Cronobacter sakazakii Genes Contributing to Persistence in Low-Moisture Dairy Matrices

Kaitlin Ash Hartmann
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Life Sciences Commons

BYU ScholarsArchive Citation
Hartmann, Kaitlin Ash, "Cronobacter sakazakii Genes Contributing to Persistence in Low-Moisture Dairy Matrices" (2020). Theses and Dissertations. 8466.
https://scholarsarchive.byu.edu/etd/8466

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
Cronobacter sakazakii Genes Contributing to Persistence in Low-Moisture Dairy Matrices

Kaitlin Ash Hartmann

A thesis submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Master of Science

Bradley J. Taylor, Chair
   Joel Griffitts
   Frost Steele

Department of Nutrition, Dietetics, and Food Science Brigham Young University

Copyright © 2020 Kaitlin Ash Hartmann

All Rights Reserved
ABSTRACT

Cronobacter sakazakii Genes Contributing to Persistence in Low-Moisture Dairy Matrices

Kaitlin Ash Hartmann
Department of Nutrition, Dietetics, and Food Science, BYU
Master of Science

Cronobacter sakazakii is a gram-negative opportunistic pathogen known to survive in dry environments and food matrices, such as infant formula. This foodborne bacterium can cause fatal human infections of the blood, central nervous system, and gastrointestinal tract; it is also problematic in wounds and urinary tract infections. Preterm infants and immunocompromised individuals are in higher risk categories related to necrotizing enterocolitis, neonatal sepsis, and meningitis due to this organism. Therefore, there is a need for increased understanding of how this bacterium is able to persist in thermally treated low-moisture products that do not support growth. The objective of this research is to identify genes and mechanisms in C. sakazakii that contribute to its resistance to desiccation and survival in low-moisture food matrices, including powdered infant formula. C. sakazakii sequence type 4 (ST4) is of particular interest as it is often the cause of neonatal infections originating from contaminated feedings of powder infant formula. The method chosen to explore these genetic patterns is massively parallel transposon insertion sequencing (Tn-seq). The E. coli strain MFDpir was used to facilitate transposon insertional mutagenesis to create a library of mutated C. sakazakii. Three different C. sakazakii ST4 isolates of different origins (clinical, environmental, and infant formula-derived) were selected for this study. Once transposon mutagenesis occurred with the aid of E. coli MFDpir, the three mutant libraries were subjected to desiccation stress in a closed system equilibrated to 11.3% relative humidity. The surviving mutant genomes were analyzed with Tn-seq. The sequencing data revealed that, while transposition events did occur successfully within the genomes of each of the selected C. sakazakii isolates, these events were not dense enough to draw biological conclusions nor statistical inferences concerning which genes contribute to this organism’s uncanny desiccation tolerance. However, we concluded that the Tn-seq method is a promising tool with this organism of interest, despite incomplete results in this first round of experimentation.

Keywords: low-moisture foods, neonatal infection, microbial genetics, foodborne pathogen, Cronobacter sakazakii
ACKNOWLEDGEMENTS

I cannot express enough thanks to my committee for their consistent support and encouragement throughout my graduate journey. I especially thank Dr. Brad Taylor, my committee chair, for being brave enough to take me on as his first-ever graduate student. I’m grateful we could learn this process together and I’m tremendously appreciative for his advice and encouragement. I couldn’t have asked for a better mentor. I also thank Dr. Joel Griffitts who gave me invaluable advice and graciously allowed me to use his laboratory and resources. And thanks to Dr. Frost Steele for his involvement and support in this project.

I cannot neglect to especially thank Alex Benedict for his mentorship. Many hours were sacrificed to tutor me in molecular biology techniques. Nothing but kindness and patience were shown as I frequently returned seeking advice. This project would not have been possible without him. Thank you to all other lab mates, both in the Griffitts lab and the Food Microbiology lab, for your curiosity and friendship that helped fuel many long hours. I am deeply grateful to Lizzy Porter for her tireless work on the programming aspects of this project.

Deep appreciation goes to BUILD Dairy for providing the funding for this project, in addition to providing me with opportunities to present my research and get involved with the dairy community. Finally, to my husband, who joined me on the last and most difficult leg of this research, thank you for keeping me going daily.
TABLE OF CONTENTS

TITLE PAGE ................................................................................................................................... i
ABSTRACT .................................................................................................................................... ii
ACKNOWLEDGEMENTS ........................................................................................................... iii
TABLE OF CONTENTS ............................................................................................................... iv
LIST OF TABLES ......................................................................................................................... vi
LIST OF FIGURES ...................................................................................................................... vii
INTRODUCTION .......................................................................................................................... 1
  Background and Literature Review ......................................................................................... 1
  Research goals and experimental design ............................................................................... 11
MATERIALS AND METHODS .................................................................................................. 16
  Bacterial Strains, media, and growth conditions ................................................................. 16
  Tn-seq experimental setup ................................................................................................... 16
  Application of desiccation stress to mutant libraries ......................................................... 17
Arbitrary PCR ............................................................................................................................ 18
Genomic DNA Extraction ....................................................................................................... 20
Fragmentase digestion .............................................................................................................. 25
Zymo clean-up ......................................................................................................................... 27
C-tailing of genomic DNA fragments ..................................................................................... 27
First-round PCR ....................................................................................................................... 27
Second-round PCR ................................................................................................................... 29
Illumina Sequencing ................................................................................................................ 29
Tn-seq data extraction and visualization

RESULTS AND DISCUSSION

C. sakazakii mutant libraries successfully constructed

Desiccation stress method established and successfully applied

Tn-seq data yields incomplete results

Future Work

CONCLUSIONS

REFERENCES

APPENDIX A

E. coli MFDpir Ptet3 Sequence

APPENDIX B

Tn-seq analysis on a Mac
LIST OF TABLES

Table 1: Strain information for recipient and donor organisms used in this study ....................... 12
Table 2: List of all primers and their sequences used in this study ............................................... 21
Table 3: Sample number key for Genomic Prep methodology and incubation periods for fragmentase digestion ................................................................................................................... 26
Table 4: Sample number key for genomic preparation methodology and second round PCR primers utilized ............................................................................................................................. 30
Table 5: *C. sakazakii* transposants enumerated after selecting for mutants on LB agar supplemented with kanamycin .................................................................................................................. 32
Table 6: Tn-seq output data from 12 samples of treated *C. sakazakii* mutant libraries, as compared to reference genome of *C. sakazakii* Sp291 ................................................................. 39
LIST OF FIGURES

Figure 1: Survival of *Enterobacter sakazakii* and other Enterobacteriaceae after drying and rehydration ................................................................. 7

Figure 2: General desiccation resistance mechanisms ........................................ 8

Figure 3: The intracellular trehalose content of exponential and stationary phase *Enterobacter sakazakii* 1387-2 grown in BHI broth with and without drying at 25°C ............................................. 9

Figure 4: Biparental mating system and transposon insertional mutagenesis in *C. sakazakii* ...... 15

Figure 5: Desiccation stress setup ..................................................................... 19

Figure 6: Tn-seq workflow. .................................................................................. 28

Figure 7: Plate counts of *C. sakazakii* 023, 024, and 036 before desiccation, and post-desiccation with filter paper and post-desiccation plate spread ...................................................... 34

Figure 8: Gel results of crude DNA preparation .................................................. 34

Figure 9: Gel results of genomic DNA extraction ............................................... 35

Figure 10: Gel results of final DNA preparation. ............................................... 35

Figure 11: Hypothetical diagrammatic representation of anticipated Tn-seq output, visualized with IGV. ................................................................. 37

Figure 12: Screenshot of IGV output for desiccated sample of *C. sakazakii* 023 with reference genome *C. sakazakii* SP291 .......................................................................................... 37

Figure 13: Average number of transposants for strains 023, 024, and 036, under rich medium (control) and desiccated conditions ............................................................................. 38

Figure 14: Concentrations of microorganisms at key stages of the research .......... 42
INTRODUCTION

Background and Literature Review

*Cronobacter* is a newly defined bacterial genus, consisting of seven gram-negative, non-spore-forming, yellow-pigmented species that are closely related to the genera *Enterobacter* and *Citrobacter* (1). These bacteria are opportunistic pathogens and one species, *Cronobacter sakazakii*, is of particular concern because of its direct association with infections from contaminated powder infant formula (PIF) (2). PIF and follow-on formulae are predominantly dairy-based non-sterile nutritional or medical food products produced under specialized hygienic conditions. While *Cronobacter spp.* have been isolated from a wide range of environments, its natural habitat remains unknown, though evidence suggests a plant-based niche (3).

The first *Cronobacter* infections were reported in 1961, when two neonates died of serious systemic infections. Pigmented bacteria, later described as *Cronobacter spp.*, were isolated from the cerebrospinal fluid and liver of the two infants, (4). Since that time, *C. sakazakii* infections have been consistently reported, though infections are rare. The CDC’s official statistical report of infections from this organism states: “It seems like sickness from *Cronobacter* does not happen often. CDC usually hears about 4 to 6 babies getting it each year. The problem is that the CDC does not hear about everyone who gets it, so we do not have a full count. *Cronobacter* probably happens more often in older adults than babies, and usually adults do not get as sick as babies when they get *Cronobacter*” (5). Additionally, a national FoodNet survey conducted in 2002 approximated that, “the annual incidence of invasive *Cronobacter* infection at one per 100,000 infants aged <1 year and at 8.7 per 100,000 low birth weight infants (<2,500 g [5.5 lbs])” (6).
Most neonatal infections occur when contaminated PIF is reconstituted, warmed, held, and then ingested by infants. This process creates an ideal environment for recovering many cells and promoting growth of potentially pathogenic bacteria. Furthermore, some nosocomial infections have been traced to *C. sakazakii* adhered to feeding tubes (7). While less common today, *C. sakazakii* infections are often serious, with mortality rates as high as 80% (8). The pathogenicity of *C. sakazakii* enables it to invade brain capillary endothelial cells, leading to meningitis (9). In addition, the bacterium is prone to colonize the gut of premature infants, which can lead to necrotizing enterocolitis, subsequent intestinal perforation, and ultimately systemic infection (10). If the infant survives the infection, he or she may be subject to long-term sequelae like delayed neurological development, hydrocephalus, and permanent neurological damage, consequential of the swelling and inflammation of the brain associated with meningitis (11).

While *C. sakazakii* is a devastating opportunistic pathogen, some specific strains exhibit a higher frequency of infection than others. Most serious clinical meningitis cases caused by *Cronobacter spp.* in neonates during the previous thirty years in six countries were caused by a single sequence type (ST): *C. sakazakii* ST4 (12). We suspect strains of this ST exhibit a higher rate of neonatal infection than other STs because of an enhanced pathogenicity and unusual capacity to resist desiccation in harsh PIF production environments.

Research teams have proposed various biological pathways ST4 organisms employ that increase pathogenicity and persistence in low-moisture environments. *Cronobacter spp.*, most noteworthy ST4 strains, have been shown to produce capsular material that aid its attachment to plant materials and biofilm formation on a variety of surfaces. This capsule is composed of various different components, like cellulose, K-antigen, *Enterobacteriaceae* common antigen, and colonic acid (13). These compounds, and potentially others, are suspected to contribute
directly to the organism’s pathogenicity in high-risk individuals like infants. In fact, *C. sakazakii* capsular material is produced at such high efficiencies that processes and purification strategies are ironically the subject of patents for use as a thickening agent in food products (14, 15). Additionally, this capsular material contributes to its ability to adhere to plant material, which may explain its ability to persist for long periods of time in dry ingredients blended in the final stages of PIF production, including lactose, corn syrup solids, and starches. In addition to its capsular material consisting of capsular polysaccharides, high molecular weight glycopolymers, colanic acid, bacterial cellulose, and various polysaccharide antigens that may increase its pathogenicity (13), *C. sakazakii*, including ST4, has been shown to utilize exogenous sialic acid as a carbon source in order to grow. This has been proposed to have significance in commercial PIF product and clinical settings. Sialic acid is deliberately added as an ingredient in some PIF products. Sialic acid is claimed to enhance brain development in infants in concert with other specialized ingredients including choline, milk fat globule membrane (MFGM), and docosahexaenoic acid (DHA). Sialic acid is also found naturally in breastmilk, mucin, and gangliosides, which all have associations with *Cronobacter* infection (16).

