5-29-2019

Novel Characterization of the Role of Orthologous XAP5 in Caenorhabditis elegans

Nabor Vazquez
Lawrence University

Novel Characterization of the Role of Orthologous XAP5 in *Caenorhabditis elegans*

Nabor Vázquez

A Thesis Submitted in Candidacy for Honors at Graduation from Lawrence University

May 2019

ABSTRACT

Cilia are one of the oldest and most well conserved cellular organelles. Cilia provide an essential role in cellular locomotion, fluid regulation, and are a site for signal transduction pathways involved in sensation. A new study suggests that XAP5 is a transcription factor in a unicellular organism, *Chlamydomonas reinhardtii*, which regulates gene expression needed for proper cilium assembly. Our study investigates the conservation of the role of XAP5 in a multicellular system, *Caenorhabditis elegans*. Alignments between protein, coding region, and promoter sequences for XAP5 orthologs from related species show a good conservation in DNA and protein sequences. As part of the million-mutation project, we obtained a strain, VC40591, which carries a mutation in a gene corresponding to *xap-5* (CE*xap5*). To remove unwanted mutations we "backcrossed" our strains twice to replace mutant chromosomes with wild type chromosomes. Backcrossed strains containing CE*xap5(gk709587)* showed chemosensory and dwelling defective patterns, similar to *C. elegans* with known ciliary deficiencies. Furthermore, a dye-filling assay showed an irregular pattern of dye filling in the tail and sometimes in both the head and tail, at multiple stages of development. However, CExap5 mutant worms typically showed dye filling in the adult stage. Difference at multiple life stages suggest a developmental problem that must be further studied. Overall, our results indicate that CE-XAP5 confers a loss-of-function phenotype consistent with a cilia deficiency, which might suggest a conservation in function with its orthologous protein.

Table of Contents

INTRODUCTION

Living organisms have identical genetic material stored in virtually all cells (sperm, eggs, and white blood cells in mammals), but not all cells express the same proteins from their genes. Differential gene expression is due to the specificity of different regulatory proteins expressed in different cell types. The fundamental idea that deoxyribonucleic acid, DNA, carries genetic material is clearly established. However, our understanding of the identity and mechanism of action of unique regulatory proteins, such as transcription factors that regulate the transcription of DNA to an RNA intermediate is incomplete. This area of research is of interest to the scientific community because of its medical applications and potential to further our understanding of cell differentiation. Transcriptional misregulation by a mutation in any one of several transcription factors can lead to cancer, inflammation, neurological disorders, and cardiovascular diseases (Lee and Young, 2013). The goal of this research project is to study a gene that encodes functional proteins responsible for the formation of cilia.

Cilia are important hair-like extensions of some cells that allow the cell to sense internal and external stimuli. A cilium is a complex structure that is assembled by numerous proteins. Since multiple proteins are involved, many genes must initially be activated for the correct assembly of cilia. *xap5* is a gene recently discovered in a unicellular organism, *Chlamydomonas reinhardtii,* that encodes a transcription factor and regulates certain target genes necessary for structure and function of cilia. My study looks at the orthologous gene of *xap5*, C47E8.4, in a multicellular organism, *Caenorhabditis elegans,* to establish the function of this gene by assessing its mutant phenotypes. Our goal is to determine whether C47E8.4 plays a role in cilia development (hereafter called *CExap5* for *C. elegans xap5*).

Caenorhabditis elegans as model organisms

C. elegans, 1.5 mm in length, are free-living nematodes first used as a model organism by Sydney Brenner in 1965 (Riddle et al., 1997; Figure 1). In the areas of genetics, cell biology, and neurobiology, *C. elegans* are a powerful tool because of their fast life cycle, transparency, self-fertilization, mapped nervous system, and invariant cell lineage. In nature, *C. elegans* are found in temperate regions feeding off bacteria in compost piles and often rotting vegetation (Caswell-Chen et al., 2005). In a laboratory, scientists feed *C. elegans* bacteria, *E. coli*, on small, agar plates (Corsi et al., 2015). The plates can be incubated at 20°C to produce normal growth of the worms or delay growth at 15°C; in both cases, worms can live up to 3 weeks (Corsi et al., 2015).

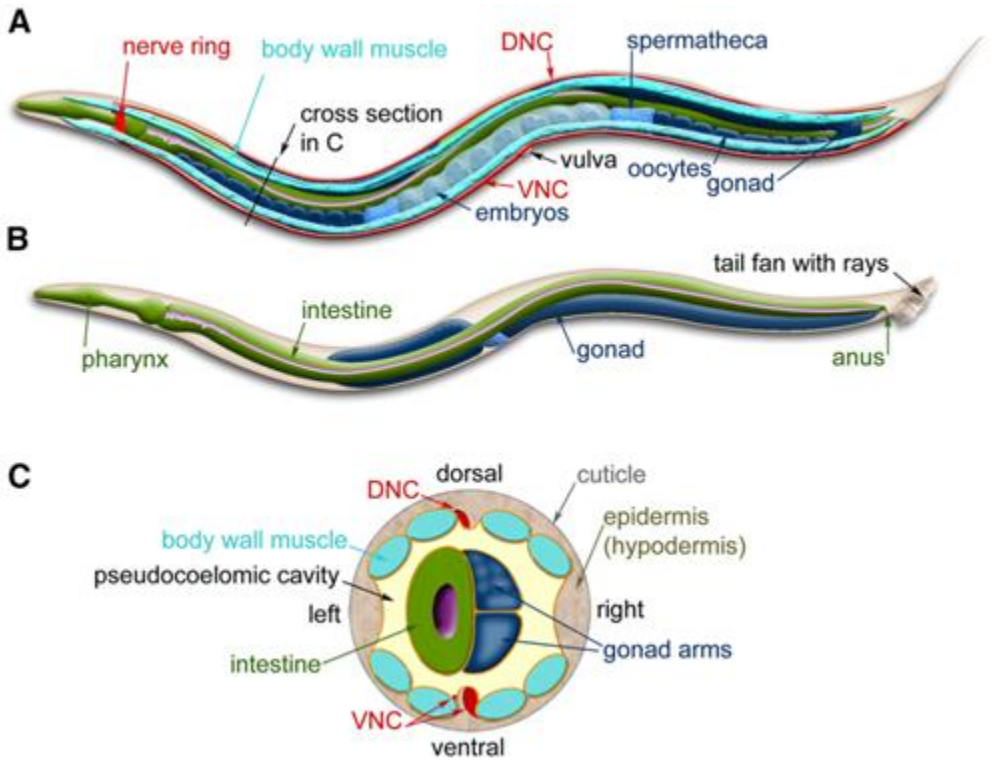

Figure 1. Anatomy of *C. elegans*. Side views of an adult hermaphrodite (A) and male (B). Additionally, a cross section of an adult worm is presented (C). Structures are annotated within the diagram. Figure adapted from Corsi et al. (2015).

C. elegans typically take four days to fully develop; their initial stage begins as an embryo developing first *in utero* and completing development outside the mother (Figure 2). Development continues after hatching into the larval stages L1, L2, L3, and L4; each larval stage is followed by a molt (Corsi et al., 2015). After the final molt, the worm reaches a reproductive, adult stage. In conditions of long periods without food or other environmental stressors, *C. elegans* have developed an evolutionary defense mechanism that allow them to transition developmentally from the L1 stage to a dauer state (Corsi et al., 2015; Figure 2). During the transition period, the cuticle of the worm begins to extend and cover its oral opening, which allows the worm to survive approximately four months without eating. Dauer larvae can decipher food signals and pheromones released by neighboring worms through a signaling pathway that alerts the worm that it has reached a stable environment (Butcher et al., 2007). Then the worm initiates a recovery pathway to proceed directly into the L4 stage of development (Figure 2).

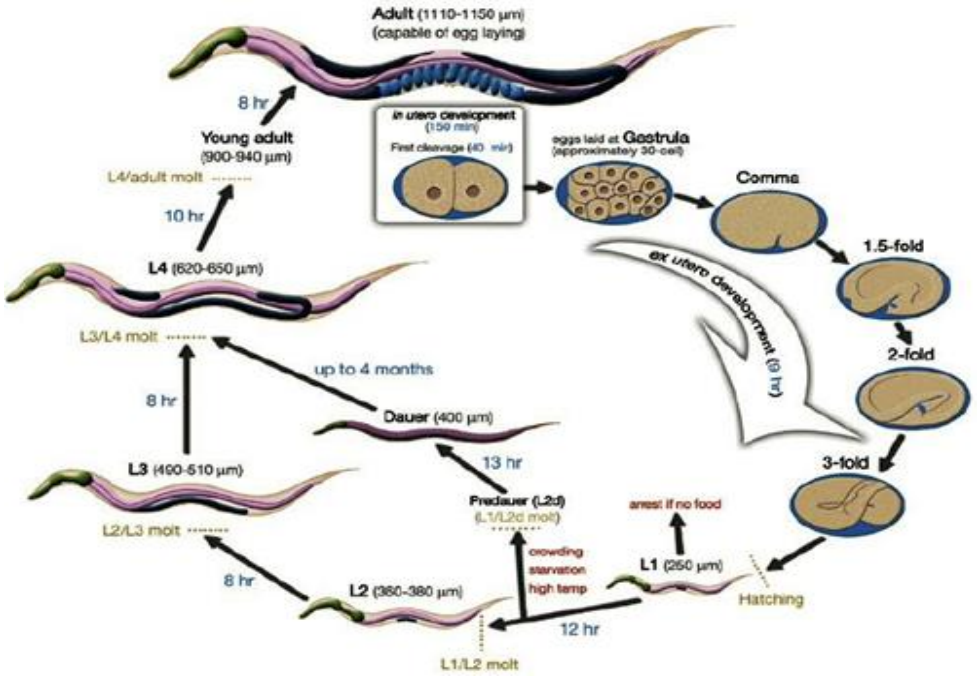

Figure 2. Life cycle of *C. elegans.* Cycle begins at fertilization (0 mins) and continues to the adult hermaphrodite. Blue numbers in the diagram indicate the time a worm typically spends at a certain stage at 22°C. Stages are bolded, and marked next to it are the typical length of the worm in that stage. Yellow text is the transition period from one stage to another. *in-utero* describes development within the uterus and *ex-utero* describes development of egg outside an organism. Figure adapted from (Altun and Hall, 2019).

In addition to its low maintenance and fast life cycle, an advantageous characteristic of working with *C. elegans* is their self-fertilization. One can put a single worm on a bacteria- and agar-coated petri dish and in a matter of days there will be a population with identical genetic material inherited in Mendelian ratios, due to the worm being a hermaphrodite. *C. elegans* self-fertilizes its haploid sperm and haploid egg cells, better known as oocytes, in its gonad. The gonad of *C. elegans*, ovotestis, produces sperm first via meiosis, and then oocytes, which are fertilized by sperm stored in its spermatheca (Corsi et al., 2015). Most commonly *C. elegans* are found as hermaphrodites in the wild, but they can also exist as male at a frequency of less than 0.2% (Corsi et al., 2015). Males and hermaphrodites have five autosomal chromosomes, denoted as I-V, and one sex chromosome. Hermaphrodites develop when an embryo has a pair of sex chromosomes, denoted as XX, while males only have one sex chromosome present, denoted as

XO, or the X/autosome ratio is 1:2 (Herman, 2005). A technique known as heat shock, long exposure to 30°C, causes non-disjunction of the chromosomes in meiosis to produce males (Corsi et al., 2015). Males are needed to mate with hermaphrodites for genetic manipulation to produce offspring with combinations of traits of interest for the researcher.

