

Genome Sequence of *Cronobacter sakazakii* BAA-894 and Comparative Genomic Hybridization Analysis with Other *Cronobacter* Species

Eva Kucerova¹, Sandra W. Clifton³, Xiao-Qin Xia², Fred Long², Steffen Porwollik², Lucinda Fulton³, Catrina Fronick³, Patrick Minx³, Kim Kyung³, Wesley Warren³, Robert Fulton³, Dongyan Feng³, Aye Wollam³, Neha Shah³, Veena Bhonagiri³, William E. Nash³, Kymberlie Hallsworth-Pepin³, Richard K. Wilson³, Michael McClelland^{2*9}, Stephen J. Forsythe¹⁹

1 School of Science and Technology, Nottingham Trent University, Nottingham, United Kingdom, **2** Vaccine Research Institute of San Diego, San Diego, California, United States of America, **3** Genome Center at Washington University, Washington University School of Medicine, St. Louis, Missouri, United States of America

Abstract

Background: The genus *Cronobacter* (formerly called *Enterobacter sakazakii*) is composed of five species; *C. sakazakii*, *C. malonaticus*, *C. turicensis*, *C. muytjensii*, and *C. dublinensis*. The genus includes opportunistic human pathogens, and the first three species have been associated with neonatal infections. The most severe diseases are caused in neonates and include fatal necrotizing enterocolitis and meningitis. The genetic basis of the diversity within the genus is unknown, and few virulence traits have been identified.

Methodology/Principal Findings: We report here the first sequence of a member of this genus, *C. sakazakii* strain BAA-894. The genome of *Cronobacter sakazakii* strain BAA-894 comprises a 4.4 Mb chromosome (57% GC content) and two plasmids; 31 kb (51% GC) and 131 kb (56% GC). The genome was used to construct a 387,000 probe oligonucleotide tiling DNA microarray covering the whole genome. Comparative genomic hybridization (CGH) was undertaken on five other *C. sakazakii* strains, and representatives of the four other *Cronobacter* species. Among 4,382 annotated genes inspected in this study, about 55% of genes were common to all *C. sakazakii* strains and 43% were common to all *Cronobacter* strains, with 10–17% absence of genes.

Conclusions/Significance: CGH highlighted 15 clusters of genes in *C. sakazakii* BAA-894 that were divergent or absent in more than half of the tested strains; six of these are of probable prophage origin. Putative virulence factors were identified in these prophage and in other variable regions. A number of genes unique to *Cronobacter* species associated with neonatal infections (*C. sakazakii*, *C. malonaticus* and *C. turicensis*) were identified. These included a copper and silver resistance system known to be linked to invasion of the blood-brain barrier by neonatal meningitic strains of *Escherichia coli*. In addition, genes encoding for multidrug efflux pumps and adhesins were identified that were unique to *C. sakazakii* strains from outbreaks in neonatal intensive care units.

Citation: Kucerova E, Clifton SW, Xia X-Q, Long F, Porwollik S, et al. (2010) Genome Sequence of *Cronobacter sakazakii* BAA-894 and Comparative Genomic Hybridization Analysis with Other *Cronobacter* Species. PLoS ONE 5(3): e9556. doi:10.1371/journal.pone.0009556

Editor: Niyaz Ahmed, University of Hyderabad, India

Received: December 6, 2009; **Accepted:** February 14, 2010; **Published:** March 8, 2010

Copyright: © 2010 Kucerova et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by National Institutes of Health (NIH) grants R01AI52237, R01AI075093, and R01AI073971, and by a grant from Nottingham Trent University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mcclelland.michael@gmail.com

⁹ These authors contributed equally to this work.

Introduction

Cronobacter spp. (formerly *Enterobacter sakazakii*) are Gram-negative, motile, non-sporeforming, peritrichous rods of the *Enterobacteriaceae* family. *Cronobacter* is a ubiquitous organism present in a wide range of environments, including water, soil, and a variety of processed foods and fresh produce [1]. The bacterium has been isolated from factory production lines including powdered infant formula factories and households [2] as well as from a wide range of clinical samples including cerebrospinal fluid, blood, bone marrow, sputum, urine and faeces [3]. The organism

is an opportunistic pathogen of humans that can cause infections in all age groups. However, low birth weight neonates are most at risk. In this host group *Cronobacter* has been associated with outbreaks of necrotizing enterocolitis, meningitis and septicaemia. Infections with these presentations result in exceptionally high mortality rates ranging from 40 to 80 percent [4]. In recent years, some outbreaks of bacterial infection in neonatal intensive care units (NICU) have been traced to powdered formula contaminated with *Cronobacter* [5–8].

Cronobacter was defined as ‘yellow-pigmented *Enterobacter cloacae*’ until 1980, when it was designated a new species, *Enterobacter*

sakazakii, by Farmer *et al* [9]. Analysis of both partial 16S rDNA and *hsp60* sequences showed that *E. sakazakii* isolates formed at least four distinct clusters, and it was proposed that clusters 2, 3, and 4 could be unique species [10]. Based on DNA-DNA hybridization and phenotyping, *Enterobacter sakazakii* was subsequently proposed to be re-classified into a new genus *Cronobacter*, composed of five distinct species: *Cronobacter sakazakii*, *C. malonaticus*, *C. turicensis*, *C. muytjensii* and *C. dublinensis* [11]. Due to their close relatedness *C. sakazakii* and *C. malonaticus* are difficult to distinguish by 16S rDNA sequence analysis. However, multilocus sequence typing (MLST) differentiates between the two species, and also reveals a strong clonal nature of the organism [12]. Previous studies on '*E. sakazakii*' will therefore be difficult to interpret unless the strains are re-examined and re-classified according to the current taxonomic structure.

Cronobacter strains vary in their virulence, as determined by epidemiological studies and in-house mammalian tissue culture [6,13,14], but their virulence mechanisms are unknown. The bacteria can attach to intestinal cells and survive in macrophages [13], but the specific receptors involved remain to be determined. To date, only strains from *C. sakazakii*, *C. malonaticus* and *C. turicensis* have been associated with neonatal infections. Recently it was shown that the disruption of tight junctions significantly enhances association of *C. sakazakii* with Caco2 cells [15]. Some reports suggest a similarity between the tropism of *Cronobacter* and *Citrobacter koseri* for invasion and infection of the central nervous system [16,17]. It was noted that brain abscesses due to *Cronobacter* and *Citrobacter koseri* were morphologically similar and may be due to similar virulence mechanisms [18]. The first putative *Cronobacter* virulence factors were enterotoxin-like compounds produced by four of eighteen strains [19]. The genes encoding the putative toxin have yet to be identified, however.

Here, we present the genome sequence of *C. sakazakii* strain BAA-894, isolated from powdered formula associated with a NICU outbreak [7], and use that sequence for comparative genomic hybridization (CGH) analysis of physiological and virulence related traits across the *Cronobacter* genus. Due to the severity of infant infection, a better understanding of the genomic variation between *Cronobacter* spp. is needed, and will be of interest to manufacturers of powdered infant formula, regulatory bodies,

as well as those studying the evolution and diversity of pathogenicity.

Results and Discussion

Cronobacter sakazakii BAA-894 Genome

The complete sequencing of the genome of *C. sakazakii* BAA-894 revealed that it was composed of 1 chromosome (4.36837 Mb, 57% GC) and 2 plasmids (pESAK2 31 kb, 51% GC, pESAK3 131 kb, 56% GC); (Genbank accessions CP000783-5). A largely automated annotation of the genome resulted in the identification of 4,392 genes, covering 87% of the chromosome, 38 genes covering 83% of pESAK2 and 127 genes, covering 87% of pESAK3. The genome is aligned to many other enterobacterial genomes in the Enterix [20,21] server (<http://enterix.cbcb.umd.edu/enteric/enteric.html>).

