DETECTION OF SINGLE NUCLEOTIDE MUTATIONS UNIQUE TO THE WISCONSIN OUTBREAK STRAINS OF ELIZABETHKINIA ANOPHELIS

By

DAVID PAYNE

Bachelor of Science in Cellular and Molecular Biology

Oklahoma Christian University

Oklahoma City, Oklahoma

2018

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 2020

DETECTION OF SINGLE NUCLEOTIDE

MUTATIONS UNIQUE TO THE WISCONSIN

OUTBREAK STRAINS OF ELIZABETHKINIA

ANOPHELIS

Thesis Approved:

Dr. Peter R. Hoyt

Thesis Adviser

Dr. Charles Chen

Dr. Patricia Jane Ayoubi Canaan

Dr. Robert L. Burnap

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Peter R Hoyt, who encouraged me to begin this analysis and has provided much invaluable instruction on the use of bioinformatic software in investigating biological questions. I also thank my committee members Dr. Charles Chen, Dr. Patricia Canaan, and Dr. Robert Burnap for their suggestions as to how to refine my work.

I would also like to thank the other members of the Hoyt lab, Dr. Hong Hwang and Rita Flores, as they provided key assistance at several critical junctures of my analysis.

Additionally, I would like to thank Jesse Schafer, who not only provided the necessary computing power by allowing us access to the Cowboy and Pete supercomputers, but also lent a tremendous amount of assistance by installing the software necessary for this research and by teaching me the proper use of the software.

For funding sources, I would like to thank the National Science Foundation for the grants NSF-MRI-BIO-1626257, NSF-MRI-OCI-1126330, and NSF-MRI-OCI-1531128, which were crucial for making this research a possibility. I would also like to thank the Oklahoma State University Department of Biochemistry and Molecular Biology for providing the required funds for completing graduate school, as well as the Oklahoma State University Genomics Core Facility, which provided the computing hardware used for much of this research.

.

Acknowledgements reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: DAVID PAYNE

Date of Degree: DECEMBER 2020

Title of Study: DETECTION OF SINGLE NUCLEOTIDE MUTATIONS UNIQUE TO THE WISCONSIN OUTBREAK STRAINS OF ELIZABETHKINGIA ANOPHELIS

Major Field: BIOCHEMISTRY AND MOLECULAR BIOLOGY

Abstract: An unusually large outbreak of the emerging opportunistic pathogen Elizabethkingia anophelis occurred in southeastern Wisconsin starting in late 2015. The strains involved in this outbreak seemed to spread more easily than usual then typical E. anophelis strains, and treatment appeared to be more difficult. The genetic homogeneity of the strains involved in this outbreak allows for comparative genomic analysis aimed at uncovering abnormalities within these strains. This study aims to detect single nucleotide mutations potentially linked to the severity of this outbreak. Using core genome analysis and variant calling, sixty-four single nucleotide mutations unique to the Wisconsin outbreak strains were detected and analyzed for potential functional effects. Several of these mutations were found to potentially affect genes involved in virulence-related processes such as antibiotic resistance, biofilm formation, and oxidative stress response. While the effects of these mutations cannot be fully elucidated in this research, targeted in vitro or in vivo analyses of these mutations may further our understanding of the epidemiology of the Wisconsin Elizabethkingia outbreak.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION 1

II. METHODS 5

III. RESULTS 8

IV. DISCUSSION 12

i. Oxidative and Osmotic Stress Response Genes 12
ii. Biofilm Formation Genes 13
iii. General Virulence Genes 14
iv. Antibiotic Resistance Genes 15
v. DNA Maintenance Genes 15
vi. Other Genes 16
vii. Non-coding Region Mutations 16
viii. Final Conclusions 17

REFERENCES 18

LIST OF TABLES

Table Page

1. Genes with unique SNP mutations in the Wisconsin outbreak strains 9
2. BPROM analysis of the noncoding region at 128165-128597 bp 10

CHAPTER I

INTRODUCTION

Elizabethkingia anophelis is an emerging pathogenic threat that has virulence mechanisms that remain incompletely understood. It is a gram-negative bacillus bacterium ubiquitously found in soil, water, and is notably present within the midgut of the *Anopheles gambiae* mosquito. While human contact with the bacterium typically does not result in disease, clinical patients who are suffering from underlying medical conditions or humans otherwise immunocompromised may be infected (Lau et al, 2016). Once infection occurs, it is very difficult to treat patients due to intrinsic and broad antibiotic resistances present in this species. This leads to infections with uncertain, and sometimes high, mortality rate. *E. anophelis* infections tend to occur in nosocomial environments, and typically affect neonates or older patients with underlying health conditions.

The first confirmed case of *E. anophelis* infection occurred in Bangui, Central African Republic, in an eight-day old infant delivered via caesarean section and was immediately intubated to treat asphyxia upon birth (Frank et al, 2013). The infectious strain was resistant to several beta-lactam antibiotics, as well as a broad spectrum of other antimicrobials (Frank et al, 2013). Treatment with a combination of ampicillin and gentamicin appeared to reduce the infection, the patient died shortly afterwards due to unknown causes (Frank et al, 2013). Approximately five years later, another case occurred in Bangui. Although the genomic profiles for the strains in these infections were very similar, no epidemiological link between them was found (Breurec et al, 2016).

E. anophelis infections are relatively rare and have not been analyzed by the scientific community as extensively as other pathogens. However, this bacterium is clearly established to be a public health threat. There is a general consensus that mortality associated with opportunistic infections by this pathogen is as high as twenty to fifty percent and improved understanding of the pathogenic biochemistry is needed to combat the infections when they occur (Pereira et al, 2013). There have been notable outbreaks in Singapore (Teo et al, 2013) and in Wisconsin, where over sixty patients were affected (Perrin et al, 2016). It is currently unclear if *E. anophelis* infections are becoming increasingly prevalent, or whether this is an illusion caused by increased awareness of *E. anophelis* has led to higher detection rates.