An annotated neuronal network map and invariant cell lineage in *C. elegans* has proved to be invaluable to study development and neurobiology. It is known that in hermaphrodites there are 959 somatic cells; in contrast, males have 1031 (Herman, 2006). Research of the developmental process has provided insight at the molecular and cellular level for programs like embryogenesis and cell death. The closely annotated neuronal interactions, the connectome, in *C. elegans* have set an important path for future neuronal studies (Figure 3). Hermaphrodites have 302 neuronal cells; from these cells, there are 60 known ciliated, sensory neurons (Inglis, 2016). Ciliated neurons are of interest in this project as we are investigating whether CE*xap5* is responsible for regulating genes involved in cilia functions.

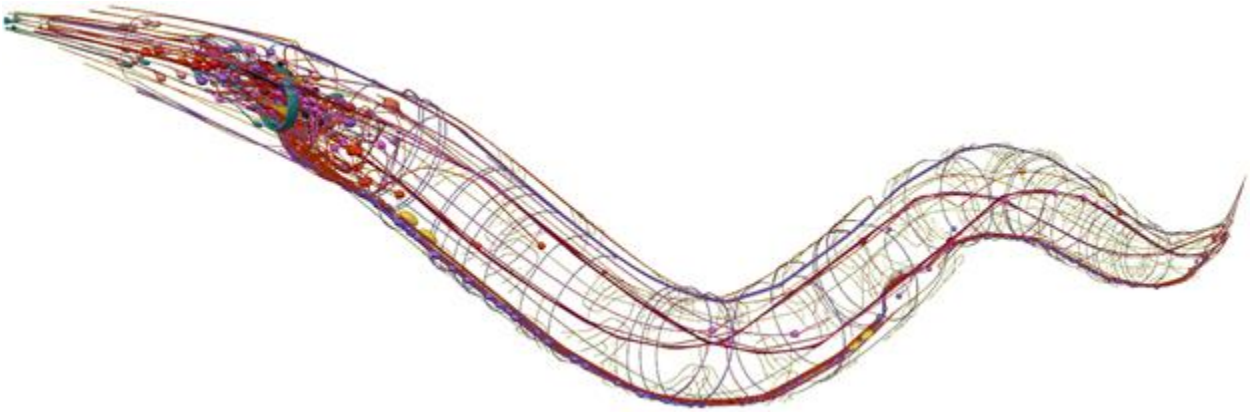

Figure 3 *C. elegans* connectome. Full construction of a map of the nervous system in a hermaphrodite. Figure adapted from http://openworm.org/

A cilium is a dynamic structure

A cilium is a small, microtubule-based organelle found at the surface of virtually all eukaryotic cells (Satir and Christensen, 2008; Ishikawa and Marshall, 2011). Evidence indicates that the last eukaryotic organism had cilia more than 800 million years ago, which demonstrates evolutionary conservation over a long period due to the great functional importance of this organelle (Piasecki et al., 2010). Cilia are important because they have an essential role in cellular locomotion, fluid regulation, and are a site for signal transduction pathways (Choski et al., 2014). Cilia typically have morphologically conserved structures but differ in their size, shape, number per cell, motility patterns, and beating pattern (Choski et al., 2014). Modifications in cilia have occurred through evolution to repurpose the specialization of the organelle for different cell types.

The oldest known cellular organelle is cilia, which was first discovered by Anthony van Leeuwenhoek in 1675 (Francis, 1932). The early observations of Anthony and another scientist Kurt Zimmerman, in the latter 19th century, showed that cilia exist in two states, immotile and motile (Zimmerman 1898; Beales and Jackson 2012). In 1954 two electron microscopist, Fawcett and Porter, began their observations of the ultrastructure of cilia using *Chlamydomonas reinhardtii*, a green alga (Fawcett, 1954). Even though a select group of scientists were interested in studying cilia it took many years for cilia to receive the appreciation they deserved by the scientific community. At first, immotile cilia were categorized as vestigial structures, which shifted the focus for microscopists to devote their early years researching only motile cilia. Motile cilia were considered exciting at first, because their motility helped propel cells and move substances.

An important discovery in 1957 changed how cilia were viewed, a combination of results from Porter, in 1956, and Eduardo De Robertis, suggested that immotile cilia mediated sensory signal transduction between cells (Choski et al., 2014; Satir and Christensen, 2008). Additionally, a breakthrough moment in cilia research happened in 2003 when the Anderson lab concluded that cilia are necessary for the hedgehog signaling pathway (Anderson and Bangs, 2017). This finding caught the attention of many developmental biologists and shed light on the importance of cilia. The organelle we once believed to be a vestigial structure continuous to provide important insights to this day. *Chlamydomonas* remains a popular organism for studying the function and structure of cilia, while other model organisms were adapted at a later point in the investigation of cilia function including *Danio rerio*, *Drosophila melanogaster*, and *C. elegans* (Adams, 2010).

To better understand the different functions of immotile and motile cilia, one must analyze the ultrastructural makeup of the organelle. Near the surface of the cell lies a modified centriole, composed of nine sets of triplet microtubules and other proteins that attach to the plasma membrane. The modified centriole is the nucleation site for the microtubules that will become the cytoskeleton for the cilium. Microtubules begin to extend outwards from the centriole into the membrane to form a section of the cilium known as the basal body (Figure 4). The basal body is connected to the plasma membrane by fibers that are in an intermediate section known as the transition zone, which is a regulated area that permits molecules needed to construct the cilium to enter and exit. As the appendage of the cilium begins to expand due to the extension of axonemal microtubules, the bilayer lipid modifies itself to continuously enclose the organelle (Ishikawa and Marshall, 2011). Cilia have conserved a complex motor mechanism,

intraflagellar transport (IFT), that assists in the continuous elongation of cilia by trafficking the necessary proteins for proper assembly along the axoneme (Figure 4).

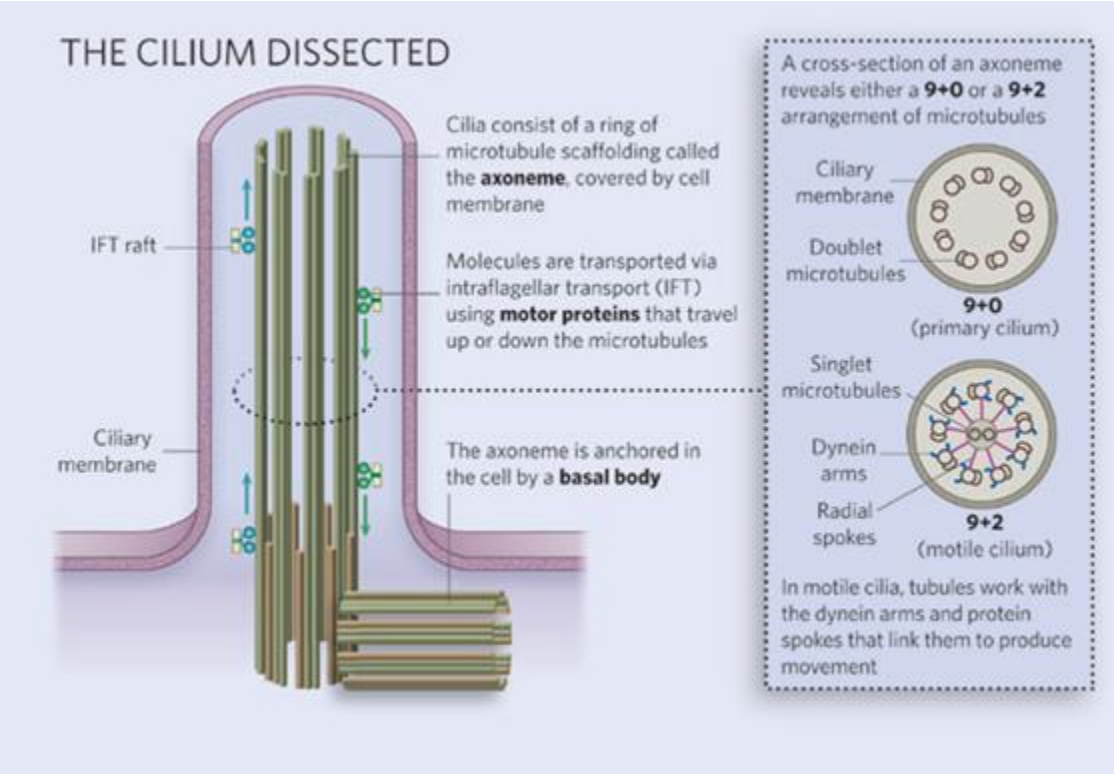

Figure 4. An analysis of a cilium. A microtubule-based organelle is anchored to a modified centriole, a basal body. Intraflagellar transport (IFT), is a system that has motor proteins that 'walk' on microtubules to assist in the assembly and maintenance of a cilium. Primary cilium consist of nine outer doublet microtubules and zero central microtubules, 9 + 0. Motile cilium consist of two central microtubules and nine outer doublet microtubules, 9 + 2. Additionally, a motile cilium has inner and outer dynein and radial spokes that provide motility. Figure modified from Adams (2010).

Differences in the composition of the cytoskeleton in both types of cilia provide an explanation for the motility of motile cilia. The cytoskeleton of a cilium is referred to as the axoneme. In motile cilia, the axoneme is made up of nine outer microtubule doublets and two microtubule singlets in the center, referred to as 9+2 axoneme (Figure 4). Each microtubule, in a doublet, has typically 11-13 protofilaments that form the wall for the circular doublet microtubules; protofilaments are shared between the A tubule and B tubule (Figure 5). A tubule

and B tubules are different as they have a slightly different spatial arrangement and tubulin subunits, but both facilitate the motor-bending proteins that attach to the microtubule (Marshall and Nonaka, 2006; Figure 5). Motile cilia have dynein motors that attach to the A tubule of each doublet; movement of these motor proteins is powered by ATP hydrolysis (Marshall and Nonaka, 2006; Choski et al., 2014; Figure 4). Another unique structure motile cilia have are proteins known as radial spokes, they run inward from the outer doublets and serve to link them to the central pair of microtubules (Marshall and Nonaka, 2006). The axoneme of an immotile cilium lacks the two microtubule singlets, radial spokes, and dynein arms leading the cilium (Adams et al.,1981). Cilia are complex organelles that are assembled by more than 300 unique proteins that allow it to have its specialized function (Piasecki et al., 2010). Since the axoneme is found in the external environment of the cell's cytoplasm, protein synthesis in the cilium is limited.

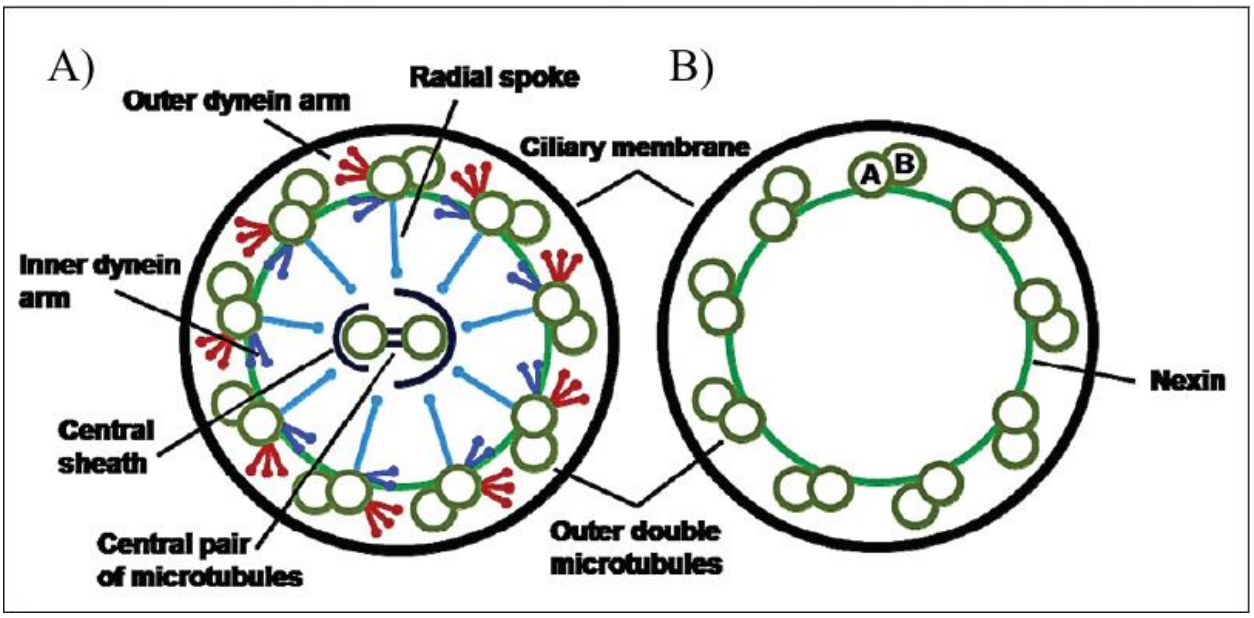

Figure 5 Cross section of a motile (A) and an immotile (B) cilium. 9 outer doublet and 2 central pairs, 9 + 2, is shown in motile cilium. 9 outer doublet and 0 central pairs, 9 + 0, is shown in motile cilium. Unique features of motile cilium include radial spokes, dynein arms, and a central pair of microtubules. Shared features include ciliary membrane and outer, A and B, doublet microtubules. Different components of motile and immotile cilium are labeled to its corresponding location. Figure legend and image modified from Venkatesh (2017).