Genome Cluster Analysis

In order to compare closely related genomes to the sequence of *C. sakazakii* BAA-894, we designed a set of 384,030 50-mer oligonucleotides that tiled the whole genome in both strands at an average density of about one oligonucleotide every 12 bases. An array was then manufactured by Roche NimbleGen (www.nimblegen.com).

Genomic diversity of 10 strains of *Cronobacter* representing the five different recognized species of this genus (Table 1) was analyzed by CGH on this tiled DNA microarray against the sequenced strain *C. sakazakii* BAA-894. *Cronobacter* genes were classified as present, absent or of intermediate status, as defined in the Materials and Methods section. The raw data is deposited in GenBank GEO, accession number GSE19308.

To determine the presence or absence of genes, the median \log_2 ratio of the genome relative to the reference strain for all the oligonucleotides in that gene was calculated. Then GACK analysis [22] was used, which sets a floating threshold for presence and absence of genes for every hybridization (see below). Using Cluster and Treeview softwares [23], *Cronobacter* strains formed two distinct phylogenetic clusters. All *C. sakazakii* strains formed one cluster (Figure 1). *C. malonaticus*, *C. turicensis*, *C. dublinensis* and *C. malonaticus* formed a second, separate cluster. Within *C. sakazakii*, strains 701

Table 1. Bacterial strains used in this study.

Organism	Organism ID	Comment	Source	% 16S rDNA sequence difference from <i>C. sakazakii</i> BAA-894	MLST sequence type	Reference
<i>C. sakazakii</i> strain BAA-894	Csak658	Genome sequenced strain.	Powdered formula	-	1	[7,12]
<i>C. sakazakii</i> ATCC 29544 ^T	Csak1	Species type strain.	Child's throat	0.38	8	[11,12]
<i>C. sakazakii</i> ATCC 12868	Csak2		Unknown	0.57	3	[12]
<i>C. sakazakii</i> strain 20	Csak20		Clinical	0.57	4	[12]
<i>C. sakazakii</i> strain 701	Csak701	Fatal NEC III case	Peritoneal fluid	0.47	4	[6,12]
<i>C. sakazakii</i> strain 767	Csak767	Fatal meningitis case	Trachea	0.47	4	[6,12]
<i>C. sakazakii</i> strain 696	Csak696	NEC II case	Stools	0.57	12	[6,12]
<i>C. malonaticus</i> LMG 23826 ^T	Cmal1212	Species type strain.	Breast abscess isolate	0.57	NA	[11]
<i>C. turicensis</i> LMG 23827 ^T	Ctur1211	Species type strain. Meningitis case	Blood isolate	1.80	NA	[11]
<i>C. muytjensii</i> ATCC 51329 ^T	Cmuy3	Species type strain.	Unknown	2.94	NA	[11]
<i>C. dublinensis</i> LMG 23823 ^T	Cdub1210	Species type strain.	Environmental. Milk powder production plant	3.69	NA	[11]

doi:10.1371/journal.pone.0009556.t001

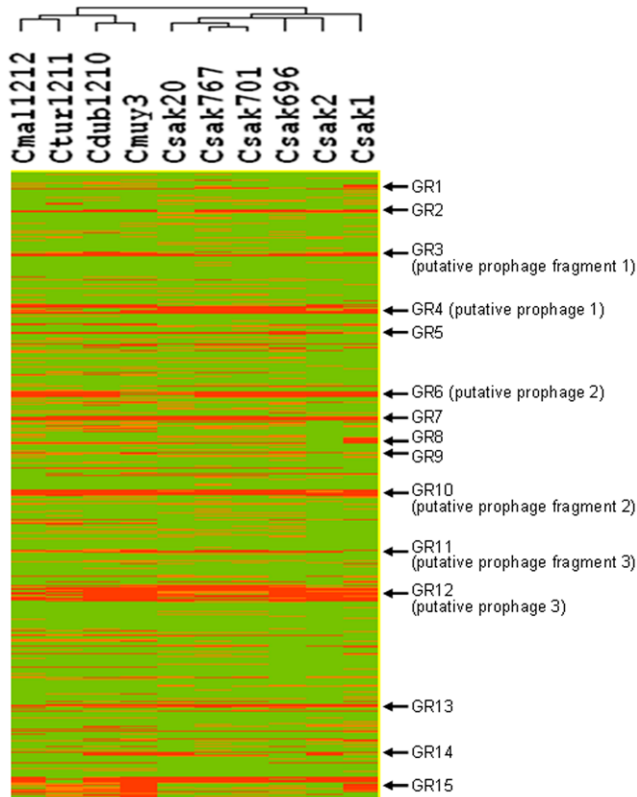


Figure 1. Whole genome clustering analysis showing the relationship between 10 *Cronobacter* isolates. Genomic regions GR1–GR15 are marked. Clustering analysis was performed using Gene Cluster (EisenSoftware). Hierarchical clustering was performed using the average linkage method on the trinary matrix based on the CGH analysis (1 for presence, 0 for uncertain and –1 for absence/divergence of a gene). For description of strains refer to Table 1. doi:10.1371/journal.pone.0009556.g001

and 767 were the most closely related and clustered together with strain 20. Previously, strains 701 and 767 were shown to belong to the same pulse field gel electrophoresis restriction digestion type [6]. Although the clinical details of the source of *C. sakazakii* strain 20 are unknown, the strain belongs to MLST sequence type 4 (as do 701 and 767), which is a stable clone of *C. sakazakii* isolated from both powdered infant formula and clinical sources [12]. *C. sakazakii* strain ATCC 29544^T (species type strain) formed a separate branch within the *C. sakazakii* cluster. The remaining *Cronobacter* species formed sub-clusters: *C. malonaticus* clustered with *C. turicensis* and *C. dublinensis* grouped with *C. muytjensii*. The tree remained identical when adjacent genes were collapsed into a single phylogenetic character if they had the same pattern of presence and absence.

Core Genome

Of the 4,382 unique annotated gene sequences represented on the microarray, 54.9% (2404) were common to all *C. sakazakii* strains and 43.3% (1899) were common to all five *Cronobacter* species. The vast majority of these shared genes are predicted to encode cellular essential functions such as energy metabolism, biosynthesis, DNA, RNA and protein synthesis, cell division and membrane transport. The proportion of genes absent from test strains compared with *C. sakazakii* BAA-894 ranged from 10.3% (453) in *C. sakazakii* strain 20 to 17.1% (751) in *C. muytjensii* (Table 2). In total, 5.1% (224) of BAA-894 genes were absent in all

Table 2. Number of absent or divergent and intermediate genes in *Cronobacter* strains when compared to the sequenced *C. sakazakii* BAA-894.

Strain	Absent genes (%) ¹	Intermediate genes (%) ¹
<i>C. sakazakii</i> ATCC 29544 ^T	582 (13.3)	449 (10.2)
<i>C. sakazakii</i> ATCC 12868	461 (10.5)	314 (7.2)
<i>C. sakazakii</i> strain 20	453 (10.3)	580 (13.2)
<i>C. sakazakii</i> strain 701	521 (11.9)	569 (13.0)
<i>C. sakazakii</i> strain 767	546 (12.5)	649 (14.8)
<i>C. sakazakii</i> strain 696	497 (11.3)	593 (13.5)
<i>C. malonaticus</i> LMG 23826 ^T	522 (11.9)	608 (13.9)
<i>C. turicensis</i> LMG 23827 ^T	495 (11.3)	591 (13.5)
<i>C. muytjensii</i> ATCC 51329 ^T	751 (17.1)	522 (11.9)
<i>C. dublinensis</i> LMG 23823 ^T	682 (15.6)	549 (12.5)
Absent in at least one <i>C. sakazakii</i>	835 (19.1)	
Absent in at least one <i>Cronobacter</i>	1081 (24.7)	

¹The numbers in brackets indicate the percentage of absent genes out of 4,382 annotated genes analyzed by comparative genomic hybridization. doi:10.1371/journal.pone.0009556.t002

C. sakazakii strains, and 3.1% (137) genes were absent in all *Cronobacter* strains (Table 2). Even though *C. muytjensii* and *C. malonaticus* are classified as separate species, the proportion of absent genes was only 11.3% and 11.9%, respectively, when compared to *C. sakazakii* BAA-894. This is in concordance with the previous 16S sequence comparison studies which showed that all *Cronobacter* strains are closely related (Table 1).