Muytjens et al. conducted one of the first studies to determine the prevalence of *C. sakazakii* in PIF using samples from 35 countries. They found that 14% of samples analyzed (from 13 countries) contained the pathogen (17). In 2017, Fei et al. affirmed that ST4 strains were generally the most desiccation-resistant sequence types (18, 19). Given the reported correlation between high desiccation tolerance and pathogenicity in infants, there is a critical need to understand the mechanisms that enable *C. sakazakii* ST4 to persist in low-moisture foods (LMF) and processing environments. LMF are foods (or food ingredients) with a water activity less than 0.85. These matrices do not support growth of bacteria because the water is tightly
bound and generally unavailable to participate in or enable microbial metabolism (20). PIF, depending on the formulation and processing, generally exhibits a range of 0.150 to 0.225 aw. Additional fundamental research contributing to the limited body of knowledge will potentially inform strategies to prevent further infections globally. Of note, this organism is particularly devastating in the vulnerable premature (constituting >10% of live births) and newborn infant populations. In our study, we focused on the development and application of Tn-seq methods to identify genes and mechanisms in *C. sakazakii* ST4 that contribute to its resistance to desiccation and survival in LMF matrices including PIF.

Though *Cronobacter* spp. have been shown to survive well in LMF, they do not demonstrate exceptional abilities to survive high temperature treatments in liquid (21). Dairy ingredients used in PIF are pasteurized and/or thermally treated before spray-drying using time and temperature parameters that render the reconstituted, liquified, and blended compositions known as “mixes” containing ingredients free of pathogens, including *C. sakazakii*. Based on the use of well-established and validated “kill steps” in the food safety plans of this highly-regulated industry and controls on in-coming goods, the risk of infection has decreased dramatically, though some infections related to PIF still occur. Today, the root causes for the presence of *C. sakazakii* is poor equipment design or sanitation practices that introduce moisture or humidity. Alternatively, the causes are related to the introduction of bacteria after thermal processing during dry blending or packaging (22). Because contaminations and subsequent infections continue to happen, Farber et al. stressed that new approaches need to be developed to control foodborne pathogens in LMF. In a recent review, his team summarized novel methods where additional research is needed to evaluate their effectiveness in lowering contamination. Such methods include high-pressure processing, nonthermal plasma, ultraviolet light, pulsed light,
irradiation, oxidizers, disinfectants, radio-frequency, microwave heating, and even superheated steam drying (23).

*C. sakazakii* strains have been isolated from widely dispersed locations in PIF production facilities, including employee shoes, external roofs, powder processing and air treatment areas, suggesting that distribution is assisted by air movement and production personnel (24). Numerous PIF recalls due to *Cronobacter spp.* contamination have occurred in recent years all over the world, the most recent recall occurring in Canada. Loblaw Companies Ltd recalled their President’s Choice brand PIF, though no reported illnesses came about from the contaminated formula (25). Relatedly, a major recall was issued in the United States in 2011 after a 10-day old infant died of *C. sakazakii* infection, which was traced back to the PIF consumed by the infant (26). In the early 2000s, the organism was the cause of several infant deaths in North America and Europe. Because of the increasing frequency of cases and isolates from PIF samples, the World Health Organization and FAO organized a meeting in 2008, chaired by Dr. Jeffrey Farber, in which the world’s leading experts on *Cronobacter spp.* gathered to assess risk and establish microbial criteria for PIF. In this meeting, a call was given for relevant data and additional research on *Cronobacter spp.* as it relates to PIF and infant health (27). Our research is in response to this call, focusing on the persistence of *C. sakazakii* in desiccated conditions.

While other *Enterobacteriaceae* like *Salmonella spp.* have been isolated from PIF, *Cronobacter* is isolated more frequently (28) and has shown a marked ability to survive desiccation and osmotic stressors better than *Salmonella spp.*, the primary organism of public health concern in a variety of LMF products. Manufacturers are cognizant that *Salmonella spp.* has caused infection, outbreaks and product recalls due to its strong xerotolerance, in comparison
to other *Enterobacteriaceae*. However, *C. sakazakii* has shown greater desiccation tolerance than *Salmonella spp.*, as shown in Figure 1, reported in a study performed by Breeuwer et al. (21).

Bacteria such as Salmonella employ a wide range of mechanisms to survive in dry environments including LMF matrices, as illustrated in Figure 2. Generally, bacteria undergo changes in their cell membrane; downregulate flagellar motility and glycolysis; upregulate fermentative enzymes, catabolic energy pathways, and DNA repair and protection proteins; take up manganese ions to protect against oxidative stress; accumulate shock-response proteins; produce sugars and other osmoprotectant molecules, and form biofilms (29). From a variety of studies, *C. sakazakii* has been shown to produce extracellular polysaccharides, form a biofilm, and accumulate trehalose (7, 16, 30, 31) One study showed that the distinct yellow pigmentation of *C. sakazakii* may contribute to its ability to withstand desiccation stresses. In contrast, colorless mutants showed an increased ability to withstand osmotic stress and an increased susceptibility to desiccation (32).

Some *C. sakazakii* mechanisms for xerotolerance have been studied using qualitative and quantitative analysis methods to measure desiccation stress response products. Breeuwer et al. (21), used high-performance liquid chromatography to measure trehalose produced before and after desiccation stress and found that the bacteria that underwent desiccation stress contained six times more trehalose than the unstressed bacteria, see Figure 3 extracted from this study. In addition, they determined that trehalose production could not be detected in other *Enterobacteriaceae* exposed to the same conditions, suggesting trehalose production to be a specialized mechanism in *Cronobacter spp.* Trehalose, or \(\alpha-D\)-glucopyranosyl-(1\(\rightarrow\)1)-\(\alpha-D\)-glucopyranoside, is extremely stable at high temperatures, and acidic and dry conditions. It has been shown to have protective effects on plasma membranes and proteins and does not react
**Figure 1** Survival of *Enterobacter sakazakii* and other Enterobacteriaceae after drying and rehydration. Cell suspensions (50 µL) were dried in air at 25C (a) and 45C (b) and incubated at this temperature for 46 days. The cells were rehydrated by addition of 500 µL TS, and plated immediately on TPAP agar.
Figure 2 General desiccation resistance mechanisms.

- Changes in cell membrane:
  - Cis → Trans isomerization of fatty acids (keeps membrane in liquid crystalline phase)
  - Monoenoic → Cyclopropane fatty acids (reduces permeability to H+)

- Production of trehalose (osmoprotectant)
- Upregulation of fermentative enzymes
- Downregulation of glycolysis
- Mn2+ uptake (oxidative stress protection)
- Accumulation of shock-response proteins
- Biofilm formation (retains water and protects neighboring cells)
- DNA repair and protection proteins upregulated
- Upregulation of catabolic pathways
- Flagellar motility downregulated
**Figure 3** The intracellular trehalose content of exponential and stationary phase *Enterobacter sakazakii* 1387-2 grown in BHI broth with and without drying at 25°C. The error bars represent the standard deviation of four measurements. Figure extracted from Breeuwer et al (21), with minor modifications.
readily as it is a nonreducing sugar. These characteristics have been proposed to contribute to C. *sakazakii*’s ability to persist in LMF. Another study used a proteomic approach to determine stress response in *C. sakazakii*, showing down-regulation of core metabolic pathways, an increase in proteins serving a structural or protective role, and significant down-regulation of motility apparatus and the formation of filamentous cells (33).

While these and similar studies provide valuable information on the desiccation resistance of *C. sakazakii*, our analysis of published information indicated that few studies have studied the stress responses of this bacterium on a genetic level using state-of-the-art tools in molecular biology. As there is a dearth of evidence to describe how *Cronobacter* spp. are able to resist desiccation better than other Enterobacteriaceae on a genetic level, our study focused on determining the genes that enable ST4 strains of *C. sakazakii* to survive desiccation and persist in PIF due to its notable xerotolerance.

Massively parallel transposon insertion sequencing (Tn-seq) is a powerful tool that can reveal much about an organism’s genetic requirements. A collaborative research project between Brigham Young University and the University of Florence successfully used this tool to not only determine give insight on the phenotypic response of the soil bacterium *S. meliloti* of gene deletions, but to also identify all the essential genes this organism required for growth under laboratory conditions (34). Another powerful feature of this tool is, once a robust mutant library of a bacterium has been created, Tn-seq analysis can be applied repeatedly under different conditions to help understand survival and response mechanisms in a variety of scenarios and environments. This could lead to enhanced strategies to eliminate or minimize *C. sakazakii* proliferation.

Recent studies have shown that the method of Tn-seq is successful in yielding valuable data in *Cronobacter malonaticus*, an organism closely related to *C. sakazakii* that has also shown to cause life-threatening disease in infants. Zhang et al. described a system in which an *E. coli*
The donor organism imparted kanamycin resistance onto the recipient organism, *C. malonaticus*, upon reception of a transposon through biparental mating. Osmotic stress conditions were applied to the mutated *C. sakazakii* and transposon insertion sites were mapped (35). The mapping revealed several candidate genes involved in osmotic tolerance, with key functions such as the transport of potassium, cell inner membrane structure, and biofilm formation. A similar study used transposon mutagenesis technique to study oxidative stress tolerance in *C. malonaticus*, where multiple genetic responses to oxidative stress were identified, in addition to reporting a relationship between biofilm-forming ability and oxidative tolerance (36). Additionally, a study identified genes associated with biofilm formation in *C. sakazakii* using Tn-seq (7). These genes that were identified have functions related to cellulose biosynthesis, flagellar structure, capsular components, and other genes involved in basic cellular process and virulence.

**Research goals and experimental design**

Specific genes that contribute to the ability of *C. sakazakii* to survive thermal and osmotic conditions have been identified previously (7, 32, 37, 38). We designed our study to focus specifically on genes that contribute to xerotolerance. We used transposon mutagenesis and transposon sequencing (Tn-seq) to perform a genetic scan of specific mutants whose genetic responses allow for survival post-exposure to desiccation stressors. In addition, these tools were selected for their potential to determine all essential genes in *C. sakazakii* ST4. To our knowledge, we were the first research team to use these tools and advanced techniques to gain a greater understanding of *C. sakazakii* ST4 xerotolerance. In our study, three strains of *C. sakazakii* ST4 (shown in Table 1) were subjected to transposon mutagenesis.
**Table 1** Strain information for recipient and donor organisms used in this study.