Today enough evidence demonstrates that categorizing cilia as motile and immotile is an oversimplification. There are exceptions to the classifications of what entails a motile and immotile cilium. For example, 9+0 motile cilia are found in embryonic nodes that help create fluid movement for left-right body axis specification. In addition, olfactory sensory neurons have a 9+2 structure although they are immotile cilia (Jenkins et al., 2009). Recent evidence also suggests that some motile cilia may also have a sensory function (Mao et al., 2018). Motile cilia primarily serve a mechanical function for the cell; they help move fluids and play a role in locomotion. The cells in our respiratory tract are lined with motile cilia that move dust particles and viscous fluids, like mucus, to avoid complications for the air flow when we breathe (Perkins et al., 1986). Sperm is another example of a cell type that uses motile cilia to move. Motile cilia

undulate in a metachronal, planar beat, a rhythm that helps perform their function efficiently (Reiter and Leroux, 2017).

In contrast, the primary function of immotile cilia is to send and receive signals to and from the external or internal compartments of the cell. Immotile cilia can sense (light, odorants, dissolved chemicals, or signaling molecules, that begin a cascade of biochemical reactions to relay the message to its proper location (Falk et al., 2015). Immotile cilia are found as single appendages in multiple cell types such as neurons, Schwann cells, and epithelial cells in most organs such as kidneys (Adams 2010; Satir et al., 2010). It was previously believed that *C. elegans* lacked cilia, however, recent studies show that certain sensory neurons have a single, immotile cilium (Adams, 2010). Cells that are ciliated in most mammals are typically immotile, but other organisms like *C. elegans* only contain immotile cilia. Immotile cilia are known for their sensory roles in model organisms but scientists noticed that counterparts of genes that are responsible for cilia maintenance, formation, and morphogenesis in *C. elegans* can be found in humans (Bae and Barr, 2008). The diversity of cilia arises due to the different specialization of function due to differential gene expression in different cells types.

Cilia are excellent receivers of signals from environmental stimuli. Modified immotile cilia in the rods and cones of the mammalian retina are responsible for receiving light and transducing that signal to the neurons where the brain interprets the external stimuli (Berbari et al., 2009). Cilia act as a backbone for photoreceptors as they connect the cell body of the sensory cell, rod or cone, to the dendrites of neurons (Figure 6). Therefore, the modification in cilia are essential in its specialization, mediating proteins from the cell body to the dendrites of a neuron. Defects in photoreceptors appear in Bardet Biedl Syndrome and nephronophthisis (Bae and Barr, 2008); both disorders are caused by deficiencies in cilia and are known as ciliopathies. As in

mammalian eyes, mammalian noses (and counterparts in other organisms) have immotile cilia that are enriched by G protein coupled receptors (GPCRs) and are able to bind volatile chemicals that are then recognized by our brain as a specific odor (Berbari et al., 2009). Neurons extend their terminal ends to embed 15-20 cilia from the terminal end into the olfactory epithelium. Once a volatile chemical is perceived by the cilium, GPCRs are able to enhance the signal and initiate a signaling cascade that will lead to the depolarization of the neuron, which is then interpreted as a specific odor.

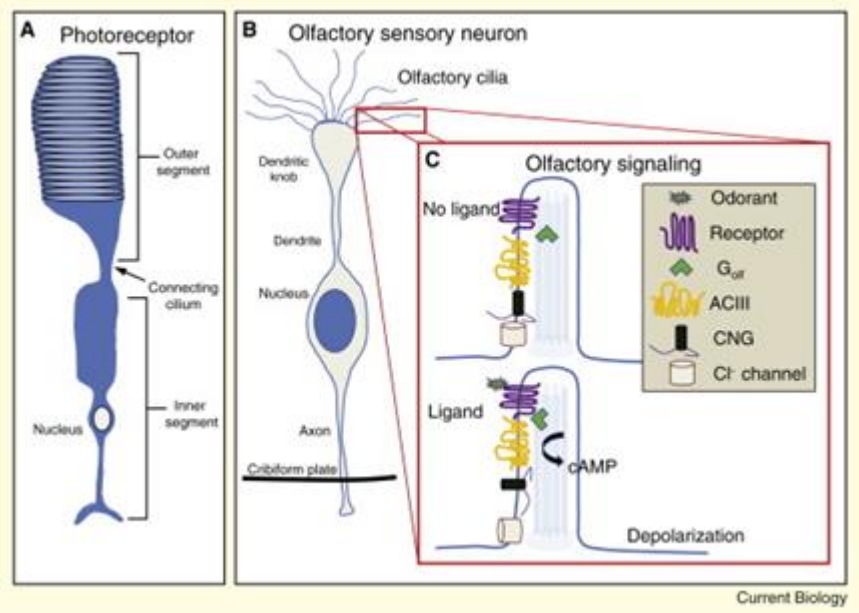

Figure 6. Cilia modification leads to diverse functional specialization. Modified cilium in a photoreceptor mediates a signal cascade between the inner and outer segment of the photoreceptor. Olfactory sensory neurons contain G proteins to enhance olfactory signal. Figure modified from (Berbari et al., 2009).

Chemical cues also change the locomotion, developmental process, and physiological process of an organism. For example, *C. elegans* has developed the ability to distinguish chemicals; 5% of the *C. elegans* genome is composed of these recognizable chemotaxis genes (Bae and Barr, 2008). Using chemotaxis assays, dye-filling, and osmotic avoidance, geneticists have developed an understanding of the deficiencies in cilia in mutant worms (Bae and Barr, 2008). Chemosensory mutant worms do not recognize harmful chemicals, while other mutants

do not respond when physical pressure is applied with an eyelash in the nose touch assay (mechanosensation). Three groups of neurons are responsible for eliciting a response in the nose touch assay: two ASHs, four QLQs, and two FLPs. ASH is a nociceptive master (nose touch, high osmolarity, and noxious chemicals) and has an exposed cilium while the other neurons do not.

Immotile cilia are also important in cell to cell communication. Three pathways that have been studied include Hedgehog (Hh), Wingless (Wnt), and PDGFRα signaling (Veland et al., 2009; Berbari et al., 2009). In vertebrates and invertebrates Hedgehogs (Hh) are important proteins that are secreted and have a functional role in developmental processes, particularly for limb formation in humans and other mammals. Researchers discovered that the hedgehog protein localizes in cilia and mutants that did not have cilia were not able to exhibit hedgehog-directed developmental processes. Interestingly, when studying the conservation of this process in *D. melanogaster* they observed that hedgehog proteins were not localized in cilia, which shows evolutionary divergence between species (Berbari et al., 2009)

In multicellular organisms, researchers identified a gene-specific transcription factor family, Regulatory Factor X (RFX), that is needed to produce ciliary diversity in various cell types. RFX proteins have a highly conserved winged- helix DNA-binding domain that is required for the regulation of sensory, ciliary genes (Piasecki et al., 2010). These proteins bind to a DNA sequence motif that is shared among the ciliary target genes called an x-box. In humans, there are seven RFX transcription factors that regulate transcription for tissue-specific genes in organs like the brain, testis, pancreatic tissues, kidneys, and various immune tissues (Swoboda et al., 2000; De Stasio et al., 2018). In *C. elegans* there is only one RFX transcription factor ortholog, DAF-19, first identified by Peter Swoboda in 2000 (Swoboda et al., 2000). *daf-19*

encodes nested isoforms that have pleiotropic effects when this gene is removed (De Stasio et al., 2018; Figure 7). *daf-19a/b* is localized in non-ciliated neurons, meanwhile *daf-19c* is localized to ciliated neurons and is responsible for regulating ciliogenesis (Swoboda et al., 2000). A mutation in *daf-19* will result in a dauer-constitutive phenotype in *C. elegans*, meaning it will cause the worms to be in a state of dauer without a stressor activating this phenotype (Swoboda et al., 2000). *daf-19c* is able to regulate and control the regulation and specification of cilia in at least IL2 neurons (Swoboda et al., 2000). Functional specialization of one type of cilium in *C. elegans* has been demonstrated to be associated with an isoform of *daf-19, daf-19m*. *daf-19m* provides male specific cilia their sensory properties (Wang et al., 2010). *daf-19m* is only found in male worms and is able to control cilia specialization in a *daf-19c* independent manner in other neurons, PKD. *daf-19m* facilitates gene expression necessary for activating PKD and without disrupting ciliogenesis in IL2 neurons. *daf-19m* is regulated by its internal promoter and enhancer elements that lie within the introns of the larger *daf-19* gene. A mutation of *daf-19, daf-19(n4132),* shows that *daf-19m* is required for male mating and specialization of the PKD sensory neurons in males (Wang et al., 2010).

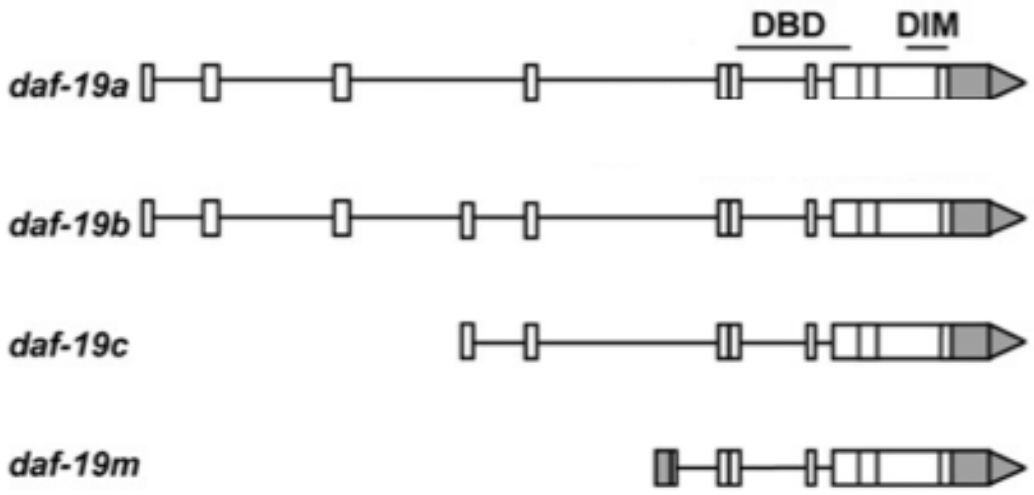

Figure 7. *daf-19* nested isoforms. A comparison of the known *daf-19* isoforms highlighting the similarities and differences in exon use. *daf-19a* is responsible for synaptic maintenance, *daf-19b* function is unknown, *daf-19c* is responsible for ciliogenesis, and *daf-19m* is responsible for male cilia specification. DBD = DNA binding domain; DIM = dimerization domain. Figure modified from (Wang et al., 2010).