C. sakazakii Unique Genes

CGH analysis showed there were 21 genes present in all five *C. sakazakii* strains which were absent in *C. malonaticus*, *C. muytjensii*, *C. turicensis* and *C. dublinensis*. The genes unique to *C. sakazakii* strains were in two separate clusters of proteins involved in pilus assembly (ESA_02540–ESA_02542 and ESA_02796–ESA_02799), pilin FimA proteins (ESA_02541, ESA_02542, ESA_02796 and ESA_02799), porin PapC (ESA_02797) and the chaperone PapD (ESA_02798). The genes unique to *C. sakazakii* also included proteins for the phosphotransferase system (ESA_03303 and ESA_03305), a putative sialic acid transporter (ESA_03611), N-acetylneuraminase lyase (ESA_03612) and RelB from a toxin/antitoxin system (ESA_00257).

Invasion of Brain Microvascular Endothelial Cells

Because *Cronobacter* is associated with often fatal cases of neonatal meningitis, the status of genes identified in other organisms as associated with invasion of brain microvascular endothelial cells (BMEC) (*ibeA*, *ibeB*, *yijP* and *ompA*) in the sequenced isolate was of particular interest [24–27]. The gene encoding OmpA was present in all tested strains. This protein is associated with the invasive ability of neonatal meningitic *E. coli*. While genes *ibeA* and *yijP* produced no match in the reference strain *C. sakazakii* BAA-894, *ibeB* (synonymous to *cusC*) was found. *CusC* belongs to a cluster of genes encoding a copper and silver resistance cation efflux system which allows bacteria to invade BMEC [28]. The complete cation efflux system *cusA* (ESA_04242), *cusB* (ESA_04241), *cusC* (ESA_04239), *cusF* (ESA_04240) and its

regulatory gene *cusR* (ESA_04238) was present in strains associated with neonatal infections (*C. sakazakii* ATCC 29544^T, 701, 767, 696, *C. turicensis* and *C. malonaticus*), and absent in the other tested strains.

Other Physiological Traits

The presence of genes conferring physiological traits commonly associated with *Cronobacter* spp. was examined. Seventy genes involved in desiccation resistance [29], the metalloprotease *zpx* which causes rounding of Chinese hamster ovary (CHO) cells [30] and yellow pigment production genes [31] were present on the arrays. All these genes were present in all 10 *Cronobacter* strains tested.

Comparison of *C. sakazakii* Neonatal Intensive Care Unit (NICU) Outbreak Strains with *C. sakazakii* Type Strain ATCC 29544^T

The genes that were shared by the three strains associated with *C. sakazakii* outbreaks in NICUs (BAA-894, 701 & 767) were compared with the *C. sakazakii* species type strain ATCC 29544^T, which showed decreased virulence properties compared to strains 701 and 767 in tissue culture studies [14]. One hundred and forty-four genes present in the three NICU strains were absent in the type strain, 66 (46%) in clusters of consecutive genes based on the annotation of BAA-894. In most of these clusters, genes encoding proteins associated with resistance to different forms of stress were identified, including multidrug efflux systems, genes involved in resistance to oxidative stress, and those with a putative function in resistance to metals. The complete list of genes present in NICU outbreak strains *C. sakazakii* BAA-894, 707 and 767 and absent in the *C. sakazakii* type strain ATCC 29544^T is in Table S1; genes of interest are listed below.

Category I. Genes encoding proteins associated with resistance to different forms of antibiotics: (i), a transcriptional regulator (ESA_01938) from the TetR family of protein repressors that control the level of susceptibility to hydrophobic antibiotics and detergents; (ii), a homologue of the CpmG protein involved in carbapenem resistance (ESA_pESA3p05435); (iii), a protein conferring resistance to antimicrobial peptides Mig-14 (ESA_pESA3p05439); and (iv), a transcriptional regulator (ESA_pESA3p05448) involved in tetracycline resistance.

Category II. Genes encoding multidrug efflux systems: (i), a cationic drug transporter (ESA_01940) from the family of proteins that confer resistance to a wide range of toxic compounds; (ii), genes for complete bacterial ABC-transport systems involved in active transport across the cytoplasmic membrane (ESA_01944–ESA_01946); and (iii), a variety of genes encoding multidrug efflux components located on the plasmid pESA3 in BAA-894 (Table S1).

Category III. Genes involved in resistance to oxidative stress and genes with a putative function in resistance to metals: (i), a redox-sensitive transcriptional activator SoxR (ESA_00115); (ii), a glutathione S-transferase (ESA_00116); (iii), ADP-ribose pyrophosphatase involved in oxidative stress protection (ESA_pESA3p05446); (iv), an arsenate reductase (ESA_pESA3p05485); and (v), a predicted transcriptional regulator involved in mercury resistance ESA_pESA3p05463.

Category IV. Other genes of interest include: (i), putative adhesins which are recognized as virulence factors in enteric bacteria [32] (ESA_00983–ESA_00986); (ii), the universal stress protein UspA (ESA_01955) which can enhance the rate of cell survival during prolonged exposure to stress conditions [33]; (iii), a gene encoding a Type VI secretion lysozyme-related protein (ESA_02735); (iv), a gene for a predicted virulence SciE-type

protein (ESA_02736) which affects the ability of bacteria to enter eukaryotic cells [34]; and (v), genes involved in pilus assembly (ESA_03515 and ESA_03516).

C. sakazakii Plasmids

The sequenced strain *C. sakazakii* BAA-894 contains two plasmids; pESA2 (31 kb) and pESA3 (131 kb). Thirty-eight genes were annotated on pESA2 and 127 genes on pESA3.

The copy number of the plasmids was estimated from the median hybridization signals of oligonucleotides representing the plasmid compared to the sequenced genome. The ratio was (1:1.1:8.6) for the chromosome versus pESA2 versus pESA3. Thus, pESA2 exists as low copy, and pESA3 appears to be a moderate copy number plasmid.

The genes on pESA2 were absent in all other strains tested except *C. turicensis*, which had 19 (61.3%) genes present, and *C. sakazakii* 696, which had 4 (12.9%) genes present. The results for genes on pESA3 are summarized in Table 3.

Note that it is possible that some or all of the detected genes are on the chromosome in other strains. In addition, genes on a multicopy or medium copy plasmid may require a different degree of divergence to be identified as absent or divergent by comparative hybridizations. Plasmid profiling was performed on the *Cronobacter* strains analyzed by comparative hybridization. A plasmid of a size similar to pESA2 (31 kb) was detected in *C. sakazakii* 696 and in *C. turicensis*, which is in accordance with our CGH results. A large plasmid similar in size to pESA3 (131 kb) was visible in *C. sakazakii* strains ATCC 29544^T, ATCC 2868, 20, 696, 701, 767 and *C. malonaticus* (Table 3).

Genomic Regions Absent in Some Strains of *Cronobacter*

Genes that were absent in more than half of the *Cronobacter* strains relative to the sequenced strain *C. sakazakii* BAA-894 were selected for further analysis. These genes form 15 clusters of contiguous genes (based on the annotation of the reference genome). These are shown on Figure 2, where the number of strains in which a particular gene was classified as absent is plotted against the gene locus. The clusters were designated as regions GR1–GR15.

Table 3. pESA3 genes present in *Cronobacter* strains when compared to the sequenced strain of *C. sakazakii* BAA-894.