While the *anophelis* species could be responsible for the vast majority of human *Elizabethkingia* infections (Lau et al, 2016), earlier studies have often mistaken it for species such as *meinigoseptica* because 16S rRNA analysis and more primitive forms of MALDI-TOF are incapable of species-level distinction for *Elizabethkingia* (Lau et al, 2016). This introduced much confusion into early understanding of how *E. anophelis* could spread.

Transmission mechanisms in clinical outbreaks are not fully understood. Because this bacterium is found in the mosquito midgut and salivary glands, it is possible that infection may occur through mosquito bites (Onyango et al, 2020). However, solid evidence that a mosquito-borne transmission to humans has not been established. In contrast, strong evidence has been found for human-to-human transmission resulting in neonatal meningitis after birth (Lau et al, 2015). Additionally, community spread is well documented between clinical settings among vulnerable populations, as is the case with the Wisconsin outbreak (Perrin et al, 2016). Recently, a study regarding the Singapore ICU outbreak conclude the bacteria were found to be spread via contaminated aerators at hand wash stations (Yung et al, 2018). In the Wisconsin outbreak, however, no method of spread was conclusively identified.

One of the biggest obstacles to treating patients infected with *E. anophelis* is a general uncertainty concerning which treatments would be most effective against this infection because of

resistance to a broad range of antibiotics. *E. anophelis* genomes include several types of multiple efflux pumps and many copies of beta-lactamase-like loci (Kukulta et al, 2014). Unfortunately, it is not yet clear if all these genes are active under some or all conditions. This results in high variability in resistance phenotypes among Elizabethkingia strains, and optimal treatment options have been difficult to determine using the relatively low number of strains analyzed (Han et al, 2016). The high levels of resistance variation both between and within species suggests that further understanding of the differential genomic properties of each strain needs to be investigated to determine the most rational treatment option. A study of 67 *E. anophelis* strains found very high resistance levels to all antibiotics tested, except for minocycline and levofloxacin (susceptibility at 100% and 58.2% susceptibility respectively) (Lin et al, 2018). Another study that analyzed 115 *E. anophelis* strains found that the most promising antimicrobial treatments were minocycline, doxycycline, and piperacillin/taxobactam and 100%, 90%, and 88% susceptibility rates respectively (Jian et al, 2019). While efflux pump variability could explain some of these variations, it is obvious that more clarification is needed.

In addition to multiple numbers and types of antibiotic resistance genes, *E. anophelis* bacteria are likely to resist host immunity and antimicrobial treatments due to its capsular polysaccharide layer. The exact composition of the capsular polysaccharide layer varies widely between various strains (Breurec et al, 2016). These genes allow *E. anophelis* bacteria to form strong biofilm layers that allow it to persist against sanitation and treatment in hospital settings. Furthermore, *E. anophelis* can use a wide variety of siderophores in order obtain iron from host cells, tolerate oxidative stressors such as hydrogen peroxide treatment, and participate in biofilm formation to enhance its resistance to sanitation and antimicrobial measures (Chen et al, 2020). In one report, a strain of *E. anophelis* found in the Singapore outbreak showed several mechanisms for resistance to hydrogen peroxide, an oxidant frequently used to treat hospital water sources (Li et al, 2015).

One of the largest and most serious of the *E. anophelis* outbreaks occurred in southeastern Wisconsin in 2015-2016, infecting over sixty patients (Perrin et al, 2016). Despite extensive analyses, the origin and spread of the outbreak remains unclear (Perrin et al, 2016). The *Elizabethkingia anophelis* outbreak in Wisconsin appeared to be derived from a single ancestral strain, as all strains involved in the outbreak shared close genetic similarity (Perrin et al, 2016). These strains had a disrupted *mutY* gene, causing an increase in mutation rate and therefore increasing the strain's general adaptability (Perrin et al, 2016), but a precise mechanism for spread remains unclear. These strains also have a cluster of capsular polysaccharide genes identical to those in strains from the Singapore outbreak, suggesting biofilm formation involvement (Perrin et al, 2016). While adaptability and biofilm formation are sufficient for explaining the unusual severity of the Wisconsin outbreak, undiscovered factors may contribute to the clonal spread and rapid infectivity of these strains. Previous studies have examined single nucleotide polymorphisms (SNPs) that occurred over the course of the outbreak (Perrin et al, 2016). To address additional genomic elements related to this outbreak, this report focuses on differences between the Wisconsin outbreak strains as a group compared to other *E. anophelis* strains. We find several single base mutations common to the Wisconsin outbreak strains that have the potential for generating atypical pathogenicity.

CHAPTER II

METHODS

This analysis used the genomes of 259 strains of *Elizabethkingia anophelis*, with five of these consisting of strains associated with the Wisconsin outbreak. When possible, the genomes were downloaded as pre-assembled fasta sequences either from NCBI or the Pasteur Institut, with NCBI being used if the genome was available on both. When pre-assembled sequences were not available, paired-end fastq reads were obtained from the Sequence Read Archive (Leinonen et al, 2010). These were then converted into a fasta assembly with Spades version 3.9.0, using the default settings for paired-end reads (Bankevich et al, 2012).

In order to find SNPs, a core genome alignment of the genomes was performed using Parsnp version 1.2 (Treangen et al, 2014). The core genome consists of DNA sequences found in each of the strains analyzed. As these sequences are ubiquitous throughout a large number of strains, these tend to be more essential genes for the species' survival (Treangen et al, 2014). A core genome alignment is therefore thought to select for genes most essential for *E. anophelis* survival. The reference genome used was NZ_CP016373.1, which is also known as "3375". This genome belongs to the first *E. anophelis* strain to be fully sequenced and annotated. Parsnp's output alignment was in XMFA format, which was converted to the proper VCF format using HarvestTools version 1.2 (Treangen et al, 2014). The VCF format consists of a table in which each SNP location is a row and each genome is a column. A value of '0' in a cell indicates that a

specific SNP in a given genome matches the SNP of the reference genome, while a '1' indicates a mutation relative to the reference genome. The VCF file was manually formatted in Microsoft Excel to list only SNPs where each of the Wisconsin outbreak strains had a '1' values, while all other strains had a '0' value.