Transcriptional regulation

Genes, functional segments of DNA, can be regulated to encode the formation of a functional protein. A gene is comprised of two functional parts, the control region and coding regions, which are typically located upstream and downstream of the transcription start site, respectively (Figure 8). As eukaryotic organisms evolved, gene regulatory mechanisms became more complex than those of prokaryotic cells. The complexity of eukaryotic cells gave rise to the compartmentalization of cellular processes, thereby separating the processes of transcription and translation into two different cellular compartments. The genetic material is encapsulated in the nucleus by a double membrane and is transcribed to produce RNA copies in the nucleus, mRNA then must be transported to the cytoplasm to be translated by ribosomes.

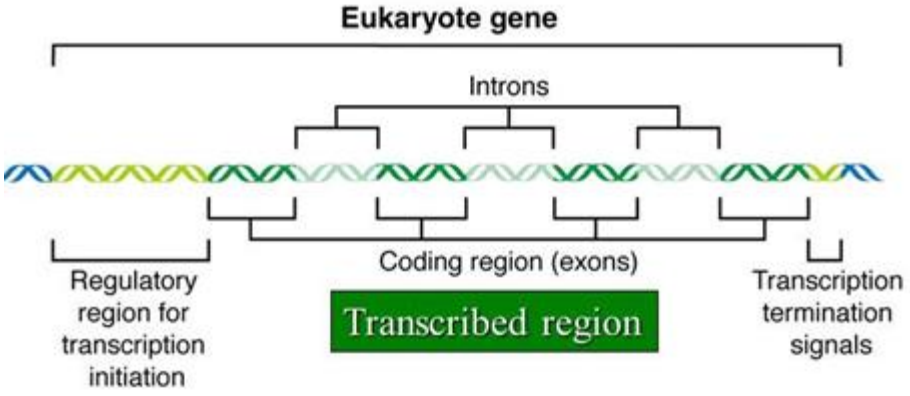

Figure 8. Gene component of eukaryotic DNA. Eukaryotic gene has a regulatory region for transcription to initiate transcribing the coding region and it has a region to terminate transcription. Unique regions of eukaryotes are introns, non-functional segments of the DNA, and exons, functional segments of DNA.

In eukaryotic organisms, the *cis*-acting control regions include the promoter where proteins like RNA polymerase and general transcription factors bind to a specific DNA sequence via hydrogen bonds and electrostatic interactions. General transcription factors are responsible for the activation of a gene by recruiting a RNA polymerase to the promoter. Once RNA polymerase is bound to the control region it will read the DNA sequence in a 3' to 5' direction to identify the start codon, which it will recognize as the starting point for transcription and then will transcribe the coding region of DNA to create ribonucleic acid (RNA). Large genomes with vast numbers of genes are necessary to construct complex, multicellular eukaryotic organisms. Due to the complexity of eukaryotic genomes there are additional regulatory elements that interact with transcription factors to enhance transcription, called enhancers, or to repress transcription, including DNA sequences called silencers.

Enhancer and silencers are *cis*-acting DNA regions found hundred to thousands of bases away from a gene's promoter that indirectly assist the RNA polymerase to binding to the promoter of the gene. In cases where the enhancer or silencer is far away, DNA is able to loop or bend so proteins bound to the enhancer or silencer can interact with general transcription factors

(Figure 9). Activators are specialized transcriptional factors that bind to the enhancer DNA binding site and repressors are DNA binding proteins that bind to the silencer region to decrease the transcription of DNA (Figure 9).

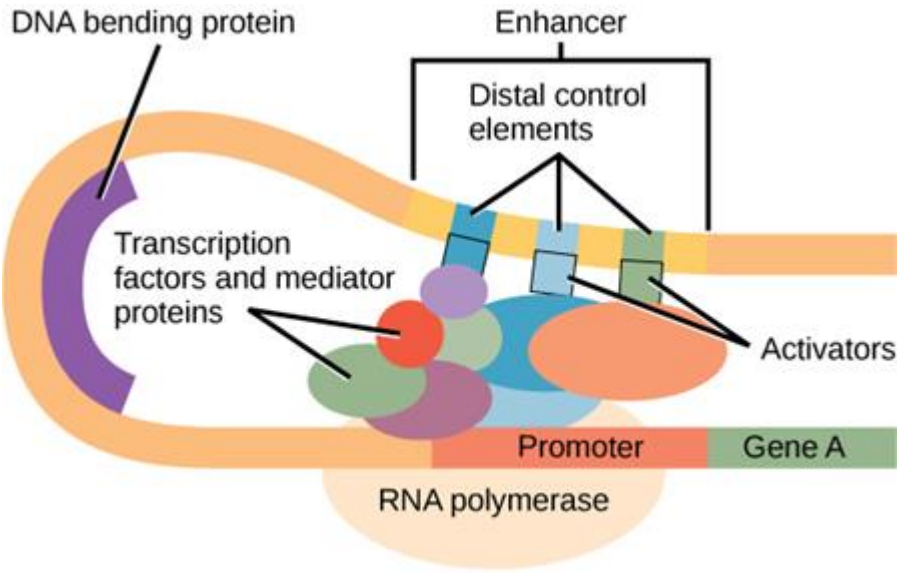

Figure 9. Regulatory proteins in transcription. Multiple proteins, transcription factors, mediator proteins, activators, DNA bending protein, mediate the facilitation of an RNA polymerase binding to the promoter. Enhancer, DNA sequence, interacts with these proteins to promote transcription. Figure adapted from lumenlearning.com.

Transcription is further regulated by specialized transcription factors. These transcription factors can bind to the *cis*-acting control region of specific genes to directly regulate their transcription. They also can bind distal to the promoter and affect the regulation of a gene by indirectly strengthening the binding abilities of other regulatory proteins. Lastly, activators and repressors can directly or indirectly modify histones that in turn alter the coiling of DNA around the histone proteins of the nucleosomes. Transcription factors bind to various overlapping DNA sequences, and they therefore may compete for control region binding (Todeschini et al., 2014).

Most transcription factors belong to large families with conserved protein motifs and various paralogs with similar DNA binding sites. As a response to these challenges eukaryotic

cells have evolved to have a combination of specialized transcription factors and a high cooperativity between proteins to provide a clear discrimination of the control region it's attaching to. Specialized transcription factors such as DAF-19 regulate transcription via several mechanisms. The inability for the DAF-19 mutant protein to regulate cilia-specific genes also leads to a deficiency in the proper assembly of cilia. This deficiency in cilia can be translated to applicable disorders in humans like Polycystic Kidney disease (PKD), Bardet-Biedel syndrome (BBS). PKD is a genetic disorder where the misregulation of cilia leads to the formation of growing malignant cysts in the kidneys. BBS is also a genetic disorder that affects many parts of the body and results in vision loss, obesity, or polydactyly. Disorders like these show the importance of cilia, not only in unicellular *C. reinhardtii*, but also in multicellular eukaryotic organisms including humans.

A group is always stronger than an individual; a phenomenon that can be demonstrated by the actions of humans, wolves, ants, and even transcription factors. With a transcription factor bound to the control region, the signal to recruit different proteins involved in the assembly of RNA polymerase will be disproportionate. A weak signal will lead to mismatches in how the RNA polymerase binds to the control region and will lead to low transcription levels. On the other hand, if multiple transcription factors are bound together they will provide a stronger signal; this behavior is known as transcriptional synergy. Dimerization between transcription factors increases their cooperativity to create a stronger signal in recruiting other proteins to its specific location. This lowers the probability of a transcription factor or other proteins binding randomly to the genome. Discrimination in gene specific transcription factors can be noted in steroids, or other nuclear factors that can discriminate between vitamin D receptors, estrogen receptors, and retinoic receptors (Todeschini et al., 2014).

Current project

For many years the regulation of ciliary genes was a mystery for the scientific community. Interestingly, transcription factors, such as DAF-19, can regulate ciliary genes in multicellular organisms but not in unicellular organisms. For example, *C. reinhardtii* does not contain the genes to encode RFX transcription factors known to regulate ciliogenesis in worms, flies, humans, and other animals (Piasecki et al., 2010). The reason for this is that cilia evolved before RFX transcription factors. A study reports that cilia are present in all eukaryotic supergroups, including the last eukaryotic common ancestor, but RFX transcription factors are only found in animals (Piasecki et al., 2010; Figure 10). Eukaryotes are categorized into five supergroups: Excavata, Chromalveolata, Rhizaria, Archaeplastida, and Unikonta (Piasecki et al., 2010). Within these five groups, *C. reinhardtii* is categorized in the group of Archaeplastida (Figure 10). Cilia are conserved in the eukaryotic lineage dating back approximately 800 million years. The cell's natural selection for this organelle during these millions of years demonstrates the great importance of cilia functions in animals.

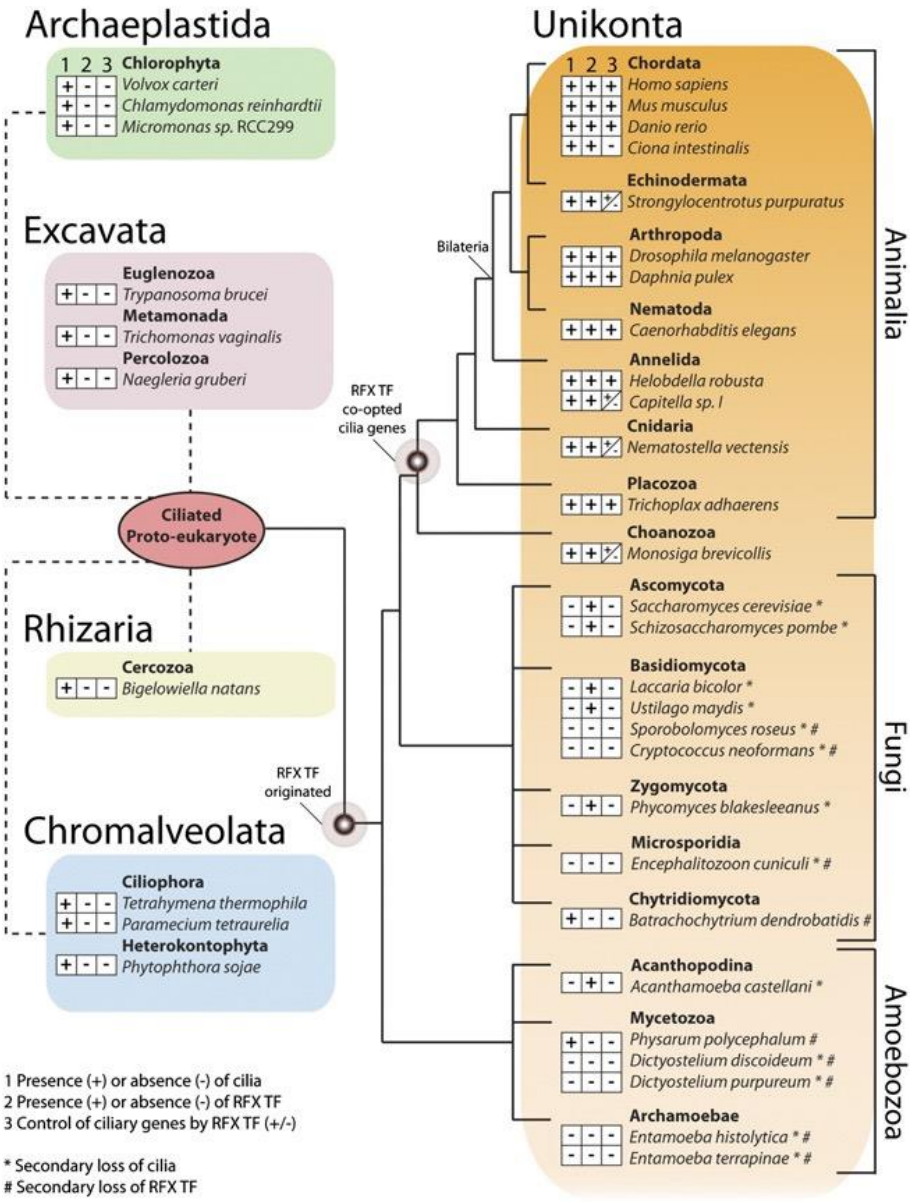

Figure 10. A phylogenetic analysis depicting the presence of cilia and the regulation of RFX transcription factors on ciliary genes, among the five eukaryotic super groups. For each organism, there are three boxes with either + or – signs. First box shows the presence (+) or absence (-) of cilia in an organism. Second Box, shows the presence (+) or absence (-) of RFX transcription factors in an organism. Third Box, shows the presence (+) or absence (-) of x-box promoter motifs, necessary binding sequence for RFX transcription factors for ciliary regulation. Figure legend adapted from Piasecki et al. (2009).