Strain	Present genes (%) ¹	Plasmid profile ²
<i>C. sakazakii</i> ATCC 12868	103 (88.8)	+
<i>C. sakazakii</i> 696	81 (69.8)	+
<i>C. sakazakii</i> 767	77 (66.4)	+
<i>C. sakazakii</i> 20	73 (62.9)	+
<i>C. sakazakii</i> 701	61 (52.6)	+
<i>C. malonaticus</i> LMG 23826 ^T	57 (49.1)	+ (3)
<i>C. sakazakii</i> ATCC 29544 ^T	51 (44.0)	+ (3)
<i>C. turicensis</i> LMG 23827 ^T	35 (30.2)	–
<i>C. dublinensis</i> LMG 23823 ^T	19 (16.4)	–
<i>C. muytjensii</i> ATCC 51329 ^T	0 (0.0)	–

¹Of the 127 annotated genes on pESA3, 116 were analyzed by comparative genomic hybridization. The other genes did not have a sufficient number of probes that passed the filters for manufacture on the array.

²Plasmid of a size similar to pESA3 (131 kb) was detected by gel electrophoresis of plasmid isolations.

³The detected plasmid was smaller in size than pESA3 (110 kb).

doi:10.1371/journal.pone.0009556.t003

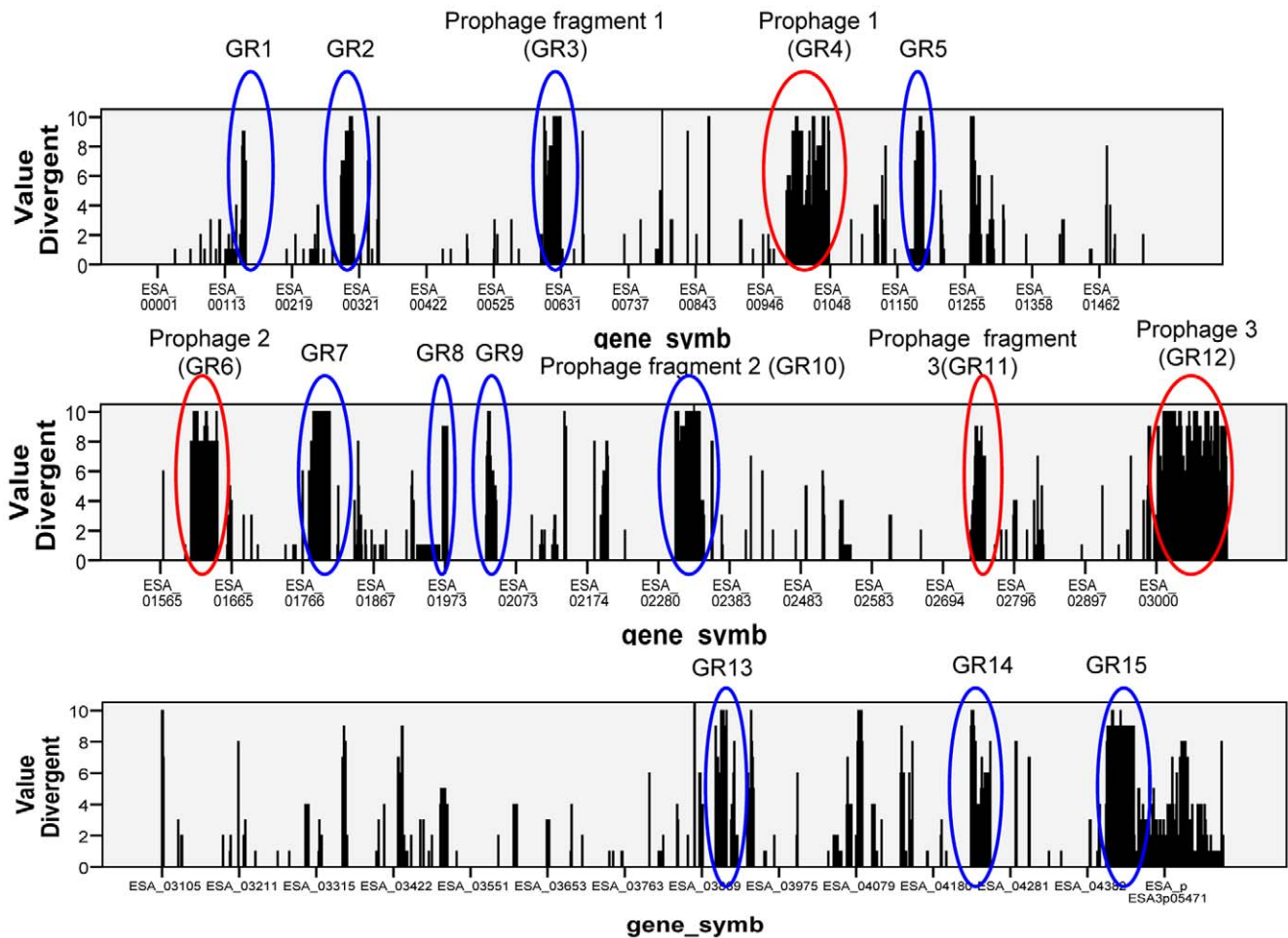


Figure 2. Distribution of the 15 variable regions across the *C. sakazakii* BAA-894 genome. Each column represents a gene classified as absent by CGH analysis in at least one strain. The height of the columns indicate the number of strains (out of 10) in which the gene was found to be absent. The major variable regions (blue) and prophages (red) are indicated in order of their appearance in the genome of *C. sakazakii* BAA-894. doi:10.1371/journal.pone.0009556.g002

Prophages

Of the 15 clusters, three putative prophage genomes and one prophage fragment were identified by Prophinder [35], and two additional regions are probable prophage fragments based on the presence of phage protein homologues identified by BLASTX (Table 3).

In the three prophage gene clusters (prophages 1, 2 and 3), genes encoding close homologues of known phage genes involved in integration, lysis and termination as well as head and tail structure were identified based on amino acid identity searches in IMG-JGI (<http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>). The average GC content of the sequenced *C. sakazakii* BAA-894 genome is 56%, the GC content of prophages 1, 2 and 3 was 53, 49 and 51%, respectively. The complete list of annotated putative prophage genes is available Table S2. In addition, Figure S1 shows the status of all putative prophage genes in the 10 *Cronobacter* strains.

Prophage 1 (GR4; ESA_00990–ESA_01052). In the 46-kb putative prophage 1 (Figure 3A), 30/63 (48%) hypothetical proteins were similar to known phage proteins. Homologues of phage genes involved in integration, lysis, head morphogenesis, tail assembly and phage regulation were identified. Prophage 1 also contains a gene (ESA_00997) encoding a protein homologous to the *eae*-like adhesion protein associated with the attaching and

effacing phenotype. The *eae* gene is carried by some other bacteriophages of enteric pathogens: *Salmonella* phage epsilon34, *E. coli* O157:H7 bacteriophage PhiV10, enterobacterial phage P22 and enterobacterial phage epsilon15.

Prophage 2 (GR6; ESA_01608–ESA_01644). Putative prophage 2 (Figure 3B) contains 37 genes, out of which 25 (68%) were homologous to known phage proteins. Prophage 2 contains several lambdoid phage genes encoding the following proteins: repressor CII (ESA_01613), replication proteins O and P (ESA_01614 and ESA_01615), the antitermination protein Q (ESA_01622), small and large subunits of the phage terminase (ESA_01632 and ESA_01633) as well as head and tail morphogenesis proteins. Head morphogenesis genes (ESA_01635–ESA_01637) were similar to head proteins of bacteriophage HK97 from the family of lambda phages. Two gene clusters have very low average GC content; 33% (ESA_01616–ESA_01620) and 44% (ESA_01627–ESA_01631). Both these clusters contain hypothetical proteins that showed no similarity with known phage proteins or functions.