The reference genome was annotated in RAST in order to create a GenBank file which identified the positions of the coding sequences (Aziz et al, 2008). This allowed each nucleotide containing a SNP unique to the Wisconsin outbreak strains to be characterized as either a coding sequence or non-coding sequence mutation.

The genome of one of the Wisconsin outbreak strains, known as NZ_CP014805.2, or "CSID_3015183678" according to CDC designation (Perrin et al, 2017), was also annotated using RAST. For each SNP found to be within a coding region, the coding sequence was compared between the reference strain and the Wisconsin outbreak strain in BLASTn using the default settings (Altschul et al, 1990). The alignments were viewed in the BLASTn output, and CDS feature was then turned on so that the output displayed the change in amino acid, if any, caused by the nucleotide mutation.

Each protein containing an amino acid substitution due to a SNP mutation unique to the Wisconsin outbreak strains was further analyzed to predict protein function. The first method to predict protein function was to analyze the amino acid sequence using NCBI's Conserved Domain Search (Marchler-Bauer and Bryant, 2004). Conserved Domain Search quickly identifies functional domains within a fasta sequence, and in some cases, can determine gene identity. If no annotation was found via Conserved Domain Search, the amino acid fasta sequence was analyzed via Phyre2 version 2.0 in order to find protein identity (Kelley et al, 2015).

When the SNP associated with the Wisconsin outbreak strains was located in a non-coding region, this region was extracted from the genomes for both the reference strains and a Wisconsin outbreak strain. These non-coding regions were analyzed in Bacterial Promoter Prediction (BPROM) software from Softberry (Solovyev and Salamov, 2015), which detects

bacterial promoters as well as the TATA box, the Pribnow box, and various locations for transcription factor binding. BPROM also determines the confidence for the assignment of a sequence region as a promoter, promoter element, or transcription factor binding site with a proprietary score. BPROM also assigns the promoter an LDF score based on the likelihood that the query sequence would cluster with known promoter sequences rather than known non-promoter sequences (Solovyev and Salamov, 2011), (Yona et al, 2018). The overall expression score was then calculated by adding the scores of each promoter element and multiplying by the LDF score (Yona et al, 2018).

CHAPTER III

RESULTS

The core genome alignment produced by Parsnp represented 2,123,293 bp, or 53%, of the 4,006,214 bp reference genome NZ_CP016373.1. This core genome is conservative and represents an underestimate of the complete core genome because it was necessary to use several incomplete genome assemblies in the alignment. Single Nucleotide Polymorphisms were detected at 128,930 of the core genome positions, resulting in a SNP occurrence rate of roughly 6.07%.

All 128,930 SNP sites were analyzed to select for the ones in which the Wisconsin outbreak strains, and only these strains, had a mutant allele. This resulted in 64 total SNPs. Out of the 64 SNPs, 61 were found to occur within coding sequences, while 3 were found in non-coding regions. Out of the 61 SNPs in coding regions, 20 (33%) of them resulted in changes in amino acid codon mutations, while 41 conserved the presumptive amino acid sequence.

For each SNP associated with a unique amino acid mutation in the Wisconsin outbreak strains, the gene identity, gene function, and the amino acid substitution were characterized (Table I). Notably, several gene functions were identified that could likely to be linked to the increased virulence or robustness of the Wisconsin outbreak strains.

Gene	Amino acid substitution	SNP position on ref	Potential Function
AcrR	R → I	17,265	Drug resistance
TonB-dependent receptor	G → D	169,503	Oxidative stress response
CsuC*	G → R	173,594	Biofilms
RhaT	N → I	565,711	Drug resistance
Hypothetical	E → K	612,933	Unknown
RibA	A → S	981,752	Oxidative stress response
ManA	M → V	1,273,551	Biofilms
GH3	A → I	1,338,518	Plant immunity
CaiD	A → S	1,379,513	Biofilms
BKR_SDR	V → I	1,634,994	Lipid metabolism
CcmH	A → T	1,654,765	Biofilms
CcmH	A → V	1,654,772	Biofilms
Hypothetical	A → T	1,822,646	Unknown
ComEC	L → M	1,966,480	DNA uptake
TauE	V → I	2,593,634	Sulfite export
WecH	M → I	2,777,407	Biofilms
FGAM synthase	L → I	3,321,920	Purine synthesis
SAICAR synthase	A → V	3,325,625	Purine synthesis
TFIIIC/TPR-repeat lipoprotein/YbgF	K → N	3,482,241	Transcription factor
PD-(D/E)XK nuclease	E → K	3,661,883	DNA repair

Table 1: **Genes with unique SNP mutations in the Wisconsin outbreak strains**. Nucleotide position numbers refer to the reference sequence NZ_CP016373.1. The protein annotation denoted with an asterisk was derived from Phyre2. All other annotations were derived from NCBI.

For the three SNPs found within non-coding regions of the genome, BPROM was used to determine whether the SNP changed the likelihood of promoter expression within the non-coding sequence. BPROM assigns a putative promoter an LDF score based on the extent to which it clusters with known promoter sequences, as well as assigning individual scores for each predicted promoter element. The scores of each promoter element are multiplied by the LDF score to determine the expression score, which estimates the likelihood of promoter expression (Yona et al, 2018). For the mutation at 128,501, a five percent increase (862 --> 903) in expression score was found for the Wisconsin outbreak strains, suggesting this region could have increased promoter expression in these strains (Table 2). For the other two mutations in non-coding regions, no difference in putative promoter region activity was suggested by BPROM data.

