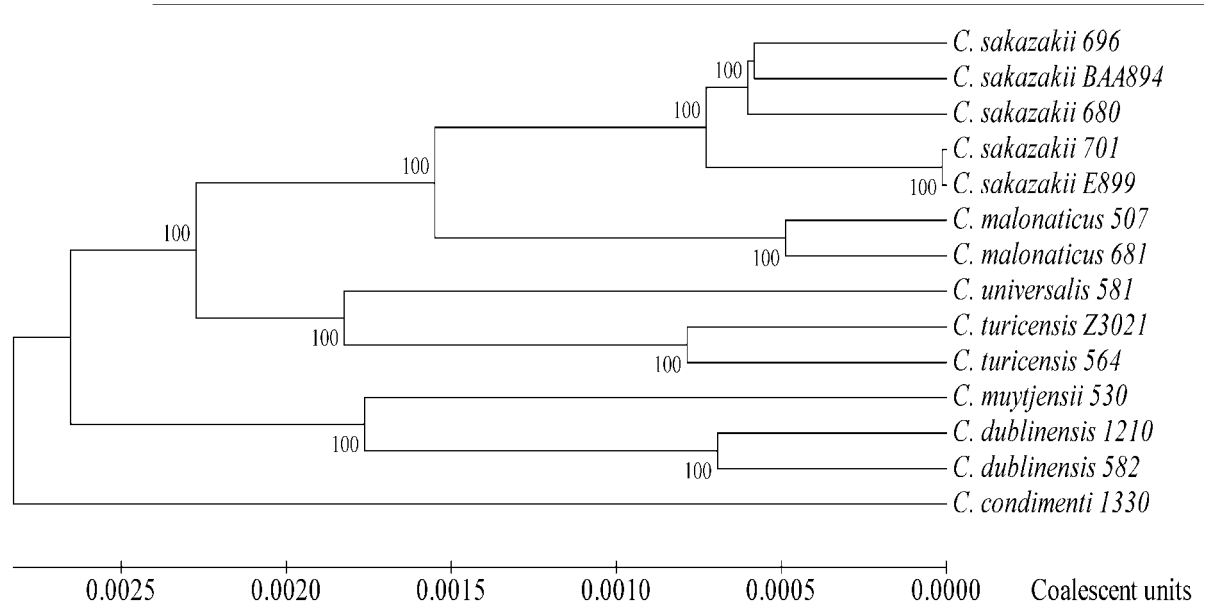
Appendix Fig. 5 Maximum-likelihood tree of the *infB* alleles (441 bp) of the *Cronobacter* MLST dataset. The numbers at the end of each branch indicate the allele numbers. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage.



Appendix Fig. 6 Maximum-likelihood tree of the *ppsA* alleles (495 bp) of the *Cronobacter* MLST dataset. The numbers at the end of each branch indicate the allele numbers. The tree is drawn to scale using MEGA5, with 1000 bootstrap replicates. The numbers at the nodes indicate the bootstrap values expressed in percentage.



Appendix Fig. 7 Clonal Frame (CF) analysis of the fourteen *Cronobacter* spp. genomes based on randomly selected 99 core loci. The consensus tree was obtained from 2 CF runs with 200,000 MCMC iterations. The numbers on the nodes represent the bootstrap support in percentage. Published in Joseph *et al.* (2012c).



Appendix Table 1 PCR identification of the serotypes of a selection of *C. sakazakii* isolates

Strain No.	ST	*C. sakazakii* serogroup
553	4	2
767	4	2
6	4	2
20	4	2
1240	4	No PCR product
1242	4	No PCR product
721	4	2
1219	4	2
1220	4	2
1221	4	2
1222	4	2
1223	4	2
1224	4	2
1225	4	2
1231	4	2
558	4	2
557	4	2
4	15	2
555	1	1
658	1	1
1019	1	1
12	1	1
1241	1	No PCR product
1	8	1
5	8	1
680	8	1
520	12	2
580	18	3
1249	31	2

Appendix Table 2 DNA-DNA (mean % ± SD) relatedness of *Cronobacter condimenti* sp. nov. and *Cronobacter universalis* sp. nov. with other species of the genus. Data from Iversen *et al*. (2008) and Joseph *et al*. (2012a).

