Undergraduate Journal of Experimental Microbiology and Immunology (UJEMI) Vol. 25:1- 11

Generating a sucralose-resistant *Escherichia coli* strain and variant calling pipeline for whole genome sequence analysis of point mutations

Victoria Lui, Camilla Wielunski, Alvin Wong, Timothy Yaroshuk

Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY The gut microbiome is central to many physiological functions, therefore perturbations to microflora composition may pose a significant health concern. Sucralose is a non-caloric artificial sweetener that has been known to have bacteriostatic effects on the gut microbiota populations, and specifically been shown to reduce *Escherichia coli* colony forming units. Discovering the bacteriostatic mechanism of sucralose may allow for the prediction of specific bacteria affected as well as any potential physiological implications. Furthermore, *E. coli* grown in sucralose was found to eventually obtain quinolone antibiotic resistance, suggesting that the mechanism for sucralose and quinolone resistance potentially converges. Therefore, this study aims to determine the mutations conferring sucralose resistance in *E. coli.* Given the relationship to quinolone resistance and the stability of the resistance, the mutations were hypothesized to reside within genomic DNA around gene clusters involved in DNA replication and aminoglycoside transferases. A stable sucralose-resistant strain was successfully generated from *E. coli* MG1655, and a variant calling pipeline for whole genome sequence analysis of point mutations was developed. The pipeline was tested on the wildtype and 11 non-essential mutations were identified. Therefore, here we present a system that could potentially be applied to the sucralose-resistant strain generated here to determine in-frame mutants that may be contributing to the resistance.

INTRODUCTION

Non-caloric artificial sweeteners (NASs) are alternative sweeteners with no nutritional value and are widely used as a sugar substitute in a variety of food products and soft drinks (1). Sucralose, an analogue of sucrose, is commonly used as a NAS, however it has been documented in several studies to have bacteriostatic effects (2; Dalkilic, Cheng, Avasthi, and Chi, manuscript in publication). The majority of NASs pass through the human gastrointestinal (GI) tract without digestion and come into direct contact with the gut microbiota which comprises of hundreds of bacterial species and is crucial to many physiological functions in the host (3). Therefore, it is posited that NAS consumption may lead to changes in the microbiota composition as demonstrated by an *in vivo* increase in *Bacteroidetes* and a decrease in *Clostridiales* (3). Additionally, sucralose was found to significantly increase *Firmicutes* populations *in vivo* (4). Furthermore, a study found that sucralose reduced the number of *Escherichia coli* colonies grown in a dose-dependent manner, suggesting that consumption of NASs affects gut microbiota composition through a selective bacteriostatic effect (4). Sucralose-related changes in the gut microbiome may lead to unknown negative impacts on essential physiological functions, which in turn could result in detrimental impacts on the host (4).

While certain sucralose-related changes may be indirect, the bacteriostatic effects of sucralose towards *E. coli* and *Enterobacter aerogenes* provides evidence of a direct mechanism of sucralose-bacteria interaction. Understanding the bacteriostatic mechanisms of sucralose may be significant as it may allow us to predict which specific bacterial species are affected. One study suggests that sucralose is a competitive inhibitor of invertase which decreases the microbiome's ability to metabolize sucrose (5). Indeed, an invertase inhibition assay showed that both the initial and overall reaction rates of invertase were inhibited when suspended in solutions with sucralose (5). However, sucralose has been shown to inhibit

Published Online: September 2020

Citation: Victoria Lui, Camilla Wielunski, Alvin Wong, Timothy Yaroshuk. 2020. Generating a sucralose-resistant *Escherichia coli* strain and variant calling pipeline for whole genome sequence analysis of point mutations. UJEMI 25:1-11

Editors: Daniela Morales and Julia Huggins, University of British Columbia

Address correspondence to: https://jemi.microbiology.ubc.ca/

growth in *E. coli* even though *E. coli* does not metabolize sucrose, suggesting that a different bacteriostatic mechanism is in play (2). Another study shows that *E. aerogenes* and *E. coli* grown in tryptic soy broth (TSB) with 0.7 mM of sucralose could eventually grow in TSB with 150 mM of sucralose, suggesting that the bacteria can adapt to grow in high sucralose concentrations and acquire sucralose resistance (2). As the exact effects of sucralose on *E. coli* are not yet known, it is important to better understand the effects of *E. coli* sucralose resistance. The potential for mutations in gut microbiota like *E. coli* due to sucralose-resistant growth may have detrimental effects on the overall health of the GI tract.