Prophage 3 (GR12; ESA_03025–ESA_03102) was the largest (47 kb) putative prophage identified (Figure 3C). Thirty-four genes (39%) genes had close homology to known phage proteins or functions. The rest of the annotated genes are conserved proteins of unknown functions or hypothetical proteins. Similarly to prophage 2, several regulatory genes characteristic for lambdoid

Prophage fragment 2 (GR10). The region ESA_02304–ESA_02339 is likely to represent another prophage remnant. The region mostly contains hypothetical proteins with unknown functions. However, eight genes showed some degree of homology to proteins of phage origin. The cluster is flanked by a gene homologous to phage methyltransferase (ESA_02304) and a gene containing a site-specific recombinase domain that is found in putative integrases/recombinases of mobile genetic elements of diverse bacteria and phages (ESA_02339). It also includes genes homologous to phage lysozyme (ESA_02309), a phage tail component (ESA_02311), a putative phage tape measure protein (ESA_02313), another unspecified phage protein (ESA_02316), a major capsid protein (ESA_02319) and a phage portal protein (ESA_02320). GR10 might be another remnant of a prophage that has previously integrated into *C. sakazakii* BAA-894 genome. The presence of putative integrase flanking the cluster suggests introduction of this cluster into the genome by horizontal gene transfer. This cluster was absent in all *Cronobacter* strains except the genes encoding phage lysozyme (ESA_02309) and a hypothetical protein (ESA_02310), present in *C. sakazakii* strains 2 and 20, probably as a part of different prophages.

Prophage fragment 3 (GR11; ESA_02740–ESA_02755). The gene cluster of the 8 kb putative prophage fragment 3 comprises 16 hypothetical proteins, 10 of which (63%) may be associated with phage functions. As it lacks genes for head and tail morphogenesis as well as phage regulatory genes, it is likely to be a non-functional phage remnant.

Two thirds of all gamma-proteobacteria and low GC Gram-negative bacteria harbor prophages [39]. There is an increasing body of evidence that phages play a pivotal role in the diversification of bacterial species. Some phages can carry additional cargo genes, which are not required for the phage cycle and are suspected or proven virulence factors [39]. Such genes are typically located near prophage ends, downstream of phage tail genes or next to Q or N-like antiterminator genes [40]. Putative prophages identified in this study contain genes which are not similar to any other known prophage genes. Moreover, prophage 1 contains a gene encoding a protein homologous to the *eae*-like adhesion protein, which is a recognized virulence factor in enteropathogenic *E. coli* associated with the attaching and effacing phenotype [41]. It is hypothesized that most prophages are lost from bacterial genomes shortly after their acquisition [39]. Hence, some of the cargo genes carried by the prophages remaining in the chromosome of *Cronobacter* are possible virulence factors or fitness factors that increase the survival of the bacterium in its host. Further research into these putative virulence factors is warranted.

Non-Phage Regions That Differ among the Strains

The complete list of genes belonging to the variable non-phage regions and their presence in tested strains is available in Table S2.

GR1 (ESA_00140–ESA_00145) is a small cluster of type VI secretion system genes. ESA_00142 shares a conserved region with a family of IcmF-related proteins proposed to be involved in increased *Vibrio cholerae* adherence to epithelial cells [42]. ESA_00143 is a secretion protein belonging to the VC_A0110 family; mutations in proteins of this family are associated with impaired virulence [43]. ESA_00145, a secretion lipoprotein from a VC_A0113 family was present in *C. sakazakii* strains 701 and 767 associated with fatal outbreaks. The rest of GR1 was absent in all tested strains except *C. sakazakii* 696.

GR2 (ESA_00292–ESA_00310) mostly contains uncharacterized conserved proteins. It also contains the gene encoding a protein from a family of beta-lactamases (ESA_00299).

GR5 (ESA_01179–ESA_01189) contains a cluster of proteins involved in cell wall biogenesis and nucleotide sugar metabolism. GR5 corresponds to the *C. sakazakii* O-antigen gene locus used to distinguish the two *Cronobacter* serotypes O:1 and O:2 [44]. DNA microarray analysis revealed that GR5 is highly divergent; its genes were not sufficiently similar to be detected by microarray hybridization in any other *Cronobacter* strains. The O-antigen locus contains two more genes (homologues of ESA_01177 and ESA_01178) which were present in all strains except *C. sakazakii* 696.

GR7 (ESA_01775–ESA_01804). Most genes in GR7 were predicted to be involved in tellurite and stress resistance. It contains homologues of tellurite resistance proteins TerA, TerC, TerD, TerY and TerZ. The cluster contains two putative transposase genes, which suggests that the cluster was acquired by horizontal transfer. GR7 was found to be carried on plasmid pK29 of *Klebsiella pneumoniae* strain NK29, plasmid pEC-IMPQ of *Enterobacter cloacae*, plasmid R478 of *Serratia marcescens* and plasmid pAPEC-O1-R of *Escherichia coli* APEC O1, which is further evidence of horizontal gene transfer. As the gene cluster was entirely absent from all other *Cronobacter* strains in our study, the reference strain BAA-894 probably acquired the tellurite resistance cluster recently.

GR8 (ESA_01970–ESA_01976) contains seven genes encoding pilus assembly proteins. Fimbriae (or pili) enable bacteria to colonize the epithelium of specific host organs and are therefore considered major virulence factors [45]. This cluster of genes was absent in all *Cronobacter* strains except *C. sakazakii* strain 20.

GR9 (ESA_02032–ESA_02041) contains genes encoding hypothetical proteins and four proteins involved in Type VI secretion system (ESA_02037–ESA_02040). These four genes were present in *C. sakazakii* strains 1 and 20 and absent in all other strains.

GR13 (ESA_03887–ESA_03912) is a cluster of 16 hypothetical proteins without homology to known proteins or functions. This cluster was absent in all other *Cronobacter* strains.

GR14 (ESA_04248–ESA_04255) is a cluster of genes involved in copper resistance (Table 4). This cluster was found in *C. sakazakii* strain 1 and 696, as well as *C. turicensis* and *C. malonaticus*.

GR15 (ESA_pESA3p05493–ESA_pESA3p05505) involves genes located on *C. sakazakii* plasmid pESA3 and includes components of type IV and type VI secretion pathways (Table S2) as well as a gene encoding an outer membrane protein from the OmpA family (ESA_pESA3p05495). GR15 was absent from all strains except *C. sakazakii* strains 1 and 696.

Most of the described regions contain suspected or proven virulence factors. The genes in GR1, GR9 and GR15 are involved in a type VI secretion system, a newly described mechanism for protein transport across the cell envelope of Gram-negative

Table 4. Putative prophages identified in the genome of *C. sakazakii* BAA-894 by Prophinder [35] or BlastX.

Designation	Prophinder ID	Size (kb)	Coordinates
Prophage 1 (GR4)	48917	45.7	969835–1015482
Prophage 2 (GR6)	48912	26.7	1561010–1587672
Prophage 3 (GR12)	48916	46.9	2974743–3021625
Prophage fragment 1 (GR3)	48908	8.1	2693304–2701447
Prophage fragment 2 (GR10)	BlastX	33.4	2237406–2270858
Prophage fragment 3 (GR11)	BlastX	8.2	2693304–2701466

doi:10.1371/journal.pone.0009556.t004

bacteria that can increase adherence to epithelial cells [46,47]. GR3 contains four genes (ESA_00614–ESA_00617) that are homologous to a restriction-modification gene cluster (api49–api52) in the *Yersinia pseudotuberculosis* pathogenicity island (YAPI) [38]. As these genes were present in strains 701 and 767 isolated from two neonates that died as a result of infection by *Cronobacter* during an outbreak in France [8,17], and are absent in all other strains tested by CGH, these genes may be important virulence factors contributing to the pathogenicity of *Cronobacter*.