	***C. universalis* sp. nov. NCTC 9529^T**	***C. condimenti* sp. nov. LMG 26250^T**
***C. sakazakii* ATCC 29544^T**	55.5% ± 1.0	40.3% ± 7.7
***C. malonaticus* CDC 1058-77^T**	60.1% ± 1.3	53.0% ± 14.4
***C. muytensii* ATCC 51329^T**	53.1% ± 6.6	42.0% ± 9.3
***C. dublinensis* LMG 23823^T**	45.9% ± 2.0	54.2% ± 8.7
***C. turicencis* LMG 23827 ^{T}**	55.0% ± 3.3	47.9% ± 5.9
***C. universalis* sp. nov. NCTC 9529^T**		50.7% ± 7.6

Appendix Table 3 Key tests for phenotypic differentiation of *Cronobacter condimenti* sp. nov. and *Cronobacter universalis* sp. nov. from other species of the genus *Cronobacter*, as published in Joseph *et al*. (2012a).

Characteristic	1	2	3*	4*	5*	6*	7*	8*	9*
Motility	-	v(-)	+(+)	v (+)	+(+)	+(+)	+(+)	+	+
Carbon utilization:									
Dulcitol	-	+(+)	-(-)	-(-)	+(+)	+(+)	-(-)	-	-
Indole	+	-(-)	-(-)	-(-)	-(-)	+(+)	+(+)	v	+
Malonate	+	+(+)	-(-)	+(+)	v(+)	+(+)	+(+)	-	-
1-0-Methyl α-D-glucopyranoside (AMG)	+	+(+)	+(+)	+(+)	+(+)	-(-)	+(+)	+	+
Melezitose	-	+(+)	-(-)	-(-)	+(+)	-(-)	+(+)	-	-
Turanose	-	-(-)	+(+)	+(+)	+(+)	v(+)	+(+)	v	-
Inositol	-	+(+)	v(+)	v(-)	+(+)	+(+)	+(+)	+	-
Lactulose	-	+(+)	+(+)	+(+)	+(+)	+(+)	+(+)	+	-
Putrescine	-	-(-)	+(+)	v(+)	+(+)	+(+)	+(+)	+	v
***Cis*-aconitate**	-	v(-)	+(+)	+(+)	+(+)	+(+)	+(+)	+	+
***Trans*-aconitate**	-	-(-)	-(-)	+(+)	-(-)	+(+)	+(+)	+	+
4-Aminobutyrate	-	-(-)	+(+)	+(+)	+(+)	v(+)	+(+)	+	+
Maltitol	-	+(+)	+(+)	+(+)	+(+)	-(-)	+(+)	+	-
Palatinose	-	v(-)	+(+)	+(+)	+(+)	v(+)	+(+)	+	+

Cronobacter species: 1 - *C. condimenti* sp. nov. (1330^T) ; 2 - *C. universalis* sp. nov. (n=4); 3 - *C. sakazakii* (ATCC 29544^T); 4 - *C. malonaticus* (LMG 23826^T) ; 5 - *C. turicensis* (LMG 23827^T); 6 - *C. muytjensii* (ATCC 51329^T); 7 - *C. dublinensis* subsp. *dublinensis* (LMG 23823^T) ;8 - *C. dublinensis* subsp. *lactaridi* (CDC 5960-70^T); 9 - *C. dublinensis* subsp. *lausannensis* (NCTC 9844^T).

*Results from Iversen *et al*. (2007 & 2008), except for type strains (shown in parenthesis) which were obtained in this study. (V: 20-80 % variable results.)