Although the bacteriostatic mechanism of sucralose and conversely its resistance mechanism are unknown, *E. coli* BW25113 grown in sucralose was found to eventually obtain quinolone antibiotic resistance, suggesting that mechanisms of sucralose resistance and quinolone resistance may share genetic similarities (6). One of the main pathways of quinolone resistance is through mutations within the highly-conserved *gyrA* gene and *parC* genes, which are responsible for encoding DNA gyrase and topoisomerase IV, respectively (7). However, Dalkilic *et al.* (manuscript in publication) discovered that there was no *gyrA* mutation in *E. coli* resistant to high sucralose concentrations (which will be referred to as HS *E. coli* henceforth whereas no sucralose is NS). Although potential *parC* mutations were not investigated, *parC* should be mutated in conjunction with *gyrA* for the occurrence of notable quinolone resistance (Dalkilic *et al*., manuscript in publication, 7). Therefore, *parC* mutation most likely could not have been a contributing factor to the observed sucralose resistance. Additionally, according to Dalkilic *et al.* (manuscript in publication) growth of HS *E. coli* in high sucralose media after growth in TSB without sucralose was maintained, indicating that the mechanism of resistance to quinolone and sucralose is likely dependent on a permanent mutation rather than transient gene expression. Despite the observation that mutations in *gyrA* are not implicated in sucralose resistance, mutations in other genes involved with DNA interactions may be a key area of focus to elucidate the resistance to sucralose and conversely its mechanisms. Another mechanism of resistance to quinolones include mutations in aac(6′)-Ib-c (8). This variant of an aminoglycoside transferase is only capable of acetylating aminoglycosides, but aminoglycoside transferase mutants are capable of acetylating quinolones (8). Therefore, genes encoding similar classes of enzymes may also be mutated and confer sucralose resistance, although cautious observation of the entire genome should be conducted to widen the scope of any future studies.

The goal of our research is to elucidate potential mutations conferred in a sucralose-resistant *E. coli* strain via whole genome sequencing. We hypothesize that sucralose-resistance mutations exist in the bacterial genome within gene clusters that encode for proteins that interact with DNA and aminoglycoside transferases because the literature has shown that stable resistance in sucralose-resistance *E. coli* strains also potentiate eventual quinolone resistance. The anticipated experimental goals resulted in the generation of a sucralose-resistant *E. coli* MG1655 mutant and a variant calling pipeline for whole genome sequence analysis of point mutations that was not applied to the resistant strain but instead tested as a proof of concept on the wildtype as an indication that this could be a potential pipeline used to achieve the research goals highlighted here.

METHODS AND MATERIALS

Growth conditions. *E. coli* sub-strain K-12 MG1655 was used in this study to generate the HS strain. Sucralose was purchased from Tokyo Chemical Industry (Catalog #S0839), Tryptic Soy Broth (TSB) was purchased from Sigma-Aldrich (CASO Broth; Catalog #1.05459.0500), and Select Agar powder was purchased from Invitrogen (Catalog #30391-023). Due to the solubility of sucralose, the highest concentration of 271 mM sucralose was solubilized in TSB and made into a stock sucralose solution. Stock sucralose was added to the media for sterilization by autoclave. Cells were grown in TSB media and 1.5% Tryptic Soy Agar (TSA) plates supplemented with various concentrations of sucralose. TSB media was prepared by dissolving TSB powder in deionized water to make a stock solution for sterilization by autoclave. The appropriate amount of sterilized stock sucralose solution was added to the media for desired concentrations of sucralose. Agar media was prepared by mixing TSB powder and Select Agar powder in deionized water. TSA plates were prepared

by adding the appropriate amount of 271 mM sucralose stock solution and added to the agar media for sterilization by autoclaving. Agar plates of desired concentrations were then poured for usage.

Minimum inhibitory concentration (MIC) assay. To study the sucralose-resistance mutations, *E. coli* MG1655 was cultured in TSB media at varying levels of sucralose to determine its maximum sub-inhibitory concentration (SIC) of sucralose. A liquid culture of a single *E. coli* colony was grown overnight at 37℃. Culture tubes containing 1 mL of TSB and varying levels of sucralose were prepared at every 25 mM increment between 0 mM and 225 mM sucralose. Each tube with its corresponding sucralose concentration was inoculated with a loop of bacteria from the overnight liquid culture and incubated at 37℃ and 200 rpm overnight. Turbidity and OD_{600} readings were observed the next day to determine the sucralose SIC and MIC of the wildtype (WT) strain.

There was clear growth determined through visual observations of turbidity at 125 mM, indicating that sucralose concentration to be the SIC; however, OD_{600} readings showed that there was minimal growth at 150 mM, and no growth at 175 mM, suggesting that 150 mM is the SIC (Table A1). Nevertheless, the MIC assay was only conducted with an n=1 and used to find approximate, appropriate sucralose concentrations to work with. Therefore, two pathways for generation of a sucralose-resistant mutant were devised; one accepting 125 mM as the SIC and one with 150 mM as the SIC (HS-MG1655-3 and HS-MG1655-1, respectively). An additional pathway (HS-MG1655-2) was added branching off from the one using 125 mM as the SIC; colonies grew readily on the MIC plate (150 mM) therefore they were transferred to the next concentration up (175 mM).