<table>
<thead>
<tr>
<th>Bacterium</th>
<th>Sequence Type</th>
<th>Source</th>
<th>Strain ID</th>
<th>CFSAN ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. sakazakii</em></td>
<td>4</td>
<td>Cornell University, ILSI North America Collection</td>
<td>F6-023</td>
<td>CFSAN080726</td>
<td>Clinical human isolate</td>
</tr>
<tr>
<td><em>C. sakazakii</em></td>
<td>4</td>
<td>Cornell University, ILSI North America Collection</td>
<td>F6-024</td>
<td>CFSAN080727</td>
<td>Infant formula isolate</td>
</tr>
<tr>
<td><em>C. sakazakii</em></td>
<td>4</td>
<td>Cornell University, ILSI North America Collection</td>
<td>F6-036</td>
<td>CFSAN080739</td>
<td>Environmental isolate</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>n/a</td>
<td>Dr. Joel Griffitts, BYU</td>
<td>pJG972</td>
<td>n/a</td>
<td>Donor organism</td>
</tr>
</tbody>
</table>
In order to determine potential genetic contributions during desiccation stress of the listed isolates of *C. sakazakii* ST4, three research objectives were devised.

1. Create at least a 500,000-membered library of *C. sakazakii* transposon mutants for each of the isolates outlined in Table 1, facilitated by an engineered donor strain of *E. coli*, MFDpir.

2. Develop methods to apply desiccation stress under conditions that are highly controlled to the mutant libraries in a manner such that at least 50% of the cells exposed to the desiccation stress survive in order to be genetically analyzed.

3. Identify essential genes and genes that contribute to desiccation resistance in low-moisture foods with the aid of transposon sequencing (Tn-seq), a form of massively parallel DNA sequencing.

For the first objective, the mutant libraries were created using an engineered donor strain of *E. coli*, MFDpir from the molecular biology laboratory of Dr. Joel Griffitts. The donor strain contains the genes necessary for plasmid transfer (i.e. from *E. coli* to *C. sakazakii*) and harbors a transposon-containing plasmid. Replication of this transposon-containing plasmid in MFDpir is regulated by the *pir* gene, found uniquely in MFDpir. Once the transposon-containing plasmid has been successfully transferred to the *C. sakazakii* recipient, the pir-dependent plasmid is no longer able to replicate and insertional mutagenesis can occur, with the transposon being inserted at a low frequency (no more than once per cell) in a random location in the genome. In a library of 500,000 mutants, this means that there is one insertion every 9 base pairs. These interruptions enable analysis via massively parallel sequencing to determine the essentiality and function of certain genes in *C. sakazakii*. The wild-type *C. sakazakii* strains, acting as the recipients, received a transposon-containing plasmid (pJG972) from the engineered *E. coli* MFDpir (donor),
received from the Griffitts Lab with the support of Alex Benedict, a PhD candidate trained in the required molecular techniques. See Figure 4 for a detailed explanation of the plasmid system.

MFDpir has special properties that facilitate the biparental mating process. A series of genes found in the organism’s genome, known as *tra* genes, enable the formation of mating pores through which the plasmid is transferred. In addition, the donor strain contains a *pir* gene (for which it was named) that encodes the Pi protein that pJG972 needs to replicate, ensuring that antibiotic resistance conferred in recipient (*pir*-negative) cells occurs only in cases where the transposon has hopped into the genome. For selection purposes, the donor *E. coli* strain lacks the *dapA* gene, which enables us to use this property as a selection tool, as this organism requires supplementation of diaminopimelic acid (DAP) to grow in media.

The second objective enables the determination of which several thousand *C. sakazakii* genes contribute to desiccation resistance. Desiccation stress was applied to the *C. sakazakii* transposants before analyzing their genomes with massively parallel DNA sequencing similar to the method described by Breeuwer et al. (21). Their team exposed cells to desiccation stress by transferring cells in stationary phase to culture plates, and keeping the plates in a 25°C incubator for air drying. The mean relative humidity (RH) of the incubator was 20.7%. Lithium chloride and other salts can be used to alter the atmospheric RH, if different conditions are desired (39). Under these conditions, the samples were dry in approximately one hour. Conditions adapted from Breeuwer’s experiment were modified to yield survival rate of >50%.

For objective three, after the samples were exposed to desiccation stress, the surviving cells were reconstituted and their genomes analyzed. We attempted to identify genes contributing to desiccation resistance by comparing the number of and types of modifications in sequence reads for specific insertions.
Figure 4 Biparental mating system and transposon insertional mutagenesis in *C. sakazakii*. A) Two-plasmid system found in *E. coli* MFDpir used to facilitate biparental mating and transposon insertional mutagenesis in *C. sakazakii*. The plasmid pJG972 contains key features that enable transposon insertional mutagenesis to occur, including transposase and the transposon itself that contains a kanamycin resistance gene that enables us to select for transposants. The second plasmid, pJG944, contains the gene tetR (tet repressor) that prevents transcription of transposase in the donor organism, which prevents transposase events in the donor organism. B) The plasmid pJG972 is transferred to the recipient organism through mating pores created by tra genes. Once the plasmid has been transferred, the plasmid cannot be replicated, since the recipient organism lacks the pir gene. The transposon from the plasmid can then insert into a single location in the genome of the recipient.
MATERIALS AND METHODS

Bacterial Strains, media, and growth conditions

Three wild-type Cronobacter sakazakii strains, 023, 024, and 036, along with the donor organism E. coli MFDpir pJG972 used in this work are described in Table 1. See Appendix A for the DNA sequence of the specialized plasmid of the donor organism. All strains used in this study were grown at 37˚C. The growth medium for all bacterial strains is Luria-Bertani (LB) medium, consisting of 10 g Bacto Tryptone, 5 g Bacto Yeast Extract, 50 g NaCl, 1 mL 2N NaOH for buffering per 1 L ddH2O. Diaminopimelic acid (DAP 50 μg/mL) was added to this liquid growth medium for growth of E. coli MFDpir pJG972, which is necessary due to this organism’s inability to synthesize DAP independently. For growth on solid medium, 12 g/L Standard Methods Agar was added. The solid media was supplemented with kanamycin (50 μg/mL) or DAP (50 μg/mL) as required.

Tn-seq experimental setup

Transposon mutagenesis was accomplished in the three wild-type C. sakazakii strains in parallel. A 500-mL overnight flask culture of each of the three C. sakazakii strains were grown, along with three 1000-mL overnight flask cultures of E. coli MFDpir pJG972. The cultures were pelleted and washed and suspended in LB to a final OD600 value of approximately 60. Portions of each suspension were combined in a ratio of 2:1, donor to recipient, respectively. To further facilitate these bi-parental matings and increase mobilization of the transposon delivery vector into the C. sakazakii recipient strains, these cell mixtures were plated on LB supplemented with 50 μg/mL DAP, 200 μL per plate with approximately twenty plates of each of the three recipient/donor biparental mating mixtures. The plates were allowed to dry and were then incubated at 37˚C for 6 hours. Directly after incubation, the mating mixtures were collected by
harvesting each plate with 2 mL of LB with 20% glycerol. The collected cells were homogenized by shaking for 30 minutes at 200 RPM. The homogenous mixture was divided into aliquots and stored at -80°C.

In order to select for transposants, the aliquots of the mating mixes were thawed and plated at a density of 1.5x10⁴ cfu/plate onto 150-mm plates on LB supplemented with kanamycin (Km). Approximately 9x10⁵ colonies were selected from each of the three C. sakazakii strains; these were harvested with 2mL LB with 20% glycerol and homogenized by shaking for 30 minutes at 200 RPM. The three different C. sakazakii mutant libraries (023, 024, 036) were aliquoted and stored at -80°C.

**Application of desiccation stress to mutant libraries**

A loopful of each of the three C. sakazakii mutant libraries was inoculated into 10-mL tubes of sterile LB broth and incubated overnight at 37°C to revive cells after cold storage. A 0.1-mL sample of each culture was spread onto LB agar plates with Km (50 μg/mL) and incubated for 24 hours at 37°C. The cells were harvested using LB broth and collected into individual 50-mL centrifuge tubes. Samples of each strain underwent serial dilution and spread onto plates and incubated for 18 hours. All three strains were found to be at an approximate concentration of 10⁸ CFUs/mL. 0.1-mL samples of each C. sakazakii slurry at concentration of 10⁸ CFUs/mL were added to empty petri plates (60 mm x 15 mm) and spread onto the plates to create a thin even layer. This was done in duplicate for each of the three strains.

Once the strains were spread, the lidless petri plates were placed in a large desiccator jar equilibrated to 11.3% relative humidity (RH) at 22°C using a saturated lithium chloride solution placed in the bottom of the desiccator jar. The saturated salt solution was prepared by slowly adding 138 grams of powdered LiCl to 162 mL ddH₂O with constant agitation at room
temperature (22°C) until a thick slurry formed. The desiccator jar was sealed, and the relative humidity and temperature were monitored to ensure stability of RH and temperature. The samples were dried in the sealed jar for 24 hours. Simultaneously, a control of each of the mutant libraries of the 3 strains accompanied the desiccation treatment. These controls consisted of a 1-mL sample of each of the three mutant library slurries placed in 3 separate sealed microcentrifuge tubes and left at room temperature (22°C) for 24 hours. The setup for the desiccation stress is shown in Figure 5.

After the 24-hour drying period, 10 mL of sterile LB broth was added to each of the petri plates containing the samples that underwent desiccation, and lids were placed on the plates before shaking at 100 RPM at 22°C for 20 minutes to resuspend the desiccated cells. Samples of each underwent a serial dilution, were plated on LB + kanamycin (50 μg/mL), and enumerated after 24 hours to ensure sufficient survival of the mutant library exposed to the desiccation stress. It was determined that a reduction of less than one log was satisfactory in order to proceed to the sequencing stage. Simultaneously, the remainder of the 10 mL of each strain and treatments of suspended cells was cultured overnight at 37°C to make a saturated culture. The 10 mL of each were divided into aliquots of approximately 3.3 mL, centrifuged at 9,184 RCF (10,000 RPM) for 5 minutes. The supernatant was removed, and the pellets stored at -80°C.

**Arbitrary PCR**

In order to confirm that transposition events occurred within the chromosomes of the mutant libraries created and used in experimentation, an arbitrary PCR and subsequent sequencing were performed on two randomly selected colonies from each of the three mutant
Figure 5 Desiccation stress setup with desiccator jar and lithium chloride saturated solution for relative humidity regulation.
libraries (023, 024, 036). The standard procedure for colony lysis for PCR preparation was used. Each of the colonies selected was cultured overnight at 37°C and was centrifuged in 50 mL centrifuge tubes, the supernatant was discarded, and the pellet was transferred to a microtube. 50 μL of PCR lysis buffer (72.15 mL sterile ddH2O, 375 μL of 1 M Tris pH 8.0, 600 μL of 0.25 M EDTA, and 1.875 mL of 20% Triton X-100) was added to each of the 6 tubes of pelleted cells and boiled for 2 minutes, followed by vortexing. 1 μL of each sample of the boiled cells was added to 19 μL the first round PCR solution, consisting of the following: 11.9 μl dH2O, 2 μl Taq buffer, 0.5 μL 10mM dNTPs, 0.1 μL Taq polymerase, 1.5 μL 10 μM 714A primer, and 3.0 μL 10 μM oKJ001 primer. All primers (and their sequences) used in this study are listed in Table 2. This 20 μL-solution was transferred to PCR tubes and the PCR tubes were placed in the PCR machine (Applied Biosystems’ Proflex™ Base, Ref# 4483636) and the following settings were programmed: 1. 94 1:30, 2. 94 :15, 3. 33 0:15, 4. 70 :30, 5. 94 :15, 6. 43 :15, 7. 70 :30, 8. 70 1:00, 9. 4 infinity. END.