A paramount discovery shows XAP5 is responsible for recruiting RNA polymerase II to genes involved in cilium assembly in *C. reinhardtii* (Li et al., 2018). *xap5* is highly expressed in various human fetal tissues, and XAP5-like protein expression is high during spermatogenesis (Mazzarella et al., 1997; Zhang et al., 2011). In one study, an analysis of the cDNA corresponding to *xap5* showed an abnormal repeating segment of nucleotides, CGG, that is exhibited by some genetic disorders (Mazzarella et al., 1997). This study suggested *xap5* is a candidate disease gene that encodes a DNA binding protein or a transcription factor (Mazzarella et al., 1997). In another study, an RNA-mediated down regulation of XAP5 caused embryonic death in *C. elegans*, which suggests that this protein is important in development (Piano et al., 2002; Tryon and Harmer, 2008). XAP5 is also important in non-ciliated organisms like plants and yeast. A study performed on *Schizosaccharomyces pombe*, a yeast, shows that XAP5 is important in chromatin regulation while in the plant *Arabidopsis thaliana* XAP5 regulates the circadian clock and photomorphogenesis (Tryon and Harmer, 2008; Anver et al., 2014). XAP5 is highly conserved among eukaryotes, though it has diverse biological functions in different organisms.

XAP5 protein's new classification as a transcription factor in *C. reinhardtii* allows for a new understanding of the regulation of ciliary development. Our current research implements reverse genetics; we will be studying Ce*x*ap5 with a presumed loss-of-function allele to eventually determine whether cilia development or function is impaired. *C. reinhardtii* does not have an orthologous gene to encode a ciliary transcription factor found in multicellular organisms. However, *C. elegans* does have an ortholog of CE*x*ap5, which allows us to study whether the function of the XAP5 protein is conserved between unicellular and multicellular organisms.

Due to how recent XAP5 was discovered there is no known information connecting the protein's function between multicellular and unicellular organisms. The data that will be presented will establish a foundation for further investigation in an area that is not well understood. We have obtained a strain of worms, VC40591, that was created by researchers in the reverse genetics core facility at University of British Columbia. This strain contains many mutations, including one in the CE*x*ap5 ortholog. Using a genetic method known as backcrossing, we will be mating the VC40591 worms to a wild type worm in order to isolate the mutation of interest and "clean" the other mutations from its chromosomes. The mutation of interest is a nonsense mutation, which will change the genetic code to encode an early-termination of the protein during synthesis. The result of this mutation will be an improper formation of CE-XAP5.

We would expect a mutation in the orthologous gene, CE*xap5*, to alter the ciliary development causing phenotypic changes in the worm. *C. elegans* that have ciliary deficiency have dwelling, mechanosensation, and chemosensory complications due to their inability to sense external stimuli. There are also dye filling-defective worms that take in lipophilic dye but do not absorb it into their ciliated sensory neurons. The difference can be easily visualized under a microscope; wildtype worms will fluoresce in particular regions of the head and tail while ciliary-deficient worms will not. Multiple assays will be completed to verify the presence of proper cilia development in *C. elegans.* If the evidence presented does not support the data, further experiments will be done to completely rule out the possibility of a function conservation between unicellular and multicellular protein, XAP5.

MATERIALS AND METHODS

Strains:

To assess the role of CExap5**,** we worked with four strains: VC40591, wild type (N2), CB3223 (*che-13(e1805))*, and CB7272, a marker strain. We selected VC40591 because it has a point mutation in the CExap5 gene, named allele *gk709587*, created as a part of the million-mutation project. However, VC40591 possesses multiple other mutations in its genome. For the purpose of this project we needed to remove the other mutations in order to study the relationship between the CExap5 mutation and the behavior of the worm. We implemented a known method, backcrossing, to replace mutant chromosome with wild type chromosomes. The wild type N2 strain has functional genes on all chromosomes, which is why we selected this strain to mate to in our backcrossing and as control in our experiments. The CB3223 (*che-13(e1805))* strain was used as a negative control in our experiments because it has nonfunctional cilia. The wild type and *che-13* strain were good controls to which we could compare mutant strains because we saw how the presence or absence of cilia, respectively, affects the behavior of the worm. CB7272 is the marker strain that has an integrated set of fluorescent markers that are homozygous dominant on each chromosome, with the exception of chromosome III, which has recessive alleles for the gene *dpy-17(e164)* (Figure 11).

A. CB7272 integrated chromosomes:
   - I. A green fluorescent protein, GFP, tag on Chromosome I is expressed in the muscle nuclei in body wall.
   - II. GFP tag on Chromosome II is expressed only in pharynx muscles.
   - III. Chromosome III expresses a short and dumpy phenotype.
   - IV. A red fluorescent tag, DsRed, expressed from Chromosome VI is localized in epidermis.
   - V. Chromosome V has a red fluorescent tag, mCherry, that localizes in pharynx muscles.

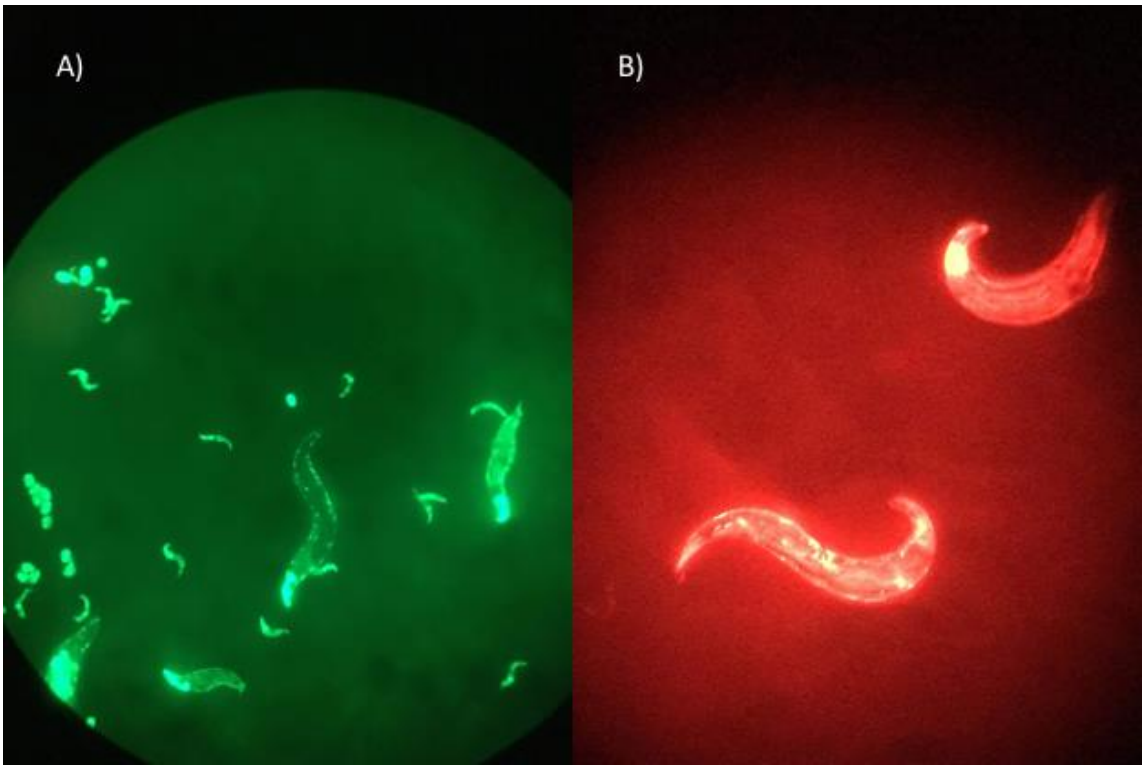

Figure 11. CB7272 demonstrates multiple phenotypes. The phenotypes this strain expresses are a dumpy phenotype (A,B), green fluorescence in pharynx and muscle nuclei (A), and red fluorescence in pharynx and epidermis.

Construction of backcrossed strains

To exchange wild-type chromosomes for the VC40591 chromosomes exposed to mutagens in the million-mutation project, we mated an N2 wild type male to the marker strain, CB7272, to produce fluorescent offspring. We screened for offspring that were male and that displayed all the dominant fluorescent markers. The selected males were then mated with the VC40591 worms, which contain mutant CE*xap5*. This mating ensured the offspring carried the mutant CE*xap5* in chromosome V. The offspring of this mating had the probability of producing half *xap5*/red pharynx and half *xap5*/normal pharynx (from N2 worms) from chromosome V. To differentiate between genotypes, we picked worms that had a red pharynx and put them on

separate agar plates. We predicted that each single worm would yield half of their F2 offspring that would be fluorescent in the pharynx and the other half of the population would be non-fluorescent in the pharynx. Half of the population that was not fluorescent was due to the genotype being homozygous recessive, *xap5*/*xap5*. We continued by selecting homozygous *xap5* worms and mated them with the F1 fluorescent males, offspring of N2 male x CB7272, which would count as the second back cross. We would repeat the same mating to produce more back crosses.

Mutation analysis

We used the database from WormBase (www.wormbase.org) to obtain the gene sequence from the CE*xap5* mutant, *gk709587* allele, and the corresponding wild type gene sequence from chromosome V. We used EMBOSS Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) to compute an optimal global sequence alignment between the CE*xap5*(*gk709587*) and CE*xap5*. The output from this alignment allowed us to identify the position of the nucleotide change in the CE*xap5*(*gk709587)* sequence.

Protein, coding region, and promoter sequence comparison

The *xap5* orthologous protein sequences for *C. elegans* and related nematodes were taken from WormBase. The protein sequence for *H. sapiens* was provided by Ensembl database (https://useast.ensembl.org/index.html), and the *Chlamydomonas*' protein sequence was found in BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The sequences attained were put into Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) to perform a simultaneous multiple sequence alignment.

In the same way, DNA sequences from the coding region were attained and aligned to analyze the similarity between organisms. Additionally, the coding sequences of CE*xap5* and

Chlamydomonas were put into local alignment tool, Water (https://www.ebi.ac.uk/Tools/psa/emboss_water/), to visualize most similar regions.

We used Legacy Gbrowse image (https://www.wormbase.org/tools/genome/gbrowse/c_elegans) to manually select approximately 1000 bp region upstream from the start site of CE*xap5* and its orthologous nematodes. These sequences were presumed to include the promoter sequence and other regulatory sequences. Then, we compared all the predicted promoter sequences using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Motif

DNA and protein sequences were put into MEME (http://meme-suite.org/tools/meme) to locate any known DNA or protein functional sequence motifs. XAP5-like protein sequences and DNA promoter sequences of all organisms in this study were assessed in MEME. MOTIF (https://www.genome.jp/tools/motif/) was used to help identify motifs from the consensus sequence motifs provided by MEME.

