2016

Attempted Identification of the Calcium-Dependent Sodium Channel Gene in Paramecium tetraurelia using RNAi

Gwendolyn Kuzmishin
The College of Wooster, gkuzmishin16@wooster.edu

The College of Wooster

Attempted Identification of the Calcium-Dependent Sodium Channel Gene in *Paramecium tetraurelia* using RNAi

A THESIS SUBMITTED TO THE FACULTY OF THE COLLEGE OF WOOSTER
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF BACHELOR OF ARTS

By

Gwendolyn B. Kuzmishin

Dean Fraga _____ Advisor

James West _____ Second Reader

Wooster, Ohio
March 25, 2016

Table of Contents

ABSTRACT 2

INTRODUCTION 3

Gene Duplications Can Drive the Evolution of Novel Genes 3

Paramecium tetraurelia As a Model Organism for Studying Gene Duplication Events 6

Ion Channels As a Model For Studying the Fates of Dupicated Genes 7

Knocking Down Genes Can Help Determine Function 12

Goals of This Project 13

MATERIALS AND METHODS 14

Identification of Conserved Sodium Channel Domains 14

Identification of Candidate Sodium Channel Genes in *Paramecium* 14

Molecular Biology 15

Paramecium Cultures 16

RNAi Treatment 17

Behavioral Testing 18

RESULTS 19

Sodium Channel Genes Possess Motifs That Are Conserved Across Species 19

The Paramecium Genomes Contains Genes That Resemble Sodium Channels 21

GSPAT00010323001 Knockdown 25

GSPAT00017334001 Knockdown 28

DISCUSSION 31

LITERATURE CITED 39

SUPPLEMENTARY INFORMATION 43

ACKNOWLEDGEMENTS 58

Abstract

Ion channels play an important role in the swimming behavior of *Paramecium tetraurelia.* Electrophysiological experiments indicate that a calcium-dependent sodium channel exists in *Paramecium*, but a gene has not been identified for this channel. It is hypothesized that this channel evolved from a voltage-dependent sodium channel after a whole genome duplication event. In this paper, 13 candidate genes were identified from the *Paramecium* genome based on their similarity to a reference human voltage-dependent sodium channel. Bioinformatics analysis of the selectivity filter of these putative ion channel genes suggested that at least 8 of the candidate genes were sodium channels and 4 were likely calcium channels. Functional analysis of 2 of the candidate genes thought to be sodium channel genes, via RNAi knockdowns and behavioral testing, led to the identification of a calcium channel.

Introduction

Gene Duplications Can Drive the Evolution of Novel Genes

Genome evolution is driven by spontaneous mutations, or changes, to the genome. Single amino acid changes within a protein coding region are sufficient to slightly change the function of a gene (Tanaka et al., 2011). But how do genes with completely new functions arise evolutionarily? One way that new functions can arise is through gene duplication events (Zhang, 2003).

A gene duplication event occurs when a genome acquires an extra copy of a gene. This additional gene copy can arise from unequal crossing over, retroposition, or chromosomal and genome duplications (Zhang, 2003). All three of these mechanisms have been observed to occur in model organisms from across all three domains of life (Zhou et al., 2008). After one of these gene duplication events occurs, there are four possible outcomes of a duplicated gene: silencing, retention, subfunctionalization, or neofunctionalization (Qian and Zhang, 2014). These four different evolutionary fates are made possible by the fact that each copy of a duplicated gene will individually and independently acquire mutations. Over time, this results in the divergence of the two duplicated genes, with at least one copy maintaining its original function (Zhang, 2003). This leaves the other duplicated gene free to acquire new mutations without threatening the phenotypic function of the original gene. How these accumulated mutations are selected for or against dictates which outcome the duplicated gene will present: silencing, retention, neofunctionalization, or subfunctionalization.

Gene Silencing

After a duplication event, it is possible that possessing multiple copies of a gene will confer no net advantage or even be deleterious to an organism (Harrison et al., 2002). Because of this, there is no selection against any mutations that may occur on the copied version of the gene, resulting in the accumulation of many mutations (Zhang, 2003). These many mutations can serve to either prevent the gene from being expressed or to render the gene functionless. Genes that have lost their function are often referred to as "pseudogenes"(Sisu et al., 2014). A pseudogene forms as a result of a gene duplication event when at least one copy of a gene is undergoing purifying selection, thus guaranteeing that the original function is not lost to the organism (Qian and Zhang, 2014).

Retention of Function

It has been found that, in some cases, having multiple copies of a gene can be beneficial to an organism (Magadum et al., 2013). There are two proposed reasons for why this might be: the gene may provide functional redundancy or the additional gene may result in a greater number of transcripts (Magadum et al., 2013). When the overall fitness of an organism is improved upon by the duplication of the gene, both genes will undergo purifying selection or concerted evolution in order to maintain the function of both copies of the gene (Zhang, 2003). For example, genes related to gene expression and DNA replication in *Arabidopsis thaliana* have been found to exist in duplicate because the additional copies confer an advantage to the organism (Seoighe and Gehring, 2004). Therefore, many duplicated genes related to gene expression in *Arabidopsis* have been conserved.

Subfunctionalization

After a gene duplication event, it is possible for both copies of a gene to accumulate some mutations, but not abolish the phenotype or be rendered completely dysfunctional (Woolfe and Elgar, 2007). Mutations that render a portion of a gene functionless will not be selected against as long as the corresponding portion in the other copy of the gene is still functional (Zhang, 2003). This can happen multiple times for different segments of a gene such that two duplicated genes can split the duties, or function, of the original gene. This sharing of function is known as subfunctionalization (Magadum et al., 2013). Subfunctionalization has been observed to contribute to the evolution of the *hox* genes of Zebrafish (Kleinjan et al., 2008). In simpler terms, subfunctionalization occurs when the function of one ancestral gene is distributed amongst many genes.

Neofunctionalization

Occasionally advantageous mutations can accumulate on the duplicated copy of a gene (Zhang, 2003). When this happens, the mutations will be selected for rather than against. Overtime, these mutations can result in a completely novel function that is advantageous for the organism, resulting in neofunctionalization (Magadum et al., 2013). A well-known and well-characterized example of neofunctionalization is the red and green opsin genes in humans. These genes are involved in color reception in the human eye and are believed to have evolved from a gene duplication event (Yokoyama and Yokoyama, 1989).

In order to increase our understanding of how evolution can be driven by gene duplication events, it is helpful to study organisms that have undergone gene or genome duplication events. One such organism is *Paramecium tetraurelia*.

Paramecium tetraurelia as a Model Organism for Studying Gene Duplication Events

Paramecium tetraurelia is a simple eukaryotic model organism that is well characterized and easy to work with in the lab (Sperling, 2011). Paramecium stands out as an excellent model organism because its entire genome has been sequenced and is available online, along with annotations (Kung et al., 2000). Analysis of its genome has revealed that *Paramecium* has undergone at least three whole genome duplications (WGDs) (Beisson et al., 2010). These WGDs provide *Paramecium* with multitudes of genetic material that can evolve and change rather quickly (Sperling, 2011). Comparing and analyzing these duplicated genes can help us better understand how genes evolve to have new functions as a result of gene duplication events.

In addition to having undergone three whole genome duplications, *Paramecium* also possess a multitude of behavioral phenotypes that make them ideal for studying ion channel function (Beisson et al., 2010). Swimming behavior is affected by a *Paramecium's* extracellular environment (Hinrichsen et al., 1992). *Paramecium* are roughly shaped like elongated eggs, with a relatively thin anterior end and thicker posterior. When swimming in a forward motion, *Paramecium* swim very smoothly in the direction of the anterior end. When swimming backwards, the *Paramecium* move in the direction the posterior end is facing while its anterior end oscillates back and forth. Dancing behavior occurs when the *Paramecium* rotates in place without moving much towards any direction. Depending on

Wendy Kuzmishin 3/22/2016 11:14 PM

Comment [1]: Come up with better transition…need to state somehow that *Paramecium* contain lots of families of genes and one that has already been studied is Calcineurin…This whole paragraph is good, just need to work on flow here,

Maybe also include those last few sentences that Dr. Fraga added on pg. 10 of my edits

ion identity and concentrations of the environment, *Paramecium* will exhibit forward swimming, backwards swimming, or dancing behavior, sometimes called avoidance behavior. For example, when stimulated by a testing solution containing 30 mM potassium chloride, calcium enters the cell, thus depolarizing it. This depolarization results in a ciliary reversal that leads to backwards swimming behavior (Fraga et al., 2010). It has been found that the length of backwards swimming behavior is proportional to the amount of time that the calcium level is elevated within the *Paramecium* (Fraga et al., 2010). This observable behavior, along with other behaviors associated with other ionic environments, can help determine the function and identity of certain unidentified channel genes (Fraga et al., 2010).

The combination of observable phenotypes and multitudes of genetic information from three WGD's makes studying families of ion channels in *Paramecium* ideal for increasing our understanding of the fates of duplicated genes.

Ion Channels as a Model for Studying the Fates of Duplicated genes

The voltage-dependent ion channels are a family of transmembrane proteins that are responsible for selectively allowing ions to permeate the cell membrane (Zakon, 2012). These channels play a fundamental role in all cells; they are responsible for many important biochemical functions such as signaling cascades and amplification, cell polarization, cell excitability, and transfer of information from the environment to the cell (Anderson and Greenberg, 2001). Because these functions are so fundamental, ion channels are found in the membranes of most cell types. In mammals, these channels are responsible for nerve, muscle, and synapse electrical signaling. The existence of ion channels, therefore,

makes it possible for animals to possess sophisticated biological systems such as the nervous system (Yu and Catterall, 2003).

Although ion channels are diverse in their specific functions, all ion channels share the same basic structure and function (Zakon, 2012). All voltage-dependent ion channels selectively allow ions across a membrane via a mechanism called "gating." Gating involves the opening or closing of the channel pore via a conformational change, as shown in Figure 1 (Alberts et al., 2002).

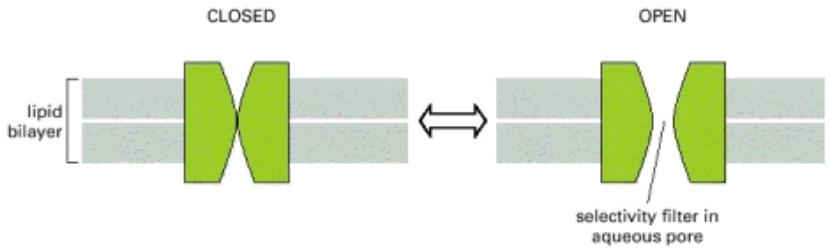

Figure 1. Voltage-dependent ion channels respond to electrical stimuli. An electrical stimulus will induce a conformational change in the channel protein that results in an opening of the pore. The open pore will allow certain ions to pass through at nearly he rate of diffusion. (Alberts et al., 2002)

Once the pore is opened, specific ions are able to pass through the membrane at nearly the rate of diffusion (Hille, 1992). All voltage-dependent ion channels are able to accomplish this function because they contain the same basic structural elements: a transmembrane pore, a selectivity filter, and a voltage-sensitive gate (Zakon, 2012). The selectivity filter plays an important role in all ion channels: it dictates which ions selectivity permeate the membrane. It has been shown that single amino acid changes to the selectivity filter are enough to change ion channel specificity (Heinemann et al., 1992). Therefore, it is possible

to make predictions about an ion channel's function and structure by comparing its amino acid sequence to those of various other channels.

Ion Channel Structure and Function

Five broad classes of ion channels have been identified in eukaryotes: voltage-dependent potassium, sodium, and calcium channels, and calcium-dependent potassium and sodium channels (Anderson and Greenberg, 2001). While these channels all have different biological functions, they are all composed of the same basic structure: a six segment transmembrane domain, with the fifth and sixth segments separated by a pore loop (Anderson and Greenberg, 2001). All functional ion channels contain at least 4 of these domains. Each of the domains in a potassium channel is expressed from a different protein (Anderson and Greenberg, 2001). Therefore, a functional potassium channel could be compromised of a homotetramer or a heterotetramer. In calcium and sodium channels, this basic building block, or domain, is repeated four times within a single protein, as shown in Figure 2 (Zakon, 2012). The presence of four structurally identical domains in calcium and sodium channels suggests the possibility of gene duplication early on in the evolutionary history of the voltage gated ion channels.