Generation of HS strains. To study the sucralose-resistance mutations in the bacteria, an HS strain was generated by culturing the *E. coli* MG1655 in TSB media at the SIC of sucralose as described above in the MIC assay section and incubated at 37℃ overnight. The culture was plated onto TSA plates at the SIC sucralose and incubated at 37℃ overnight. Colonies formed on plates were then plated onto TSA plates at the MIC of sucralose and incubated at 37℃ overnight. Colonies formed on the MIC of sucralose were potential HS strains that were further tested for stability prior to sequencing.

Stability testing of HS strains. In order to ensure that the HS strain and its high sucralose-resistance observed was stable, colonies grown on TSA plates with 150 mM sucralose were picked and transferred to plates with 175 mM sucralose. The resulting colonies were transferred to a no sucralose (0 mM sucralose) plate and incubated at 37℃ overnight. Colonies formed on the plate were picked and transferred back onto TSA plates with 150 mM sucralose and incubated at 37℃ overnight. Transferred colonies that formed at the MIC were determined to be stable HS strains.

DNA extraction and Oxford nanopore technology (ONT) MinION sequencing. Two bacterial isolates from the same colony were selected for ONT sequencing based on genomic DNA (gDNA) quality and concentration. Phenolic contamination of purified product could lead to poor data collection; therefore, phenolic-based extraction was avoided. Therefore, gDNA extraction of WT *E. coli* was performed using Invitrogen PureLink® Genomic DNA Mini Kit (Catalog #K1820-01) as per manufacturer's instructions. gDNA extraction of HS strain isolates was performed using the BioBasic ONE-4-ALL Genomic DNA Mini-Preps (Cat #BS88503) kit. *E. coli* was first pelleted and resuspended in the digestion buffer with proteinase K. Lysis buffer was then used to enhance proteinase K activity and digest cells. Mixing the lysate with the binding buffer and ethanol enabled the DNA to bind to the silica-based membrane. Wash buffer was used to elute contaminants. The elution of gDNA was used with a low salt elution buffer that removed the salt bridges formed between the DNA and silica into a concentrated product. Vortexing steps were replaced with gentle inverting to prevent shearing of the gDNA in consideration of subsequent whole genome sequencing. All gDNA samples were quantified using Qubit Fluorometer and Qubit dsDNA HS Assay Kit (Catalog #Q32851, Q32854)

prior to sequencing. Library preparation was done using an Oxford Nanopore Rapid Sequencing Kit (Catalog #SQK-RAD004) as per manufacturer's protocol. ONT Sequencing was performed using the 120-pores Flongle Chip Flow Cells adapter. The Flongle Chips that we used however, were compromised of only about 40-60 active sequencing pores.

Whole genome sequencing (WGS) analysis. With the gDNA that was prepared and sequenced, variant calling analysis was performed (Supplemental 4 Fig. A1). The reference genome was obtained via NCBI for the MG1655 U00096.3 strain (9). The reference genome (fasta) and all fastq files from the nanopore sequencing runs were uploaded onto the orca-wg.bcgsc.ca server. Although gDNA from the WT was prepared twice, and thus sequenced twice, all the fastq files from the "fastq_pass" folder generated by the nanopore for both runs were concatenated into 1 fastq file for downstream analysis. Since these reads generated were long reads with a high error rate, minimap2 was used to align the reads to the reference (10). Samtools sort was used to sort the SAM file prior to variant calling (11, 12). Samtools mpileup was used for variant calling to create a BCF file (11, 12). Bcftools call was used to convert the BCF file to a VCF file (11, 12). The VCF file CHROM column contents were replaced with "Chromosome". The VCF file was transferred from the server to a local computer where snpEff (13) and snpSift (14) were used to annotate VCF file for various information such as gene name, amino acid substitution, and variant impact on protein. RStudio was used to generate a final table of single nucleotide polymorphisms (SNPs) and their respective amino acid changes. Code to generate the final table is attached as Supplemental 1A-1C.

RESULTS

Generation of sucralose-resistant *E. coli* strain HS-MG1655-1B. Three sucralose-resistant strains, HS-MG1655-1B, HS-MG1655-3E, and HS-MG1655-3L, were successfully generated. Resistance strain generation and relative success is described using percent of colonies retained from the previous passage during stability testing as shown in Table 1. No resistance strains came out of HS-MG1655-2 because the colonies failed stability testing and were terminated; namely, there were no colonies retained after transfer from 0 mM to 175 mM sucralose TSA. The three strains generated all successfully passed stability testing, but only HS-MG1655-1B was selected to move forward in preparation for sequencing. Between passage 3 and 4 (0 mM and 175 mM sucralose), 75% of colonies were retained. In contrast, 100% of colonies for HS-MG1655-3 strain were retained at equivalent passages (Table 1). Lower retention of colonies during stability testing increases the confidence that the retained colonies may have stable resistance characteristics and that the sucralose resistance observed in the lost colonies were due to transient effects.