Genotyping

We used Benchling (https://benchling.com) to analyze the CE*xap5* and CE*xap5*(*gk709587)* sequences to determine which region we wanted to amplify, or copy, using polymerase chain reaction (PCR). The region selected was influenced by the location of the cut sites of a restriction enzyme. We noticed that the mutation altered one restriction cut site for enzymes BtsCI or FOKI, which allowed us to choose an enzyme that would produce different size DNA fragments between the mutant and wild type gene following PCR amplification and FOKI digest. Wild type CE*xap-5has* three restriction cut sites for BtsCI or FokI and the mutant has two restriction cut sites for BtsCI or FokI in the area of the *gk709587* mutation. Either restriction enzyme can be used, but we selected FokI to digest the DNA in this project. We

designed a forward and reverse primers flanking the mutation in CE*xap5* that, when used in PCR, will produce a 720 bp fragment of DNA. To find the correct temperature at which the primers will attach to the DNA, we used New England Biolabs TM calculator (https://tmcalculator.neb.com/#!/main). Our forward primer contains 18 nucleotides and will anneal to template DNA at 56°C. Our reverse primer contains 24 nucleotides, and anneals at 58°C, thus an annealing temperature of 57°C was used for PCR.

Forward Primer: 5'AAGTACTATTTTCGCTGG 3'
-Guanine to Cytosine ratio of 39%
Reverse Primer: 5'AAAAATTTGAGAACTTTCGAGAAA 3'
-Guanine to Cytosine ratio of 25%
Restriction Enzyme Sites: BtsCI or FokI

Worm lysis

Two different methods were used to lyse, or open, worms and extract their DNA. In our first method, we added 0.5µl of, 20mg/ml, Proteinase K to 100 µl lysis buffer. We took 3µl aliquots from the lysis buffer Proteinase K mixture and added it into the cap of a 0.2ml PCR tube. Then we quickly put 3-4 worms into each cap to prevent the lysis mixture from evaporating. Once worms were put into cap, we centrifuged tubes and stored tubes at -80°C for a minimum of 10 mins. PCR tubes were placed in the thermocycler for 1 hour at 60°C and 95°C for 15 mins.

During the time the lysis was occurring, we prepared a Promega master mix consisting of forward and reverse primer, $MgCl_2$, dNTPs, Taq Polymerase, and buffer. 1µl of lysed DNA was then added to an appropriate amount of master mix aliquot. After centrifuging the mixture, the preset PCR protocol below was initiated in the thermocycler to amplify our desired gene.
For our second method, we acquired a worm lysis kit from Nemametrix to expedite and efficiently amplify CE*xap5* without having to further purify our DNA. We mixed 2 µl of 5x lysis

buffer A and 1µl of 10x lysis buffer B with 7µl of water. We placed the worms and initiated the lysis protocol as above. However, following the lysis protocol we diluted the product by adding 90 µl of PCR grade water. Following this step, we added the master mix and initiated PCR protocol in the same way as the first method.

Lysis protocol:
60.0°C- 10:00 min (Initial Denaturing)
95.0°C- 15:00 min (Denaturing)
8.0 °C - Storage in thermocycler

PCR protocol
95.0°C- 10:00 min (Initial Denaturing)
95.0°C- 00:30 min (Denaturing)
54.6°C- 00:30 min (Annealing)
72.0°C- 00:30 min (Initial Extension)
72.0°C- 10:00 min (Extension)
8.0 °C - Storage in thermocycler

Confocal microscopy preparation:

Agar plates, containing specific worm populations, were tilted at 30 degrees and washed with 1mL of M9 buffer. The fluid from the plate was collected into an Eppendorf tube and centrifuged. Additional washing of the worm pellet is circumstantial depending on the condition of the plates. In the meantime, we melted less than 500 µl of 2% agarose in a 95°C heat block, and added 15µl of 1M sodium azide to eventually paralyze worms when placed into pad. The mixture was added to the center of the imaging slide to form our 'worm pads.' To keep the pads a constant thickness, we modified slides by wrapping tape at the center to provide a fixed height. We set our modified slides adjacent to the sides of the slide we were pouring on. After quickly dispensing about 250 µl of the agarose onto the slide we quickly placed another slide perpendicular on top of all three slides and added gentle pressure to make sure we had an ideal thickness for the agarose pad. To keep the gel from desiccating we placed all slides into a humid

worm pad chamber. Pads were made fresh for each imaging session. We loaded 3µl of the worm pellet, directly onto an agarose pad once we were ready to image.

The microscope used to image *C. elegans* was a Leica TCS SP5 II confocal microscope with LAS AF SP5 II software. We used a 40x objective lens with a numerical aperture of 1.3. The selected immersion medium is oil, 40x/1.3. Settings were formatted to fit a DiI parameter, and we used a 543-laser corresponding to DiI. The selected scanhead for DiI was HyD4 with a spectral range of 653-695 nm. Images were also collected using Differential Interference Contrast (DIC) to see the outline and interior of the worm. This allowed optical modifications to highlight any outstanding contrast in the cellular structure.

Dye filling assay:

This assay was used to stain the amphid (head) and phasmid (tail) neurons of *C. elegans* that are accessible to the dye. The extent of dye-filling was qualitatively assessed using a fluorescent microscope or imaged using the confocal microscope. The dye-fill assay allows us to visually determine whether *C. elegans* has functional neuronal cilia accessible to the environment. For example, when comparing wild type, N2, strain to a deficient cilia strain, *che-13*, there is a noticeable difference between where the dye is localized. For both the worms, dye will appear to be fluorescent in the intestinal tract from worms "drinking" the dye. However, since *che-13* has a cilia deficiency, the ciliated neurons will not be able to absorb the dye, thus fluorescing neither in its head nor its tail. In contrast, wild type has proper ciliated neurons that can absorb the dye which is why we can see its amphid and phasmid neurons fluoresce under a fluorescent microscope. Using both strains, N2 and *che-13*, as controls we can properly compare other strains to assess their cilia function.

First, we washed the worm population from plates as for the confocal preparation. After washing, we mixed 500 microliters of M9 buffer with 5 microliters of the DiI, 2mg/ml, in an Eppendorf tube. We quickly covered the tube with aluminum foil and placed it in the shaker, because the exposure of light on dimethlyformamide can alter its chemical properties. Lastly, worms are transferred from tube to a plate, 8-12 hours before imaging.

Drop assay:

A drop assay was used to assess *C. elegans'* response to water soluble repellents. We selected 10-15 L4 worms from a well-fed plate to put in the center of a non-streaked plate. Worms were undisturbed for 15 minutes, which allowed them to acclimatize to their new environment. After 15 minutes, we applied a chemical to its posterior end, near its tail, just enough so it engulfed the worm to activate its sensory neurons. Applying too much caused the worm to swim in the fluid and inaccessible to test. We modified a capillary tube by heating the end and elongating the tip, which reduced the diameter of the opening to an appropriate size so when chemical was applied it would be enough to engulf worm. Once chemical was properly applied, we waited two minutes to apply the next drop, thereafter, and repeated these seven times for one trial. A total of three trials were recorded for each strain. We recorded whether worms moved forward or backwards after chemical was applied. If the worm responded, backwards or forward, within four seconds then it was measured as a positive response, for the avoidance index. To quantify the behavior of the worm to the chemical we divided the number of positive responses by the total number of trials. The A.I will range from 0-1, lower value showing a weaker response and a higher value showing a stronger response.

A. Repellents used in this assay**:**
   i. 1M glycerol- Low pH avoidance
   ii. M acetic acid- Osmotic avoidance
   iii. Absolute ethanol- Treatment control

Roaming assay:

This experiment demonstrated the type of roaming or dwelling behavior different strains manifested. We single picked individual worms and placed each at the edge of the bacteria lawn of a streaked NGM plate. These worms were removed two hours after they were placed on the plate. We placed the plate on a laminated grid containing 3 mm x 3 mm squares to count squares crossed. The number of squares crossed by a worm track was divided by the number of squares that contained food when recorded.

Wormtracker:

Worm Tracker is a software that uses a camera linked to a microscope to record a single moving worm on a petri dish (http://worm-tracking.sourceforge.net/). For the best quality videos we used low peptone plates (Islam, 2018), and prepared our own bacteria to add to the plates (Ricky's Wormtracker Protocol). For every recording we measured our bacteria in a spectrophotometer to obtain an optimal density between 0.3 to 0.4 in 600 nm before spreading the bacteria on the low peptone agar. Bacteria were allowed to grow on the plate for 45 minutes. Worms selected for imaging were at an L4 stage, and were taken from well-fed populations. The settings for the recording was set at 9000 frames, 15 frames per seconds, resulting in a total of 10-minute videos. After recording, we continued to "batch process" the videos. During this process the program applies a skeleton outline for each worm recorded, and repeats this step for all videos. After the batch processing ended we ensured that the proper skeleton was applied because sometimes the software does not apply the skeleton correctly. We needed to manually adjust the outline of the skeleton, if improperly done by the computer. We proceeded to recompress the videos, which joined the trajectories between the worm and its skeleton. Once the

trajectories were joined it allowed us to insert the compressed files into an R script to analyze our data.

RESULTS

In a recent study, a group of scientists identified XAP5 as an important transcription factor that recruits RNA Polymerase II to genes required for the assembly of cilia in *Chlamydomonas* (Li et al., 2018). Before this finding, the identity of specific transcription factors that regulate ciliary genes in *Chlamydomonas* was a mystery. Although we will not study XAP5 in *Chlamydomonas*, we have identified an orthologous *xap5* in the genome of our model organism, *C. elegans*. The intention of this project is to examine the conservation of XAP5 and its function in *C. elegans*. Our experiments were designed to provide clarity to the role of CE*xap5*.

Bioinformatics

To better understand the structure and the position of CE*xap5* in chromosome V we used Legacy Gbrowse image. Analysis of the structure and position of CE*xap5* revealed that it shares a promoter with its neighboring gene, *daf-21*/hsp-90/C47E8.5. This finding showed the individual promoter was responsible for transcriptionally regulating both divergently transcribed genes (Figure 12). Although both genes are transcribed in opposite directions relative to the chromosome, they conserve the same directionality (5’ to 3’) because they were transcribed from different DNA strands of the double helix. We also noticed that there were multiple isoforms for both genes, but the 5’ untranslated region of the *daf-21*.a isoform stood out because it spans from the start site to the third intron of CE*xap5*, suggesting that control regions for CE*xap5* would lie within the transcribed portion of *daf-21*.