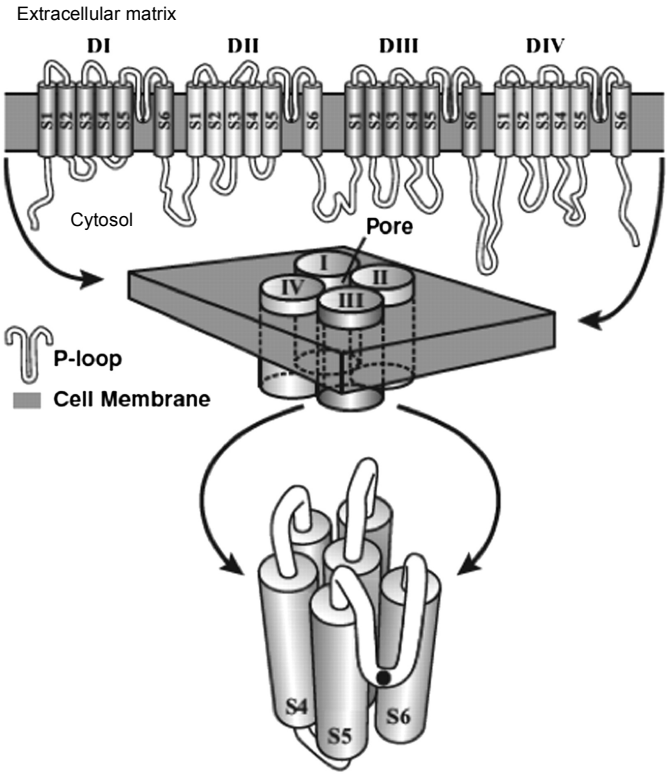

Figure 2. Basic Structure of voltage-dependent ion channels. The five broad classes of voltage-dependent ion channels share the same basic structure shown above. There are four identical domains; each consists of six transmembrane segments. Segments five and six are separated by a pore-loop and segment four has been implicated as playing an important role in selectivity. The presence of four identical domains in calcium and sodium channels suggests the possibility of a gene duplication event in the early evolutionary history of voltage-dependent ion channels. Figure adapted from Zakon, 2012.

Evolution of Voltage-Dependent Ion Channels

Sequence alignments of ion channels from all domains suggest that all voltage-dependent ion channels evolved from a simple transmembrane protein (Anderson and Greenberg, 2001). This transmembrane protein consists of two transmembrane segments separated by a pore-forming loop (shown in blue in Figure 3).

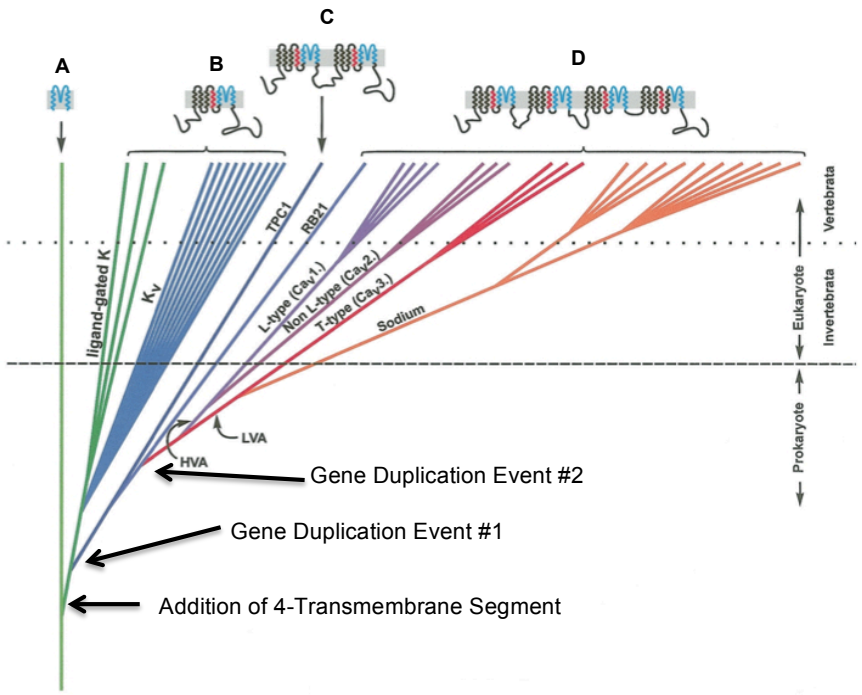

Figure 3. The evolution of voltage-dependent ion channels from a two-transmembrane segment protein. The voltage-dependent ion channels evolved from a simple two-transmembrane (**A**) protein via a series of gene duplication events. The two-transmembrane protein underwent an addition of a four-transmembrane segment to form a six transmembrane domain (**B**). This domain then underwent a gene duplication (Gene Duplication #1) to form a two-domain protein (**C**). This two-domain protein then underwent a second gene duplication event (Gene Duplication #2) resulting in the 4-domain voltage-dependent ion channels (**D**). Figure is adapted from Anderson and Greenberg, 2001.

This simple protein can be found in simple eukaryotes, including the inward rectifier family (Anderson and Greenberg, 2001). This basic two transmembrane motif then underwent an addition of four transmembrane segments to form the basic domain that currently exists in voltage-dependent ion channels today (Figure 3) (Alberts et al., 2002). Phylogenetic studies reveal that a two-domain ion channel protein evolved from this six transmembrane segment domain. It is thought that this two-domain protein evolved via a gene duplication

and fusion event (Figure 3). This two-domain protein is known to exist as a calcium channel in some lysozymes and endosomes (Galione et al., 2009). Comparison of this two-domain protein to modern voltage-dependent calcium channels (which possess four-6 transmembrane segment domains) suggests that voltage-dependent ion channels evolved via a second gene duplication of the 2-domain protein (Figure 3) (Anderson and Greenberg, 2001).

Most of what is known about the evolution of voltage-dependent ion channels has been deduced from phylogenetic studies and sequence alignments (Anderson and Greenberg, 2001). While these approaches are useful, gaps in the evolutionary history of ion channels still exist. Therefore, taking a more direct biological approach by looking at ion channels *in vivo* can reveal much about the evolution of structure and function of these channels and thus also provide us with a tool to better understand how proteins evolve to have new functions. One way to study a family of genes is to selectively knockdown one gene and observe any resulting changes to phenotype.

Knocking Down Genes Helps To Determine Function

Knocking down genes, or almost completely inhibiting their expression, can provide insights into the function of a particular gene (Beisson et al., 2010). If knocking down a particular gene results in a change in an observable phenotype relative to the control, then it is highly likely this observed change is the result of the knocked down gene. It has been previously shown that Ribonucleic Acid interference (RNAi) is an effective tool for knocking down genes in *Paramecium* (Fraga et al., 2010; Galvani and Sperling, 2002; Kissmehl et al., 1996). In this technique, double stranded RNA (dsRNA) for the candidate

gene of interest is introduced to *Paramecium* via *E. coli* containing a pL4440 expression vector (Fraga et al., 2010; Galvani and Sperling, 2002). Upon consuming the dsRNA, the *Paramecium's* posttranscriptional gene regulation machinery will degrade the dsRNA into small fragments. The dsRNA fragments then enter the RNA induced silencing complex (RISC) where the fragment is used to aid in degrading any free mRNA in the cytosol that matches the fragment sequence (Meister and Tuschl, 2004; Galvani and Sperling, 2002) In effect, RNAi utilizes natural mechanisms within the cell to halt translation of mRNA of interest into protein.

Goals of This Project

The goal of this project is to study sodium channel diversity in *Paramecium* by identifying novel sodium channels. Specifically, the goal is to identify the gene for a calcium-dependent sodium channel; electrophysiological experiments indicate that this channel exists, but a gene has yet to be identified (Hinrichsen and Saimi, 1984). This goal will be accomplished by identifying candidate genes for a calcium-dependent sodium channel in *Paramecium*. Candidate genes will be knocked down via RNAi and any changes to behavioral phenotype will be observed via backwards swimming tests. Because it is known that ion channels have evolved via a series of gene duplication events, it is hypothesized that the gene for the calcium-dependent sodium channel in *Paramecium* evolved via neofunctionalization or subfunctionalization and that it is homologous to the voltage-dependent sodium channel from which its gene was likely duplicated.

Materials and Methods

Identification of Conserved Sodium Channel Domains

The amino acid sequences for eight sodium channel alpha subunits were acquired from the National Center for Biotechnical Information (NCBI) protein database (Altschul et al., 1997). Sequences were acquired for the following species: *Arcobacter butzleri* (gi: 339961374), *Bacillus halodurans* C-125 (gi: 754209882), *Heterololigo bleekeri* (gi: 287449), *Homo sapiens* (gi: 320461722), *Cyanea capillata* (gi: 994814), *Mus musculus* (gi: 840654), *Archaeon* GW2011_AR11 (gi; 735016477), and *Phytophthora nicotianae* (gi: 970634281). The sequences were aligned using CLUSTALW (Matrix: BLOSUM) on MacVector (Pairwise Alignment: Open Gap Penalty – 3.5; Extend Gap Penalty – 0.75l; Multiple Alignment: Open Gap Penalty – 15; Extend Gap Penalty – 0.3; Delay Divergent – 25%). Conserved motifs were identified and the alignment was annotated using data from the literature (Payandeh and Minor, 2015; Payandeh et al., 2011; Widmark et al., 2011; Goldin, 2002; Hille, 1989).

Identification of Candidate Sodium Channel Genes in *Paramecium*

The *Homo sapiens* voltage-dependent sodium channel (gi: 320461722) was compared to the *Paramecium* genome on the online Paramecium database (http://paramecium.cgm.cnrs-gif.fr/). This allowed for identification of highly conserved ion channels. Genes were chosen based on protein similarity as determined by the bit score E value. The 13 identified candidate genes were then aligned using CLUSTALW in MacVector as previously described. The literature and the annotated alignment were used to confirm that the candidate genes contained motifs that are highly conserved in sodium channel genes. Within each individual candidate gene, an additional alignment was created

in order to look at only the selectivity filter. MacVector was used as previously described to align the four domains within a single candidate gene. Residues of interest within the selectivity filter were compared to the literature in order to further analyze the candidates (Tang et al., 2014; Yue et al., 2002; Stary et al., 2008; Zhen et al., 2005; Payandeh et al., 2011).

The output from the alignment of candidate genes was then used to build a phylogenetic tree on an online phylogeny generator, phylogeny.fr (http://www.phylogeny.fr/) (Dereeper et al., 2008). This phylogenetic tree was built using PhyML, which is based on the maximum likelihood principle, and visualized using TreeDyn (Dereeper et al., 2008; Edgar, 2004; Castresana, 2000; Guindon et al., 2010; Anisimova and Gascuel, 2006; Chevenet et al., 2006). From this phylogenetic tree, four candidate genes were chosen for further study.

Molecular Biology

The DNA sequences for the genes of interest were downloaded from the *Paramecium* database and primers were designed for cloning (Table 1)(Fraga et al., 2010).

Candidate Gene	Primer Name	Primer Sequence
GSPAT00017334001	PtChan17334001F	5' - ATTCTAGAAGAGGAAAAACAGAGGTAAGAGAGAGAAAGG
	PtChan17334001R	5' - ATAAGCTTTCTAACTACTCGGGGCAGGTATATTG
GSPAT00010323001	PtChan10323001F	5' - ATGAGCTCTGTCCAAAAGTAGAATCGGGAAATTTTA
	PtChan10323001R	5' – ATAAGCTTGTCCGAATTCCTCTTCCAAGTCTTC
GSPAT00024320001	PtChan24320001F	5' - ATTCTAGAGAAGCACCAAGTAGCTGGACTGAC
	PtChan24320001R	5' - ATCTCGAGTTGCTGACACTCTAGCAGAAGAAATG
GSPAT0001733301	PtChan33001F	5' - ATGAGCTCACTGCAGGTGTTAAAAAGCGATTAACTG
	PtChan33001R	5' - ATAAGCTTGTAGTGGACTACCCGTCGTGTATAAGTT

Table 1. Primer Sequences of Cloned Candidate Channel Genes. All of the primer sequences used both Forward (F) and Reverse (R) primers for cloning of candidates into the pL4440 plasmid as described below. The underlined portion of each primer represents the restriction site that was added for cloning.

Each fragment of interest was amplified by PCR using AccumePrime Pfx DNA polymerase as described by the manufacturer (Thermo Fischer Scientific, MA). Each PCR reaction contained 1X AccuPrime Pfx Reaction mix, 0.3 mM of each primer, approximately 100 ng of template DNA, distilled H_2O and 1 unit of AccuPrime Pfx DNA polymerase. Each reaction went through an initial denaturation (95 °C for 1 min), 35 cycles of denaturation (95 °C for 15 s), annealing (55-72 °C depending on primer pair for 30 sec), and extension (~ 68 °C for 1 min). PCR products were purified via gel electrophoresis (1.5% agarose gel, 1X TAE, 90 V for approximately 1 hour). The fragments of interest were inserted into the pL4440 vector using the appropriate restriction sites. The ligated plasmids were then transformed into competent HT115 cells, which were made previously using the calcium chloride method, and plated on LB Ampicillin (50 µg/mL) plus tetracycline (12.5 µg/mL) plates. Minipreps were prepared from a single colony using the Qiagen Miniprep Kit as described by the manufacturer (Qiagen, CA). DNA from the minipreps was sent for sequencing at the Molecular Cellular Imaging Center (Ohio State University OARDC, Wooster, Ohio) for confirmation using the M713 Forward primer.

Paramecium Cultures

Wild-type stock cultures of *Paramecium tetraurelia* (51S strain) were maintained in wheat grass media containing wheat grass buffer (1.17 g Tris base, 0.84 g Na_2HPO_4 dibasic, 0.315 g Na_2HPO_4 monobasic, 0.105 g Na_2EDTA) and stigmasterol solution (15 mg in 3 mL

100% EtOH) and inoculated with *Klebsiella aerogenes* at 27 °C as previously described (Kissmehl et al., 1996; Haynes et al., 2003; Fraga et al., 2010).