Cronobacter virulence factors have not been extensively studied, although it is known that *Cronobacter* species vary in their virulence with respect to invasion of intestinal cells, survival in macrophages and serum resistance [13,14]. GR 5 (ESA_01179–ESA_01189) encodes the lipopolysaccharide (LPS) genes. Characterisation of LPS structure and consequently O-antigen can be important in developing identification schemes based on serotyping, and has a role in virulence and serum resistance of the organism. The LPS is one of the few structural features of *Cronobacter* which has been investigated and it is known that it varies across the *Cronobacter* spp. In *C. sakazakii* and *C. malonaticus*, the LPS are composed of various branched polymers, whereas they are unbranched in *C. mytjensii*. In *C. sakazakii* BAA-894 [48] it is a branched polymer of pentasaccharide units composed of 2-acetamido-2-deoxy-D-galactose, 3-(N-acetyl-L-alanyl-amido)-3-deoxy-D-quinovose, D-glucuronic acid, and D-glucose. *C. sakazakii* strain 767 is also a branched polymer but of a repeating heptasaccharides composed of 2-acetamido-2-deoxy-D-glucose, and D-galacturonic acid, L-rhamnose, and D-glucose [49]. *C. malonaticus* LPS [50] is also a branched pentasaccharide unit of 2-amino-2-deoxy-D-glucose, 2-amino-2-deoxy-D-galactose, 3-deoxy-D-manno-oct-2-ulosonic acid, D-galactose and D-glucose residues. Whereas, *C. mytjensii* LPS [51] is a linear unbranched pentasaccharide polymer of 2-acetamido-2-deoxy-D-galactose, 2-acetamido-2-deoxy-D-glucose, 2-acetamido-3-deoxy-D-quinovose, L-rhamnose and D-glucuronic acid. These considerable differences correspond with the lack of sequence conservation in GR5 as revealed in the microarray analysis. The individual genes encoding these differences in enzymology have yet to be assigned.

Comparison to Other Enterobacterial Genera

The sequenced genome *C. sakazakii* BAA-894 was compared to the genomes of *Citrobacter koseri* BAA-895, *Klebsiella oxytoca* VJSK009, *E. coli* K12 MG1655 and *Salmonella enterica* Typhimurium strain LT2, representing some of the most closely related genera to *Cronobacter*. Using a threshold of identity of >85% in a 100 base window, 334 genes were present in all *Cronobacter* but absent or diverged in the four members of other genera (manuscript in preparation). These genes included a cluster of type VI secretion genes (ESA_03943 - ESA_03948) which might be involved in virulence, and a putative palatinose operon (ESA_02709 - ESA_02715). Alpha-glucosidase activity, which has been linked to palatinose metabolism, is considered as one of the major biochemical traits that distinguish *Cronobacter* from other related *Enterobacteriaceae*.

Summary and Conclusions

Using a whole-genome ~384,000 oligonucleotide tiling microarray, we analyzed the genomic content of isolates representing all five *Cronobacter* species by CGH. A dynamic determination of cut-offs GACK [22] was used to minimize the number of incorrectly categorized genes. Among 6 strains of *C. sakazakii* 2,404 genes (54.9%) represented a core shared genome. Of these genes 1,899 (43.3%) were also in the core genome when compared to four other *Cronobacter* species. CGH highlighted a copper/silver

resistance cluster associated with invasion of BMEC, which were unique to the three *Cronobacter* species associated with neonatal infections, as well as efflux pumps and adhesins unique to *C. sakazakii* strains from NICU outbreaks.

The main genetic features that distinguished the *Cronobacter* strains were putative prophages and several other gene clusters, a pattern of divergence typical among bacteria [52]. A few of the regions present in the sequenced strain and absent in some other *Cronobacter* strains are found in only a few other *Enterobacteriaceae*. For example, GR 7 is found in four of the hundreds of *Enterobacteriaceae* genomes that have been sequenced, which indicates that this region may have been horizontally acquired, possibly from a source outside of the *Enterobacteriaceae*. We have shown that gene acquisition via integration of phages and other mobile elements and specific gene-loss play a major role in *Cronobacter* evolution and diversity. Fifteen clusters of genes including three putative prophages and three putative prophage fragments that were absent in more than half tested strains were identified. In most of them, putative virulence genes were identified.

Future studies will focus on the expression of virulence related genes and their role in the pathogenicity of *Cronobacter* species, particularly the mechanisms of neonatal infection.

Materials and Methods

Strains and Culture Conditions

Cronobacter strains were selected which represented the five recognized species, and included those from reported clinical cases (Table 1). All *Cronobacter* strains were stored at -80°C in Nutrient Broth (Oxoid, UK) with 10% glycerol, subcultured on Trypticase Soy Agar (Oxoid, UK) and checked for purity. Overnight Trypticase Soy Broth (Oxoid, UK) cultures were used for DNA extraction.

DNA Extraction

Total genomic DNA was isolated using the QIAGEN Genomic-tip 100/G and Genomic DNA Buffer Set (www.1.qiagen.com) with extended cell lysis (1 h) and two additional washes of the precipitated DNA. The DNA samples were checked for fragmentation on 0.8% agarose gels and checked for protein and RNA content by spectrophotometry.

Sequencing of *C. sakazakii* ATCC BAA-894 and Assembly

The complete genome of *C. sakazakii* BAA-894, a strain isolated from a powdered formula used during an NICU outbreak [7], was sequenced using the whole genome shotgun method, supplemented with end sequencing of a fosmid library. Sonicated and size-fractionated DNA was cloned into plasmid vectors (pOTw13). Subclones and fosmids were sequenced using dye-primer and dye-terminator chemistry on ABI 3730 sequencing robots. Using the PCAP assembly software program [53], 51,289 sequence reads, representing 6.2 fold coverage, were assembled. Part of the 6.2X coverage included 1.19X fosmids. Under-represented areas, gaps, and ambiguities were then addressed by performing automated sequence improvement [54] using directed sequencing from the subclones (plasmids and fosmids). Following the auto-finish process, correcting misassembled regions, resolving ambiguous bases, and filling the remaining gaps by additional directed sequencing and PCR, completed the finishing process. This yielded a product with a final estimated accuracy of 99.99%.

Annotation

AceDB was the primary annotation database. The identification of protein-coding genes used a combination of GeneMark, Glimmer 2.0 and Glimmer 3.0; an evidence-based approach was used to

prioritize genes for inclusion into a final gene set. Genes missed by the two *ab initio* gene predictors were identified using BlastX.

C. sakazakii ATCC BAA-894 Microarray Design and Comparative Genome Hybridization Analysis (CGH)

A 384,030 probe oligonucleotide tiling DNA microarray was designed which comprised the complete genomic sequence of *C. sakazakii* ATCC BAA-894. Probes were designed at an average of less than 12 base spacing on alternating strands, leading to an average of over 100 50-mer oligonucleotide probes per annotated gene.

Every possible 50-base probe from both strands of the ESA genome and two plasmids was tested for the ability to be manufactured by NimbleGen. Those that required too many NimbleGen cycles were shortened. Resulting candidate probes that were less than 35 bases long were thrown out. The remaining 9,061,350 potential probes had an average melting temperature (T_m) of 74 degrees Celcius. Probes that had T_m above the average were shorted, down to a minimum of 35 bases. The probes were selected from the pool of 9,061,350 potential probes by selecting the best probe at 11.375-base increments, alternating between strands each time. The resulting 386,802 candidate probe sequences were analyzed to remove any probes that covered the same region as another probe due to duplications in the genome and the resulting 384,030 unique sequences were chosen for the array. Mappings between the unique probe sequences and their genome/plasmid/gene positions were stored in a separate file.

Sample labeling, CGH and data normalization were performed according to the method described at http://www.nimblegen.com/products/lit/cgh_userguide_v5p1.pdf.