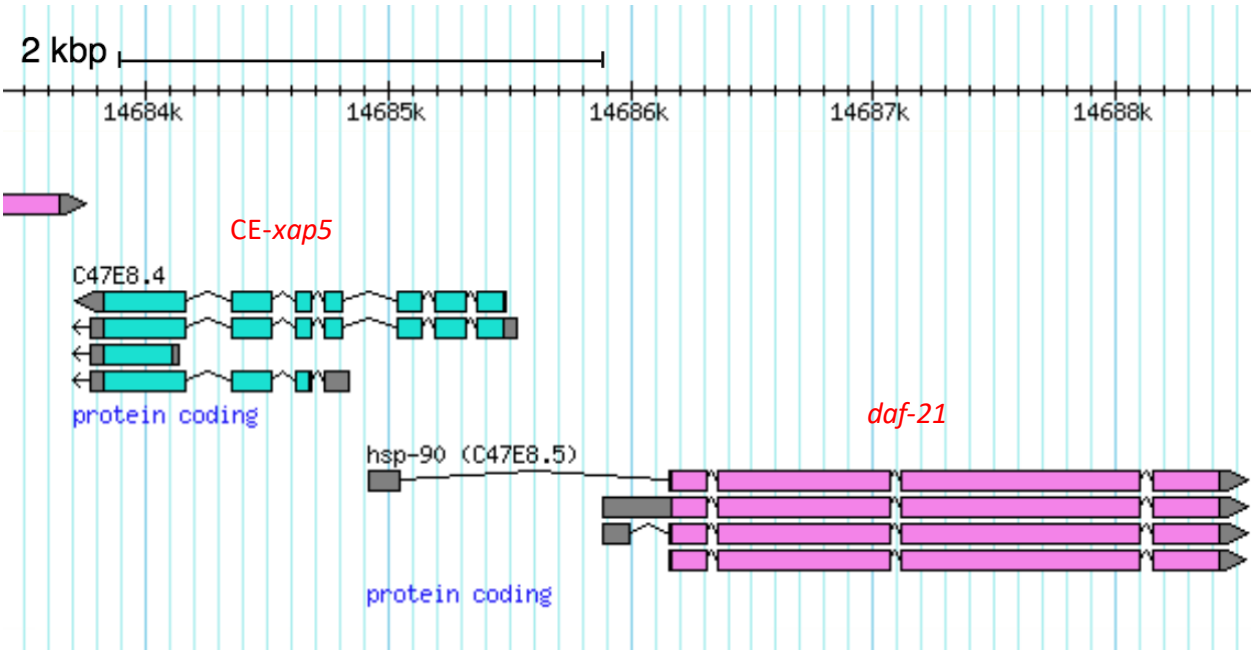

Figure 12. CE*xap5* and *daf-21* are regulated from an single promoter. Legacy Gbrowse image was used to visualize the structure and position of both genes. Arrows at the end of the gene show direction of transcription. Gray boxes at both ends of genes represent the untranslated regions. Turquoise boxes, in CE*xap5*, and pink, *daf-21*, are exons and gaps in-between represent introns.

To demonstrate that there is evolutionary conservation between the *Chlamydomonas* XAP5 and CE-XAP5, we compared the protein, coding, and promoter sequences. Otherwise, future experiments would not have sufficient evidence to predict a similar role of XAP5 in *Chlamydomonas* and *C. elegans*. The results from our bioinformatics analysis indicated that XAP5 is conserved among the species investigated. Between the four *Caenorhabditis* species for which a sequenced genome exists, there was a strong protein similarity. The protein sequence with the best match to *C. elegans* was that of *C. remanei* with 93.87% similarity (Table 1A). At first glance, the percent identities between nematode proteins and that of *Chlamydomonas* appeared as low scores, however, they were sufficient to predict similar function between both organisms. It has been suggested that protein sequence alignments that have a score of 40% or higher identity is the criteria for orthology in the overall sequence alignment (Rost, 1999). Orthology is less certain within the range of 20-35% amino acid identity; this range is known as

the twilight zone. For this reason, many comparisons in the twilight zone are false negatives, which provided us with the comfort to speculate that there is homology between the proteins we are studying.

The percent identity of the coding region (DNA sequence) was slightly lower than that of the amino acid sequence within the worm lineage (Table 1B). This result was not surprising because there are multiple combination of nucleotides (codons) that encode the same amino acid. The highest percent identity shared with the *C. elegans* coding region was that of *C. remanei* with 81.65% identity. The lowest DNA sequence similarity between nematode genes occured in our analysis of the control region (Table 1C), but we still observed a conservation within the promoter sequences. The highest percent similarity to the *C. elegans* CE*xap5* promoter was 57.90% compared to that of *C. remanei.* Thus, in every category the highest similarity occurred between *C. elegans* and *C. remanei* sequences. A phylogeny constructed from a parsimony jackknife analysis using DNA sequences supported the close relatedness between *C .elegans and C. remanei* of the *Caenorhabditis* species investigated in this analysis (Figure 13). This analysis added to our understanding by showing that there is high similarity in protein and coding sequences between organisms, including *Chlamydomonas*, that suggests an evolutionary conservation in regulatory and expression patterns between all organisms.

A. Protein

C. brenneri	100.00					
C. briggsae	94.79	100.00				
C. elegans	92.94	92.33	100.00			
C. remanei	96.01	95.71	93.87	100.00		
Chlamydomonas	35.02	35.60	35.60	35.02	100.00	
H. sapiens	47.19	46.88	47.19	47.50	45.51	100.00

B. Coding Sequence

C. brenneri	100.00					
C. briggsae	81.96	100.00				
C. elegans	79.41	79.51	100.00			
C. remanei	81.55	83.69	81.65	100.00		
Chlamydomonas	45.25	45.97	45.14	45.97	100.00	
H. sapiens	49.94	50.28	50.40	52.20	58.45	100.00

C. Promoter

C. brenneri	100.00			
C. briggsae	54.40	100.00		
C. elegans	49.3	46.80	100.00	
C. remanei	54.70	57.90	50.70	100.00

Table 1. Multiple sequence alignment reveals an evolutionary conservation of protein, coding, and promoter sequences between related organisms. Clustal Omega was used to provide the identity matrix and percent identity between sequences. Gold highlight indicates the sequence with the highest similarity between all organisms and yellow highlight indicates the highest similarity to *C. elegans* sequences.

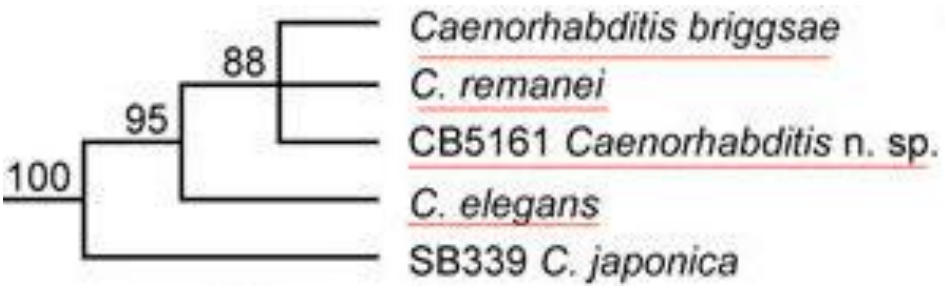

Figure 13. Phylogeny of several *Caenorhabditis* species. Numbers on branches represent percentage of 1,000 jackknife values. (Adapted from Wormbook.org)

To further establish the relationship between XAP5 orthologs, we used MEME software to search for amino acid sequence motifs. When examining the *Caenorhabditis* group, we saw five motifs that were positioned in the same region of the protein and were nearly identical; the size of the amino acid abbreviations in Figure 14 indicates conservation of sequence (Figure 14). The motifs found in *H. sapiens* and *Chlamydomonas* match only some motifs found in *Caenorhabditis* family. In *H. sapiens* the second motif was not present, but the other four were in the same order and were very similar in sequence. *Chlamydomonas* shared the last three motifs, in the carboxy half of the protein, with all five species. We noticed that the *H. sapiens* and *Chlamydomonas* sequences are longer at the amino end of the protein compared to those of the *Caenorhabditis* family. Furthermore, the data from the consensus sequences revealed that there was small variability in amino residues in most consensus sequences (Fig. 15). The sequence logo visually represented the variability and conservation of the aligned protein (or DNA) sequences by displaying the frequency with which each amino acid was found at a particular position within the motif. An E-value was also provided from the outcome, which represented the probability of a sequence alignment occurring by chance. We saw that the appearance of all of the motifs were statistically significant and their E-value ranged from 1.5E-147 to 2.6E-073 (Figure 15). The statistical significance suggested these motifs were well-conserved and were not detected by chance. From these sequence logos, we can also see that there were predominantly 1-3 amino acids that were highly repeated in the motifs. The second and last consensus sequences exhibit a high frequency of amino acid conservation across the motif, as indicated by the single amino acid abbreviation in most positions (Figure 15 b,e).

With the information gathered from MEME, we used a motif scanner to find similar domains from proteins in databases. MotifFinder identifies proteins that share domains with the query protein or domain. The program showed that most known protein consensus sequences had a low E-Value when compared to a XAP5 domain (Appendix A). From this we knew that the motif sequences we obtained were unique to XAP5 in their structure because they are not conserved in other proteins. For example, when we input [CTAENLMFVKEDLIIPHFYTFQDFIVTKAMGKSGPLFNFDVHDDVRLLSD] the program provided outputs for dihydrolipoyllysine-residue acetyltransferase component, but with an E-value of only 0.85. Due to a moderate E-value we did not find any reason to further analyze any of the other domains that were part of the output.

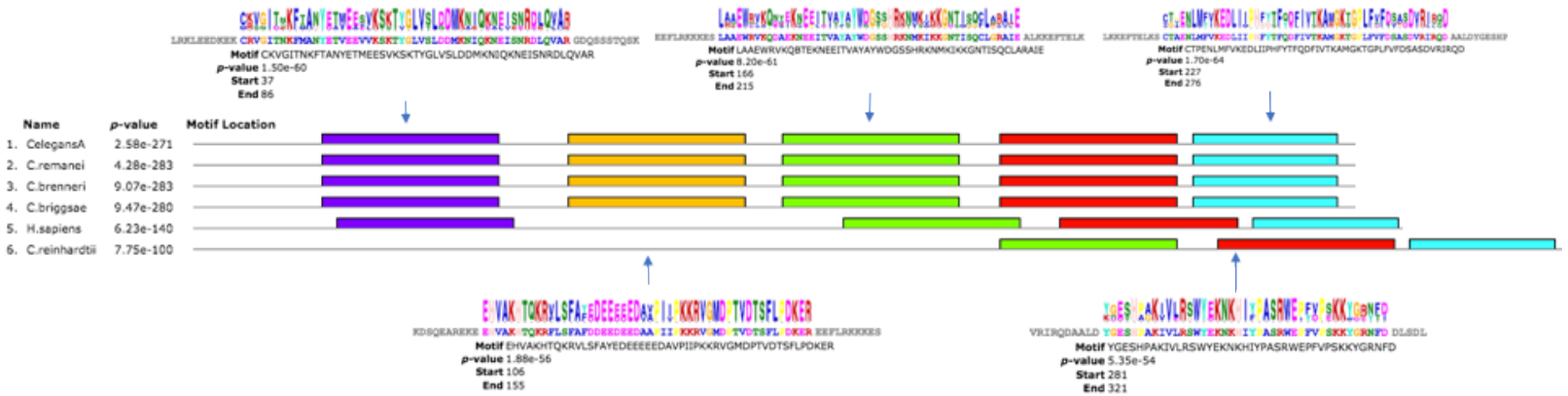

Figure 14. Conserved motifs in XAP5 protein sequences. Protein sequences were put into MEME and the output is as shown. Different color boxes show different motifs conserved in each species. Above each colored box is the consensus sequence of each motif, with its statistical data. P-values indicate the probability of obtaining this motif from the data by chance. All nematodes share the same motifs in identical positions within the protein. In humans and in *Chlamydomonas*, three of the five motifs are present in the same relative order and spacing. Humans are missing the second motif and *Chlamydomonas* is missing the first two motifs.

Figure 15. Protein consensus sequence logos. For each motif found in MEME, a consensus sequence is calculated. These graphs represent the amino acid position on the x-axis and the y-axis depicts the information of amino acid residue in bits. The frequency of the amino acid for the position at a given position is reflected by the height of the amino acid abbreviation. Colors are used to differentiate between categories of amino acids. For example, burgundy letters represent positively charged amino acids.

A nonsense mutation was identified in CE*xap5* (*gk709587*) by comparing the sequence of CE*xap5* (*gk709587*) and CE*xap5*. A cytosine is replaced with a thymine, thus leading to a premature stop codon at the amino acid in the 189th position. This mutation would prevent the complete translation of XAP5 mRNA and therefore stop protein production. In *C. elegans*, nonsense mediated mRNA decay will remove mRNAs with pre-mature stop codons, thus it is reasonable to consider the *gk709587* to be a null allele. There was a clear distinction in the protein sequence between CE*xap5* and CE*xap5* (*gk709587*); in the mutant the sequence stopped at a tryptophan residue, but the in wild type allele the sequence was continuous after the tryptophan residue (Figure 16 a,b).