RNAi Treatment

Individual colonies of verified putative calcium-dependent sodium channel genes were verified via DNA sequencing, individual colonies were used to prepare an overnight culture in LB ampicillin (50 μg/mL) plus tetracycline (5 μg/mL). Cultures were incubated overnight at 37 °C and 225 RPM. Once cultures reached an optical density of approximately 0.4 at 595 nm, cultures were induced with IPTG (125 μg/mL) for four hours (Timmons et al., 2001). After incubation, bacteria were pelleted via centrifugation (3000 RPM for 2 minutes) and resuspended in an equal volume of wheat grass media to make a 10X feeding stock. A 1X feeding stock was prepared from the 10X feeding stock with wheat grass media containing ampicillin (50 μg/mL) and IPTG (125 μg/mL). Approximately 100 *Paramecium* cells were starved in Dryls solution (1 mM NaH_2PO_4 , 1 mM Na_2HPO_4, 2 mM $Na_3Citrate \cdot 2H_2O$, 1.5 mM $CaCl_2 \cdot 2H_2O$) for approximately 1.5 hours, and then, placed into 5 mL of 1X feeding in a sterile 70 mm Petri dish. Petri dishes were placed in a humidity chamber at 27 °C overnight. 24 hours after the initial feeding, 5 mL of fresh 1X feeding stock was added to each Petri dish and incubated at 27 °C for another 24 hours. Behavioral testing was completed at 48- and 72- post initial feeding (Fraga, *et al*., 2010). Control cells were fed with transformed HT115 cells containing empty pL4440 vector according to the same method described above.

Behavioral Testing:

Behavioral tests were carried out for each knockdown and control 48- and 72-hours post RNAi treatment. *Paramecium* were placed in a neutral resting solution for two minutes (3 mM HEPES (pH 7.2), 1 mM KCl, and 1 mM $CaCl_2$) and then placed into either a testing solution containing 30 mM KCl or Na/TEA solution (10 mM NaCl and 5mM TEA) (Fraga et al., 1998). Backwards swimming time was measured for each *Paramecium* upon entry into the testing solution. Approximately 25 cells from each condition were tested in both the KCl and Na/TEA solutions, for a total of 50 control cells and 50 experimental cells, at each time point. The average backwards-swimming time was then determined for each experimental knockdown and control. A Mann-Whitney U test was run via SPSS to determine if any observed differences in backwards swimming time were statistically significant, since the data were not normal (Frey, 2015; Tokunaga, 2015).

Results

Sodium Channel Genes Possess Motifs That Are Conserved Across Species

A multiple sequence alignment of the α subunit of sodium channel genes from multiple species was created to identify the most highly conserved amino acid residues in sodium channels (highlighted in yellow in Figure 4). Residues highlighted in grey are relatively well-conserved, but some variation is observed. The colored bars above the sequence represent the six transmembrane segments that make up the α domain. The blue highlighted portion represents a series of highly conserved charged residues followed by two hydrophobic residues; a motif that is found in the S4 segment of most ion channels, regardless of ion specificity. In the alignment of sodium channels shown here, arginine (R) is the charged residue, but other sodium channels could have histidine (H) or lysine (K) at these positions. The highlighted asparagine (N) represents one of the most highly conserved residues in the entire sodium channel domain (Payandeh and Minor, 2015; Payandeh et al., 2011).

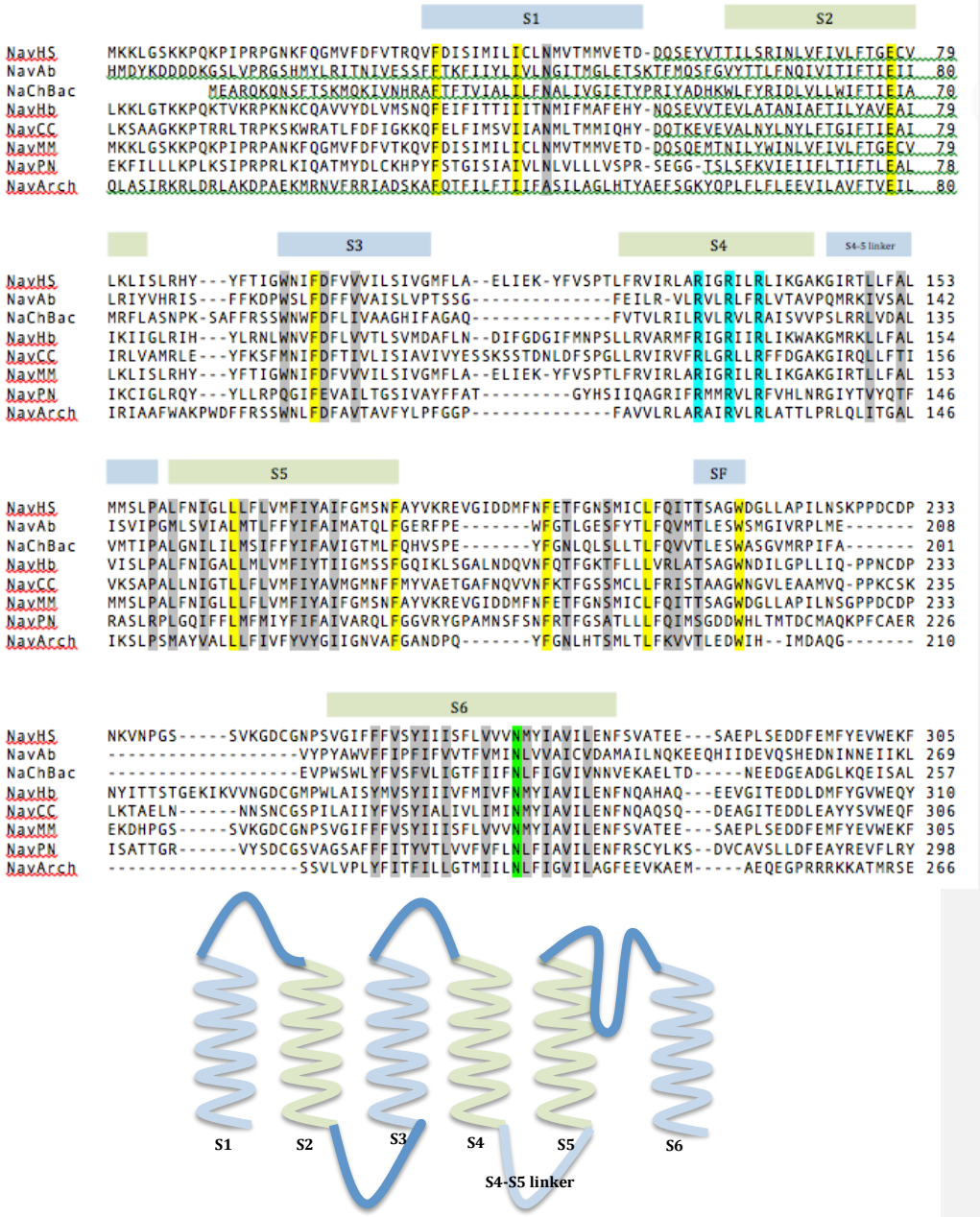

Figure 4. Sequence alignment of sodium channel genes from multiple species.

Sodium channel genes from multiple species were aligned: NavHS (*Homo sapiens),* Nav Ab (*Arcobacter butzleri),* NaChBac *(Bacillus halodurans* C-125), NavHb (*Heterololigo bleekeri)*, NavCC (*Cyanea capillata*), NavMM (*Mus musculus)*, NavPN (*Phytophthora nicotianae),* and NavArch *(Archaeon* GW2011_AR11). The most highly conserved residues are shown in yellow. Relatively well-conserved amino acids are shown in grey. The blue amino acid residues represent conserved repeated charged residues followed by two hydrophobic residues. The green asparagine (N) residues are highly conserved in all sodium channels. The bars above the sequence correspond to the six transmembrane-segment that makes up a single domain of a sodium channel (as shown in the diagram below the alignment).