DNA from the sequenced strain, *C. sakazakii* BAA-894, was used as the internal array control. For within-array normalization, a LOWESS method [55] was used as spatial correction and QSpline [56] was used to correct for dye bias. The raw data is deposited in GenBank GEO (accession number GSE19308).

Data Visualization by WebArrayDB

The CGH plotter available at WebArray (www.webarraydb.org/webarray/index.html) was used to create CGH plots. This used the \log_2 intensity ratio microarray data to calculate the median \log_2 intensity ratios for each *C. sakazakii* BAA-894 gene. WebArrayDB is a database system and online cross-platform analysis suite for analysis of microarray data, which allows storage of the data in the repository and their online analysis [57].

Dynamic Cut-Off Determination by GACK

Each gene was represented by tens or hundreds of separate oligonucleotides. These measurements were condensed into a single median ratio for each gene. The data were further analyzed using the dynamic cut-off determination tool GACK [22]. The normal probability density function was calculated from the characteristics of the main peak of the data distribution. This was used to calculate the estimated probability of presence that gives a statistical validation of the gene assignment. Each CGH experiment was attributed a specific set of cut-offs to minimize the number of falsely assigned genes.

Cronobacter genes were classified according to the most stringent settings of the ternary output of GACK as present, intermediate or

absent. Genes classified as 'present' had a \log_2 intensity ratio (test/reference) greater than the cut-off value corresponding to the 100% estimated probability of presence (EPP) calculated by GACK. The 'absent' genes had a \log_2 ratio inferior to the cut-off value corresponding to 0% EPP, and can include genes with sufficient sequence divergence. The 'intermediate' category included genes whose status could not be assigned with certainty. The EPP function and cut-offs were determined separately for hybridization data of each strain (Table S3).

Prophage Identification

ACLAME Prophinder was used for prophage identification [35]. This was accessed at <http://aclame.ulb.ac.be/Tools/Prophinder/>

16S rDNA Sequence Analysis

Partial 16S rDNA sequence analysis was done as previously described [10].

Plasmid Profiling

Plasmid DNA was isolated according to the method described in [58].

Supporting Information

Figure S1 Absence/presence status of prophage genes in *Cronobacter* strains. Red indicates absence/divergence of a particular gene, orange indicates uncertain status and green indicates presence of a gene.

Found at: doi:10.1371/journal.pone.0009556.s001 (0.15 MB TIF)

Table S1 Selected genomic regions present in NICU outbreak strains *C. sakazakii* 707 and 767 and absent in *C. sakazakii* type strain ATCC 29544T.

Found at: doi:10.1371/journal.pone.0009556.s002 (0.16 MB DOC)

Table S2 Non-phage genomic regions absent in more than half of the tested strains.

Found at: doi:10.1371/journal.pone.0009556.s003 (0.40 MB DOC)

Table S3 Cut-off values used for the assignment of absent, intermediate or present gene status.

Found at: doi:10.1371/journal.pone.0009556.s004 (0.04 MB DOC)

Acknowledgments

We thank Yollande Pui Cheng for technical assistance and Liliana Florea of Bioinformatics advise.

Author Contributions

Conceived and designed the experiments: SWC RKW MM. Performed the experiments: EK LAF CF PJM KK WW RSF DF NS VB WEN. Analyzed the data: EK SWC XQX AW MM SF. Contributed reagents/materials/analysis tools: SWC XQX FL SP KHP MM. Wrote the paper: EK MM SF.

References

- Friedemann M (2007) *Enterobacter sakazakii* in food and beverages (other than infant formula and milk powder). Int J Food Microbiol 116: 1–10.
- Kandhai MC, Reij MW, Gorris LG, Guillaume-Gentil O, van Schothorst M (2004) Occurrence of *Enterobacter sakazakii* in food production environments and households. Lancet 363(9402): 39–40.
- Gurtler JB, Kornacki JL, Beuchat LR (2005) *Enterobacter sakazakii*: a coliform of increased concern to infant health. Int J Food Microbiol 104: 1–34.
- Lai KK (2001) *Enterobacter sakazakii* infections among neonates, infants, children, and adults. Medicine 80: 113–122.