Sequencing of WT identified 11 non-essential mutations. Data from nanopore whole genome sequencing resulted in a low number of bases sequenced for both runs (refer to Supplemental Table A2 for list of gDNA preps sequenced and their DNA quality). The first run resulted in 47.87 Mbp sequenced, which is approximately 10.3X coverage for a reference genome size of 4.641Mbp (9). Quality control was performed for both runs with the buffer provided with ONT to scan the initial number of active sequencing pores available. The number of active pores (pores that Nanopore uses for sequencing) initially available for this run was only between 40 and 50 which was expected for the compromised chips that were used. The second run resulted in 75.3 Mbp, equivalent to approximately 16.2X coverage for the same reference genome. The number of active pores initially available for this run was only between 50 and 60, which was higher than the first run. Together, both runs resulted in an estimated total of 26.5X coverage, which is below the 30X coverage threshold used by some literature sources (15, 16). Both runs experienced a decreasing number of active pores as the run progressed, which reduces the opportunity for sequencing to occur and decreases the coverage possible (Supplemental 2). Sequencing run 1 (16 hours) was shorter than sequencing run 2 (24 hours) though sequencing run 2 still had more active sequencing pores during the first 16 hours (Supplemental 2). Despite the limitations, a variant calling pipeline was developed with the data generated.

TABLE. 1 Generation of the sucralose-resistance (HS-MG1655) strains. Resistant mutant generation from parent strain WT-MG1655 resulted in 3 resistance strains (HS-MG1655). Before passage 1, sucralose MIC assay was used to determine MIC and sub-inhibitory concentration (SIC). Cells from the MIC assay were streaked onto TSA plates of the same concentration. From that plate, 16 colonies were transferred to an equivalent concentration TSA plate (Passage 1). Each passage was incubated at 37^{o}C for approximately 24 hours. At passage 5, select colonies from passage 4 of HS-MG1655-1 and HS-MG1655-3 were inoculated into TSB and incubated at 37^{o}C for 16 hours at 200 rpm. Percent retained refers to the percentage of colonies (in reference to the previous passage) that continued to grow.

	Passage #	1	2	3	4	5	Final Strains
HS-MG1655-1	[Sucralose] (mM)	150	175	0	175		HS-MG1655-1B
	% Retained	100	50	100	75		
HS-MG1655-2	[Sucralose] (mM)	125	150	175	0	175	No strains
	% Retained	100	81.2	69.2	66.7	0	generated
HS-MG1655-3	[Sucralose] (mM)	125	150	0	150		HS-MG1655-3E
	% Retained	100	81.2	100	100		HS-MG1655-3L

Concatenating the fastq files and putting them through the variant calling pipeline, a list of 11 potential single nucleotide polymorphisms (SNPs) resulting in missense mutation were identified along with their respective amino acid substitution (Table 2). Quality filtering and depth filtering were not employed in order to maximize the identification of potential mutation candidates, especially since the sequence coverage was low and filtering could potentially bias the number of mutations generated. All genes have an average quality score of 6.77, which is the Phred scale score equivalent to negative 10 multiplied by the logarithm of the probability of error, and this in turn is equivalent to a 21% chance of error (Supplemental 3). Quality scores range from 3.78 to 14.03. All genes were referenced with Uniprot (17) and EcoCyc (18) for gene description and essentiality, respectively, and were predicted to be non-essential, which was as expected since the sequenced strain is WT and is not expected to have mutations that affect its growth characteristics.

DISCUSSION

In this study, we aimed to generate a sucralose-resistant mutant using *E. coli* MG1655 and determine the mutations conferring sucralose resistance using whole genome sequencing. Given the literature observations of a stable resistance phenotype and the correlation between sucralose resistance and quinolone resistance, we hypothesize that sucralose resistance arises from genomic mutations within genes involved or associated with DNA replication (Dalkilic *et al*., manuscript in publication, 6, 7, 8). We successfully generated three sucralose-resistant strains of *E. coli* (derived from *E. coli* MG1655) using a method based on previous literature, sequenced the WT genome using ONT, and performed variant calling to determine random mutations within the WT.