A

Element	Position Relative to Promoter	Nucleotide Sequence	Score
-10 box	-15	GGTTATAAT	90
-35 box	-35	TTAACC	21
rpoD17	-55	AACTAAAC	7
argR	-13	TTATAATT	14
LDF Score: 6.53			
Overall expression score: 862			

B

Element	Position Relative to Promoter	Nucleotide Sequence	Score
-10 box	-15	G**T**TTATAAT	**83**
-35 box	-35	TTAACC	21
rpoD17	-55	AACTAAAC	7
rpoD17	-14	TTTATAAT	**9**
argR	-13	TTATAATT	14
LDF Score: **6.74**			
Overall expression score: **903**			

Table 2: **BPROM analysis of the noncoding region at 128165-128597 bp.** A) BPROM analysis of the noncoding region from basepairs 128165-128597 on reference sequence NZ_CP016373.1. B) BPROM analysis of the homologous noncoding region found in the Wisconsin outbreak strains. Elements and scores differing from the reference are indicated in bold.

CHAPTER IV

DISCUSSION

Our data shows there are several virulence-related genes with single base mutations that are unique within the core genomes of *Elizabethkingia anophelis* Wisconsin outbreak strains and are likely to have an effect on gene function. These SNP-containing genes should be considered targets for biochemical and *in vitro* analyses to explain the abnormally large spread of the *E. anophelis* Wisconsin outbreak. As described below, the functions of these genes suggest that these SNP mutations could change the response to external stressors. Additional work is needed to confirm this *in vitro* or *in vivo* by testing the ability of strains (with and without these mutations) to survive treatments such as antibiotic regimens or oxidative stress (e.g. surface cleaners). Specific genes and putative effects are discussed below.

i. Oxidative and Osmotic Stress Response Genes

The RibA gene is involved in vitamin B12 synthesis and plays a role in responding to oxidative stress (Koh et al, 1996). It also enhances hemolysis and iron acquisition (Fassbinder et al, 2000), which is crucial for pathogens due to the low concentrations of soluble iron maintained by hosts for homeostatic purposes (Wooldrige and Williams, 1993). A gain of function mutation in the RibA gene would benefit *E. anophelis* colonizing hosts and resisting antiseptic measures.

TonB-dependent receptors, which also allow for the uptake of rare nutrients, are numerous in *Elizabethkingia anophelis*, with some strains containing nearly 60 TonB-dependent receptors (Kukutla et al, 2014). In certain cases, gain-of-function mutations in these receptors modify the amount of iron a bacterium can uptake, aiding in resistance against oxidative stress (Chatterjee and O'Brian, 2018). It could also be possible that a loss of function mutation could resist oxidative stress in certain cases, as a knockout of the TonB gene has been found to increase survival against the oxidative stressor hydroxyurea. (Davies et al, 2009).

CcmH is a gene involved in the maturation of cytochrome c, a compound involved in electron movement during oxidative phosphorylation. Cytochrome c enhances tolerance towards oxidative stress and increased virulence in bacteria (Naylor and Cianciotto, 2004). A gain-of-function mutation in CcmH could increase the rate of cytochrome c maturation, thereby increasing the ability of the Wisconsin outbreak strains to survive oxidative stressors. However, because bacteria tend to decrease respiration under stressful conditions, an upregulation of oxidative phosphorylation seems unlikely.

ii. Biofilm Formation Genes

The CaiD gene, encoding an carnitinyl-CoA hydratase, is involved in the metabolism of carnitine, a compound involved in resistance to osmotic stressors (Meadows and Wargo et al, 2015). A mutation in the CaiD gene leading to increased levels of intracellular carnitine could be significant because disinfection of surfaces potentially containing *E. anophelis* is commonly done using osmotic stressors such as hydrogen peroxide. Increased resistance to osmotic stressors would cause difficulties for eliminating *E. anophelis* reservoirs.

The CsuC gene is involved in biofilm formation through pili assembly, and disruption of this gene reduces biofilm formation (Lannan et al, 2016), (Tomaras et al, 2003). Because CusC has a key role in the formation of biofilms, a gain-of-function mutation in this gene could increase the ability of *Elizabethkingia anophelis* to resist several disinfectants.

The ManA gene promotes biofilm formation by shunting glucose away from the glycolytic pathway and towards pathways involved in the formation of exopolysaccharides and lipopolysaccharides. Experimental evidence has confirmed that disruption in the ManA gene can interfere with biofilm formation (Amos et al, 2011). A gain-of-function mutation in this gene could increase pathogenesis and biofilm formation.

iii. General Virulence Genes

The WecH gene is involved in the addition of O-acetyl groups to antigens, which can trigger a host immune response (Kajimura et al, 2006). Loss or decrease in the function of this gene in the Wisconsin outbreak species could logically decrease host immune response. Conversely, antigen O-acetylation can increase virulence in certain types of bacteria, so a gain-of-function mutation in WecH could increase virulence (Kajimura et al, 2006). However, the latter case is less likely because it has been reported in only two species of bacteria, both of which are genetically distant to *E. anophelis* (Kajimura et al, 2006).

RhaT is an L-rhamnose and proton symporter. The monosaccharide rhamnose is a cell wall component that is found within the O-antigen (Santhanam et al, 2017). This sugar is regarded as essential for the pathogenicity of some fungi (Santhanam et al, 2017). Although similar studies are lacking in gram-negative bacteria, L-rhamnose has been found to promote virulence and biofilm formation in *Flavobacterium*, a gram-negative genus closely related to *Elizabethkingia* (Lange et al, 2017). *Listeria monocytogenes*, a gram-positive bacterium, has been found to upregulate rhamnose synthesis while invading host cells (Lobel et al, 2012). Additional studies on gram-positive organisms show that the presence of rhamnose in the cell wall promotes drug resistance and intracellular survival (Carvalho et al, 2018). Together, these studies suggest mutations in the RhaT gene that increase the rate of L-rhamnose intake could therefore increase the virulence and robustness of the Wisconsin outbreak strains.