The *Paramecium* Genome Contains Genes that Resemble Sodium Channels

Thirteen candidate genes were identified from the *Paramecium* genome as resembling a known human sodium channel gene on he basis of relatively low E values.. The candidate genes were then aligned with CLUSTALW in MacVector and the alignment was annotated to highlight residues that were highly conserved (Figure 5). The candidate gene alignment was compared against the alignment of known sodium channels (Figure 4) to ensure that all candidate genes possessed motifs that are characteristic of all sodium channels. This comparison confirmed that all the candidates were likely ion channels, but did not necessarily which, if any, were sodium channels. Since many of the residues that are highly conserved in sodium channels are also highly conserved in calcium channels, further analysis was needed to evaluate the candidate genes. As shown in figure 6, the domains of each candidate gene were aligned against one another in order to analyze the selectivity filter as a whole. As seen in Table 2 below, the selectivity filter analysis revealed that the candidate genes likely represent both calcium and sodium channels.

```
GSPATP000171050  VKSIYFKLFMTTLIIFNVALFIYVKTNNRKD-----------TDQIEQVITVFFLIEVSF  50
GSPATP000173340  ----TFQLILMVTIILNSLCLALEQ------------------DQYEIIFIICYSVEAIF  38
GSPATP000193790  ----FFQLILMVIIILNSLCLALEV------------------DQYEIIFIICYSVEAIF  38
GSPATP000104430  ----IFEFFTLLVIIFNSIMLALDDPTTDIQ--------TPFQDLMDIIFLAVYTFEAVL  48
GSPATP000345100  ----IFEIITIFIIIFNSVMLAMDDPTTEEQ--------TDFANLMDQIFLIYYTIEAIL  48
GSPATP000173330  ----IFEVITLIMIVFNSVMLAIDDPTTNVQ--------SPFQNLTDLIFLAYYTFEAVL  48
GSPATP000103230  ----LFELMTLLMIIFNSAMLAIDDPTTNVQ--------TSFQDLTDIIFLAYYTAEAVL  48
GSPATP000334140  ----LFELMTLLMIIFNSAMLALDDPTTDVQ--------TSFQDLTDIIFLAYYTAEAVL  48
GSPATP000297160  ----LFLQVIRIITFLNIIAVCLYDYQSR--INQEDDNRNRISITLEIFCNIIFGLECLG  54
GSPATP000139240  ----WFDKVITLIIVLNTISLAMQNYDFRVNGTSNQSELTYVRNQMEYFFTSVFLLEFIL  56
GSPATP000056360  ----WFDRFIITVIFLNSILLATQDYSWRETDSEAPNSWTDS---FEYIFTAIFIIEFLL  53
GSPATP000243200  LIFSWFDRFIITVILLNSILLATQDYSWRETNSEAPSSWTDS---FEYIFTAIFIIEFLL  57
GSPATP000213240  LTTKRLSKIIVRSRIYQYLLTFVTLGNIILFIYEHLNHDNQNLFAINFGFLWLYVVDIVL  60
```

```
GSPATP000171050  RIIASGVILNRKSFFRSPLNIYDFTLVVLTTLNLYRPDV--IILDLSPLRMITLLNYLGN 108
GSPATP000173340  KMLALGLYIKRDAYFRDFWNVIDFLILVFQYLPFFVNIK-FNFNALRTIRILRSLDAGQS  97
GSPATP000193790  KMLALGLYSKRDSYFRDFWNVIDFVILVFQYLPYFVNIQ-FNFNALRTIRILRSLDAGQS  97
GSPATP000104430  KIIAMGFVFNSNTYLRDLWNVLDFTVIVTAYIPYFVNNSGLQLSALRSLRVLRPLRTISS 108
GSPATP000345100  KIIAQGLIWPKKAYLRETWNILDFSVIITAYLPYFVSSNSLNLNSLRSFRVLRPLRTISS 108
GSPATP000173330  KIVAQGFIIPKKSYLRDTWNILDFSVIITAYIPYFLASNSVNLNALRSFRVLRPLRTVSS 108
GSPATP000103230  KIVALGFILPKKAYLKDTWNILDFSVIVTAYIPYFLSSNSVNLNALRSFRVLRPLRTVSS 108
GSPATP000334140  KIVALGFIFPKKAYLKDTWNILDFSVIVTAYIPYFLASNSVNLNALRSFRVLRPLRTVSS 108
GSPATP000297160  SIIIKGFVLGKNAYLKSWWNVVNFLTFIATWSILGDIKS--ENQLVHILRVIRLFRTLRL 112
GSPATP000139240  KILAMGFLFEKHTYLRDGWNVLDFAVVGTSILSLVGSFN---LSAIRTIRILRPLRSIKS 113
GSPATP000056360  KMIAMGFLLDKQTYLRDGWNFIDFIVVITGIISLFISAR---VSAIRIIRIMRPLRSINS 110
GSPATP000243200  KMIAMGFMLDKQTYLRDGWNFIDFIVVITGIISLFISAR---VSAIRIVRIMRPLRSINS 114
GSPATP000213240  QLFAFGLFRSKFIFLGRVIDAFLIVMFYYHLAYTET------VVNITPLRMMRLLKQLGA 114

GSPATP000171050  VLKG--LSVMLKALISSLKFLLEALIIVGLFSLFFGIFGVLLFQNLFNYRCQFENGDETE 166
GSPATP000173340  IPA---LKVILTSLFQSVYQLKDALIVLVFSYSIFAIVALLLFGGGLKRRCVNLETGIPR 154
GSPATP000193790  IPA---LKVILTSLFQSVYQLKDALIVLVFSYSIFAIVALLLFSGGLKRRCVNLETGIPR 154
GSPATP000104430  VKS---LRSIMMTLFASFAELGNSLIILGFTYTIFAIAGLQLFSGYAKLRCFDENGIIQM 165
GSPATP000345100  IKS---LRMILLALFASLAQLRDAAIVLIFFYTIFAIAGVQLFSGYLKRRCFSDIGITLM 165
GSPATP000173330  IKA---LRTILLALFASIAQLRDAAVVLMFFYSIFAIAGVQLFSGYLKRRCIGEESGITW 165
GSPATP000103230  IKA---LRTILLALFASIAQLRDAVVVLIFFYSIFAIAGVSLFSGYLKRRCIGEMSGITW 165
GSPATP000334140  IKA---LRTILLALFASIAQLRDAVVVLIFFYSIFAIAGVSLFSGYLKRRCIGEMSGITW 165
GSPATP000297160  IQDSKILKKQFDAFIGSFSRLGPILIPLLFVVLYYSIIGLHLFMGITEQRCRETPEPVNE 172
GSPATP000139240  VPG---LRILVASLLDSLPNLGNVLVFLLYLLVIFGILGLQLFAGAYENRCRFTQFPVNG 170
GSPATP000056360  LKQ---MKVLIITLLDSLPALGNVVIFLLFIIILFGILGLQIFMGALENRCRMTEHPVGN 167
GSPATP000243200  LKE---MKVLIITLLDSLPALGNVVIFLLFIIILFGILGLQIFMGALENRCRETEYPDGN 171
GSPATP000213240  IFEG--LQRMIEALIESLKFIFESTAVVILFVLFFASLGNSLFMSLFNERCLP---SEED 169

GSPATP000171050  GWIQCNE-------------------------------------------------NQCP 177
GSPATP000173340  NDEVC-------------------------------------------------SQCM 163
GSPATP000193790  NNEVC-------------------------------------------------SECQ 163
GSPATP000104430  SGYDTDEPFCNGS-----------------------------------------CP 180
GSPATP000345100  TSIEIDAFCSADSD-----------------------------------CPT------- 182
GSPATP000173330  VSEEI-LFCADDNN-------------------------------------CPFPEQTIYN 188
GSPATP000103230  ISDEI-LFCADDNN-------------------------------------CPFPEDTIYN 188
GSPATP000334140  ISDEI-LFCADDNN-------------------------------------CPFPEDTIYN 188
GSPATP000297160  DWKAVDDIKFLCG-----------------------------------IWECSENSYC 195
GSPATP000139240  TWPADPLITHLCSS-----------------------------------DSDCPDNEFC 194
GSPATP000056360  EWVAS-NYTKLCQD-----------------------------------SSNCPDGTYC 190
GSPATP000243200  IWKAS-NYTKLCQD-----------------------------------SSNCPAGTYC 194
GSPATP000213240  EWIQCIQ-------------------------------------------------GECP 180

GSPATP000171050  EGMSCSYSDYTPKP---------PTSFNNIIYSVGQILRTITMDDWSWVMFFTMRIFN-P 227
GSPATP000173340  GEFDCVKQIANPFN--------DLINFDTFGWSLIQVFICTLLESWSQILDYTIAAYN-E 214
GSPATP000193790  GEFECVKQIANPFN--------SLINFDTFGWSLIQVFICTLLESWSQILDYTVAAYN-E 214
GSPATP000104430  ENYICGKMMDNPIN--------GLLSFDDFGLAFLQVFIITTLEGWTTIMQTIMTTFS-Q 231
GSPATP000345100  -DYLCGKQTSNPQN--------DLINFDTFGWAFLQVFVITTLEGWAQIQEAVILTFS-Q 232
GSPATP000173330  ENFICGKQIANPQN--------NLINFDTFGYAFLQVFIITTLEGWTQIQTAVMLTFS-Q 239
GSPATP000103230  ENFICGKQIANPQN--------DLVNFDTFGYSFLQVFIITTLEGWTQIQTAVMLTFS-Q 239
GSPATP000334140  ENFICGKQIANPQN--------DLVNFDTFGYSFLQVFIITTLEGWTQIQTAVMLTFS-Q 239
GSPATP000297160  GSLADYNLPRNRTENVIEQFGFGFIRFDDFFYSLFVVFTFLNVTGWSGTTFMFWRAMTTY 255
GSPATP000139240  GNPINYNLPNQDTPIPELSY--SYTTFDNILTATFTIFQALTTEGWSTMVTQLIDAVNPI 252
GSPATP000056360  GNPNTYNLPQQESDDETFNY--GYTNFNNIISATFTIFQALTTEGWTKLIFIYQEALSTA 248
GSPATP000243200  GNPNTYNLPQQESDDENFNY--GYTNFNNIISATFTIFQALTTEGWTVLIFIYQEALSTA 252
GSPATP000213240  DGKTCVMTNKAFNI---------PTNFNNIIYSFGLMLKTVTMDDWSWVMYYTIRAFH-P 230
```

```
GSPATP000171050  WIWIYYLLIIFVGGFFGFNLVIAVLKTHYAEVAQETVQEEIKQQMMARLKEKQENPERDL 287
GSPATP000173340  SVYIYFFMVVMVGGFFLVNLTLAIIKLNFSNN---QKLIISP-------------VLEES 258
GSPATP000193790  AVYIYFFMVVMVGGFFLVNLTLAIIKLNFSNN---QKFVIPP-------------VIEES 258
GSPATP000104430  FSVFYFLLCVIIGAFFLVNLTLAIIKINFSKA---ESQEPIK-------------PDDVS 275
GSPATP000345100  FVIFYFLLVVIVGAFFLVNLTLAVIKLNFKPEKIEEELKLIQ-------------EDIEF 279
GSPATP000173330  FVVLYFIIVVLVGAFFLVNLTLAIIKLNFKPEKIQEELAQIK-------------EEIEE 286
GSPATP000103230  YVVLYFIIVVIVGAFFLVNLTLAIIKLNFKPEKIQEELAQIK-------------EEIEE 286
GSPATP000334140  YVVLYFIIVVIVGAFFLVNLTLAIIKLNFKPEKIQEELAQIK-------------EEIEE 286
GSPATP000297160  ITAFYFVSLIFILAYLLSNLLLASFYESFLVLSTIKQNKSID-------------NEKEV 302
GSPATP000139240  LVYVYFNLLIVIGSFFTINLILAVINDSFLKNQSERRQELLK-------------EEQQR 299
GSPATP000056360  IVYIYFLLLIFLGSFFVVNLILAVINDSFMAAQMRSSIKNSQ-------------ASASS 295
GSPATP000243200  IVYIYFLLLIFLGSFFVVNLILAVINDSFIATQMRASLKNSQ-------------TSASS 299
GSPATP000213240  IVWIYYVLIIFVGGFFGFNLPIAVFKTHFSEMQFRSISNQD-------------EKVEA 276
```

Figure 5. Sequence alignment of candidate genes. Candidate genes were aligned and then compared to known sodium channels. The alignment represents only one domain of each candidate gene; this same basic sequence was repeated four times in each candidate. The most highly conserved residues are shown in yellow. Relatively well-conserved amino acids are shown in grey. The blue amino acid residues represent conserved repeated charged residues followed by two hydrophobic residues, a motif that is commonly found in the S4 segment of ion channels. The green asparagine (N) residues are highly conserved in all sodium channels. The red residues correspond to those residues that are thought to be a part of the selectivity filter.

```
GSPATP000173340 I   YFRDFWN        GSPATP000103230 I   YLKDTWN        GSPATP000056360 I   LTTEGWT
GSPATP000173340 II  TLLESWS        GSPATP000103230 II  TTLEGWT        GSPATP000056360 II  LTNESWN
GSPATP000173340 III FTISNWS        GSPATP000103230 III LTLENWN        GSPATP000056360 III STTEGWI
GSPATP000173340 IV  STLLDWP        GSPATP000103230 IV  STLEGWP        GSPATP000056360 IV  STGEDWN

GSPATP000193790 I   YFRDFWN        GSPATP000334140 I   YLKDTWN        GSPATP000243200 I   LTTEGWT
GSPATP000193790 II  TLLESWS        GSPATP000334140 II  TTLEGWT        GSPATP000243200 II  LTNEAWN
GSPATP000193790 III FTISNWS        GSPATP000334140 III LTLENWN        GSPATP000243200 III STTEGWI
GSPATP000193790 IV  STLLDWP        GSPATP000334140 IV  STLEGWP        GSPATP000243200 IV  STGEDWN

GSPATP000104430 I   YLRDLWN        GSPATP000297160 I   LNVTGWS        GSPATP000213240 I   VTMDDWS
GSPATP000104430 II  TTLEGWT        GSPATP000297160 II  SASINYE        GSPATP000213240 II  MTNDDWF
GSPATP000104430 III LTLENWN        GSPATP000297160 III ATSENWI        GSPATP000213240 III STFDGWG
GSPATP000104430 IV  STQEGWP        GSPATP000297160 IV  ASSESHW        GSPATP000213240 IV  VSGDDFR

GSPATP000345100 I   YLRETWN        GSPATP000139240 I   LTTEGWS        GSPATP000171050 I   ITMDDWS
GSPATP000345100 II  TTLEGWA        GSPATP000139240 II  LTSESWN        GSPATP000171050 II  LRANEFN
GSPATP000345100 III LTLENWN        GSPATP000139240 III STTEGWL        GSPATP000171050 III STFDGWG
GSPATP000345100 IV  STLEGWP        GSPATP000139240 IV  STGEQWD        GSPATP000171050 IV  ASGDDWR

GSPATP000173330 I   YLRDTWN
GSPATP000173330 II  TTLEGWT
GSPATP000173330 III LTLENWN
GSPATP000173330 IV  STLEGWP
```

Figure 6. Selectivity filter sequence alignment for candidate genes. The four domains of each candidate gene were aligned and the portion of the alignment containing the selectivity filter is shown above (each candidate gene name is truncated above). Each grouping represents one candidate gene and the roman numerals correspond to one of the four domains in a given candidate. The highlighted green Tryptophan residues (W) are among the most highly conserved residues in the selectivity filter in both sodium and calcium channels. The highlighted blue residues correspond to those residues that are thought to contribute to ion specificity. To determine ion specificity, the blue residues should be read downwards, from domain I to domain IV. Eukaryotic sodium channels are believed to contain one of the following motifs: EKEE, EEKE, DKEA, DEKA, EEMD. Calcium channels are believed to have the following motif: EEEE (Tang et al., 2014; Yue et al., 2002; Stary et al., 2008; Zhen et al., 2005; Payandeh et al., 2011).

Table 2. Summary of the characteristics of candidate genes.

Candidate Gene	Voltage Sensor Residues*	S6 Asparagine (N) Residue	Selectivity Filter Residues#	Prediction@
GSPATP00017105001	Yes	Yes	DNDD^	Likely not a sodium or calcium channel
GSPATP00017334001	Yes	Yes	DESL	Sodium Channel
GSPATP00019379001	Yes	Yes	DESL	Sodium Channel
GSPATP00010443001	Yes	Yes	DEEE	Sodium Channel
GSPATP00034510001	Yes	Yes	EEEE	Calcium Channel
GSPATP00017333001	Yes	Yes	DEEE	Sodium Channel
GSPATP00010323001	Yes	Yes	DEEE	Sodium Channel
GSPATP00033414001	Yes	Yes	DEEE	Sodium Channel
GSPATP00029716001	Yes	Yes	TIEE^	Likely not a sodium or calcium channel
GSPATP00013924001	Yes	Yes	EEEE	Calcium Channel
GSPATP00005636001	Yes	Yes	EEEE	Calcium Channel
GSPATP00024320001	Yes	Yes	EEEE	Calcium Channel
GSPATP00021324001	Yes	Yes	DEEE^	Likely not a sodium or calcium channel

* Since it is predicted that the calcium-dependent sodium channel evolved from a sodium channel, it is expected that the channel of interest would contain some residues resembling the voltage sensor.
# EKEE, EEKE, DKEA, DEKA, EEMD are indicative of a sodium channel. EEEE is indicative of a calcium channel.
@ Predictions are shaped mostly by the identity of the residues in the selectivity filter.
^ One domain does not contain the highly conserved Tryptophan (W) residue.
Highlighted candidate genes were chosen for experimental analysis.

After the sequence alignments were completed, a phylogenetic analysis of the 13 candidate genes was carried out. The phylogenetic tree in Figure 7 reveals that the candidate genes can be separated into three main groups, with homology between the genes being greater within a group. The first group of candidates (GSPAT00021324001 and GSPAT00017105001) does not appear to cluster with either the known calcium channel (CavL_HS) or the known sodium channel (NavHS). This grouping corresponds with the

Wendy Kuzmishin 3/24/2016 2:19 PM
Comment [2]: This corresponds with the selectivity filter analysis…SF does not resemble either a Ca or Na channel.

prediction made in Table 2 for these two candidate genes (blue in Figure 7). The second group of candidate genes appears in a cluster that suggests that they are most similar to known calcium channels (green in Figure 7). This grouping also corresponds to the prediction made in Table 2. The third group of candidates (towards bottom of tree) clusters most closely to the known sodium channel gene, suggesting these genes are similar to a sodium channel (red in Figure 7).