In order to generate a stable sucralose-resistant strain derived from WT *E. coli* MG1655, we utilized the bacteriostatic effects of sucralose to pressure and select for resistance strains (2, Dalkilic *et al*., manuscript in publication). A MIC assay was performed to determine the SIC and MIC. Secondly, as per the methods described, the generation was ultimately performed on sucralose-containing TSA to assess the percent retention and to assure culture clonality for mutant generation. Colonies were transferred from the SIC to MIC, to solid media containing no sucralose, and back onto MIC. The stability testing was devised as a way to test the literature observations of a stable resistance and support our hypothesis of genomic mutations conferring this resistance (Dalkilic *et al*., manuscript in publication). If resistance is maintained after growth in media without the selective pressure then transferred back onto the highest sucralose concentration, it is a strong indicator of a permanent mutation, which was seen in our HS-MG1655-1 and HS-MG1655-3 strains and in a similar study performed by Dalkilic *et al.* (manuscript in publication). HS-MG1655-2

failed stability testing and resistance was lost after transfer back onto 175 mM, even though the HS-MG1655-1 pipeline also used 175 mM as the MIC. Given these observations, it suggests that resistance may not always be a result of permanent genomic mutation, rather a result of transient gene expression and controlled by transcriptional regulation. Additionally, HS-MG1655-3 demonstrated 100% retention of colonies after stability testing whereas HS-MG1655-1 demonstrated 75% retention. Given that some colonies were lost in HS-MG1655-1, it further indicates the possibility of resistance resulting from transient effects. Nonetheless, the presence of genomic mutations conferring sucralose-resistance should be confirmed via sequencing.

Compared to other methods of generating stable sucralose-resistant strains of *E. coli*, our method was similar in regard to the technique of transferring cultures grown in intermediate concentrations of sucralose to higher concentrations, but varied in that it is performed using solid media and does not take weeks to accomplish. For example, Dalkilic *et al.* generated their stable resistance mutant in *E. coli* DH5α in 21 days by transferring culture grown in 1 mM to 10 mM then to 100 mM sucralose, monitoring for growth to an OD_{600} similar to their WT control in 0 mM sucralose (manuscript in publication). Similarly, Corder and Knobbe successfully generated their sucralose-resistant *E. coli* by transferring culture grown at 0.7 mM sucralose to 150 mM sucralose and culturing for 38 days (2). In contrast, our method only required 5 passages after transferring from liquid to solid culture, and also included a stability test. It is possible that our method took a shorter amount of time because we utilized a sucralose MIC assay to define more appropriate sucralose conditions that would better select for and encourage the proliferation of resistant *E. coli*. The MIC assay results when determined using visual observation were similar to reported MIC in the literature equivalent to 157 mM but not when considering our OD_{600} values (Table A1), indicating the need for replicates in the future (6). We conducted our mutant generation on solid media which may support more robust growth and tolerance of stressful conditions induced by the high sucralose concentration in comparison to liquid culture conditions. When bacteria are grown in liquid culture, it is more exposed to the nutrients - and in this case the sucralose - available in the media, and therefore more susceptible to their effects. Additionally, surface growth on solid media has been shown to bolster culture resilience via quorum sensing and protection from desiccation and antagonists (19).

Since the MinION provides long-read sequencing, it was paramount that our gDNA was not fragmented or sheared in addition to being of high quality. Additionally, by loading the maximum amount of high-quality gDNA (400 ng in 3.75 µL; or 106.7 ng/µL) onto ONT for whole genome sequencing as per the manufacturer's protocol, it would theoretically provide the highest number of reads and the best sequencing results. Therefore, an acceptable gDNA purification should have a concentration of at least 106.7 ng/µL with an A260/280 reading between 1.8-2.0. As shown in Appendix B, our Nanodrop reading for the WT-MG1655 using the PureLink® Genomic DNA Kit determined that three of the four preps were below the desired gDNA concentration; two of the four preps' A260/A280 were outside the accepted allowable 1.8-2.0 range; no prep satisfied both defined conditions of an acceptable prep. Therefore, the purification of gDNA using the PureLink® Genomic DNA Kit failed. Due to the previously encountered challenges, BioBasic ONE-4-ALL Genomic DNA Mini-Preps was used for the gDNA purification of the HS-MG1655-1B strain in hopes of meeting the criteria for gDNA concentrations and quality. The Nanodrop showed that the gDNA concentration was 963.9 ng/µL, which was high, but the A260/A280 was 2.12 and outside the 1.8-2.0 range (Table A2). In addition, subsequent Qubit determination, a more accurate reading for DNA concentration, showed the gDNA concentration was only 14.6 ng/µL, suggesting that there was serious contamination in the purified product and that the gDNA isolation and purification failed as well.

One of the issues we encountered with the PureLink® protocol was the lysis time. The kit protocol called for a lysis time ranging from 30 minutes to 4 hours. On the one hand, for all our PureLink® preps, we used a 4-hour lysis time, which was the maximum allowed time to minimize the chance of incomplete lysis. However, given the results, the lysis time could have been too long and led to the denaturation of gDNA. On the other hand, the BioBasic protocol clearly indicated that the incubation time for the lysis was 30 minutes.

TABLE. 2 List of missense mutations found in 11 predicted non-essential genes with a predicted moderate impact on protein. VCF file generated from combined sequence data was annotated with snpEff (13) and relevant data was extracted and organized with snpSIFT (14) and Rstudio, respectively. Gene descriptions were gathered from UniProt (197) and essentiality data was gathered from EcoCyc (18). All genes were predicted to be non-essential (18) and were predicted to only have moderate impact on protein by snpEff (13). For full data, refer to Supplemental 3.