TPR-repeats are found in a diverse variety of proteins, many of them being virulence factors and/or putative vaccine targets (Cerveny et al, 2013). Proteins with this motif are most

likely prokaryotic transcription factors involved in stress response, although the specific pathways and effectiveness of this stress response have not been clearly determined (Cerveny et al, 2013). Therefore, it would be worthwhile to study whether a gain-of-function mutation in these transcription factors may enhance the ability of *E. anophelis* strains to survive stressors, and if so, what biochemical mechanisms cause increased stress response.

iv. Antibiotic Resistance Genes

The AcrR gene represses the expression of the multidrug exporter AcrAB. Mutations in the AcrR gene can cause an increase in AcrAB expression, thereby enhancing resistance to antimicrobials (Watanabe and Doukyu, 2012). AcrR mutations have been associated with increased antibiotic resistance, ethanol tolerance, virulence, and biofilm formation (Webber et al, 2005), (Luhe et al, 2012), (Subhadra et al, 2018).

The TauE gene is a sulfite transporter. This gene could be involved in the export of sulfonamide-derived antibiotics and methicillin because sulfites have structural similarities to sulfonamides (Grossoehme et al, 2011). This makes the TauE gene an important target for evaluating increased antibiotic resistance in the Wisconsin outbreak strains.

v. DNA Maintenance Genes

RibA is involved in the protection of DNA from G:C → T:A, particularly in the presence of oxidative stress (Kobayashi et al, 1998). Interestingly, these are the same transversions thought to be greatly increased in the Wisconsin outbreak strains due to the disruption of the MutY gene (Perrin et al, 2016). It would be important to study whether the RibA mutation found in the Wisconsin outbreak strains can work synergistically with the MutY disruption to increase *E. anophelis* survival and adaptability.

PDDEXK nucleases are a diverse array of genes that interact with DNA. They may perform different functions such as DNA repair, replication, or restriction (Steczkiewicz et al, 2012). Due to the wide diversity of PDDEXK nucleases, it is difficult to predict the effect of the Wisconsin outbreak specific mutation in *E. anophelis*. Additional studies should elucidate any

role of PDDEXK nucleases for increased protection from DNA damage caused by antimicrobial treatment and/or the MutY disruption.

The SAICAR and FGAM synthases are key enzymes in the purine biosynthetic pathway. A gain-of-function mutation in one or both of these enzymes could cause an increased rate of purine synthesis, facilitating DNA repair and/or replication in the Wisconsin outbreak strains. The presence of multiple DNA-repair type proteins in this study may have biochemically testable implications in whether *E. anophelis* used DNA repair for improved vigor.

vi. Other Genes

The ComEC gene allows bacteria to uptake DNA for transformation. Because the Wisconsin outbreak strains did not gain genes during the course of the outbreak (Perrin et al, 2017), mutations in this gene likely had no direct effect. However, if the ComEC mutation immediately predated the Wisconsin outbreak, it could have driven the intake of the ICE element responsible for disruption the MutY gene (Perrin et al, 2016).

GH3 functions in *E. anophelis* are unknown, although GH3 is a plant-associated gene mediating resistance to bacteria such as the gram-negative *Xanthamonas* (Hui et al, 2019). It would be interesting to pursue functional aspect of this plant-like antimicrobial gene when present in a virulent gram-negative bacterium.

The BKR_SDR gene, is a short chain dehydrogenase beta-keto-ACP reductase. This gene has not been implicated in any pathogenic mechanisms. However, a mutation in this gene could affect the lipid composition of the membrane (Zhang and Rock, 2008), thus these changes may affect *Elizabethkingia anophelis* survival in particular with additional mutations described in our report.

vii. Non-coding Region Mutations

In addition to the mutations in protein coding regions, the non-coding SNP mutation found at nucleotide site 128501 (Δ^{128501}) on the *E. anophelis* reference sequence resides upstream of multiple antibiotic resistance and membrane protein genes. Our *in silico* studies suggest this

mutation may decrease the chance of polymerase binding to the TATA box in the Wisconsin outbreak strains, while simultaneously providing an additional rpoD17 resulting in a slightly increased promoter expression score for the Wisconsin outbreak strains. *In vitro* studies on this promoter, however, will be required to confirm this.

viii. Final Conclusions

In summary, this analysis has detected several base-pair mutations specific to the Wisconsin outbreak strains of *Elizabethkingia anophelis*, most of which are targets for *in vitro* or *in vivo* functional studies. Using only mutations from the core genome permits analyzing the most evolutionarily resistant regions, which should enrich for functional importance. It is worth noting that, the vast majority of the strains analyzed (236 out of 259 strains) were incomplete, so our analyses should be considered an underestimate of functionally relevant genes. The Wisconsin outbreak strains of *Elizabethkingia anophelis* may have additional functionally relevant single base mutations that were not discovered by this method. Biochemical validation of gene/protein function associated with these genes provides incentive to complete the genome sequencing of additional strains. Our study provides multiple single base mutations unique to the Wisconsin outbreak strains which should provide targeted studies based on the presence of these mutations in genes associated with pathogenic mechanisms. Each of these mutations are worth individual investigation to determine whether they lead to an increase in the ability of *E. anophelis* to establish pathogenesis or to resist antibiotics and disinfectants. Analysis of these mutations will shed further light on the genomic changes responsible for the unusual virulence of the *E. anophelis* strains associated with the 2015-2016 Wisconsin outbreak.

REFERENCES

Lau SK, Chow WN, Foo CH, Curreem SO, Lo GC, Teng JL, Chen JH, Ng RH, Wu AK, Cheung IY, Chau SK, Lung DC, Lee RA, Tse CW, Fung KS, Que TL, Woo PC. Elizabethkingia anophelis bacteremia is associated with clinically significant infections and high mortality. Sci Rep. 2016 May 17;6:26045. doi: 10.1038/srep26045. PMID: 27185741; PMCID: PMC4868968.