Once the first round PCR was completed, 1 μL of each first round PCR solution was added to 19 μl of the second round PCR solution, consisting of the following: 13.4 μl dH2O, 2 μL Taq buffer, 0.5 μL 10mM dNTPs, 0.1 μL Taq polymerase, 1.5 μL 10 μM 714B primer, and 3.0 μL 10 μM oKJ003 primer. This 20 μL-solution was transferred to PCR tubes and the PCR tubes were place in the PCR machine and the following settings were programmed: 1. 94 1:30, 2. 94 :15, 3. 54 0:15, 4. 70 :30, 5. 70 1:00, 6. 4 infinity. END. From here, the samples were labeled and put on ice and were sent to Eton Bioscience Inc. for sequencing.

**Genomic DNA Extraction**

The aforementioned pellets containing the transposants exposed to the treatments were removed from cold storage and thawed on ice before using the Quiagen DNA Extraction Kit
Table 2 List of all primers and their sequences used in this study.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
<th>Research Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>714A</td>
<td>CCGCTCCCGATTCG</td>
<td>Arbitrary PCR</td>
</tr>
<tr>
<td></td>
<td>AGCGCATCGCCTTTCTATCG</td>
<td></td>
</tr>
<tr>
<td>714B</td>
<td></td>
<td>Arbitrary PCR</td>
</tr>
<tr>
<td></td>
<td>GCCACGCCTCGACTAGTACNNNNNNNNNN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGGCG</td>
<td></td>
</tr>
<tr>
<td>oKJ001</td>
<td></td>
<td>Arbitrary PCR</td>
</tr>
<tr>
<td></td>
<td>GCCACGCGTCGACTAGTAC</td>
<td></td>
</tr>
<tr>
<td>oKJ003</td>
<td></td>
<td>Arbitrary PCR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1TN714</td>
<td>CAGACGTCGCTCTTCCGATCgggggggggg</td>
<td>1st round PCR</td>
</tr>
<tr>
<td>1OLIGOG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AATGATACGGCGACCACCGAGATCTACACT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTTTCCCTACACGACGCTCTTCCGATCTNN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NNGAGATGTGTATAAGAGACAG</td>
<td></td>
</tr>
<tr>
<td>2TNA</td>
<td>AATGATACGGCGACCACCGAGATCTACACT</td>
<td>2nd round PCR</td>
</tr>
<tr>
<td></td>
<td>CTTTCCCTACACGACGCTCTTCCGATCTNN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NNGAGATGTGTATAAGAGACAG</td>
<td></td>
</tr>
<tr>
<td>2TNB</td>
<td>AATGATACGGCGACCACCGAGATCTACACT</td>
<td>2nd round PCR</td>
</tr>
<tr>
<td></td>
<td>CTTTCCCTACACGACGCTCTTCCGATCTNN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NNGAGATGTGTATAAGAGACAG</td>
<td></td>
</tr>
<tr>
<td>2TNC</td>
<td>AATGATACGGCGACCACCGAGATCTACACT</td>
<td>2nd round PCR</td>
</tr>
<tr>
<td></td>
<td>CTTTCCCTACACGACGCTCTTCCGATCTNN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NNGAGATGTGTATAAGAGACAG</td>
<td></td>
</tr>
<tr>
<td>2TND</td>
<td>AATGATACGGCGACCACCGAGATCTACACT</td>
<td>2nd round PCR</td>
</tr>
<tr>
<td></td>
<td>CTTTCCCTACACGACGCTCTTCCGATCTNN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NNGAGATGTGTATAAGAGACAG</td>
<td></td>
</tr>
<tr>
<td>2INX01</td>
<td>CAAGCAGAAGACGGCATACGAGATTGAGAT</td>
<td>2nd round PCR</td>
</tr>
</tbody>
</table>
GTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

2INX02

2nd round PCR

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

2INX03

2nd round PCR

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

2INX04

2nd round PCR

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

2INX05

2nd round PCR

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

2INX06

2nd round PCR

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

2INX07

2nd round PCR

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

2INX08

2nd round PCR

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

2INX09

2nd round PCR

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

CAAGCAGAAGACCGCATACGAGATACGTC
GGTGACTGGAGTTCAGACGTGTGCTCTTTCCG
ATC

2INX10

2nd round PCR
CAAGCAGAAGACGGCATACGAGATTTGGTA
GTGACTGGAGTTCAGACGTGTGCTCTTCCG
ATC

2nd round PCR

CAAGCAGAAGACGGCATACGAGATGGTTCC
GTGACTGGAGTTCAGACGTGTGCTCTTCC
GATC

2nd round PCR

CAAGCAGAAGACGGCATACGAGATGTTCG
GTGACTGGAGTTCAGACGTGTGCTCTTCC
CGTGAATGCTGTCAGACGTGTGCTCTCC
GATC

2nd round PCR
The pellets were resuspended using 960 μL of T10E1 (73.95 mL ddH2O, 750 μL 1 M Tris pH 8.0, 300 μL .25 M EDTA), and then each of the resuspensions were divided into two microtubes, with approximately 480 μL/tube. The following describes the protocol for each of the 24 tubes that were created in the previous step. 5 μL of 20 mg/ml proteinase K was added and mixed well using a vortexer. 25 μL of 10% SDS was added and mixed well using the micropipettor. The tubes were then incubated on a heat block at 50°C for 25 minutes and were intermittently vortexed. After the incubation, 80 μL of 5M NaCl and 500 μL of CHCl₃ were added before shaking vigorously. The tubes were then microcentrifuged for 25 minutes at 16,002 RCF (13,200 RPM). Using a micropipettor, the bottommost layer (CHCl₃) was carefully removed before microcentrifuging the tubes again for 3 minutes at 16,002 RCF (13,200 RPM).

Next, approximately 500 μL of the aqueous phase was transferred to a new tube, taking great caution to not disturb the proteinaceous upper layer. 1 μL of RNase (20 mg/ml) was added and mixed well with micropipettor before incubating at 30°C for 20 minutes in a standing incubator with a pre-warmed microtube rack. After the incubation, 500 μL isopropanol was added. At this stage, the DNA precipitate was faintly visible and became even more prominent after 5 minutes on ice. The tubes were then microcentrifuged for 2 minutes at 16,002 RCF (13,200 RPM). The supernatant was removed, leaving a small pellet of DNA. The tubes were microcentrifuged under the same conditions for a second time and a micropipettor was used to carefully remove any remaining supernatant from the pellet.

100 μL of prewarmed T10E1 was added to each of the tubes with the DNA pellets in order to dissolve the pellets. At this stage, this was considered to be the crude DNA prep. We chose to check a 4-μL sample of each tube on a 1% agarose gel to ensure that we extracted sufficient DNA to move on to subsequent steps of the DNA prep. From the gel, we were able to
select the best twelve samples to move forward with on subsequent steps. We selected one of the two samples from each treatment based on DNA quantity as determined by subjective brightness.

In order to complete the following steps of DNA prep, the DNeasy PowerLyzer Microbial Kit (Qiagen 12255-50) was used. To each of the twelve samples selected, 300 μL of buffer SB was added to 100 μL of the crude DNA preparation. This mixture was then applied to individual columns contained in the kit and microcentrifuged for 1 minute at 16,002 RCF (13,200 RPM). The flowthrough was discarded and 300 μL of CB was applied to the column before microcentrifuging for 1 minute at 16,002 RCF (13,200 RPM), discarding the flowthrough. The columns were then centrifuged again for 30 seconds to remove residual buffer and the columns were then transferred to a new microtube. 100 μL of warm 2.5 mM Tris 8.0 was applied to each column and the columns were allowed to incubate for 1 minute before microcentrifuging for 1 minute at 16,002 RCF (13,200 RPM). The flowthrough is the extracted genomic DNA. A 4-μL sample of each tube was checked on a 1% agarose gel to ensure presence of DNA.

**Fragmentase digestion**

16 μL of the genomic DNA of each of the 12 samples were put into new tubes and 2 μL of 10X fragmentase v.2 buffer was added and mixed well with the micropipettor. The remainder of the genomic DNA was stored at -40°C for preservation until further experimentation. The fragmentase was pre-vortexed and then 2 μL of the fragmentase was added into each tube and mixed immediately. The tubes were then incubated at 37°C in a standing incubator with a pre-warmed microtube rack for varying incubation periods, depending on the concentration of DNA in each sample, as determined visually by brightness from the gel of the genomic DNA extraction (see Table 3 for incubation periods). After the incubation period, 10 μL of 0.25 M
Table 3 Sample number key for Genomic Prep methodology and incubation periods for fragmentase digestion.

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Strain</th>
<th>Treatment</th>
<th>Replicate</th>
<th>Fragmentase digestion incubation period (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>C. sakazakii</em> 023</td>
<td>Rich Media (Control)</td>
<td>a</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td><em>C. sakazakii</em> 023</td>
<td>Rich Media (Control)</td>
<td>b</td>
<td>9.5</td>
</tr>
<tr>
<td>3</td>
<td><em>C. sakazakii</em> 023</td>
<td>Desiccated</td>
<td>a</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td><em>C. sakazakii</em> 023</td>
<td>Desiccated</td>
<td>b</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td><em>C. sakazakii</em> 024</td>
<td>Rich Media (Control)</td>
<td>a</td>
<td>8.5</td>
</tr>
<tr>
<td>6</td>
<td><em>C. sakazakii</em> 024</td>
<td>Rich Media (Control)</td>
<td>b</td>
<td>8.5</td>
</tr>
<tr>
<td>7</td>
<td><em>C. sakazakii</em> 024</td>
<td>Desiccated</td>
<td>a</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td><em>C. sakazakii</em> 024</td>
<td>Desiccated</td>
<td>b</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td><em>C. sakazakii</em> 036</td>
<td>Rich Media (Control)</td>
<td>a</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td><em>C. sakazakii</em> 036</td>
<td>Rich Media (Control)</td>
<td>b</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td><em>C. sakazakii</em> 036</td>
<td>Desiccated</td>
<td>a</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td><em>C. sakazakii</em> 036</td>
<td>Desiccated</td>
<td>b</td>
<td>9.5</td>
</tr>
</tbody>
</table>
EDTA was promptly added to stop the reaction. The samples were then placed on ice until they were ready for the Zymo clean-up.

**Zymo clean-up**

200 μL of EndoWash was added to each of the fragmentated samples. This mixture of each sample was then applied to a spin column, microcentrifuged for 30 seconds, and the flowthrough was discarded. 400 μL of PlasmidWash was added to each column and the samples were centrifuged for 1 minute. The columns were transferred to new microtubes and 50 μL of warm 2.5 mM Tris pH 8.0 was added to each. This was left to sit for 1 minute before microcentrifuging for 1 minute at 16,002 RCF (13,200 RPM). The samples were then checked on a 1% agarose gel with a runtime of 20 minutes to check for a broad smear that ranged from approximately 500-3000 base pairs (bp) before moving forward.

**C-tailing of genomic DNA fragments**

29 μL of each of the fragmentated and Zymo-cleaned genomic DNA was combined with 4 μL of 10X TdT buffer, 4 μL of 2.5 mM CoCl2, 3 μL 9.5 mM dCTP/0.5 mM ddCTP mix (34 μL ddH2O, 4 μL 100 mM dCTP + 2 μL 10 mM ddCTP), and 0.7 μL of TdT enzyme (mixed well). This mixture in each tube was mixed well using a micropipettor and were then incubated for 30 minutes at 37°C. Another zymo clean-up was performed as described previously. See Figure 6 for a visual representation of C-tailing and preparing the *C. sakazakii* genomic DNA for sequencing.