Gene Name	Description of Gene	Amino Acid Substitution
bluF	Binds to and releases the BluR repressor from its bound DNA target in a blue light-dependent (470 nm) fashion. A shift to low temperature also triggers a BluF-mediated relief of repression by BluR, suggesting BluF may serve as a thermometer	Glu213Lys
fhuB	Part of the ABC transporter complex FhuCDB involved in iron^{3+}-hydroxamate import. Responsible for the translocation of the substrate across the membrane	Leu274Pro
queE	Catalyzes the complex heterocyclic radical-mediated conversion of 6-carboxy-5,6,7,8-tetrahydropterin (CPH4) to 7-carboxy-7-deazaguanine (CDG), a step common to the biosynthetic pathways of all 7-deazapurine-containing compounds.	Trp44Arg
rhsD	Rhs elements have a nonessential function. They may play an important role in the natural ecology of the cell.	Arg1343Lys
ruvB	The RuvA-RuvB complex in the presence of ATP renatures cruciform structure in supercoiled DNA with palindromic sequence, indicating that it may promote strand exchange reactions in homologous recombination. RuvAB is a helicase that mediates the Holliday junction migration by localized denaturation and reannealing	Ala99Thr
uidR	Repressor for the uidRABC (gusRABC) operon	Gly112Arg
waaY	Catalyzes the phosphorylation of heptose(II) of the outer membrane lipopolysaccharide core.	Val61Gly
yagM	Uncharacterized Protein	Glu164Lys
ycaM	Random transposon mutagenesis indicates that *ycaM* is required for growth in optimum conditions (rich medium at 37°C) but not in minimal medium or at low temperature (15°C) . YcaM is a member of the Glutamate:GABA Antiporter (GGA) Family within the Amino Acid-Polyamine-Organocation (APC) Superfamily; YcaM is a GadC homolog.	Glu194Lys
ydcH	Uncharacterized Protein	Lys53Asn
yoaE	YoaE is predicted to be an inner membrane protein with seven transmembrane domains; experimental topology analysis suggests the C terminus is located in the cytoplasm; potential oxidoreductase	Gly93Arg

Therefore, the lysis might not be the cause of poor gDNA preparation in that case. However, there was a shared commonality between all the preps. During the addition of pure ethanol into the lysate, there was the formation of a clearly visible precipitate. While this precipitate was expected, it was supposed to be resuspended completely through vortexing as described in the manufacturers' protocols. However, due to our restraint for long-read next-gen sequencing, vortexing the tube could not be performed in order to prevent DNA shearing which may have inhibited complete mixing and homogenization of the neutralized solution. We propose that an incomplete neutralization of the mixture was a likely problem for all the preps. Other methods that could be used to improve gDNA quality include using additional charge-switch ion-exchange chromatography to further purify the gDNA (20).

With the style of variant calling analysis we conducted, we had output a comprehensive list of genes derived from a concatenated passed fastq file of the sequence generated by nanopore sequencing (Supplemental 3). Unfortunately, the coverage was low, thus potentially contributing to low quality data. Even with the concatenation of all sequence data from both nanopore runs, coverage is still predicted to be below 30X, which could be due to low number of pores available initially, the decreasing amount of active pores as the run progressed, and low quality gDNA preparation (Supplemental 4 Table A2). We used compromised Flongle chips with only about 40-60 active sequencing pores available which reduces the opportunities for sequencing to occur, and ultimately decreases the number of bases read, resulting in a lower coverage. Additionally, the number of active pores quickly decreased (Supplemental 2), which was likely to have been a result of contaminants in the low-quality gDNA preps clogging the pores. Nonetheless, given these limitations, we continued to develop a variant calling pipeline with the data.

The quality of variants called by ONT for the WT resulting in a missense mutation can be assessed by the quality scores, mapping quality (MQ), and depth scores. Quality scores refer to the Phred score of the probability that the base called is wrong, whereas mapping quality is the Phred score of the measure of confidence that the read mapped is in the incorrect location. Depth scores refer to the number of reads at a certain position that confers to the mutant. Quality scores were on average 6.77, equivalent to a 21% chance of error, indicating a poor-quality variant (Supplemental 3). Additionally, only 2 genes had quality scores above 10, indicating that a less than 10% chance that the SNP called was incorrect (Supplemental 3). MQ of all variants called were equal to 60 which is indicative of high MQ (VCF file not attached). Depth scores varied across all the variants called with an average of 10; some researchers use depth scores greater than 10 as a filtering criterion, further indicating the variants called are of poor quality (15, 21). Therefore, it is likely that only a few of the variants listed in Table 2 were present, which nonetheless is expected for a WT strain. For this study we did not apply a filtering criterion since we were interested in maximizing the number of potential variants predicted given the low coverage and quality of the sequencing data. Quality filtering should however be employed once higher quality data is obtained to ensure the variants called are likely present in the sequence. Currently, we are unaware of any literature that explains the relationship between the mutation rate observed in WT *E. coli* and the 11 mutant genes we identified.