Frank T, Gody JC, Nguyen LB, Berthet N, Le Fleche-Mateos A, Bata P, Rafaï C, Kazanji M, Breurec S. First case of Elizabethkingia anophelis meningitis in the Central African Republic. Lancet. 2013 May 25;381(9880):1876. doi: 10.1016/S0140-6736(13)60318-9. PMID: 23706804.

Breurec S, Criscuolo A, Diancourt L, Rendueles O, Vandenbogaert M, Passet V, Caro V, Rocha EP, Touchon M, Brisse S. Genomic epidemiology and global diversity of the emerging bacterial pathogen Elizabethkingia anophelis. Sci Rep. 2016 Jul 27;6:30379. doi: 10.1038/srep30379. PMID: 27461509; PMCID: PMC4961963.

Pereira GH, Garcia Dde O, Abboud CS, Barbosa VL, Silva PS. Nosocomial infections caused by Elizabethkingia meningoseptica: an emergent pathogen. Braz J Infect Dis. 2013 Sep-Oct;17(5).

Teo J, Tan SY, Tay M, Ding Y, Kjelleberg S, Givskov M, Lin RT, Yang L. First case of E anophelis outbreak in an intensive-care unit. Lancet. 2013 Sep 7;382(9895):855-6. doi: 10.1016/S0140-6736(13)61858-9. PMID: 24012265.

Perrin A, Larsonneur E, Nicholson AC, Edwards DJ, Gundlach KM, Whitney AM, Gulvik CA, Bell ME, Rendueles O, Cury J, Hugon P, Clermont D, Enouf V, Loparev V, Juieng P, Monson T, Warshauer D, Elbadawi LI, Walters MS, Crist MB, Noble-Wang J, Borlaug G, Rocha EPC, Criscuolo A, Touchon M, Davis JP, Holt KE, McQuiston JR, Brisse S. Evolutionary dynamics and genomic features of the Elizabethkingia anophelis 2015 to 2016 Wisconsin outbreak strain. Nat Commun. 2017 May 24;8:15483. doi: 10.1038/ncomms15483. PMID: 28537263; PMCID: PMC5458099.

.

MG Onyango, AF Payne, J Stout, C Dieme, L Kuo, LD Kramer, AT Ciota. Potential for transmission of *Elizabethkingia anophelis* by *Aedes albopictus* and the role of microbial interactions in Zika virus competence
bioRxiv 702464; doi: https://doi.org/10.1101/702464

Lau SK, Wu AK, Teng JL, Tse H, Curreem SO, Tsui SK, Huang Y, Chen JH, Lee RA, Yuen KY, Woo PC. Evidence for Elizabethkingia anophelis transmission from mother to infant, Hong Kong. Emerg Infect Dis. 2015 Feb;21(2):232-41. doi: 10.3201/eid2102.140623. PMID: 25625669; PMCID: PMC4313635.

Yung CF, Maiwald M, Loo LH, Soong HY, Tan CB, Lim PK, Li L, Tan NW, Chong CY, Tee N, Thoon KC, Chan YH. Elizabethkingia anophelis and Association with Tap Water and Handwashing, Singapore. Emerg Infect Dis. 2018 Sep;24(9):1730-1733. doi: 10.3201/eid2409.171843. PMID: 30124415; PMCID: PMC6106401.

Kukutla P, Lindberg BG, Pei D, Rayl M, Yu W, Steritz M, Faye I, Xu J. Insights from the genome annotation of Elizabethkingia anophelis from the malaria vector Anopheles gambiae. PLoS One. 2014 May 19;9(5):e97715. doi: 10.1371/journal.pone.0097715. Erratum in: PLoS One. 2015;10(3):e0122154. PMID: 24842809; PMCID: PMC4026382.

Han MS, Kim H, Lee Y, Kim M, Ku NS, Choi JY, Yong D, Jeong SH, Lee K, Chong Y. Relative Prevalence and Antimicrobial Susceptibility of Clinical Isolates of Elizabethkingia Species Based on 16S rRNA Gene Sequencing. J Clin Microbiol. 2016 Dec 28;55(1):274-280. doi: 10.1128/JCM.01637-16. PMID: 27847376; PMCID: PMC5228240.

Lin JN, Lai CH, Yang CH, Huang YH, Lin HH. Clinical manifestations, molecular characteristics, antimicrobial susceptibility patterns and contributions of target gene mutation to fluoroquinolone resistance in Elizabethkingia anophelis. J Antimicrob Chemother. 2018 Sep 1;73(9):2497-2502. doi: 10.1093/jac/dky197. PMID: 29846598.

Jian MJ, Cheng YH, Chung HY, Cheng YH, Yang HY, Hsu CS, Perng CL, Shang HS. Fluoroquinolone resistance in carbapenem-resistant Elizabethkingia anophelis: phenotypic and genotypic characteristics of clinical isolates with topoisomerase mutations and comparative genomic analysis. J Antimicrob Chemother. 2019 Jun 1;74(6):1503-1510. doi: 10.1093/jac/dkz045. PMID: 30830171.

Chen S, Johnson BK, Yu T, Nelson BN, Walker ED. *Elizabethkingia anophelis*: Physiologic and Transcriptomic Responses to Iron Stress. Front Microbiol. 2020 May 7;11:804. doi: 10.3389/fmicb.2020.00804. PMID: 32457715; PMCID: PMC7221216.

Li Y, Liu Y, Chew SC, Tay M, Salido MM, Teo J, Lauro FM, Givskov M, Yang L. Complete Genome Sequence and Transcriptomic Analysis of the Novel Pathogen Elizabethkingia anophelis in Response to Oxidative Stress. Genome Biol Evol. 2015 May 26;7(6):1676-85. doi: 10.1093/gbe/evv101. PMID: 26019164; PMCID: PMC4494045.