**First-round PCR**

The c-tailed and zymo-cleaned samples were thawed on ice. 35 μL of a master mix containing 213.6 μL ddH2O, 96 μL 5X Q5 buffer (NEB M0291S), 14.4 μL 10 mM dNTPs, 6 μL
Figure 6 Tn-seq workflow. The genomic DNA is extracted, fragmented, purified, modified, and amplified. This figure was modified and adapted from an original figure by Dr. Joel Griffitts.
Q5 polymerase (NEB M0291S), 30 μL diluted (10 μM) 1TN714 primer, 60 μL diluted (10 μM) 1OLIGOG were added to 5 μL of thawed template DNA put into individual PCR tubes.

The PCR tubes were place in the PCR machine (get actual name from lab) and the following settings were programmed: 1. 96 1:00, 2. 96 :20, 3. 60 0:30, 4. GOTO 2, 18 times, 6. 72 1:00, 7. 4 Infinity, 8. END. Another zymo clean-up was then performed as described previously.

**Second-round PCR**

35 μL of a master mix containing 255.6 μL ddH2O, 96 μL 5X Q5 buffer (NEB M0291S), 14.4 μL 10 mM dNTPs, 6 μL Q5 polymerase (NEB M0291S), 36 μL diluted (10 μM) 2TNX primer, 36 μL diluted (10 μM) 2INX were added to 3 μL of thawed template DNA put into individual PCR tubes. See table below for reference to individual 2TNX and 2INX that were added uniquely to each sample (see Table 4).

The PCR tubes were place in the PCR machine (Applied Biosystems’ Proflex™ Base, Ref# 4483636) and the following settings were programmed: 1. 96 1:00, 2. 96 :20, 3. 60 0:30, 4. GOTO 2, 25 times, 6. 72 1:00, 7. 4 Infinity, 8. END. Another zymo clean-up was then performed as described previously. The samples were then checked on a 1% agarose gel with a runtime of 20 minutes to check for a smear running from 150-700 bp, with a strong peak around 400 bp.

**Illumina Sequencing**

The DNA Sequencing Center (DNASC) at Brigham Young University, under the direction of Dr. Edward Wilcox, was used to sequence the 12 samples prepared as described above. The 12 samples, containing approximately 50 μL of liquid with an unknown concentration of prepped DNA, were labeled and delivered. The DNA was quantified and pooled.
Table 4 Sample number key for genomic preparation methodology and second round PCR primers utilized.

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Strain</th>
<th>Treatment</th>
<th>Replicate</th>
<th>2\textsuperscript{nd} Round PCR Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C. sakazakii 023</td>
<td>Rich Media (Control)</td>
<td>a</td>
<td>2TNA+2INX01</td>
</tr>
<tr>
<td>2</td>
<td>C. sakazakii 023</td>
<td>Rich Media (Control)</td>
<td>b</td>
<td>2TNB+2INX02</td>
</tr>
<tr>
<td>3</td>
<td>C. sakazakii 023</td>
<td>Desiccated</td>
<td>a</td>
<td>2TNC+2INX07</td>
</tr>
<tr>
<td>4</td>
<td>C. sakazakii 023</td>
<td>Desiccated</td>
<td>b</td>
<td>2TND+2INX08</td>
</tr>
<tr>
<td>5</td>
<td>C. sakazakii 024</td>
<td>Rich Media (Control)</td>
<td>a</td>
<td>2TNC+2INX02</td>
</tr>
<tr>
<td>6</td>
<td>C. sakazakii 024</td>
<td>Rich Media (Control)</td>
<td>b</td>
<td>2TND+2INX04</td>
</tr>
<tr>
<td>7</td>
<td>C. sakazakii 024</td>
<td>Desiccated</td>
<td>a</td>
<td>2TNA+2INX09</td>
</tr>
<tr>
<td>8</td>
<td>C. sakazakii 024</td>
<td>Desiccated</td>
<td>b</td>
<td>2TNB+2INX10</td>
</tr>
<tr>
<td>9</td>
<td>C. sakazakii 036</td>
<td>Rich Media (Control)</td>
<td>a</td>
<td>2TNA+2INX05</td>
</tr>
<tr>
<td>10</td>
<td>C. sakazakii 036</td>
<td>Rich Media (Control)</td>
<td>b</td>
<td>2TNB+2INX06</td>
</tr>
<tr>
<td>11</td>
<td>C. sakazakii 036</td>
<td>Desiccated</td>
<td>a</td>
<td>2TNC+2INX11</td>
</tr>
<tr>
<td>12</td>
<td>C. sakazakii 036</td>
<td>Desiccated</td>
<td>b</td>
<td>2TND+2INX12</td>
</tr>
</tbody>
</table>
to approximately equal amounts. BluePippin equipment was used to size select for DNA fragments within the range of 250-500 base pairs in length. Samples were spiked with 10% PhiX before being loaded onto an entire two-lane flow cell that was then run on their Illumina 2500 Rapid-run Sequencing equipment. Results of the sequencing of the samples were submitted to the Office of Research Computing’s online cluster, where all sequencing files are stored.

**Tn-seq data extraction and visualization**

In order to manipulate and prepare the data received from the DNASC into a usable and interpretable form for analysis, instructions outlining Tn-seq analysis on a Mac, written in the Griffitts lab, were used. The complete instructions are found as Appendix B. Our original goal was to use the actual annotated genomes of our three *C. sakazakii* strains (023, 024, 036) to use as the reference genome for the mutant genomes we sequenced; however, annotation and preparation of those genomes is still underway. We used the fully annotated genome of *C. sakazakii* SP291 as our reference genome for analysis. Due to the highly conserved nature of *C. sakazakii* ST4 genes, we felt confident using this strain as a reference genome and the alignment proved to be sufficient for further analysis.

**RESULTS AND DISCUSSION**

*C. sakazakii* mutant libraries successfully constructed

The number of transposants enumerated after selecting for mutants on LB agar supplemented with kanamycin is found in Table 5. Though there was some variability in number of transposants among the three strains (023, 024, 036) manipulated with the donor *E. coli* MFDpir, each strain appeared to surpass our target number of transposants of 500,000. This initial indication of the size of our mutant libraries for each of the three strains suggested
Table 5 *C. sakazakii* transposants enumerated after selecting for mutants on LB agar supplemented with kanamycin.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Target Transposants</th>
<th>Achieved Transposants</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. sakazakii</em> 023</td>
<td>500,000</td>
<td>928,000</td>
</tr>
<tr>
<td><em>C. sakazakii</em> 024</td>
<td>500,000</td>
<td>920,000</td>
</tr>
<tr>
<td><em>C. sakazakii</em> 036</td>
<td>500,000</td>
<td>1,900,000</td>
</tr>
</tbody>
</table>
readiness to advance to the next stage of research. In addition, the positive results of the arbitrary PCR method conducted on all three strains showed that each mutant library did have transposon insertions in different loci on the genome, likewise indicating readiness for advancement to the next stage of research. However, our later Tn-seq results found that the number of transposants were likely not as numerous as initially recorded. This may have been due to several factors, which will be discussed in the *Tn-seq data yields incomplete results* section below.

**Desiccation stress method established and successfully applied**

The method of using a saturated lithium chloride salt solution proved effective in exposing the mutant libraries to extreme desiccation stress. The environment inside the desiccator jar maintained a consistent RH of 11.3% after approximately five hours of equilibration after the samples were introduced into the system. This method yielded a population decrease of no more than one log. See Figure 7 for plate counts before and after desiccation.

**Tn-seq data yields incomplete results**

All preliminary preparation of the DNA for Illumina sequencing was successful, as affirmed by gel electrophoresis checkpoints at various stages of the DNA preparation. During the genomic DNA extraction stage, the gel indicated that we did indeed extract sufficient DNA to move on to subsequent steps of the DNA prep, see Figure 8. Additionally, see Figure 9 for gel results of genomic DNA extraction, showing that adequate amounts of crude DNA were extracted from each sample of the samples selected after genomic DNA extraction. After the second round of PCR, the samples were again checked on a 1% agarose gel to ensure a smear running from 150-700 bp, with a strong peak around 400 bp for each sample (see Figure 10). It
Figure 7 Plate counts of C. sakazakii 023, 024, and 036 before desiccation, and post-desiccation with filter paper and post-desiccation plate spread. The error bars represent the standard error of three measurements.

Figure 8 Gel results of crude DNA preparation. Lanes marked with a sample number indicate samples chosen for subsequent steps, based on their higher concentrations. See Table 2 for sample number key.
**Figure 9** Gel results of genomic DNA extraction. See Table 2 for sample number key.

![Gel results of genomic DNA extraction](image1)

**Figure 10** Gel results of final DNA preparation.

![Gel results of final DNA preparation](image2)
was determined that the DNA prep on all treatments were satisfactory and the samples were successfully sequenced with Illumina.

We were successfully able to align the Illumina-sequenced genomes of the mutated cells of the three *C. sakazakii* strains (023, 024, 036), under the various treatments previously described, to the reference genome of *C. sakazakii* SP291. In total, data from twelve samples were successfully produced and visualized with IGV. Though we were anticipating getting visualized data similar to Figure 11, our data for all twelve samples was much sparser, see Figure 12. Given that we were not able to achieve the required transposon density for the traditional mode of analysis, we began a new approach to compare changes in transposon insertions between treatments within a specific gene. We sought to find a transposon insertion ratio of at least 1:100 among the same genes in different treatments, and ultimately the treatments were not able to be cross-compared. This was due to the vast difference in total number of transposon insertions between counterpart treatments within a strain. See Figure 13 for a comparison of the average number of transposon insertions.

Table 6 shows the Tn-seq output data from 12 samples of treated *C. sakazakii* mutant libraries, as compared to reference genome of *C. sakazakii* Sp291, and displays the incapabilities of cross-comparison. For *C. sakazakii* 023, the desiccated sample had a high magnitude of insertions, though only 27.3% of its genes had more than 100 insertions for a given gene, suggesting a low dispersion of insertions throughout the genome. The rich medium (control) samples of this strain did not have a high number of transposon insertions, with only 0.15% of its genes having more than 100 insertions. *C. sakazakii* 036 shared a similar pattern, though 44.60% of its genes had more than 100 insertions on the desiccated samples and 0.08% on the rich medium samples. Conversely, *C. sakazakii* 024, the rich medium (control) samples had a high
Figure 11 Hypothetical diagrammatic representation of anticipated Tn-seq output, visualized with IGV.

![Hypothetical diagram](image1)

- **Control**
- **Desiccated**

Figure 12 Screenshot of IGV output for desiccated sample of C. sakazakii 023 with reference genome C. sakazakii SP291.