The results as shown in Table 2 demonstrate that all variants called have a predicted moderate impact on the final protein. All of these proteins, however, are deemed non-essential for the organism's survival such as *rhsD* as predicted by EcoCyc (18). These mutations were determined for the purpose of comparing them to the sucralose resistance strain to aid in determining the significant gene mutations potentially conferring resistance. The types of mutations identified were expected since the sequenced strains were WT and were not expected to have mutations affecting their growth characteristics as observed when they grew in TSA and TSB. However, this analysis was limited to SNPs and not performed with indels in order to narrow the scope of our studies since there was a high indel error rate with nanopore sequencing (22).

There are a few limitations to sequencing itself that need to be addressed. ONT generates long reads which can map out regions of higher repeats and thus overcome the limitations of Illumina, which in turn generates short reads that can result in unresolvable loops (23). This is important for resolving gene sequences that may be responsible for

secondary metabolites (23) and are of interest in this study since they may be a candidate for sucralose resistance mechanism. ONT, while advantageous for genome assembly, is however highly inaccurate in comparison to shorter-read sequencing and increased read depth alone is insufficient to overcome the errors (23). Therefore, ONT data must be coupled with an assay whether it be PCR with Sanger sequencing, or the mutations need to be characterized with knockout and complementation experiments. Furthermore, Illumina sequencing can help polish the accuracy of the genome assembly while maintaining the contiguous nature of ONT (23).

Conclusions We established a method for generating a stable sucralose-resistant mutant *E. coli* MG1655 using the bacteriostatic effect of sucralose to select for resistance strains in a step-wise manner from SIC to MIC. Additionally, we established a variant calling pipeline for identification of SNPs and their associated amino acid substitutions in bacterial whole genome sequence data obtained via nanopore sequencing that could potentially be applied to the high resistant strain in order to identify mutations that may be contributing to sucralose resistance.

Future Directions In continuation of this study, nanopore whole genome sequencing must be performed on HS-MG1655-1B, or the other resistance mutants generated, to finally determine the genomic mutations that confer sucralose-resistance in *E. coli* MG1655. Next, the sequence obtained must be compared to the WT-MG1655 sequence following the variant calling pipeline we developed for alignment and determination of point mutations. Finally, for mutations of significance, involved genes may be individually analyzed via more accurate sequencing methods such as Illumina sequencing and complementation experiments may be performed to confirm the gene necessity and sufficiency for resistance to sucralose. It would also be interesting for future experiments to further explore the relation between quinolone and sucralose resistance given its novelty in the literature.

For future recommendations, we propose several improvements to the current methods. Firstly, the MIC assay should be performed in replicates for confident MIC determination. Upon visual observation, it was clear that the MIC is 150 mM and SIC is 125 mM but the OD_{600} readings showed minimal growth at the presumed MIC. Additionally, this was further confirmed after transfer of culture from the MIC culture onto TSA of equal sucralose concentration yielded colonies. Given the heterogeneity of our data and the vast differences between our developed method and previously reported methods, more stringent MIC assay conditions and growth parameters could yield a more defined protocol for generation of a sucralose mutant. Additionally, the stability testing could also be performed in multiple trials to ensure the maintenance of the resistance phenotype. Confirmation of a permanent mutation conferring sucralose resistance will result from whole genome sequencing and potentially mRNA sequencing and transcriptomic analysis to distinguish between mutations and transient gene expression.

Secondly, isolation of gDNA must be optimized in order to have sufficient coverage during sequencing to generate a strong consensus sequence. Although advantageous for building the whole genome sequence of an organism because of its long-read ability, one of the pitfalls of nanopore sequencing is the relatively low accuracy compared to other methods such as Illumina (23). For bacterial genomes, this can be remedied using gDNA prepared at concentrations greater than 106.7 ng/μL and free of protein, RNA, and other cyclic or phenyl-like contaminants. Other methods for gDNA extraction could be employed such as charge-switch ion-exchange chromatography 20).

Thirdly, all parameters of the variant calling analysis should be optimized should higher quality data be available. Nanopolish could potentially be used as a polishing step to improve accuracy of genome assembly (24). Currently, all indels are removed and all SNPs are kept with no quality or depth of coverage (DP) filtering done on the data set. However, this was performed since the quality of read was low with the WT sequence. If quality data become available, indels should be included for a more thorough and in-depth analysis, this would need to be confirmed with sequencing experiments with less fidelity. In addition, all quality, depth and other filtering parameters should be implemented when analyzing HS *E. coli* sequences.