Leinonen R, Sugawara H, Shumway M; International Nucleotide Sequence Database Collaboration. The sequence read archive. Nucleic Acids Res. 2011 Jan;39(Database issue):D19-21. doi: 10.1093/nar/gkq1019. Epub 2010 Nov 9. PMID: 21062823; PMCID: PMC3013647.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May;19(5):455-77. doi: 10.1089/cmb.2012.0021. Epub 2012 Apr 16. PMID: 22506599; PMCID: PMC3342519.

Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014;15(11):524. doi: 10.1186/s13059-014-0524-x. PMID: 25410596; PMCID: PMC4262987.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008 Feb 8;9:75. doi: 10.1186/1471-2164-9-75. PMID: 18261238; PMCID: PMC2265698.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10. doi: 10.1016/S0022-2836(05)80360-2. PMID: 2231712.

Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W327-31. doi: 10.1093/nar/gkh454. PMID: 15215404; PMCID: PMC441592.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015 Jun;10(6):845-58. doi: 10.1038/nprot.2015.053. Epub 2015 May 7. PMID: 25950237; PMCID: PMC5298202.

V. Solovyev, A Salamov (2011) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies (Ed. R.W. Li), Nova Science Publishers, p.61-78.

Yona, A.H., Alm, E.J. & Gore, J. Random sequences rapidly evolve into de novo promoters. *Nat Commun* **9,** 1530 (2018). https://doi.org/10.1038/s41467-018-04026-w

Koh YS, Choih J, Lee JH, Roe JH. Regulation of the ribA gene encoding GTP cyclohydrolase II by the soxRS locus in Escherichia coli. Mol Gen Genet. 1996 Jul 19;251(5):591-8. doi: 10.1007/BF02173649. PMID: 8709966.

Fassbinder F, Kist M, Bereswill S. Structural and functional analysis of the riboflavin synthesis genes encoding GTP cyclohydrolase II (ribA), DHBP synthase (ribBA), riboflavin

synthase (ribC), and riboflavin deaminase/reductase (ribD) from *Helicobacter pylori* strain P1. FEMS Microbiology Letters. 200 October. Volume 191, Issue 2.

Wooldridge KG, Williams PH. Iron uptake mechanisms of pathogenic bacteria. FEMS Microbiol Rev. 1993 Nov;12(4):325-48. doi: 10.1111/j.1574-6976.1993.tb00026.x. PMID: 8268005.

Kobayashi M, Ohara-Nemoto Y, Kaneko M, Hayakawa H, Sekiguchi M, Yamamoto K. Potential of Escherichia coli GTP cyclohydrolase II for hydrolyzing 8-oxo-dGTP, a mutagenic substrate for DNA synthesis. J Biol Chem. 1998 Oct 9;273(41):26394-9. doi: 10.1074/jbc.273.41.26394. PMID: 9756871.

Kukutla P, Lindberg BG, Pei D, Rayl M, Yu W, Steritz M, Faye I, Xu J. Insights from the genome annotation of Elizabethkingia anophelis from the malaria vector Anopheles gambiae. PLoS One. 2014 May 19;9(5):e97715. doi: 10.1371/journal.pone.0097715.

Chatterjee A, O'Brian MR. Rapid evolution of a bacterial iron acquisition system. Mol Microbiol. 2018 Apr;108(1):90-100. doi: 10.1111/mmi.13918. Epub 2018 Feb 19. PMID: 29381237; PMCID: PMC5867251.

Davies BW, Kohanski MA, Simmons LA, Winkler JA, Collins JJ, Walker GC. Hydroxyurea induces hydroxyl radical-mediated cell death in Escherichia coli. Mol Cell. 2009 Dec 11;36(5):845-60. doi: 10.1016/j.molcel.2009.11.024. PMID: 20005847; PMCID: PMC2821657.

Meadows JA, Wargo MJ. Carnitine in bacterial physiology and metabolism. Microbiology (Reading). 2015 Jun;161(6):1161-74. doi: 10.1099/mic.0.000080. Epub 2015 Mar 18. PMID: 25787873; PMCID: PMC4635513.

Kajimura J, Rahman A, Hsu J, Evans MR, Gardner KH, Rick PD. O acetylation of the enterobacterial common antigen polysaccharide is catalyzed by the product of the yiaH gene of Escherichia coli K-12. J Bacteriol. 2006 Nov;188(21):7542-50. doi: 10.1128/JB.00783-06. Epub 2006 Aug 25. PMID: 16936038; PMCID: PMC1636290.

Lannan FM, O'conor DK, Broderick JC, Tate JF, Scoggin JT, Moran NA, Husson CM, Hegeman EM, Ogrydziak CE, Singh SA, Vafides AG, Brinkley CC, Goodin JL. Evaluation of Virulence Gene Expression Patterns in Acinetobacter baumannii Using Quantitative Real-Time Polymerase Chain Reaction Array. Mil Med. 2016 Sep;181(9):1108-13. doi: 10.7205/MILMED-D-15-00437. PMID: 27612361.

Tomaras AP, Dorsey CW, Edelmann RE, Actis LA. Attachment to and biofilm formation on abiotic surfaces by Acinetobacter baumannii: involvement of a novel chaperone-usher pili assembly system. Microbiology (Reading). 2003 Dec;149(Pt 12):3473-3484. doi: 10.1099/mic.0.26541-0. PMID: 14663080.

Santhanam P, Boshoven JC, Salas O, Bowler K, Islam MT, Saber MK, van den Berg GC, Bar-Peled M, Thomma BP. Rhamnose synthase activity is required for pathogenicity of the

vascular wilt fungus Verticillium dahliae. Mol Plant Pathol. 2017 Apr;18(3):347-362. doi: 10.1111/mpp.12401. Epub 2016 Jun 10. PMID: 26996832; PMCID: PMC6638212.

Lange MD, Farmer BD, Declercq AM, Peatman E, Decostere A, Beck BH. Sickeningly Sweet: L-rhamnose stimulates Flavobacterium columnare biofilm formation and virulence. J Fish Dis. 2017 Nov;40(11):1613-1624. doi: 10.1111/jfd.12629. Epub 2017 Jun 5. PMID: 28581211.