![Screenshot of IGV output](image2)
**Figure 13** Average number of transposants for strains 023, 024, and 036, under rich medium (control) and desiccated conditions. Data are displayed on a logarithmic scale. The error bars represent the standard error of two measurements.
Table 6 Tn-seq output data from 12 samples of treated C. sakazakii mutant libraries, as compared to reference genome of C. sakazakii Sp291

<table>
<thead>
<tr>
<th>Strain</th>
<th>Treatment</th>
<th>Replicate</th>
<th>Total transposon insertions</th>
<th>Genes with transposon insertions</th>
<th>% genes with transposon insertions</th>
<th>Genes with transposon insertions &gt; 100</th>
<th>% genes with transposon insertions &gt; 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. sakazakii 023</td>
<td>Rich Media (Control)</td>
<td>a</td>
<td>5248</td>
<td>2068</td>
<td>49.4</td>
<td>3</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>8362</td>
<td>1005</td>
<td>24.0</td>
<td>9</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Average:</td>
<td></td>
<td>6805</td>
<td>1536.5</td>
<td>36.7</td>
<td>6</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Stdev:</td>
<td></td>
<td>2201.9</td>
<td>751.7</td>
<td>18.0</td>
<td>4.2</td>
<td>0.1</td>
</tr>
<tr>
<td>C. sakazakii 023</td>
<td>Desiccated</td>
<td>a</td>
<td>9786991</td>
<td>1082</td>
<td>25.9</td>
<td>1290</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>8662794</td>
<td>1748</td>
<td>41.8</td>
<td>992</td>
<td>23.7</td>
</tr>
<tr>
<td></td>
<td>Average:</td>
<td></td>
<td>9224892.5</td>
<td>1415</td>
<td>33.9</td>
<td>1141</td>
<td>27.3</td>
</tr>
<tr>
<td></td>
<td>Stdev:</td>
<td></td>
<td>794927.3</td>
<td>470.9</td>
<td>11.2</td>
<td>210.7</td>
<td>5.0</td>
</tr>
<tr>
<td>C. sakazakii 024</td>
<td>Rich Media (Control)</td>
<td>a</td>
<td>17645160</td>
<td>3183</td>
<td>76.1</td>
<td>2869</td>
<td>68.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>14950905</td>
<td>2686</td>
<td>64.2</td>
<td>2028</td>
<td>48.5</td>
</tr>
<tr>
<td></td>
<td>Average:</td>
<td></td>
<td>16298032.5</td>
<td>2934.5</td>
<td>70.2</td>
<td>2448.5</td>
<td>58.6</td>
</tr>
<tr>
<td></td>
<td>Stdev:</td>
<td></td>
<td>1905126.0</td>
<td>351.4</td>
<td>8.4</td>
<td>594.7</td>
<td>14.2</td>
</tr>
<tr>
<td>C. sakazakii 024</td>
<td>Desiccated</td>
<td>a</td>
<td>8180</td>
<td>1544</td>
<td>36.9</td>
<td>4</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>9110</td>
<td>1567</td>
<td>37.5</td>
<td>6</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Average:</td>
<td></td>
<td>8645</td>
<td>1555.5</td>
<td>37.2</td>
<td>5</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Stdev:</td>
<td></td>
<td>657.6</td>
<td>16.3</td>
<td>0.4</td>
<td>1.4</td>
<td>0</td>
</tr>
<tr>
<td>C. sakazakii 036</td>
<td>Rich Media (Control)</td>
<td>a</td>
<td>5710</td>
<td>1186</td>
<td>28.4</td>
<td>2</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>9435</td>
<td>1368</td>
<td>32.7</td>
<td>6</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Average:</td>
<td></td>
<td>7572.5</td>
<td>1277</td>
<td>30.6</td>
<td>4</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Stdev:</td>
<td></td>
<td>2634.0</td>
<td>128.7</td>
<td>3.0</td>
<td>2.8</td>
<td>0.04</td>
</tr>
<tr>
<td>C. sakazakii 036</td>
<td>Desiccated</td>
<td>a</td>
<td>23769113</td>
<td>2596</td>
<td>62.1</td>
<td>1865</td>
<td>44.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>13815922</td>
<td>2495</td>
<td>59.7</td>
<td>1794</td>
<td>42.9</td>
</tr>
<tr>
<td></td>
<td>Average:</td>
<td></td>
<td>18792517.5</td>
<td>2545.5</td>
<td>60.9</td>
<td>1829.5</td>
<td>43.8</td>
</tr>
<tr>
<td></td>
<td>Stdev:</td>
<td></td>
<td>7037968.9</td>
<td>71.4</td>
<td>1.7</td>
<td>50.2</td>
<td>1.2</td>
</tr>
</tbody>
</table>
magnitude of transposon insertions, though demonstrated low dispersion, with only 58.60% of its genes having more than 100 insertions or more. The desiccated samples for 024 were low and only 0.10% of genes had more than 100 insertions. We concluded the inconsistencies and complexities were too abundant to be able to draw confident conclusions about essential genes or genes that contribute to desiccation resistance in these strains.

We suspect that the transposon system used favored specific loci on the *C. sakazakii* genome, yielding many mutants, but not yielding the volume of diverse mutants required. Data outputs support this as, for example, only 44.9% of genes among the strains had any transposon insertions. Also, only 21.6% of genes had more than 100 insertions, which we concluded was the minimum number of transposons in a gene needed in order to perform further analysis on the gene’s candidacy for essentiality or contribution to desiccation resistance.

There are many factors that may have contributed to the low density of data points from Tn-seq. Undetected errors in our system for selecting for and enumerating mutants on kanamycin media could contribute to the low number of transposants. Quality deficiencies in the media could potentially have resulted in false positives for number of mutants. Though multiple small-scale experiments were run to ensure that the biparental matings were successful and the selective Km agar was effective in selecting for *C. sakazakii* mutants, there may have been unforeseen complexity in the scale-up of the matings and selective media enumerations.

Research conducted by Zhang et al. described a system to select for *C. malanoticus* transposants which includes an additional measure to check for true mutants. True mutants in their system would be able to grow on kanamycin-supplemented media, due to resistance obtained from the transposon, but would not be able to grow on ampicillin-supplemented media (35). Complications presented in our study when our *C. sakazakii* strains showed a level of
natural resistance to ampicillin, wherein consistent growth of all three strains was noted on LB agar supplemented with ampicillin. Because of this, we were not able to check for true mutants in a similar way. While the step of selecting for true mutants on ampicillin is not absolutely necessary, a similar additional checkpoint could be valuable in future research to help further ensure the avoidance of false positives. Further screenings can be performed in order to identify an antibiotic that can help check for true mutants in our system, such as chloramphenicol.

Additionally, and most importantly, there may have been issues with bottlenecking in earlier stages of the research. Figure 14 shows the workflow of the most important stages of the research in which maintaining certain concentrations of cells and transferring a representative portion of the mutant library are imperative to obtaining successful Tn-seq data. The most likely bottlenecking event that occurred was the stage in which a loopful of the mutant library was transferred and enriched before desiccation. In order to maintain the representative concentration of the library, at least 100 µL needed to be transferred at this step. The low quantity transferred likely led to an underrepresentation of the mutant library for each strain.

**Future Work**

The Tn-seq results show that, with some procedural modifications and additional checkpoints, the general experimental design described will be capable of identifying genes and mechanisms in *C. sakazakii* ST4 that contribute to desiccation resistance and survival in LMF matrices, including PIF. Unfortunately, the data from these experimentations were not sufficiently robust to reveal clear and valuable genetic information on this organism.
**Figure 14** Concentrations of microorganisms at key stages of the research. In order to prevent bottlenecking and poor representation of the mutant library in the final Tn-seq data, concentrations of cells must be tracked and maintained to ensure good library representation. Bottlenecking may have occurred in this research when an insufficient amount of cells was transferred from the stored aliquots of the mutant library, to enrichment, and then to the desiccation stress. 100 µl of the concentrated mutant library was needed to prevent bottlenecking.

Donor (MFDpir): >10⁹ CFU/mL

Recipient (C. sak): >10⁶ CFU/mL

Mutant Library: ~10⁶ transposants

Stored aliquots: ~10⁸ transposants/mL

Desiccation stress: ~10⁸ transposants/mL
To increase the likelihood of higher density insertions in future work with *C. sakazakii* strains (023, 024, 036), modifications are needed during the biparental mating stage. Though preliminary research was done on modifying the growth medium for the biparental matings, more research should be conducted to determine if supplements, such as calcium chloride or glucose increase mating efficiency. It would also be of interest to decrease the biparental mating time from 6 hours to 3 or 4 hours to prevent an overabundance of identical mutants caused from cellular division. Additionally, Tn-seq methods reported in literature have a multi-stage selective media process to select for true mutants (40). It would be of great interest to design a plasmid system that circumvents the native antibiotic resistance observed in order to add another checkpoint to ensure the colonies that result from the biparental mating stage are indeed transposon mutants.

Furthermore, though the step of arbitrary PCR was performed to confirm that suspect mutant colonies were indeed true mutants and that transposon mutagenesis was occurring, it would be beneficial to screen more mutants. At this stage we selected two mutants from each of the three mutant libraries created. In contrast, other researchers, including Hingston, et al., perform arbitrary PCR on twenty-five randomly selected colonies or more to confirm that transposon mutagenesis indeed occurred and that transposition events were widespread throughout the genome (40). Unfortunately, the poor dispersal of transposon insertions was not observed at this stage because our sample size was insufficient.

If future work provides data of sufficient quality to identify genes suspect in essentiality or contribution to desiccation resistance, it would be pragmatic to confirm their role in either function by doing a targeted knockout of these genes. For example, if a gene was identified that codes for a specific capsular component, this gene could be specifically knocked-out and the
organism’s ability to survive could be tested against organisms with the fully functional gene of interest. These and other methods will be explored to further the work of understanding C. sakazakii’s abilities to persist in LMF environments.

CONCLUSIONS

The first objective of this research, entailing creating a mutant library for strains C. sakazakii 023, 024, and 036 using a transposon-containing plasmid system from E. coli MFDpir, was partially successful. The biparental mating system was successful in introducing the transposon into the C. sakazakii genomes, as confirmed by limited arbitrary PCR. In principle, a sufficient quantity of mutants and transposon insertions were reached for each library. However, later sequencing revealed that the transposon mutagenesis did not yield sufficient coverage of the whole genome.

Using the mutants created from objective one, we succeeded in developing a practical method for exposing cells to extreme desiccation conditions using a saturated lithium chloride salt solution, while achieving a survival rate of no more than one log decrease, thus fulfilling objective two. This method was practical and can add value to other studies involving low-moisture foods or desiccated bacterial cells.

Our third and final objective of determining the essential genes and those genes that contribute to the ability of C. sakazakii ST4 to persist in LMF was not satisfied. The analysis after whole genome sequencing revealed that the dispersal of transposons through the genome of each strain was not widespread enough to draw conclusions about gene essentiality or desiccation resistance in our target organism. As a consequence, mechanisms related to desiccation resistance were also not able to be explored. While desired outcomes were not
achieved and associated inferences were truncated, key opportunities to improve methods, including additional checkpoints to increase success, were identified. Future work will be performed, including taking care to prevent bottlenecking of the mutant library population, to optimize these novel Tn-seq methods with *C. sakazakii* ST4 to gain a better understanding of its desiccation tolerant tendencies in order to prevent contamination in PIF and infection in neonates.