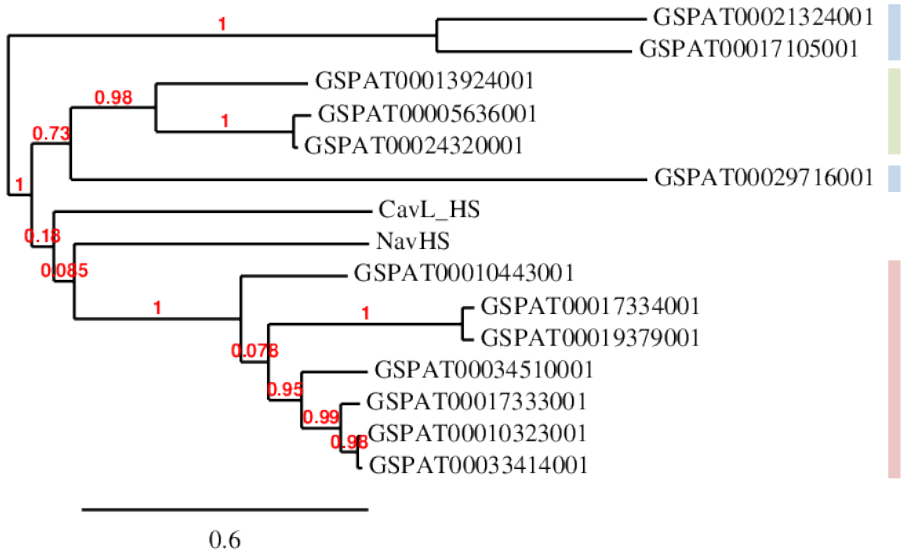

Figure 7. Phylogenetic analysis of the 13 candidate *Paramecium* genes. This phylogenetic tree was built using PhyML, which is based on the maximum likelihood principle, and visualized using TreeDyn, on http://www.phylogeny.fr/index.cgi. **(Guindon et al., 2010)**. The lengths of the branches represent the distance between the candidate genes. A scale bar, which indicates the number of changes per site, is shown on the left. The colored bars adjacent to the tree correspond to predicted ion channel identity, as determined by the alignment of the selectivity filter. Those candidate genes that were predicted to be neither a sodium nor a calcium channel gene are blue. Green represents those channels that are predicted to be calcium channels and red represents those channels that are predicted to be sodium channels. A known calcium channel (CavL_HS) and a known sodium channel (NavHS) were used as reference genes in this tree.

GSPAT00010323001 Knockdown

Knockdown Results in Decreased Backwards Swimming Behavior in KCl

Paramecium with knocked down GSPAT00010323001, exhibited significantly decreased backwards-swimming times in KCl than the empty pL4440 control (Figure 7). The mean backwards swimming time for the knockdown was 23.0 ± 4.4 s at 48 hours and 16.9± 2.2 s at 72 hours. The control had mean backwards swimming times of 31.4 ± 5.1 s and 25.2 ± 3.3 s, respectively. This observed decrease in backwards swimming time in the knockdown was determined to be statistically significant via a Mann-Whitney U test (Supplemental Information). This is not consistent with predicted sodium channel activity.

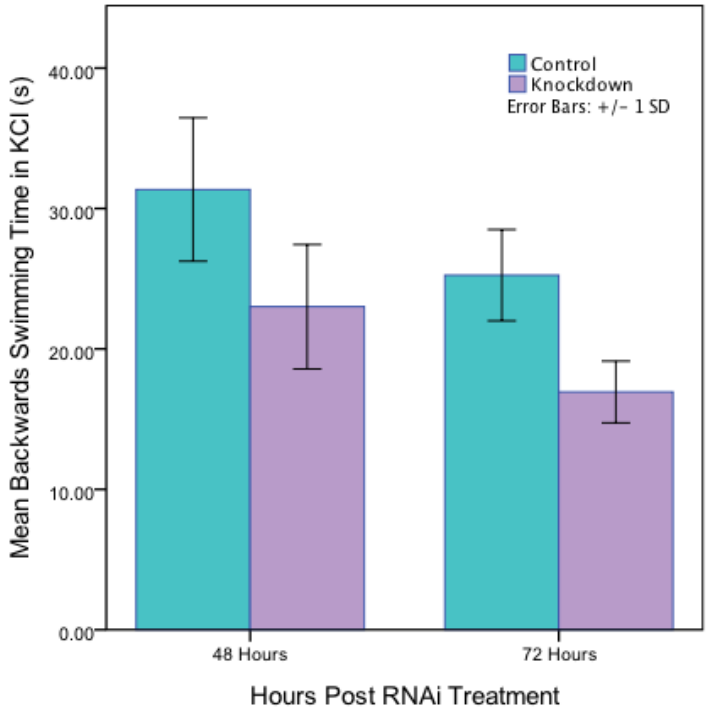

Figure 7. Average backwards swimming time of pL4440 control and GSPAT00010323001 knockdown in KCl test solution. The y-axis represents the time spent swimming backwards and the x-axis represents the time at which the test was conducted. The control (blue) is *Paramecium* fed empty pL4440 and the knockdown (purple) was fed *E. coli* with the GSPAT00010323001 knockdown construct. The error bars indicate variation to one standard deviation. The observed decrease in backwards swimming behavior between the control and knockdown is statistically significant.

Knockdown Results in No Significant Difference in Backwards Swimming Behavior in NaTEA

Paramecium with knocked down GSPAT00010323001, exhibited no significant difference in backwards swimming time compared to the pL4440 controls in Na/TEA test solution (Figure 8). The mean backwards swimming time for the knockdown was 5.0 ± 3.8 s at 48 hours and 5.6 ± 5.2 s at 72 hours. The control had mean backwards-swimming times of 4.7 ± 2.7 s and 6.1 ± 2.4 s, respectively.

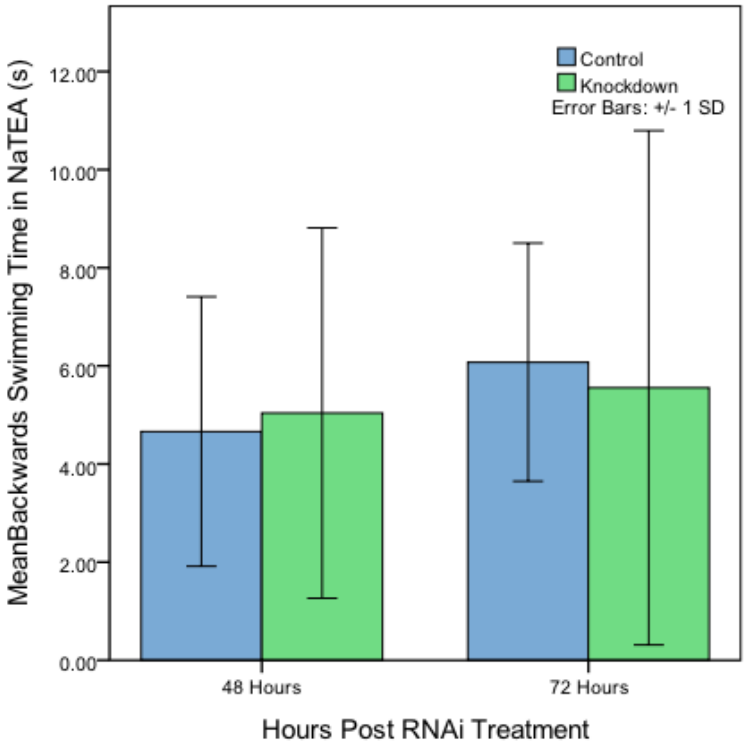

Figure 8. Average backwards swimming time of pL4440 control and GSPAT00010323001 knockdown in Na/TEA test solution. The y-axis represents the time spent swimming backwards and the x-axis represents the time at which the test was conducted. The control (blue) is *Paramecium* fed empty pL4440 and the knockdown (green) was fed *E. coli* with the GSPAT00010323001 knockdown construct. The error bars indicate variation to one standard deviation. The small observed increase in backwards swimming behavior was determined to not be statistically significant via a Mann-Whitney U statistical test.

GSPAT00017334001 Knockdown

Knockdown Results in Increased Backwards Swimming Behavior in KCl

Paramecium with knocked down GSPAT00017334001, exhibited significantly greater backwards-swimming times in 30 mM KCl than the pL4440 controls (Figure 9). The mean backwards swimming time for the knockdown was 118.0 $\pm$ 46.4 s at 48 hours and 75.5 $\pm$ 48.6 s at 72 hours. The control had mean backwards swimming times of 38.9 $\pm$ 13.9s and 50.6 $\pm$ 21.2 s, respectively. This observed increase in backwards swimming time in the knockdown was determined to be statistically significant via a Mann-Whitney U test (Supplemental Information). This is not consistent with the expected channel activity for either sodium or calcium channels.

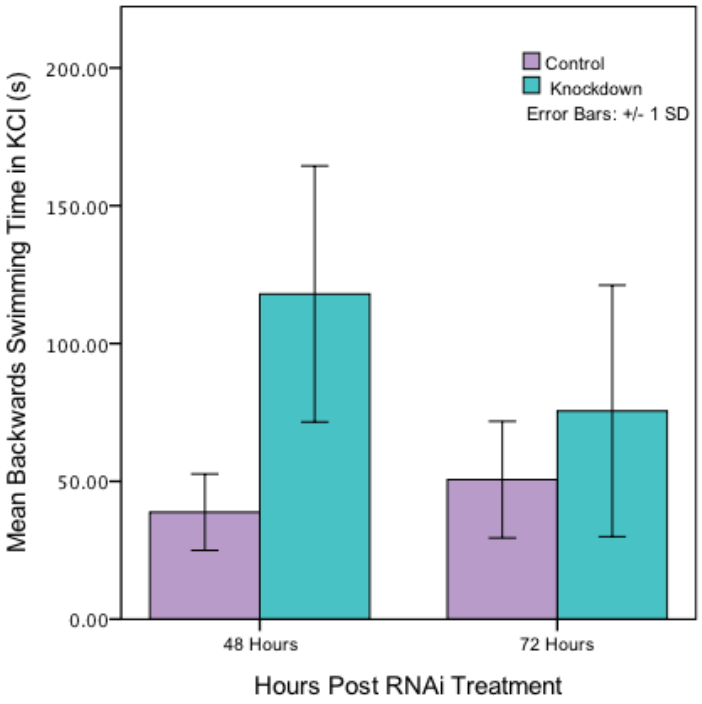

Figure 9. Average backwards swimming time of pL4440 control and GSPAT00017334001 knockdown in KCl test solution. The y-axis represents the time spent swimming backwards and the x-axis represents the time at which the test was conducted. The control (purple) is *Paramecium* fed empty pL4440 and the knockdown (blue) was fed *E. coli* with the GSPAT00017334001 knockdown construct. The error bars indicate variation to one standard deviation. The observed increase in backwards swimming behavior between the control and knockdown is statistically significant.

Knockdown Results in Increased Backwards Swimming Behavior in Na/TEA

Paramecium with knocked down GSPAT00017334001, exhibited significantly greater backwards-swimming times in Na/TEA than their empty pL4440 control counterparts (Figure 10). The mean backwards swimming time for the knockdown was 22.1 $\pm$ 7.5 s at 48 hours and 6.9 $\pm$ 3.7s at 72 hours. Whereas the control had mean backwards swimming times of 15.4 $\pm$ 8.3 s and 2.2 $\pm$ 1.0 s, respectively. This observed

increase in backwards swimming time in the knockdown was determined to be statistically significant via a Mann-Whitney U test (Supplemental Information).

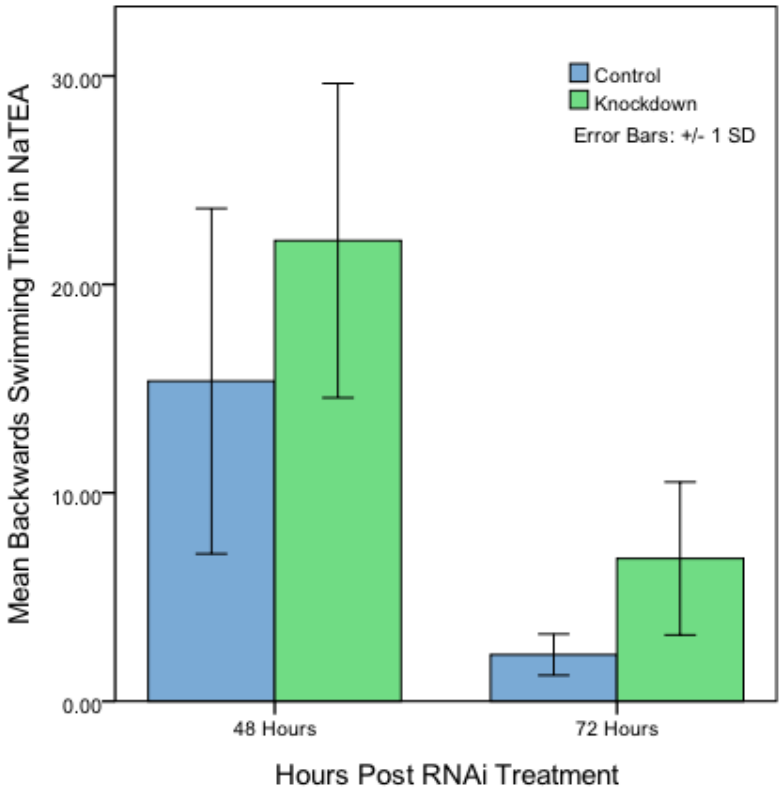

Figure 10. Average backwards swimming time of pL4440 control and GSPAT00017334001 knockdown in Na/TEA test solution. The y-axis represents the time spent swimming backwards and the x-axis represents the time at which the test was conducted. The control (blue) is *Paramecium* fed empty pL4440 and the knockdown (green) was fed *E. coli* with the GSPAT00017334001 knockdown construct. The error bars indicate variation to one standard deviation. The observed increase in backwards swimming behavior between the control and knockdown is statistically significant.