ACKNOWLEDGEMENTS

We would like to thank the UBC Department of Microbiology and Immunology for the funding and laboratory space to conduct this study. We would also like to thank Dr. Evelyn Sun and Zakhar Krekhno for their supervision and guidance during this study. We are also grateful for Dr. David Oliver and Reynold Farrera for supervising our studies as well. Finally, we would like to thank Dr. David Ng and the UBC Advanced Molecular Biology Laboratory for the sequencing support, Andrew Wilson for the bioinformatics support, and Canada's Michael Smith Genome Science Centre orca server for hosting our bioinformatic analyses.

REFERENCES

1. **Uebanso T, Ohnishi A, Kitayama R, Yoshimoto A, Nakahashi M, Shimohata T, Mawatari K, Takahashi A.** 2017. Effects of Low-Dose Non-Caloric Sweetener Consumption on Gut Microbiota in Mice. Nutrients **9**(6): 560.
2. **Corder B, and Knobbe A**. 2018. The effects of the artificial sweetener sucralose on the gut bacteria *Escherichia coli* and *Enterobacter aerogenes*. JEMI+ 4:1-9.
3. **Suez, J, Korem, T, Zeevi, D, Zilberman-Schapira, G, Thaiss, C. A, Maza, O, Elinav, E**. 2014. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. Nature **514**(7521):181+.
4. **Wang Q-P, Browman D, Herzog H, Neely GG**. 2018. Non-nutritive sweeteners possess a bacteriostatic effect and alter gut microbiota in mice. Plos One **13**.
5. **Omran A, Ahearn G, Bowers D, Swenson J, Coughlin C.** 2013. Metabolic Effects of Sucralose on Environmental Bacteria. J Toxicol **2013**:1–6.
6. **Qu Y, Li R, Jiang M, Wang X**. 2017. Sucralose Increases Antimicrobial Resistance and Stimulates Recovery of *Escherichia coli* Mutants. Curr Microbiol **74**:885–888.
7. **Bagel S, Hüllen V, Wiedemann B, Heisig P**. 1999. Impact of *gyrA* and *parC* Mutations on Quinolone Resistance, Doubling Time, and Supercoiling Degree of *Escherichia coli*. Antimicrob Agents Chemother **43**:868–875.
8. **Aldred KJ, Kerns RJ, Osheroff N**. 2014. Mechanism of Quinolone Action and Resistance. Biochemistry **53**:1565–1574.
9. **Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y.** 1997. Data from: "The complete genome sequence of Escherichia coli K-12." GenBank https://www.ncbi.nlm.nih.gov/nuccore/U00096 (accession no. U00096.3).
10. **Li H**. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**:3094–3100.
11. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup.** 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics **25**:2078-9.
12. **Li H**. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27**:2987–2993.
13. **Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM**. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly **6**:80–92.
14. **Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X**. 2012. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front in Genet **3**.
15. **Saltykova A, Wuyts V, Mattheus W, Bertrand S, Roosens NHC, Marchal K, Keersmaecker SCJD**. 2018. Comparison of SNP-based subtyping workflows for bacterial isolates using WGS data, applied to *Salmonella enterica* serotype Typhimurium and serotype 1,4,[5],12:i:-. Plos One **13**:1-23.
16. **Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, Grundman H, Hasman H, Holden MTG, Hopkins KL, Iredell J, Kahlmeter G, Köser CU, MacGowan A, Mevius D, Mulvey M, Naas T, Peto T, Rolain JM, Samuelsen Ø, Woodford N.** 2017. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. Clin Microbiol Infect **23**: 2-22.
17. **UniProt Consortium**. 2018. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research **47**.
18. **Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muñiz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP, Karp PD**. 2011. EcoCyc: a comprehensive database of *Escherichia coli* biology. Nuc Acids Res **39**.
19. **Harshey RM.** 2003. Bacterial Motility on a Surface: Many Ways to a Common Goal. Annu Rev Microbiol **57**: 249-273.
20. **Bruch T, Graalfs H, Jacob L, Frech C.** 2009. Influence of surface modification on protein retention in ion-exchange chromatography Evaluation using different retention models. Journal of Chromatography A. **1216**:919–926.

21. **Greig DR, Jenkins C, Gharbia S, Dallman TJ**. 2019. Comparison of single-nucleotide variants identified by Illumina and Oxford Nanopore technologies in the context of a potential outbreak of Shiga toxin–producing *Escherichia coli*. GigaScience **8**.
22. **Norris AL, Workman RE, Fan Y, Eshleman JR, Timp W**. 2016. Nanopore sequencing detects structural variants in cancer. Cancer Biology & Therapy **17**:246–253.
23. **Goldstein S, Beka L, Graf J, Klassen JL**. 2019. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. BMC Genomics **20**.
24. **Loman NJ, Quick J, Simpson JT**. 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods **12**:733–735.