5. Biering G, Karlsson S, Clark NC, Jonsdottir KE, Ludvigsson P, et al. (1989) Three cases of neonatal meningitis caused by *Enterobacter sakazakii* in powdered milk. *J Clin Microbiol* 27: 2054–6.
6. Caubilla-Barron J, Hurrell E, Townsend S, Cheetham P, Loc-Carrillo C, et al. (2007) Genotypic and phenotypic analysis of *Enterobacter sakazakii* strains from an outbreak resulting in fatalities in a neonatal intensive care unit in France. *J Clin Microbiol* 45: 3979–85.
7. Himelright I, Harris E, Lorch V, Anderson M (2002) *Enterobacter sakazakii* infections associated with the use of powdered infant formula—Tennessee, 2001. *J Am Med Assoc* 287: 2204–2205.
8. van Acker J, de Smet F, Muyldermans G, Bougateg A, Naessens A, et al. (2001) Outbreak of necrotizing enterocolitis associated with *Enterobacter sakazakii* in powdered milk formula. *J Clin Microbiol* 39: 293–7.
9. Farmer JJ, Asbury MA, Hickman FW, Brenner DJ, the Enterobacteriaceae Study Group (USA) (1980) “*Enterobacter sakazakii*: a new species of “*Enterobacteriaceae*” isolated from clinical specimens”. *Int J Syst Bacteriol* 30: 569–84.
10. Iversen C, Waddington M, On SL, Forsythe S (2004) Identification and phylogeny of *Enterobacter sakazakii* relative to *Enterobacter* and *Citrobacter* Species. *J Clin Microbiol* 42: 5368–70.
11. Iversen C, Mullane N, McCardell B, Tall BD, Lehner A, et al. (2008) *Cronobacter* gen. nov., a new genus to accommodate the biogroups of *Enterobacter sakazakii*, and proposal of *Cronobacter sakazakii* gen. nov., comb. nov., *Cronobacter malonatius* sp. nov., *Cronobacter turicensis* sp. nov., *Cronobacter myyjiensis* sp. nov., *Cronobacter dublinensis* sp. nov., *Cronobacter* genomospecies 1, and of three subspecies, *Cronobacter dublinensis* subsp. *dublinensis* subsp. nov., *Cronobacter dublinensis* subsp. *lausannensis* subsp. nov. and *Cronobacter dublinensis* subsp. *lactaridi* subsp. nov. *Int J Syst Evol Microbiol* 58(Pt 6): 1442–7.
12. Baldwin A, Loughlin M, Caubilla-Barron J, Kucerova E, Manning G, et al. (2009) Multilocus sequence typing of *Cronobacter sakazakii* and *Cronobacter malonatius* reveals stable clonal structures with clinical significance which do not correlate with biotypes. *BMC Microbiol* 9: 223.
13. Townsend S, Hurrell E, Forsythe S (2008) Virulence studies of *Enterobacter sakazakii* isolates associated with a neonatal intensive care unit outbreak. *BMC Microbiol* 8: 64.
14. Townsend SM, Hurrell E, Gonzalez-Gomez I, Lowe J, Frye JG, et al. (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(Pt 10): 3538–47.
15. Kim KP, Loessner MJ (2008) *Enterobacter sakazakii* invasion in human intestinal Caco-2 cells requires the host cell cytoskeleton and is enhanced by disruption of tight junction. *Infect Immun* 76: 562–70.
16. Burdette JH, Santos C (2000) *Enterobacter sakazakii* brain abscess in the neonate: the importance of neuroradiologic imaging. *Pediatr Radiol* 30: 33–4.
17. Willis J, Robinson JE (1988) *Enterobacter sakazakii* meningitis in neonates. *Pediatr Infect Dis J* 7: 196–9.
18. Kline MW (1988) *Citrobacter* meningitis and brain abscess in infancy: Epidemiology, pathogenesis, and treatment. *The Journal of Pediatrics* 113: 430–434.
19. Pagotto FJ, Nazarowec-White M, Bidawid S, Farber JM (2003) *Enterobacter sakazakii*: infectivity and enterotoxin production in vitro and in vivo. *J Food Prot* 66: 370–5.
20. Florea L, McClelland M, Riemer C, Schwartz S, Miller W (2003) EnteriX 2003: Visualization tools for genome alignments of *Enterobacteriaceae*. *Nucleic Acids Res* 31: 3527–32.
21. Florea L, Riemer C, Schwartz S, Zhang Z, Stojanovic N, et al. (2000) Web-based visualization tools for bacterial genome alignments. *Nucleic Acids Res* 28: 3486–96.
22. Kim CC, Joyce EA, Chan K, Falkow S (2002) Improved analytical methods for microarray-based genome-composition analysis. *Genome Biol* 3: RESEARCH0065.
23. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863–8.
24. Huang SH, Wan ZS, Chen YH, Jong AY, Kim KS (2001) Further characterization of *Escherichia coli* brain microvascular endothelial cell invasion gene *ibeA* by deletion, complementation, and protein expression. *J Infect Dis* 183: 1071–8.
25. Huang SH, Chen YH, Fu Q, Stins M, Wang Y, et al. (1999) Identification and characterization of an *Escherichia coli* invasion gene locus, *ibeB*, required for penetration of brain microvascular endothelial cells. *Infect Immun* 67: 2103–9.
26. Prasadarao NV, Wass CA, Weiser JN, Stins MF, Huang SH, et al. (1996) Outer membrane protein A of *Escherichia coli* contributes to invasion of brain microvascular endothelial cells. *Infect Immun* 64: 146–53.
27. Wang Y, Huang SH, Wass CA, Stins M, Kim KS (1999) The gene locus *yjipP* contributes to *Escherichia coli* K1 invasion of brain microvascular endothelial cells. *Infect Immun* 67: 4751–6.
28. Franke S, Grass G, Rensing C, Nies DH (2003) Molecular analysis of the copper-transporting efflux system *CusCFBA* of *Escherichia coli*. *J Bacteriol* 185: 3804–12.
29. Osaili T, Forsythe S (2009) Desiccation resistance and persistence of *Cronobacter* species in infant formula. *Int J Food Microbiol* 136: 214–220.
30. Kothary MH, McCardell BA, Frazar CD, Deer D, Tall BD (2007) Characterization of the zinc-containing metalloprotease encoded by *zpx* and development of a species-specific detection method for *Enterobacter sakazakii*. *Appl Environ Microbiol* 73: 4142–51.
31. Lehner A, Grimm M, Rattei T, Ruepp A, Frishman D, et al. (2006) Cloning and characterization of *Enterobacter sakazakii* pigment genes and in situ spectroscopic analysis of the pigment. *FEMS Microbiol Lett* 265: 244–8.
32. Campellone KG, Leong JM (2003) Tails of two Tirs: actin pedestal formation by enteropathogenic *E. coli* and enterohemorrhagic *E. coli* O157:H7. *Curr Opin Microbiol* 6: 82–90.
33. Nystrom T, Neidhardt FC (1994) Expression and role of the universal stress protein, *UspA*, of *Escherichia coli* during growth arrest. *Mol Microbiol* 11: 537–44.
34. Folkesson A, Lofdahl S, Normark S (2002) The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. *Res Microbiol* 153: 537–45.
35. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R (2008) Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics* 24: 863–5.
36. Lazdunski CJ (1988) Pore-forming colicins: synthesis, extracellular release, mode of action, immunity. *Biochimie* 70: 1291–6.
37. Pickard D, Thomson NR, Baker S, Wain J, Pardo M, et al. (2008) Molecular characterization of the *Salmonella enterica* serovar *Typhi* Vi-typing bacteriophage E1. *J Bacteriol* 190: 2580–7.
38. Collyn F, Billault A, Mullet C, Simonet M, Marceau M (2004) YAPI, a new *Yersinia pseudotuberculosis* pathogenicity island. *Infect Immun* 72: 4784–90.
39. Brussow H, Canchaya C, Hardt WD (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 68: 560–602.
40. Boyd EF, Brussow H (2002) Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol* 10: 521–9.
41. Kaper JB (1998) The locus of enterocyte effacement pathogenicity island of Shiga toxin-producing *Escherichia coli* O157:H7 and other attaching and effacing *E. coli*. *Jpn J Med Sci Biol* 51 Suppl: S101–7.
42. Das S, Chakraborty A, Banerjee R, Chaudhuri K (2002) Involvement of in vivo induced *icmF* gene of *Vibrio cholerae* in motility, adherence to epithelial cells, and conjugation frequency. *Biochem Biophys Res Commun* 295: 922–8.
43. Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, et al. (2006) Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the Dictyostelium host model system. *Proc Natl Acad Sci U S A* 103: 1528–33.
44. Mullane N, O’Gaora P, Nally JE, Iversen C, Whyte P, et al. (2008) Molecular analysis of the *Enterobacter sakazakii* O-antigen gene locus. *Appl Environ Microbiol* 74: 3783–94.
45. Abraham SN, Jonsson AB, Normark S (1998) Fimbriae-mediated host-pathogen cross-talk. *Curr Opin Microbiol* 1: 75–81.
46. Mougous JD, Senaratne RH, Petzold CJ, Jain M, Lee DH, et al. (2006) A sulfated metabolite produced by *stf3* negatively regulates the virulence of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 103: 4258–63.
47. Schlieker C, Zentgraf H, Dersch P, Mogk A (2005) *ClpV*, a unique Hsp100/Clp member of pathogenic proteobacteria. *Biol Chem* 386: 1115–27.
48. MacLean LL, Pagotto F, Farber JM, Perry MB (2009) Structure of the antigenic repeating pentasaccharide unit of the LPS O-polysaccharide of *Cronobacter sakazakii* implicated in the Tennessee outbreak. *Biochem Cell Biol* 87: 459–465.
49. Czerwicka M, Forsythe SJ, Bychowska A, Dziadziszko H, Kunikowska D, et al. (2010) Chemical structure of the O-polysaccharide isolated from *Cronobacter sakazakii* 767. *Carbohydrate Research*. In Press.
50. MacLean LL, Vinogradov E, Pagotto F, Farber JM, Perry MB (2009) Characterization of the O-antigen in the lipopolysaccharide of *Cronobacter (Enterobacter) malonatius* 3267. *Biochem Cell Biol* 87: 927–932.
51. MacLean LL, Pagotto F, Farber J M, Perry MB (2009) The structure of the O-antigen in the endotoxin of the emerging food pathogen *Cronobacter (Enterobacter) myyjiensis* strain 3270. *Carbohydr Res* 344: 667–671.
52. Porwollik S, Wong RM, McClelland M (2002) Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. *Proc Natl Acad Sci U S A* 99: 8956–61.
53. Huang X, Wang J, Aluru S, Yang SP, Hillier L (2003) PCAP: a whole-genome assembly program. *Genome Res* 13: 2164–70.
54. Gordon D, Desmarais C, Green P (2001) Automated finishing with autofinish. *Genome Res* 11: 614–25.
55. Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74: 829–836.
56. Workman C, Jensen IJ, Jarmer H, Berka R, Gautier L, et al. (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* vol 3: research0048.1–0048.16.
57. Xia XQ, McClelland M, Porwollik S, Song W, Cong X, et al. (2009) WebArrayDB: cross-platform microarray data analysis and public data repository. *Bioinformatics* 25: 2425–9.
58. Kado CI, Liu ST (1981) Rapid procedure for detection and isolation of large and small plasmids. *J Bacteriol* 145: 1365–1373.