Discussion

Before selecting candidate genes from the *Paramecium* genome, conserved amino acid motifs present in the α subunit of voltage-gated sodium channels were identified. It is hypothesized that the calcium-dependent sodium channel of interest evolved from a voltage-dependent sodium channel. So, the candidate genes are expected to be similar to voltage-dependent sodium channels. A sequence alignment, containing voltage-dependent sodium channels from 8 species representing all domains of life, was annotated with the help of literature and previous annotations (Payandeh et al., 2011; Payandeh and Minor, 2015; Widmark et al., 2011; Goldin, 2002). This alignment revealed multiple highly conserved motifs present in all sodium channels and also related the amino acid sequence back to the structure of the sodium channel (Figure 4). Since all the genes in the alignment contained the conserved sequences, the human sodium channel gene was chosen to search for homologous genes within the *Paramecium* genome.

The human voltage-gated sodium channel α subunit was used to search for homologous genes in the *Paramecium* database. Genes were chosen as candidates based on their E value; genes with low E values are the most similar to the reference gene. Based on low E values, 13 genes were chosen as candidates for a calcium dependent sodium channel in *Paramecium* (Table 2). The 13 candidate genes were then aligned and compared to the initial alignment of conserved sodium channel motifs (Figure 5). The annotated alignment of the candidate genes revealed that all the candidates possess motifs that are found to be conserved in voltage-dependent sodium channels. For instance, all candidates possessed an Asparagine (N) residue in the sixth transmembrane segment (S6) (Payandeh et al., 2011). All candidates also possessed the same repeating motif: one charged residue, such as

Lysine or Arginine, followed by two hydrophobic residues, such as alanine and valine. This motif of alternating charged and hydrophobic residues is characteristic of the fourth transmembrane segment (S4) and is present in most voltage-dependent ion channels. This alignment data, therefore, suggests that the candidate genes are all genes for voltage-dependent ion channels and that they are also likely sodium channels.

To further analyze the candidate genes and make predictions about ion channel type, the selectivity filter for each candidate gene was analyzed (Figure 5). The selectivity filter represents the narrowest portion of the ion channel and makes direct contacts with any ions passing through the channel (Dudev and Lim, 2014). Residues from each of the four domains of the channel protein contribute to the selectivity filter. The identity of the amino acid residues that make up this filter dictates ion channel specificity, although one residue, tryptophan, is present in all selectivity filters regardless of the other residues. It has been previously shown that sodium channels have one of the following motifs in the selectivity filter: EKEE, EEKG, DKEA, DEKA, ad EEMD. Calcium channels are known to have an EEEE motif in the selectivity filter (Tang et al., 2014; Yue et al., 2002; Stary et al., 2008; Zhen et al., 2005; Payandeh et al., 2011)**.** Analysis of the specificity filter residues (Figure 5 and Table 2), indicates that not all candidate genes are sodium channels. Four of the 13 candidate genes are predicted to be calcium channels due to the selectivity filter motif being the same as those found in calcium channels (Table 2). Three of the 13 appear to be neither calcium nor sodium channels, as they do not contain the highly conserved trytophan (W) residue in the selectivity filter. The rest of the candidates are predicted to be sodium channels. Even though some of the selectivity filter motifs do not appear to match

those expected with sodium channels, they motifs appear to be more similar to those found in sodium channels than calcium channels.

To further understand how the candidate genes were related to one another, a phylogenetic tree was created (Figure 7). The phylogenetic tree shows that the candidates can be separated into three broad groups based on homology. This suggests that the candidate genes are undergoing evolutionary selection, as they are continuing to diverge. Furthermore, the three main groupings also further support the findings from the selectivity filter analysis: that the candidate genes represent calcium, sodium, and some other (yet to be determined) channel (Table 2).

From the list of candidate genes, four were chosen for experimental analysis. Three genes were chosen because they were predicted to be sodium channels and one was chosen because it was predicted to be a calcium channel (highlighted in Table 2). The four genes chosen for further experimental analysis were GSPAT00017334001, GSPAT00010323001, GSPAT00024320001, and GSPAT0001733301. These genes were used to create knockdown constructs in pL4440 plasmid. The pL4440 plasmid is an RNAi vector that contains an IPTG-inducible T7 promoter and an Ampicillin resistance gene. This plasmid, therefore, was chosen for its ability to induce expression of dsRNA from the insert of interest. Using classic molecular biology techniques, constructs were created for each of the four selected candidate genes. Unfortunately, only two of the cloned constructs were successful, GSPAT00017334001 and GSPAT00010323001. The other two candidate constructs could not be successfully transformed in to competent *E. coli* HT115 cells.

The two knockdown constructs were used to carry out RNAi in *Paramecium*. Behavioral tests were conducted 48 and 72 hours post RNAi treatment. Data from two

timepoints was collected because RNAi is time-dependent; it can take a few days to get a successful knockdown. Two tests were also carried out on each knockdown in order to observe how backwards swimming behavior may have changed. The first test that was conducted involved the 30 mM KCl testing solution. In this test, the large amount of extracellular potassium (leading to temporary hyperpolarization) drives the opening of voltage-dependent calcium channels in the *Paramecium* membrane. Calcium is the inward charge carrier in *Paramecium* and is therefore responsible for causing depolarization upon influx of calcium into the intracellular space (Preston et al., 1992). It is this depolarization that is thought to cause the backwards-swimming behavior that is characteristic of *Paramecium.* Therefore, the more calcium channels that are open and allowing calcium to flow into the cell, the longer the observed backwards swimming behavior will be. The second test that was conducted to study changes in backwards swimming behavior involved Na/TEA solution. TEA, or tetraethylammonium, is known to inhibit voltage-dependent potassium channels. In effect, this test serves to help distinguish the functions of calcium-dependent potassium channels and sodium channels (Hinrichsen et al., 1992). It has been previously demonstrated that cells with no calcium-dependent potassium current exhibit prolonger periods of backwards swimming compared to the control (Hinrichsen et al., 1992; Fraga et al., 1998; Hinrichsen et al., 1985). Cells with no calcium-dependent sodium channel have decreased backwards swimming times in Na/TEA solution. It is important to note, however, that calcium channels are so still active during this test. So any changes to backwards swimming behavior could also be due to calcium channels, depending on whether they are active or not. The combination of these two tests, therefore, can contribute to the understanding of ion channel gene function in *Paramecium.*

Table 3. Summary of observed phenotypes in knockdowns

Gene	Backwards Swimming Time in KCl #	Backwards Swimming Time in Na/TEA #
GSPAT00017334001	Increased *	Increased *
GSPAT00010323001	Decreased *	Same

Backwards swimming time is reported as relative to the pL4440 control.
* Observed Difference was statistically significant

The goal of this thesis was to discover the gene for a calcium-dependent sodium channel in *Paramecium*. Therefore, one would expect a knockdown of this gene to result in decreased backwards-swimming behavior in Na/TEA solution and little to no change in behavior in KCl test solution. As seen in Table 3, knockdown of GSPAT00010323001 resulted in significantly decreased backwards-swimming behavior in KCl relative to the control. This is consistent with observations of inhibited calcium channels in *Paramecium* (Fraga et al., 1998; Hinrichsen et al., 1985). There was no observed difference in the backwards-swimming behavior of this knockdown in Na/TEA testing solution, which also seems to be consistent with the idea that GSPAT00010323001 is a calcium-channel. While the goal of this study was to identify a calcium-dependent sodium channel, it is not surprising that a calcium channel was found instead of a sodium channel. Sodium and calcium channels are more closely related to one another than potassium channels and both share conserved motifs that are common in all ion channels (Payandeh et al., 2011; Payandeh and Minor, 2015; Yu and Catterall, 2003). Further analysis of the selectivity filter also seems to support this. The key selectivity filter motif for this candidate gene is DEEE (Table 2) and the conserved motif for calcium channels is EEEE. Therefore, it is not

surprising that this channel has specificity for calcium (Tang et al., 2014; Yue et al., 2002; Stary et al., 2008; Zhen et al., 2005; Payandeh et al., 2011).

Knockdown of GSPAT00017334001 resulted in increased backwards swimming in both the KCl solution and in the Na/TEA solution (Table 3). This result was completely unexpected and does not align with what the behavior of an inhibited calcium-dependent sodium channel. The average backwards swimming time for the knockdown was significantly increased in both solutions compared to the control. It has been shown previously that an increase in backwards swimming behavior (90-120 s) in Na/TEA solution is associated with an inhibited calcium-dependent potassium channel (Hinrichsen et al., 1992). So, the knockdown does display behavior that is typically associated with calcium-dependent potassium channels. However, the increased backwards-swimming behavior in KCl complicates this analysis. In the KCl test solution, only calcium channels should be active, so it's likely that this observed increase in backwards swimming time in Na/TEA solution is not due to this gene being a potassium channel. It is most likely that these observed increases in backwards swimming are due to calcium and/or sodium ions not being cleared from the cell fast enough. It is difficult to classify what type of ion channel the GSPAT00017334001 gene represents, but it is likely an ion channel. Future work should focus on doing a more thorough bioinformatics analysis of this gene and repetition of the behavioral assays. It is possible that this gene is for a calcium-dependent potassium channel, but more assays and analysis need to be done to confirm this.

While this study did result in the identification of a calcium channel in *Paramecium*, this study also revealed some major drawbacks to using RNAi to knockdown genes. As seen in Figures 6-9, there is a lot of variability in the data collection. While some of this can be

attributed to the nature of working with a living organism, it is likely also due to incomplete penetrance by RNAi. Knocking down genes with RNAi involves dsRNA introduced by the plasmid being broken up into fragments by the RISC and then binding to corresponding free mRNA within the *Paramecium* cell (Arnaiz and Sperling, 2011). This means that only mRNA is being knocked down, not any translated protein. Therefore, it is likely that at least a small number of ion channels of interest were still active within the membrane. Furthermore, the RNAi procedure is very dependent on the number of *Paramecium* cells. The literature suggests that using greater than 100 *Paramecium* cells can lead to incomplete penetrance (Arnaiz and Sperling, 2011). During data collection, it was observed that some experiments contained significantly more than 100 *Paramecium* cells, which could also explain this variability. Future work, therefore, could focus on using a technique to silence these genes and not just knock them down with RNAi. It is possible that the CRISPR/Cas9 system of gene editing may work well in *Paramecium* (Sander and Joung, 2014). Future work could also utilize quantitative PCR (qPCR) to confirm knockdowns of candidate genes of interest. Quantifying the amount of mRNA of interest in the cell could reveal how well a particular knockdown worked and would also allow for the normalization of backwards-swimming behavior based on percent knockdown (Huggett et al., 2005). Not only would qPCR allow for the normalization of data, but it would also confirm that any observed changes in backwards-swimming behavior were due to the knockdown of the gene of interest and not some other experimental variable.

In addition to focusing on how the techniques utilized in this study could be improved, future work should also focus on exploring the other identified candidate genes. Of particular interest is the GSPAT0001733301 gene, which previous work in the Fraga lab

has indicated may be a calcium-dependent sodium channel. Previous bioinformatics analysis of this candidate gene found that this channel contains motifs found in sodium channels and also has calcium-binding motifs. Identifying the functions of the candidate genes may reveal more about how genes evolve to have new functions after gene duplication events.

Finally, future work should also focus on improving bioinformatics techniques. This study has shown that while bioinformatics can aid in identifying candidate genes, more robust methods are needed to make predictions about gene function. Because calcium and sodium channels are relatively closely related and share many of the same amino acid motifs, more thorough bioinformatic analysis is needed. To test predictions about key motifs, site-directed mutagenesis could be utilized to see if ion channel specificity is changed upon altering a predicted key motif. It has already been shown that changing one amino acid can alter channel specificity, so future work may exploit this idea to better identify those key motifs, especially those found in the selectivity filter.

Literature Cited

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Ion Channels and the Electrical Properties of Membranes.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Anderson, P.A.V., and Greenberg, R.M. (2001). Phylogeny of ion channels: clues to structure and function. Comp. Biochem. Physiol. B Biochem. Mol. Biol. *129*, 17–28.

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst. Biol. *55*, 539–552.

Arnaiz, O., and Sperling, L. (20111). ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate Paramecium tetraurelia. Nucleic Acids Res *39*.

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., et al. (2010). Paramecium tetraurelia: The Renaissance of an Early Unicellular Model. Cold Spring Harb. Protoc. *2010*, pdb.emo140.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540–552.

Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B., and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics *7*, 439.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. *36*, W465–W469.

Dudev, T., and Lim, C. (2014). Ion Selectivity Strategies of Sodium Channel Selectivity Filters. Acc. Chem. Res. *47*, 3580–3587.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Fraga, D., Yano, J., Reed, M.W., Chuang, R., Bell, W., Houten, J.L.V., and Hinrichsen, R. (1998). Introducing Antisense Oligodeoxynucleotides into Paramecium via Electroporation. J. Eukaryot. Microbiol. *45*, 582–588.