REFERENCES


APPENDIX A

E. coli MFDpir Ptet3 (pJG972) Sequence

CCTGCTTCGGGGGTCATATTATAGCGATTTTTTCGCTATATCCATCTCTTTTCGCACGATA
TACAGGATTTTGCCAAAGGRTGCTGTAGACTTTCTCTTCTTGTATCTCAAAGCCGCTCA
GCCGGCAGGATAGGTGAAAGTACCCACCACCGCGAGCGGGTGTTCCTTCTTCCACTGT
CCCTTATTTCGCACCTGGGCGTGCTCAACGGAAATCTTGCTCTGCAGGCTGGCCGGC
TACCGCCCGGTAAACAGATGAGGGAAGTGCTGTAGCTGAAGAGCAAGCCAACCAAGCCAACCA
GGAAAGGCCAGCCAGATGCGTCGGCGCTGGCTGGCGCTGGCGGTGGAC"
TTCAATTCCAGGGCACCGGACAGGTCGCTTTGACAAAAAGAACCGGGCGCCCCTGCG
CTGACAGCCGGAACACGGCCGCATCAGAGCAGCCGATTGTCTGGTGCCCAGTCA
TAGCCGAATAGCCTCTCCACCCAAGCGGCCAGAACCTGCGTCAATCCATCTTGT
TCAATcatGCGAAAAGCATCTCCTCTGCTCTCTTGATCAGATCTTGACAGACCTCTTGT
TCAGATCCTGCGCGCAAGAAAGCCATCCAGTTTACTTTGCAAGGCTTCTCCAAACCTT
ACCAGAGGCCGCCAGGCTGGCAATACCCTTCGCTGGCTTGTGGGGGATCCATCAAT
cttagatgtgtataagacagGTACCTTGACACTTTATCACTGATAAAACTATAATACCTTACGT
GATAAagggACGTTCatgATAACTTCTGCTCTCATCCTGCAGGCGCGACTGGGCTAAAT
CTGTGTTCCTCTCCTGGCCGCCTGGGTGATCCTCCGCGCTACTGCCCCTGGTGAAGGT
CGCCGCCAAATGGCAAAAATATTCTGGTAAATCATAAACCACATTCAATCAGAGGGTAG
TGAAAGCCGCCAGGAGGCGCTTACGGATTTTACTCCGCAATCCAAACGTTCTGCGGA
GGCGATCAGAAAGGCTGGTGGCTTGTGGGGGATCCATCGAGCTGACGAGGT
ACCTGCTGCGCCATTGAGGACACCACCTCTTTGAGTTATCGGCACAGGTCCGCAAG
AGCTTGGAAGCTGGCTCTCTATTCTAGTAAATCCCGCGGATGGGTGGTTCATCCCG
TTCTCTTCGcctgagGCCACCACATCCCGCACCCTAGGATTACTGCAATGAGGAGTGGTGG
ATGCGCCCGATGACCCCTGCGAGCTGGGATGAAAAAGGAAAGTGGCAATGGCTGGG
AGCGGCCGAGACTAGCCGGTGTTAGCATGGGCAGCATGATGAGCAGGTGGTGGTATC
TGTACAGACCATCTGAAGAACCAACGAACTGGATGGGATGCTACGATTATCATCCAACG
AAAAGGGCGTGGGTGGTAAACGCTGAAAGCTGGGCGCATCACGCTAAACAGGGGAATATCAGCTCA
ACGCGGTTGCTGGCCGAGGAGGAGTCCCAGGCTTCTCAACCGCAAGGCTGGCAATGG
APPENDIX B

Tn-seq analysis on a Mac

#### 1. Know your way around using the command-line interface ####

Open Terminal. This application is found in the utilities folder within your application folder. After you have opened Terminal for the first time, set it to stay on your dock. Within Terminal, you are in a default shell (i.e. command-line interpreter) called Bash. This environment has its own language, a little of which you must know:

On your mac, open Terminal. You will find that a prompt is awaiting your command. The prompt will look a little like this:

```
LSB3131-1LM:~ joelsg$
```

Note that it contains your homefolder name followed by a $ sign. The ~ indicates you are starting in your home directory (folder) which in my case is called joelsg. In this manual, we will refer to your prompt simply as:

```
$
```

When you type in a command at this prompt, you are to follow the command by hitting the return button on your keyboard to execute the command.

When you are in the command line interface, you should think of yourself as being physically in some specific directory (i.e. folder), with the opportunity of moving to any directory that you would like. This isn’t a new world—these are the same directories that you are used to navigating around in the graphical operating system. In the command line, your moves about your computer’s directories are accomplished by text commands instead of by mouse. There are many tasks that can be accomplished only through the command line.
While on the command line there are two complementary ways (commands) to assess where you are in your computer’s directory structure at any moment: ls and pwd. ls (list) prints the contents of the working (current) directory onto your screen. pwd (print working directory) outputs the complete path to that directory, so you can see where you are in the context of your home directory. For example, if I am in my home folder, joelsg, also defined by the tilde symbol (~), I can type in

$ ls
and it will output on the terminal screen all the directories (Documents, Desktop, etc.) and files within that directory. If I use the pwd command,

$ pwd
the path to joelsg is printed on my screen. This path is very simple because it is so high up in the directory structure: /Users/joelsg. The first / represents the root directory (the highest level of directory in my system), and the terms that follow represent the lower directories “Users” and “joelsg.” This is a good moment to mention that most operations in the command line are case-sensitive. Whenever naming files or directories, I try to use only lower-case, and I never use spaces. Instead of spaces I use the underscore symbol, _.

Here is a copy of input/output from my command line after running ls and pwd from my home folder:

LSB3131-1LM: ~ joelsg$ ls
Applications      Library      Pictures      Writing      tnseqtool
Desktop          Lists        Present       assemblies
Documents        Misc         Protocols     igv
Downloads        Movies       Public        programs
Dropbox          Music        Teaching      scripts
LSB3131-1LM: ~ joelsg$ pwd
/Users/joelsg
LSB3131-1LM: ~ joelsg$

Notice that after typing ls (return), it spits out the output and restores my prompt so I can input another command (in this case pwd).

You can move around using the “change directory” command, cd. For example, if I am in my home folder (joelsg), and I want to get to my Desktop folder, I would type the command

$ cd Desktop

Now, to prove I made it into the desktop directory, I can ls to list all the contents of the working directory (presumably the desktop, since I cd’d myself there), and I should see the contents of my desktop listed. I can also pwd, which outputs a report of the path to my current working directory.

Notice that in all of these activities you aren’t really getting any work done, but you are learning to get around your directories in the command line, and this is a valuable skill. Here are a couple more useful variations: The command

$ cd..

automatically changes you to the next directory up in the directory structure;

$ cd ~

gets you straight to the home directory.

Of course the command line can be used to make new directories or files within directories. For example, if you cd to your desktop, then type

$ mkdir testfolder
you will notice a new folder, called “testfolder” appear on your desktop. Even if your desktop were not visible through your conventional graphical interface, you could confirm the successful creation of this new folder using the ls command. You could now enter this new directory by typing

$ cd testfolder

You could then ls to see that there are no files there, and you could then pwd to confirm the path, in my case:

/Users/joelsg/Desktop/testfolder

Now command-line yourself back into the desktop directory (cd ..). Now cd back into testfolder, but this time type “cd tes” and then hit tab to autocomplete the line. The first time you do this will be a revelation because you will recognize how much time can be saved in the long run by not needing to type out long file or directory names. tab completion is a programmer’s cherished friend.

Now you know what it is like to create a directory. What about creating an actual file? Let’s make a text file using Nano, a text editor built into the bash shell environment. cd to testfolder and type:

$ nano test.txt

The next part is a little unexpected because the command line screen changes to the nano environment. Type a sentence here. Now control-x to exit. You will be asked to save changes. Type y and then hit return. Now you are back in the familiar command line environment, within
your working directory named “testfolder,” as before. But now if you ls, you will find a file in
this folder: test.txt. Would you like to read what’s in that text file? Here’s a new command:

$ less test.txt

This outputs the contents of that text file, showing a screen-sized portion of the file. Since this
file is so short, the whole text will be visible on the screen. Hit “q” to go back to your prompt.

Now go back up one directory (cd ..) You are now in Desktop. Now try to read test.txt using the
“less” command as before. The output is “no such file or directory” because test.txt is not in
Desktop. You can still read the file from the desktop, but the path to the file needs to be
specified. You need to reach out and get the file, using its path. From Desktop, type:

$ less ./testfolder/test.txt

Then the output will work. This command worked because you provided an appropriate path to
the file test.txt. In this case you used what is called a “relative path” because it expresses the path
only from the “working directory” (i.e. the directory you are currently in). In the example above,
you were in the Desktop, so you expressed the path from that point, beginning with the ./ term.
To be safe, you could also use the complete path to output the test.txt file. On my mac, this path
would be

/Users/joelsg/Desktop/testfolder/test.txt

so the command would be

$ less /Users/joelsg/Desktop/testfolder/test.txt

Now make a second file in testfolder, called “test2.txt”. Type some simple sentence to represent
the contents of this text file. Now we are going to use a command (cat) that allows you to
catenate (combine) these two text files into a single new file. We’ll call the new combined file
“combined.txt”. cd into testfolder, and type the command:
$ cat test.txt test2.txt > combined.txt

Now check your work: `ls` to ensure that the new file is there, then “less” to read the new two-line file, then “q” to get back to command line. Now delete the file “combined.txt” by typing:

$ rm com (autocomplete with tab)

Now check successful removal with “ls”. Now let’s catenate those two .txt files using a slightly different trick, the “*” wild card. So re-catenate test.txt and test2.txt by typing the following:

cat *.txt > combined.txt

Now check your work. Nice job. See how fun this is? If you needed to catenate eighteen .txt files, you can see how the wild card *.txt can save you a ton of effort.

Let’s pretend you don’t really want to keep testfolder around. While it is tempting to just go back to your conventional OS and drag it to the trash (which is an acceptable option), you could do the following. cd to Desktop and type:

$ rm testfolder

This will spit back an error, because rm (the remove function) only works for files, not directories; but the rm command has an option, -r, which allows it to delete a directory along with all its contents. So you can type

$ rm -r testfolder

to delete this folder (and the files inside of it). You could easily remove huge amounts of data from your computer using the rm -r command, so be very careful with this.

To exit this command line session, type “exit” (return), and then you can close that command line window.

End of command line basic skills.
First, install bowtie2. Google it, download the macos version; the folder ends up in downloads folder (e.g. on your desktop). Create a folder in your home folder, called “programs”. Move the bowtie2 folder into this folder. Give bowtie2 a quick testdrive. In the command line, cd to ~/programs/bowtie2-2.2.5 (notice I saved myself some characters by using ~), then type the lone command: ./bowtie2 (return). This should output a bunch of bowtie instructions, because it notices you didn’t input any index, query, or output files, so it wants to give you some help. Anyway, that was your testdrive.

Now it’s time to make it so that bowtie2 can be executable without cd’ing into your programs folder and typing in ./bowtie2. In other words, you need to add the bowtie folder to your $PATH. Seems like a weird procedure at first, but it is common practice for computer types. The “export path” procedure is as follows. In command line,

```
$ cd ~  (to get to your home folder)
$ ls -a  (to show All files/directories in your home folder)
```

You probably will not find a file called .bash_profile, so you need to create this. You do this by basically creating a text file from within the bash shell (command line) using the Nano text editor:

```
$ nano .bash_profile
```

This opens the nano text editor within the command line window. Type in:

```
export PATH=$PATH:/Users/joelsg/programs/bowtie2-2.2.5
```
(which means that when you call up a program in the command line, it should check your bowtie folder) For you, “joelsg” would be changed to whatever your home folder is called. When you have typed this out in Nano, hit control-x, then type y, then hit return. You can now verify that your .bash_profile file was successfully created. From your home folder, type

```
[note: in 2017 install, I stuck with bowtie2-2.2.5, and after setting up PATH variable, I did
$ source $HOME/.bash_profile (return)
... and to verify...
$ echo $PATH
I'm not sure all this was necessary.
]
```

$ less .bash_profile

Remember to type q to stop viewing the file using less. Now exit out of your command line session (“exit”) and start up a new Terminal window. Now, cd to your desktop or any other directory. You should be able to type in “bowtie2” and the program will run in the same manner as your testdrive above. There. bowtie2 is now ready to call up when the tnseq tool is run.