Lobel L, Sigal N, Borovok I, Ruppin E, Herskovits AA. Integrative genomic analysis identifies isoleucine and CodY as regulators of Listeria monocytogenes virulence. PLoS Genet. 2012 Sep;8(9):e1002887. doi: 10.1371/journal.pgen.1002887. Epub 2012 Sep 6. PMID: 22969433; PMCID: PMC3435247.

Carvalho F, Sousa S, Cabanes D. l-Rhamnosylation of wall teichoic acids promotes efficient surface association of Listeria monocytogenes virulence factors InlB and Ami through interaction with GW domains. Environ Microbiol. 2018 Nov;20(11):3941-3951. doi: 10.1111/1462-2920.14351. Epub 2018 Jul 31. PMID: 29984543.

Watanabe R, Doukyu N. Contributions of mutations in acrR and marR genes to organic solvent tolerance in Escherichia coli. AMB Express. 2012 Nov 13;2(1):58. doi: 10.1186/2191-0855-2-58. PMID: 23148659; PMCID: PMC3514110.

Webber MA, Talukder A, Piddock LJ. Contribution of mutation at amino acid 45 of AcrR to acrB expression and ciprofloxacin resistance in clinical and veterinary Escherichia coli isolates. Antimicrob Agents Chemother. 2005 Oct;49(10):4390-2. doi: 10.1128/AAC.49.10.4390-4392.2005. PMID: 16189130; PMCID: PMC1251554.

Luhe AL, Gerken H, Tan L, Wu J, Zhao H. Alcohol tolerance of Escherichia coli acrR and marR regulatory mutants. Journal of Molecular Catalysis. 2012, April. Volume 76.

Subhadra B, Kim J, Kim DH, Woo K, Oh MH, Choi CH. Local Repressor AcrR Regulates AcrAB Efflux Pump Required for Biofilm Formation and Virulence in *Acinetobacter nosocomialis*. Front Cell Infect Microbiol. 2018 Aug 7;8:270. doi: 10.3389/fcimb.2018.00270. PMID: 30131944; PMCID: PMC6090078.

Naylor J, Cianciotto NP. Cytochrome c maturation proteins are critical for in vivo growth of Legionella pneumophila. FEMS Microbiol Lett. 2004 Dec 15;241(2):249-56. doi: 10.1016/j.femsle.2004.10.028. PMID: 15598540.

Amos MR, Sanchez-Contreras M, Jackson RW, Muñoz-Berbel X, Ciche TA, Yang G, Cooper RM, Waterfield NR. Influence of the Photorhabdus luminescens phosphomannose isomerase gene, manA, on mannose utilization, exopolysaccharide structure, and biofilm formation. Appl Environ Microbiol. 2011 Feb;77(3):776-85. doi: 10.1128/AEM.02326-10. Epub 2010 Dec 10. PMID: 21148694; PMCID: PMC3028718.

Steczkiewicz K, Muszewska A, Knizewski L, Rychlewski L, Ginalski K. Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. Nucleic Acids Res.

2012 Aug;40(15):7016-45. doi: 10.1093/nar/gks382. Epub 2012 May 25. PMID: 22638584; PMCID: PMC3424549.

Cerveny L, Straskova A, Dankova V, Hartlova A, Ceckova M, Staud F, Stulik J. Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. Infect Immun. 2013 Mar;81(3):629-35. doi: 10.1128/IAI.01035-12. Epub 2012 Dec 21. PMID: 23264049; PMCID: PMC3584863.

Hui S, Zhang M, Hao M, Yuan M. Rice group I GH3 gene family, positive regulators of bacterial pathogens. Plant Signal Behav. 2019;14(5):e1588659. doi: 10.1080/15592324.2019.1588659. Epub 2019 Mar 22. PMID: 30900505; PMCID: PMC6512922.

Zhang YM, Rock CO. Membrane lipid homeostasis in bacteria. Nat Rev Microbiol. 2008 Mar;6(3):222-33. doi: 10.1038/nrmicro1839. PMID: 18264115.

Grossoehme N, Kehl-Fie TE, Ma Z, Adams KW, Cowart DM, Scott RA, Skaar EP, Giedroc DP. Control of copper resistance and inorganic sulfur metabolism by paralogous regulators in Staphylococcus aureus. J Biol Chem. 2011 Apr 15;286(15):13522-31. doi: 10.1074/jbc.M111.220012. Epub 2011 Feb 21. PMID: 21339296; PMCID: PMC3075698.

VITA

David Christopher Payne

Candidate for the Degree of

Master of Science

Thesis: DETECTION OF SINGLE NUCLEOTIDE MUTATIONS UNIQUE TO THE WISCONSIN OUTBREAK STRAINS OF ELIZABETHKINIA ANOPHELIS

Major Field: Biochemistry and Molecular Biology

Biographical:

Education:

Completed the requirements for the Master of Science in Biochemistry and Molecular Biology at Oklahoma State University in December 2020.

Completed the requirements for the Bachelor of Science in Cellular and Molecular Biology at Oklahoma Christian University in May 2018.

Experience:

Graduate Teaching Assistant at Oklahoma State University in Fall 2018-Fall 2020.

Published "Claire's Chemistry Adventure", a book designed to assist in teaching chemistry concepts to college freshmen.

Presented "Potential Virulence Factors in *Elizabethkingia anophelis*", regarding genes absent in potentially non-virulent *E. anophelis* strains.

Presented "Genetic Abnormalities in the Wisconsin Outbreak of *Elizabethkingia anophelis*", regarding genetic elements possibly related to increased pathogenicity in the Wisconsin *E. anophelis* outbreak.

Presented "Predicting Antibiotic Resistance in *Elizabethkingia*", regarding the use of PRAP to predict antibiotic resistance genes in *Elizabethkingia* strains.