Fraga, D., Sehring, I.M., Kissmehl, R., Reiss, M., Gaines, R., Hinrichsen, R., and Plattner, H. (2010). Protein Phosphatase 2B (PP2B, Calcineurin) in Paramecium: Partial Characterization Reveals That Two Members of the Unusually Large Catalytic Subunit Family Have Distinct Roles in Calcium-Dependent Processes. Eukaryot. Cell *9*, 1049–1063.

Frey, B.B. (2015). There's a Stat for That!: What to Do & When to Do it (Thousand Oaks: SAGE Publications, Inc).

Galione, A., Evans, A.M., Ma, J., Parrington, J., Arredouani, A., Cheng, X., and Zhu, M.X. (2009). The acid test: the discovery of two-pore channels (TPCs) as NAADP-gated endolysosomal Ca2+ release channels. Pflüg. Arch. - Eur. J. Physiol. *458*, 869–876.

Galvani, A., and Sperling, L. (2002). RNA interference by feeding in Paramecium. Trends Genet. *18*, 11–12.

Goldin, A.L. (2002). Evolution of voltage-gated Na+ channels. J. Exp. Biol. *205*, 575–584.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Harrison, P.M., Hegyi, H., Balasubramanian, S., Luscombe, N.M., Bertone, P., Echols, N., Johnson, T., and Gerstein, M. (2002). Molecular Fossils in the Human Genome: Identification and Analysis of the Pseudogenes in Chromosomes 21 and 22. Genome Res. *12*, 272–280.

Haynes, W.J., Ling, K.-Y., Saimi, Y., and Kung, C. (2003). PAK Paradox: Paramecium Appears To Have More K+-Channel Genes than Humans. Eukaryot. Cell *2*, 737–745.

Heinemann, S.H., Terlau, H., Stühmer, W., Imoto, K., and Numa, S. (1992). Calcium channel characteristics conferred on the sodium channel by single mutations. Nature *356*, 441–443.

Hille, B. (1989). Ionic Channels: Evolutionary Origins and Modern Roles. Q. J. Exp. Physiol. *74*, 785–804.

Hille, B. (1992). Ionic Channels of Excitable Membranes (Sunderland, Massachusetts: Sinauer Associates Inc.).

Hinrichsen, R.D., and Saimi, Y. (1984). A mutation that alters properties of the calcium channel in Paramecium tetraurelia. J. Physiol. *351*, 397–410.

Hinrichsen, R.D., Amberger, E., Saimi, Y., Burgess-Cassler, A., and Kung, C. (1985). Genetic Analysis of Mutants with a Reduced Ca2+-Dependent K+ Current in PARAMECIUM TETRAURELIA. Genetics *111*, 433–445.

Hinrichsen, R.D., Fraga, D., and Reed, M.W. (1992). 3'-modified antisense oligodeoxyribonucleotides complementary to calmodulin mRNA alter behavioral responses in Paramecium. Proc. Natl. Acad. Sci. U. S. A. *89*, 8601–8605.

Huggett, J., Dheda, K., Bustin, S., and Zumla, A. (2005). Real-time RT-PCR normalisation; strategies and considerations. Genes Immun. *6*, 279–284.

Kissmehl, R., Treptau, T., Hofer, H.W., and Plattner, H. (1996). Protein phosphatase and kinase activities possibly involved in exocytosis regulation in Paramecium tetraurelia. Biochem. J. *317 (Pt 1)*, 65–76.

Kleinjan, D.A., Bancewicz, R.M., Gautier, P., Dahm, R., Schonthaler, H.B., Damante, G., Seawright, A., Hever, A.M., Yeyati, P.L., van Heyningen, V., et al. (2008). Subfunctionalization of Duplicated Zebrafish pax6 Genes by cis-Regulatory Divergence. PLoS Genet *4*, e29.

Kung, C., Saimi, Y., Haynes, W.J., Ling, K.-Y., and Kissmehl, R. (2000). Recent Advances in the Molecular Genetics of Paramecium1. J. Eukaryot. Microbiol. *47*, 11–14.

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., and Ravikesavan, R. (2013). Gene duplication as a major force in evolution. J. Genet. *92*, 155–161.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature *431*, 343–349.

Payandeh, J., and Minor, D.L. (2015). Bacterial voltage-gated sodium channels (BacNaVs) from the soil, sea, and salt lakes enlighten molecular mechanisms of electrical signaling and pharmacology in the brain and heart. J. Mol. Biol. *427*, 3–30.

Payandeh, J., Scheuer, T., Zheng, N., and Catterall, W.A. (2011). The crystal structure of a voltage-gated sodium channel. Nature *475*, 353–358.

Preston, R.R., Saimi, Y., and Kung, C. (1992). Calcium current activated upon hyperpolarization of Paramecium tetraurelia. J. Gen. Physiol. *100*, 233–251.

Qian, W., and Zhang, J. (2014). Genomic evidence for adaptation by gene duplication. Genome Res. *24*, 1356–1362.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. *32*, 347–355.

Seoighe, C., and Gehring, C. (2004). Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. Trends Genet. TIG *20*, 461–464.

Sisu, C., Pei, B., Leng, J., Frankish, A., Zhang, Y., Balasubramanian, S., Harte, R., Wang, D., Rutenberg-Schoenberg, M., Clark, W., et al. (2014). Comparative analysis of pseudogenes across three phyla. Proc. Natl. Acad. Sci. *111*, 13361–13366.

Sperling, L. (2011). Remembrance of things past retrieved from the Paramecium genome. Res. Microbiol. *162*, 587–597.

Stary, A., Shafrir, Y., Hering, S., Wolschann, P., and Guy, H.R. (2008). Structural Model of the CaV1.2 Pore. Channels *2*, 210–215.

Tanaka, K., Matsumoto, T., and Suzuki, T. (2011). Identification of amino acid residues responsible for taurocyamine binding in mitochondrial taurocyamine kinase from Arenicola brasiliensis. Biochim. Biophys. Acta *1814*, 1219–1225.

Tang, L., Gamal El-Din, T.M., Payandeh, J., Martinez, G.Q., Heard, T.M., Scheuer, T., Zheng, N., and Catterall, W.A. (2014). Structural basis for Ca2+ selectivity of a voltage-gated calcium channel. Nature *505*, 56–61.

Timmons, L., Court, D.L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. Gene *263*, 103–112.

Tokunaga, H.T. (2015). Fundamental Statistics for the Social and Behavioral Sciences (Los Angeles: SAGE Publications, Inc).

Widmark, J., Sundström, G., Ocampo Daza, D., and Larhammar, D. (2011). Differential evolution of voltage-gated sodium channels in tetrapods and teleost fishes. Mol. Biol. Evol. *28*, 859–871.

Woolfe, A., and Elgar, G. (2007). Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. Genome Biol. *8*, R53.

Yokoyama, S., and Yokoyama, R. (1989). Molecular evolution of human visual pigment genes. Mol. Biol. Evol. *6*, 186–197.

Yu, F.H., and Catterall, W.A. (2003). Overview of the voltage-gated sodium channel family. Genome Biol. *4*, 207.

Yue, L., Navarro, B., Ren, D., Ramos, A., and Clapham, D.E. (2002). The cation selectivity filter of the bacterial sodium channel, NaChBac. J. Gen. Physiol. *120*, 845–853.

Zakon, H.H. (2012). Adaptive evolution of voltage-gated sodium channels: the first 800 million years. Proc. Natl. Acad. Sci. U. S. A. *109 Suppl 1*, 10619–10625.

Zhang, J. (2003). Evolution by gene duplication: an update. Trends Ecol. Evol. *18*, 292–298.

Zhen, X., Xie, C., Fitzmaurice, A., Schoonover, C.E., Orenstein, E.T., and Yang, J. (2005). Functional Architecture of the Inner Pore of a Voltage-gated Ca2+ Channel. J. Gen. Physiol. *126*, 193–204.

Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and Wang, W. (2008). On the origin of new genes in Drosophila. Genome Res. *18*, 1446–1455.

Addgene: L4440.

Supplemental Information

Statustics:

RNAi GSPAT00017374001 knockdown – KCl testing – 48 hrs

VAR00001

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	VAR00001	N	Percent	N	Percent	N	Percent
Time (S)	Control	25	100.0%	0	0.0%	25	100.0%
	Knockdown	21	100.0%	0	0.0%	21	100.0%

Descriptives

	VAR00001			Statistic	Std. Error
Time (S)	Control	Mean		38.7820	2.77039
		95% Confidence Interval for Mean	Lower Bound	33.0642	
			Upper Bound	44.4998	
		5% Trimmed Mean		38.5340	
		Median		35.6000	
		Variance		191.876	
		Std. Deviation		13.85193	
		Minimum		14.80	
		Maximum		67.56	
		Range		52.76	
		Interquartile Range		21.73	
		Skewness		.330	.464
		Kurtosis		-.706	.902
	Knockdown	Mean		117.9771	10.13439
		95% Confidence Interval for Mean	Lower Bound	96.8372	
			Upper Bound	139.1171	
		5% Trimmed Mean		118.3802	
		Median		105.9300	
		Variance		2156.823	
		Std. Deviation		46.44161	
		Minimum		48.24	
		Maximum		180.00	
		Range		131.76	
		Interquartile Range		103.96	
		Skewness		.194	.501
		Kurtosis		-1.295	.972

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	VAR00001	Statistic	df	Sig.	Statistic	df	Sig.
Time (S)	Control	.115	25	.200*	.968	25	.606
	Knockdown	.195	21	.036	.891	21	.023

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Mann-Whitney Test

Ranks

	VAR00001	N	Mean Rank	Sum of Ranks
Time (S)	Control	25	13.48	337.00
	Knockdown	21	35.43	744.00
	Total	46		

Test Statistics[a]

	Time (S)
Mann-Whitney U	12.000
Wilcoxon W	337.000
Z	-5.530
Asymp. Sig. (2-tailed)	.000
Exact Sig. (2-tailed)	.000
Exact Sig. (1-tailed)	.000
Point Probability	.000

a. Grouping Variable: VAR00001

Wald-Wolfowitz Test

Frequencies

	VAR00001	N
Time (S)	Control	25
	Knockdown	21
	Total	46

Test Statistics[a,b]

		Number of Runs	Z	Asymp. Sig. (1-tailed)	Exact Sig. (1-tailed)	Point Probability
Time (S)	Exact Number of Runs	10[c]	-4.005	.000	.000	.000

a. Wald-Wolfowitz Test

b. Grouping Variable: VAR00001

c. No inter-group ties encountered.

RNAi GSPAT00017374001 knockdown – KCl testing – 72 hrs

VAR00001

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	VAR00001	N	Percent	N	Percent	N	Percent
Time (S)	Control	25	100.0%	0	0.0%	25	100.0%
	Knockdown	25	100.0%	0	0.0%	25	100.0%

Descriptives

	VAR00001			Statistic	Std. Error
Time (S)	Control	Mean		50.6276	4.23002
		95% Confidence Interval for Mean	Lower Bound	41.8973	
			Upper Bound	59.3579	
		5% Trimmed Mean		49.1251	
		Median		44.8500	
		Variance		447.327	
		Std. Deviation		21.15012	
		Minimum		21.21	
		Maximum		109.08	
		Range		87.87	
		Interquartile Range		27.68	
		Skewness		1.147	.464
		Kurtosis		1.601	.902
	Knockdown	Mean		75.5244	9.12155
		95% Confidence Interval for Mean	Lower Bound	56.6984	
			Upper Bound	94.3504	
		5% Trimmed Mean		72.8651	
		Median		61.8400	
		Variance		2080.068	
		Std. Deviation		45.60777	
		Minimum		17.37	
		Maximum		191.18	
		Range		173.81	
		Interquartile Range		65.15	
		Skewness		.762	.464
		Kurtosis		-.009	.902

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	VAR00001	Statistic	df	Sig.	Statistic	df	Sig.
Time (S)	Control	.145	25	.188	.906	25	.025
	Knockdown	.186	25	.026	.923	25	.060

a. Lilliefors Significance Correction

Mann-Whitney Test

Ranks

	VAR00001	N	Mean Rank	Sum of Ranks
Time (S)	Control	25	21.60	540.00
	Knockdown	25	29.40	735.00
	Total	50		

Test Statistics[a]

	Time (S)
Mann-Whitney U	215.000
Wilcoxon W	540.000
Z	-1.892
Asymp. Sig. (2-tailed)	.059
Exact Sig. (2-tailed)	.059
Exact Sig. (1-tailed)	.030
Point Probability	.001

a. Grouping Variable: VAR00001

Wald-Wolfowitz Test

Frequencies

	VAR00001	N
Time (S)	Control	25
	Knockdown	25
	Total	50

RNAi GSPAT00017374001 Knockdown NaTEA Testing – 48 hrs

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	VAR00001	N	Percent	N	Percent	N	Percent
Time (S)	Control	25	100.0%	0	0.0%	25	100.0%
	Knockdown	25	100.0%	0	0.0%	25	100.0%