Now it’s time to get the TnSeq tool. Go to the following url:

https://github.com/KBoehme/TnSeq-Pipeline

Click on the “download zip” button on the right side. You now have a folder called “TnSeq-Pipeline-master” This folder can use a little rearranging, but it has three key files: TnSeq-Pipeline.py, objects.py, and (somewhere deeper in) example.config. I recommend setting up the file system within TnSeq-Pipeline-master as follows:

a. change the name of the TnSeq-Pipeline-master folder to “tnseqtool”
b. Arrange content within tnseqtool as follows: i) the TnSeq-Pipeline.py file; ii) the objects.py file; iii) the README.md file; iv) the LICENSE file; v) a new folder named “config”; vi) a new folder named “ptts”; vii) a new folder named “btrefs”; and viii) a new folder called “fastqs”.

c. The provided file, “example.config” goes into the config folder. The ptts, btrefs, and fastqs folders are populated with user-generated files that are all essential for this analysis. So let’s go into detail on these files:

The .fastq (or fastq.gz) files are given to you by the sequencing center. If you are at BYU and need to grab the fastq.gz files from the supercomputer, see Appendix B. In some cases, sequence reads from a single experimental condition may be spread over three or four files. In such cases, you will need to combine them using the cat command in the command line, just as illustrated above. An example of catenating multiple fastq.gz files is also given in Appendix C.

The fastq files should be given names that allow the associated experimental conditions to be easily discerned in the output (such as richmed1.fastq, minmed1.fastq, etc.).

The .ptt file(s) for your organism can be found at the ncbi ftp site (see instructions below).

When you save these text files, they must have the exact file names as at the ftp site (e.g. NC_003078.ptt).

2017 NOTE: the ftp system at NCBI is different now. To find files for a classical bacterial genome (e.g. S. meliloti 1021), use Safari, paste into URL box:


log in as Registerd User, Name: anonymous; Password: anonymous (use BYUSecure wifi; BYUGuest doesn’t work); then, when the

file opens, search for your organism and open that folder to get various types of genome files (fna, gbk, gff, ptt, etc.)

The bowtie ref files are generated from the .fna file(s) for your organism, also retrieved from the ftp site as described above. How do you make these bowtie ref files from the downloaded .fna file? This is done using bowtie2-build. If your .fna file(s) are on the desktop, get into Terminal, cd to the desktop. Here, type:

```
$ bowtie2-build <space> (here, drag path in by dragging .fna file into Terminal window; if multiple files for multiple replicons then place comma (no space) between paths) <space> (here type prefix that will be added to the six bowtie reference file names, e.g. `smeliloti`)
```

example:

```
```

Here, I have to drag in three .fna file paths because S. meliloti has three replicons (chromosome and two megaplasmids)

then hit return, and the process will rapidly run and spit out six bowtie reference files to the desktop. You can trash or archive the .fna files. Drag the bowtie reference files into the btrefs folder that you have already created.

Now all the basic files (fastq’s, ptt’s, and bowtierefs) are in place. You are ready to run the program. Set up your config file by simply making changes to the example config file that came with the tool. This config file is a text file like all the others you have been working with. Open it in Sublime. It will look something like this:

```
################################
[input]
```

--------------------------
Reads = ./fastqs/*.fastq.gz ; Absolute or relative path (from TnSeq-Pipeline.py)

BowtieReference = ./btrefs/1021chr ; Path to bowtie2 reference genome. Use prefix.

Ptt = ./ptts/*.ptt ; Path to PTT files of reference genome.

Out = my_example_run ; Name for the output. (For example: my_example_run-GENE.txt and my_example_run-HOPS.txt)

################################

[parameters]

Transposon = TCGAGATGTGTATAAGAGACAG ; Transposon end sequence

MinBaseOffset = 0 ; The minimum number of bases before the transposon end will start.

MaxBaseOffset = 5 ; The maximum number of bases before the transposon end will start.

Mismatches = 3 ; Mismatches allowed when finding transposon

GeneTrim = 5 ; Percent of gene length truncated on both sides of gene where hops won't be counted.

ReadLength = 20 ; Number of bases after Tn end that will be mapped.

MinimumHopCount = 3 ; Any hop site with fewer counts than this across all conditions will be disregarded.

################################

[options]

Debug = False ; Shows detailed running parameters for debugging purposes.

Normalize = Total ; [Intergenic]: Normalize based on intergenic, [Total]: Normalize based on all hops.

DeleteIntermediateFiles = True ; Delete intermediate fasta and sam files at the end of the run.

ReverseComplementReads = False ; If True, RC all reads in input fasta/fastq before searching for the Tn end.

IGVNormalize = True ; If True .igv files will have normalized values. If False, unnormalized.

IGVNegateNegStrand = False ; If True .igv files will negate all hops on the negative strand.

################################
When the config file is set up appropriately and saved, and it is placed in the config folder, you are ready to run the tool: Use Terminal to cd into the folder where the .py scripts are for the tool. According to these instructions, that folder would be called “tnseqtool.” Then type in a command similar to this:

```
joelsg$ python TnSeq-Pipeline.py
/Users/joelsg/tnseqtool/config/0614.config
```

Your path specification will be slightly different because home folder name and config file name will vary.

Then hit return.

What will happen in about 15-60 min is that the output files (GENE, HOPS, HOPS-INTERGENIC, and LOG) will populate a folder called output_files in the config folder. Also within the config folder, you will get an IGV folder with output files ready for visual analysis in IGV. The GENE, HOPS, and HOPS-INTERGENIC files (most useful of which is GENE) can be dragged into Excel to view and analyze data. The IGV file(s) will be named after the ptt file that was used as input, e.g. NC_020528.igv.

###3. Data visualization using IGV###

The IGV analysis is very powerful because it allows you to visually inspect your data. Before downloading IGV, first update Java on your machine. Then go to https://www.broadinstitute.org/igv/ and click on the Downloads button. Download requires a simple registration process that involves entering your name and email address. A Mac version of IGV is available. When the download is finished, it is recommended that you make a central
“igv” folder in your home directory. Here, you will place the application (IGV_2.3.55.app), and additional directories which you will call “genomes”, “igvfiles”, and “igvsessions”. A few other things may pop up automatically in this folder as you use IGV. Don’t mess with them.

To use IGV with your output dataset, you need to copy over the .igv files that were generated by the tnseq pipeline. If one replicon, then there will be one .igv file that encompasses as many conditions are were included in the Tn-seq analysis. If two replicons, there will be two .igv files, etc. Anyway, move these .igv files into the “igvfiles” folder that I had you create. Now, each .igv file needs to be “sorted,” which means you need to open IGV (like you would open any other GUI app) and under the Tools drop-down, select “Run igvtools.” Here, set the Command to “Sort” and browse to select your .igv file that came from the tnseq output (you should have moved it to your “igvfiles” folder). Then hit “RUN.” This operation retains the original .igv file in the ivgfiles folder, but adds a new file with the .sorted.igv suffix. This is the one that needs to be fed into IGV.

IGV needs the reference genome. This is done by first getting two files from the ftp site you used before: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/ (use Chrome). Find your organism and get the .fna and .gff files, similar to how you did this earlier with .fna and .ptt files. Move the .fna and .gff files to the /igv/genomes folder. With those in place, launch IGV and use the “Genomes” drop-down to run “Create .genome file”. You will fill in four fields: “Unique identifier” (your choice), “Descriptive name” (your choice), FASTA file (path to the .fna file you just created), and “Gene file” (path to the .gff file you just created). Then hit OK, and the proper .genome file will be created in the “genomes” folder. An additional file with the suffix .fna.fai will also show up in that folder. Don’t mess with this. For a given replicon, there will end up being 4 files in the “genomes” folder. I like to just keep them all there.
Okay, now that you have your .genome file ready in the “genomes” folder, and your .sorted.igv file ready in the “ivgfiles” folder, you are ready to visualize. Open IGV, use the “Genomes” drop-down to load the appropriate .genome, use the “File>Load from File” drop-down to load the appropriate .igv, and the board should be set. Now you have to play with settings within the IGV interface. First, go to the bottom left of the screen and go to the gene track. Ctrl-click on this and click on the “expanded” option. Then Go to your data tracks above; shift-click to select them all, and then ctrl-click, then use the “Set Data Range” option. Set this to some number between 100 and 200. Ctrl-click these data tracks again and go to “Change Track Height” to change track height to some number between 50 and 200 (depends on how many tracks you are viewing at once). The final challenge is getting your zoom set so that the window you see is covering a window of ~2000-20,000 bp. This is a nice range for viewing Tn-seq data. Now you can use your right/left arrows to scroll through the genome with all your data visualized right in front of you. At this point, save your session, and stick the session file in the folder within your igv folder, called “igvsessions”. Now you can launch this session at another time, and most of the settings will still be active. The one exception may be that you need to reset the “Set Data Range” value for your data tracks.

####Appendix: scp from the BYU supercomputer (aka Marylou)####

use the following protocol to copy files to your local machine from the BYU Supercomputer (aka marylou):

a. You need to have an account on the supercomputer, which will be accompanied by login and password.
b. In the command line on your local computer, ssh to marylou to find your files:

```
ssh joelsg@ssh.fsl.byu.edu (your username will be something other than joelsg)
```

(password will be required)

Then you will get a prompt like this:

```
-bash-4.1$
```

Here, you can ls to get the lay of the land. You will hopefully see a directory called fsl_groups. cd into this directory. Here, you should find fslg_dnasc. cd into this. Now cd to compute, and then deeper in according to instructions from sequencing center, until you get to your fastq.gz files. They may be pretty deep in. When you get to the folder that has your fastq.gz files, you can pwd to summarize the path, for example:

```
/fslhome/joelsg/fsl_groups/fslg_dnasc/compute/150714_D00723_0021_AH5MVLBCXX/Unaligned/Project_Project1/Sample_01
```

It looks messy, but it is what it is. Your fastq.gz files within the Sample_01 directory will additionally have cumbersome filenames such as:

```
1_ATCACG_L001_R1_001.fastq.gz
1_ATCACG_L001_R1_002.fastq.gz
1_ATCACG_L001_R1_003.fastq.gz
```
And Sample_01 will likely be only one of a dozen or so directories (Sample_01, Sample_02, etc.) with your fastq files. But at least these Sample_* directories are all nicely tucked away in the directory Project_Project1 (see path above). Now, the best way to get these all to your laptop in a single step is to perform a recursive scp command (scp –r) as follows: First, open a new terminal window and stay on your local machine (do not ssh to marylou). Here, cd yourself into whichever folder you want the fastq files to end up in, for example, your Desktop. Now, to get all the Sample_* directories downloaded to Desktop, type the following:

```bash
scp -r joelsg @ssh.fsl.byu.edu:/fslhome/joelsg/fsl_groups/fslg_dnasc/compute/150714_D00723_0021_AH5MVLBCXX/Unaligned/Project_Project1 ./
```

Note the only spaces in this command are:

```bash
scp <space> -r <space> (path) <space> ./
```

This command basically has three terms:

i) `scp -r`, which means secure copy (-r = recursive, to get many files in one step)

ii) `joelsg @ssh.fsl.byu.edu:/fslhome/joelsg/fsl_groups/fslg_dnasc/compute/150714_D00723_0021_AH5MVLBCXX/Unaligned/Project_Project1`, which is the complete path to the remote stuff to retrieve. Everything within the directory Project_Project1 will be downloaded.

iii) `./` which means to deliver the contents to the working directory that you are reaching out from. When you hit Return, your marylou password will probably be required.
Once I get these files on my local machine, I like to immediately give them abbreviated names:

```
sample01_001.fastq.gz
sample01_002.fastq.gz
sample01_003.fastq.gz
sample01_004.fastq.gz
```

If these are four parts to what should be collapsed into a single fastq.gz file, then cat them up:

```
cat sample01* > s1combined.fastq.gz
```

Notice the wise use of the wildcard. Also notice that I didn’t allow the combined sample name to begin with “sample01...” because I don’t want the command to attempt to concatenate the original 4 files with the output file, potentially leading to a dreaded infinite loop.

There, now you have catenated fastq.gz files on your local machine. You don’t need to be connected to marylou anymore, so you can “exit” out.