Descriptives

	VAR00001			Statistic	Std. Error
Time (S)	Control	Mean		15.3612	1.65654
		95% Confidence Interval for Mean	Lower Bound	11.9423	
			Upper Bound	18.7801	
		5% Trimmed Mean		14.9363	
		Median		14.4500	
		Variance		68.603	
		Std. Deviation		8.28270	
		Minimum		3.61	
		Maximum		35.15	
		Range		31.54	
		Interquartile Range		11.52	
		Skewness		.823	.464
		Kurtosis		.037	.902
	Knockdown	Mean		22.1036	1.50888
		95% Confidence Interval for Mean	Lower Bound	18.9894	
			Upper Bound	25.2178	
		5% Trimmed Mean		21.7728	
		Median		21.8100	
		Variance		56.918	
		Std. Deviation		7.54440	
		Minimum		10.73	
		Maximum		39.18	
		Range		28.45	
		Interquartile Range		10.32	
		Skewness		.695	.464
		Kurtosis		.263	.902

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	VAR00001	Statistic	df	Sig.	Statistic	df	Sig.
Time (S)	Control	.125	25	.200*	.936	25	.119
	Knockdown	.097	25	.200*	.946	25	.204

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Mann-Whitney Test

Ranks

	VAR00001	N	Mean Rank	Sum of Ranks
Time (S)	Control	25	19.38	484.50
	Knockdown	25	31.62	790.50
	Total	50		

Test Statistics[a]

	Time (S)
Mann-Whitney U	159.500
Wilcoxon W	484.500
Z	-2.969
Asymp. Sig. (2-tailed)	.003
Exact Sig. (2-tailed)	.003
Exact Sig. (1-tailed)	.001
Point Probability	.000

a. Grouping Variable: VAR00001

Wald-Wolfowitz Test

Frequencies

	VAR00001	N
Time (S)	Control	25
	Knockdown	25
	Total	50

Test Statistics[a,b]

		Number of Runs	Z	Asymp. Sig. (1-tailed)	Exact Sig. (1-tailed)	Point Probability
Time (S)	Minimum Possible	20[c]	-1.715	.043	.058	.027
	Maximum Possible	20[c]	-1.715	.043	.058	.027

a. Wald-Wolfowitz Test

b. Grouping Variable: VAR00001

c. There are 1 inter-group ties involving 2 cases.

RNAi GSPAT00017374001 Knockdown NaTEA Testing – 72 hours

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	VAR00001	N	Percent	N	Percent	N	Percent
Time (S)	Control	25	100.0%	0	0.0%	25	100.0%
	Knockdown	25	100.0%	0	0.0%	25	100.0%

Descriptives

	VAR00001			Statistic	Std. Error
Time (S)	Control	Mean		2.2368	.19822
		95% Confidence Interval for Mean	Lower Bound	1.8277	
			Upper Bound	2.6459	
		5% Trimmed Mean		2.1544	
		Median		1.9300	
		Variance		.982	
		Std. Deviation		.99108	
		Minimum		1.06	
		Maximum		5.15	
		Range		4.09	
		Interquartile Range		1.42	
		Skewness		1.305	.464
		Kurtosis		1.665	.902
	Knockdown	Mean		6.8528	.73437
		95% Confidence Interval for Mean	Lower Bound	5.3371	
			Upper Bound	8.3685	
		5% Trimmed Mean		6.4974	
		Median		6.0000	
		Variance		13.482	
		Std. Deviation		3.67184	
		Minimum		2.93	
		Maximum		18.26	
		Range		15.33	
		Interquartile Range		4.84	
		Skewness		1.452	.464
		Kurtosis		2.502	.902

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	VAR00001	Statistic	df	Sig.	Statistic	df	Sig.
Time (S)	Control	.210	25	.006	.881	25	.007
	Knockdown	.159	25	.104	.872	25	.005

a. Lilliefors Significance Correction

Mann-Whitney Test

Ranks

	VAR00001	N	Mean Rank	Sum of Ranks
Time (S)	Control	25	14.04	351.00
	Knockdown	25	36.96	924.00
	Total	50		

Test Statistics[a]

	Time (S)
Mann-Whitney U	26.000
Wilcoxon W	351.000
Z	-5.560
Asymp. Sig. (2-tailed)	.000
Exact Sig. (2-tailed)	.000
Exact Sig. (1-tailed)	.000
Point Probability	.000

a. Grouping Variable: VAR00001

Wald-Wolfowitz Test

Frequencies

	VAR00001	N
Time (S)	Control	25
	Knockdown	25
	Total	50

Test Statistics[a,b]

		Number of Runs	Z	Asymp. Sig. (1-tailed)	Exact Sig. (1-tailed)	Point Probability
Time (S)	Exact Number of Runs	10[c]	-4.572	.000	.000	.000

a. Wald-Wolfowitz Test

b. Grouping Variable: VAR00001

c. No inter-group ties encountered.

RNAi GSPAT00010323001 knockdown – KCl testing – 48 hrs

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	VAR00001	N	Percent	N	Percent	N	Percent
Time (S)	Control	25	100.0%	0	0.0%	25	100.0%
	Knockdown	25	100.0%	0	0.0%	25	100.0%

Descriptives

	VAR00001			Statistic	Std. Error
Time (S)	Control	Mean		31.3512	1.02108
		95% Confidence Interval for Mean	Lower Bound	29.2438	
			Upper Bound	33.4586	
		5% Trimmed Mean		31.2726	
		Median		31.6600	
		Variance		26.065	
		Std. Deviation		5.10542	
		Minimum		19.66	
		Maximum		46.05	
		Range		26.39	
		Interquartile Range		5.04	
		Skewness		.203	.464
		Kurtosis		2.630	.902
	Knockdown	Mean		23.0076	.88650
		95% Confidence Interval for Mean	Lower Bound	21.1779	
			Upper Bound	24.8373	
		5% Trimmed Mean		23.2040	
		Median		22.5600	
		Variance		19.647	
		Std. Deviation		4.43252	
		Minimum		12.32	
		Maximum		29.81	
		Range		17.49	
		Interquartile Range		5.75	
		Skewness		-.550	.464
		Kurtosis		.090	.902

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	VAR00001	Statistic	df	Sig.	Statistic	df	Sig.
Time (S)	Control	.151	25	.144	.925	25	.065
	Knockdown	.110	25	.200*	.963	25	.488

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Mann-Whitney Test

Ranks

	VAR00001	N	Mean Rank	Sum of Ranks
Time (S)	Control	25	35.60	890.00
	Knockdown	25	15.40	385.00
	Total	50		

Test Statistics[a]

	Time (S)
Mann-Whitney U	60.000
Wilcoxon W	385.000
Z	-4.899
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: VAR00001

RNAi GSPAT00010323001 knockdown – KCl testing – 72 hrs

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	VAR00001	N	Percent	N	Percent	N	Percent
Time (S)	Control	25	100.0%	0	0.0%	25	100.0%
	Knockdown	25	100.0%	0	0.0%	25	100.0%

Descriptives

	VAR00001			Statistic	Std. Error
Time (S)	Control	Mean		25.2488	.65079
		95% Confidence Interval for Mean	Lower Bound	23.9056	
			Upper Bound	26.5920	
		5% Trimmed Mean		25.3577	
		Median		25.6800	
		Variance		10.588	
		Std. Deviation		3.25394	
		Minimum		17.94	
		Maximum		30.81	
		Range		12.87	
		Interquartile Range		4.49	
		Skewness		-.651	.464
		Kurtosis		-.069	.902
	Knockdown	Mean		16.9252	.43938
		95% Confidence Interval for Mean	Lower Bound	16.0184	
			Upper Bound	17.8320	
		5% Trimmed Mean		17.0783	
		Median		17.2700	
		Variance		4.826	
		Std. Deviation		2.19691	
		Minimum		11.43	
		Maximum		19.62	
		Range		8.19	
		Interquartile Range		2.41	
		Skewness		-1.212	.464
		Kurtosis		1.025	.902

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	VAR00001	Statistic	df	Sig.	Statistic	df	Sig.
Time (S)	Control	.119	25	.200*	.954	25	.313
	Knockdown	.174	25	.049	.883	25	.008

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Mann-Whitney Test

Ranks

	VAR00001	N	Mean Rank	Sum of Ranks
Time (S)	Control	25	37.48	937.00
	Knockdown	25	13.52	338.00
	Total	50		

Test Statistics[a]

	Time (S)
Mann-Whitney U	13.000
Wilcoxon W	338.000
Z	-5.811
Asymp. Sig. (2-tailed)	.000
Exact Sig. (2-tailed)	.000
Exact Sig. (1-tailed)	.000
Point Probability	.000

a. Grouping Variable: VAR00001

RNAi GSPAT00010323001 knockdown – NaTEA testing – 48 hrs

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	VAR00001	N	Percent	N	Percent	N	Percent
Time (S)	Control	25	100.0%	0	0.0%	25	100.0%
	Knockdown	25	100.0%	0	0.0%	25	100.0%

Descriptives

	VAR00001			Statistic	Std. Error
Time (S)	Control	Mean		4.6620	.54951
		95% Confidence Interval for Mean	Lower Bound	3.5279	
			Upper Bound	5.7961	
		5% Trimmed Mean		4.3326	
		Median		4.1200	
		Variance		7.549	
		Std. Deviation		2.74754	
		Minimum		2.26	
		Maximum		13.00	
		Range		10.74	
		Interquartile Range		1.26	
		Skewness		2.434	.464
		Kurtosis		5.340	.902
	Knockdown	Mean		5.0368	.75489
		95% Confidence Interval for Mean	Lower Bound	3.4788	
			Upper Bound	6.5948	
		5% Trimmed Mean		4.6919	
		Median		3.2300	
		Variance		14.247	
		Std. Deviation		3.77446	
		Minimum		1.71	
		Maximum		15.49	
		Range		13.78	
		Interquartile Range		6.47	
		Skewness		1.306	.464
		Kurtosis		.813	.902

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	VAR00001	Statistic	df	Sig.	Statistic	df	Sig.
Time (S)	Control	.366	25	.000	.631	25	.000
	Knockdown	.330	25	.000	.783	25	.000

a. Lilliefors Significance Correction

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Time (S)	50	4.8494	3.27280	1.71	15.49	2.9075	3.5650	4.6625
VAR00001	50	.5000	.50508	.00	1.00	.0000	.5000	1.0000

Mann-Whitney Test

Ranks

	VAR00001	N	Mean Rank	Sum of Ranks
Time (S)	Control	25	28.28	707.00
	Knockdown	25	22.72	568.00
	Total	50		

Test Statistics[a]

	Time (S)
Mann-Whitney U	243.000
Wilcoxon W	568.000
Z	-1.349
Asymp. Sig. (2-tailed)	.177
Exact Sig. (2-tailed)	.181
Exact Sig. (1-tailed)	.090
Point Probability	.002

a. Grouping Variable: VAR00001

RNAi GSPAT00010323001 knockdown – NaTEA testing – 72 hrs

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	VAR00001	N	Percent	N	Percent	N	Percent
Time (S)	Control	25	100.0%	0	0.0%	25	100.0%
	Knockdown	25	100.0%	0	0.0%	25	100.0%

Descriptives

	VAR00001			Statistic	Std. Error
Time (S)	Control	Mean		6.0748	.48545
		95% Confidence Interval for Mean	Lower Bound	5.0729	
			Upper Bound	7.0767	
		5% Trimmed Mean		5.9129	
		Median		5.5500	
		Variance		5.892	
		Std. Deviation		2.42726	
		Minimum		3.02	
		Maximum		12.27	
		Range		9.25	
		Interquartile Range		3.27	
		Skewness		.914	.464
		Kurtosis		.587	.902
	Knockdown	Mean		5.5532	1.04858
		95% Confidence Interval for Mean	Lower Bound	3.3890	
			Upper Bound	7.7174	
		5% Trimmed Mean		4.8740	
		Median		3.6500	
		Variance		27.488	
		Std. Deviation		5.24292	
		Minimum		1.20	
		Maximum		24.41	
		Range		23.21	
		Interquartile Range		4.75	
		Skewness		2.288	.464
		Kurtosis		6.155	.902

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	VAR00001	Statistic	df	Sig.	Statistic	df	Sig.
Time (S)	Control	.106	25	.200*	.929	25	.084
	Knockdown	.270	25	.000	.729	25	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Acknowledgements

I have been fortunate throughout this Independent Study process to be advised by Dr. Dean Fraga, who has provided an immense amount of support both in the lab and by challenging me to think critically and independently about my work. Thanks to his guidance, I now feel comfortable approaching biological problems from a bioinformatics standpoint.

I would also like to thank my peers in the Biochemistry and Molecular Biology Class of 2016, and particularly those in the Fraga Lab, for their support and camaraderie.

I would like to thank my parents, Lauran and Eric Weber and John Kuzmishin, for teaching me to always work towards my goals and for always supporting me. Thank you to my grandparents for always being my number one fans and always encouraging me.

I would like to thank my sisters (and best friends), Lexie and Katie, for always being there and for helping me with this project. Thanks to Lexie for always being there when I had lab issues, like when my *Paramecium* died, and thanks to Katie for helping me with my statistical analysis and for bringing me food when I was in the lab late at night.

This work was supported financially by the joint funds and resources of the College of Wooster Biochemistry and Molecular Biology Program and Department of Biology. This thesis was also supported by the Henry J. Copeland Fund for Independent Study.