(A)> wild type C47E8.4a
MSRADEGRLIHLAKKREREKEDIEQQLRKLEEDKEKCRVGITNKFMANYETVEEVVKSKTYGLVSLDDMK
NIQKNEISNRDLQVARGDQSSSTQSKDSQEAREKEEHVAKHTQKRFLSFAFDDEEDEEDAAPIIPKKRVGM
DPTVDTSFLPDKEREEFLRKKKESLAAEWRVKQDAEKNEEITVAYAYWDGSSHRKNMKIKKGNTISQCLG
RAIEALKKEFTELKSCTAENLMFVKEDLIIPHFYTFQDFIVTKAMGKTGPLFVFDSASDVRIRQDAALDYGE
SHPAKIVLRSWYEKNKHIYPASRWEPFVPSKKYGRNFDDLSDL

(B) > mutant C47E8.4a (WBVar01069155: W to opal stop (189))
MSRADEGRLIHLAKKREREKEDIEQQLRKLEEDKEKCRVGITNKFMANYETVEEVVKSKTYGLVSLDDMK
NIQKNEISNRDLQVARGDQSSSTQSKDSQEAREKEEHVAKHTQKRFLSFAFDDEEDEEDAAPIIPKKRVGM
DPTVDTSFLPDKEREEFLRKKKESLAAEWRVKQDAEKNEEITVAYAYW

(C) wild type, with 250 bp flanks
aagtactattttcgctggatgagattctccatagtcaagtgcagcatcttggcgtatacgaacatcggaagccgaatcga
acacgaacagcggaccagtctttcccattgcttttgtgactataaagtcttggaacgtgtagaaatgtggtatgatgaga
tcctctttgacgaacatcagattctcggctgtgcacgatttaagctccgtgaattccttttcagagcctgaaaatgaat
atttataaatatttttgatttgttgaaaaaaaaaaacaattaaaaaatattttagaggggaagtaaagtttgtgggttt
ttgcccgaaacggtcgaatttcataaattgccttctcaaaaaatttccaaaaaaataaatacatagaacacttttttttt
attttctgaagttcattattacctcaatggctcttcctagacattgcgaaatcgtgtttccttttcttgattttcatattt
ttccgatgtgaagatccatcCcaataggcgtaggcaactgtaatttcctcgtttttctcggcatcttgcttcacacgcca
ctcggctgccaatgattctttctttctgaaaccagacttatattttttatttatttgttcaccgggatgatctacaccc
gccaaaatttgaaatgtcgttgtttttttcttctagaactcactttctcaagaactcctctcgttctttatcaggcaggaa
gcttgtgtcgacagttgggtccattccaacgcgctgcaaacatttttcgctttaatttttttctttttaattattattctt
acctttttcgggataattggtgcagcatcttcttcgtcctcttcgtcatcaaatgcaaacgataaaaatcgcttctgaaa
ttagcaatattaaaacagtttttctcgaatgttctacaaaaatgttttatatctgcgtttcctcacactctcacctgccc tcaatttttctttctcgaaagttctcaaatttttagaaatc

(D) >*CExap5(gk709587)* with 250 bp flanks
aagtactattttcgctggatgagattctccatagtcaagtgcagcatcttggcgtatacgaacatcggaagccgaatcga
acacgaacagcggaccagtctttcccattgcttttgtgactataaagtcttggaacgtgtagaaatgtggtatgatgaga
tcctctttgacgaacatcagattctcggctgtgcacgatttaagctccgtgaattccttttcagagcctgaaaatgaat
atttataaatatttttgatttgttgaaaaaaaaaaacaattaaaaaatattttagaggggaagtaaagtttgtgggttt
ttgcccgaaacggtcgaatttcataaattgccttctcaaaaaatttccaaaaaaataaatacatagaacacttttttttt
attttctgaagttcattattacctcaatggctcttcctagacattgcgaaatcgtgtttccttttcttgattttcatattt
ttccgatgtgaagatccatcTcaataggcgtaggcaactgtaatttcctcgtttttctcggcatcttgcttcacacgcca
ctcggctgccaatgattctttctttctgaaaccagacttatattttttatttatttgttcaccgggatgatctacaccc
gccaaaatttgaaatgtcgttgtttttttcttctagaactcactttctcaagaactcctctcgttctttatcaggcaggaa
gcttgtgtcgacagttgggtccattccaacgcgctgcaaacatttttcgctttaatttttttctttttaattattattctt
acctttttcgggataattggtgcagcatcttcttcgtcctcttcgtcatcaaatgcaaacgataaaaatcgcttctgaaa
ttagcaatattaaaacagtttttctcgaatgttctacaaaaatgttttatatctgcgtttcctcacactctcacctgccc tcaatttttctttctcgaaagttctcaaatttttagaaatc

Figure 16. Mutation in CE*xap5* (*gk709587*) contains a nonsense mutation in its DNA sequence, which encodes for a premature stop codon. In our alignment, the length of both sequences are identical, however, CE*xap5* (*gk709587*) has a cytosine replaced by thymine. CE*xap5* protein sequence (A) and CE*xap5* (*gk709587*) protein sequence (B) annotations were adapted from Wormbase.com. CE*xap5* sequence (C) and CE*xap5* (*gk709587*) DNA sequence(D) annotations were adapted from Wormbase.com. The yellow in DNA sequence shows the flanking sequence. The red in protein sequence shows the site of the mutation.

Behavioral Assays

To measure whether there was an alteration in the roaming or dwelling behavior phenotype conferred by mutations in the in VC40591 strain, we compared the behavior of VC40591 worms to wild type and CB3223 (*che-13)* worms. The results from our experiment suggested there was a significant difference between the percentage of bacterial lawn which VC40591 and *che-13* worms explored when compared to wild type worms (Figure 17). The wild type worms explored an average of 33% of the total area of food in two hours. However, VC40591 worms roamed over only about half as much as wild type worms in the same time frame with only 16% of the food area explored ($P<0.0001$). Initially our data showed that *che-13* worms explored, on average, only 7% of the area of the food. The results provided from this strain are misleading, because after examining multiple plates we found multiple faint tracks that were typically found on the portions of the agar lacking food. Our *che-13* worms roamed off the bacteria lawn and were found near the edges of the plates. This type of behavior did not occur with wild type or VC40591 worms were always found in the bacteria lawn. Overall, this study helped us understand that there was a dwelling behavior phenotype in VC40591. Furthermore, it provided us the sufficient evidence to assess the effect of CE*xap5* mutations on neuronal function. To do so, we needed to backcross the VC40591 strain to remove unwanted mutations.

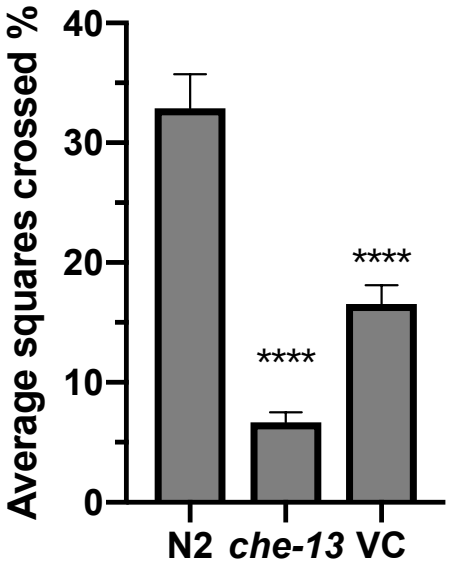

Figure 17. Roaming behavior of N2, *che-13,* and VC40591 worms. Single worms were placed near the bacteria lawn on a plate with a diameter of 5 cm, and were removed after two hours. The percent of z x z mm squares crossed by worm tracks/squares containing food is plotted. Error bars = + SEM; an average of n=36 for each strain. ****P<0.0001 compared to N2.

The assessment of the backcrossed strains showed a behavioral pattern similar to that of VC40591. The backcrossed strains exhibited a significant difference in the percentage of squares crossed compared to the roaming behavior of wild type worms. Regardless of a similar behavior, there was a higher percent of squares crossed, between 22-24%, by backcrossed worms than the VC40591 strain (Figure 18). There were three backcross strains, 1A, 2A, and 2D, that did not show a difference in behavior relative to the wild type, but showed a significant difference compared to VC40591. Two of the three strains, 2A and 2D, were expected to act differently from the rest, because we constructed this strain without any chromosomal markers in their genome. In contrast, strains 1B, 1E, 2B, and 2E all lacked the red pharynx marker from chromosome V suggesting that *CExap-5* is homozygous on chromosome V, and these strains exhibited green fluorescence in the cell body nuclei (marker for chromosome I).

We were surprised to see that backcrossed strain 1A did not roam differently than the wild type, even though this strain exhibited no fluorescence in the pharynx and had the same marker

chromosome as the rest of the backcrossed strains. Strain 1A, on average, crossed 27% of the food area, which fell in the range squares crossed by 2A and 2D, 26% and 29%, respectively. The average squares crossed decreased significantly in strains 1B, 1E, 2B, and 2E. On average, a worm in strain 1B and 1E crossed about 24% of the bacteria lawn while 2B and 2E only crossed approximately 22% of the bacteria covering the plate (Figure 19). This experiment showed that there was some over-dwelling/under-roaming behavior occurring in four of our twice-backcrossed strains. We followed this experiment with a chemotaxis assay to amplify our understanding of the neurobiological underpinnings that occurred in the dwelling/roaming phenotypes.

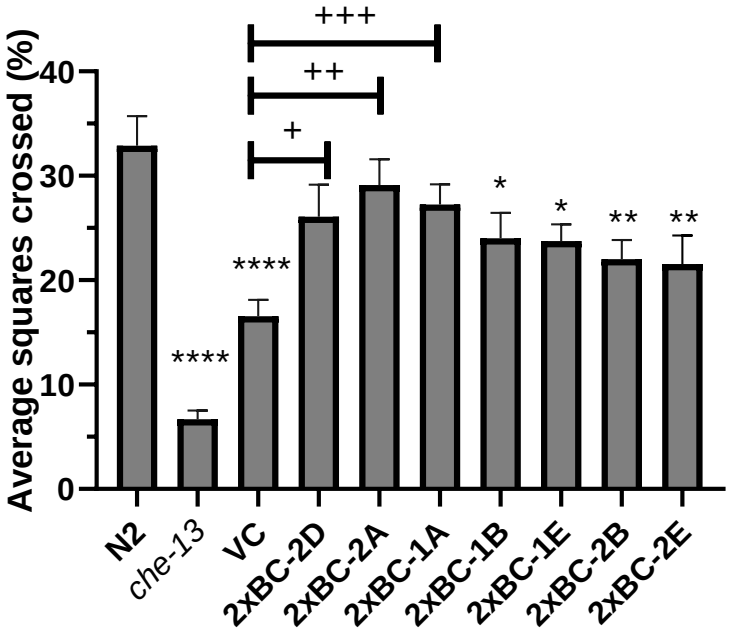

Figure 18. Roaming behavior of twice-backcrossed strains and controls. Single worms were placed near the bacteria lawn on a plate with a diameter of 5 cm, and were removed after two hours. The percent of 3 x 3 mm squares crossed by worm tracks/squares containing food is plotted. Strains 1E, 1B, 2E, 2B roam significantly less than N2 worms. Error bars = + SEM; n=29 average living worms per strain. *$P<0.05$, **$P<0.001$, ****P<0.0001 compared to N2 and $^{+}P<0.05$, $^{++}P<0.001$, *and* $^{+++}P<0.0001$ compared to VC40591.

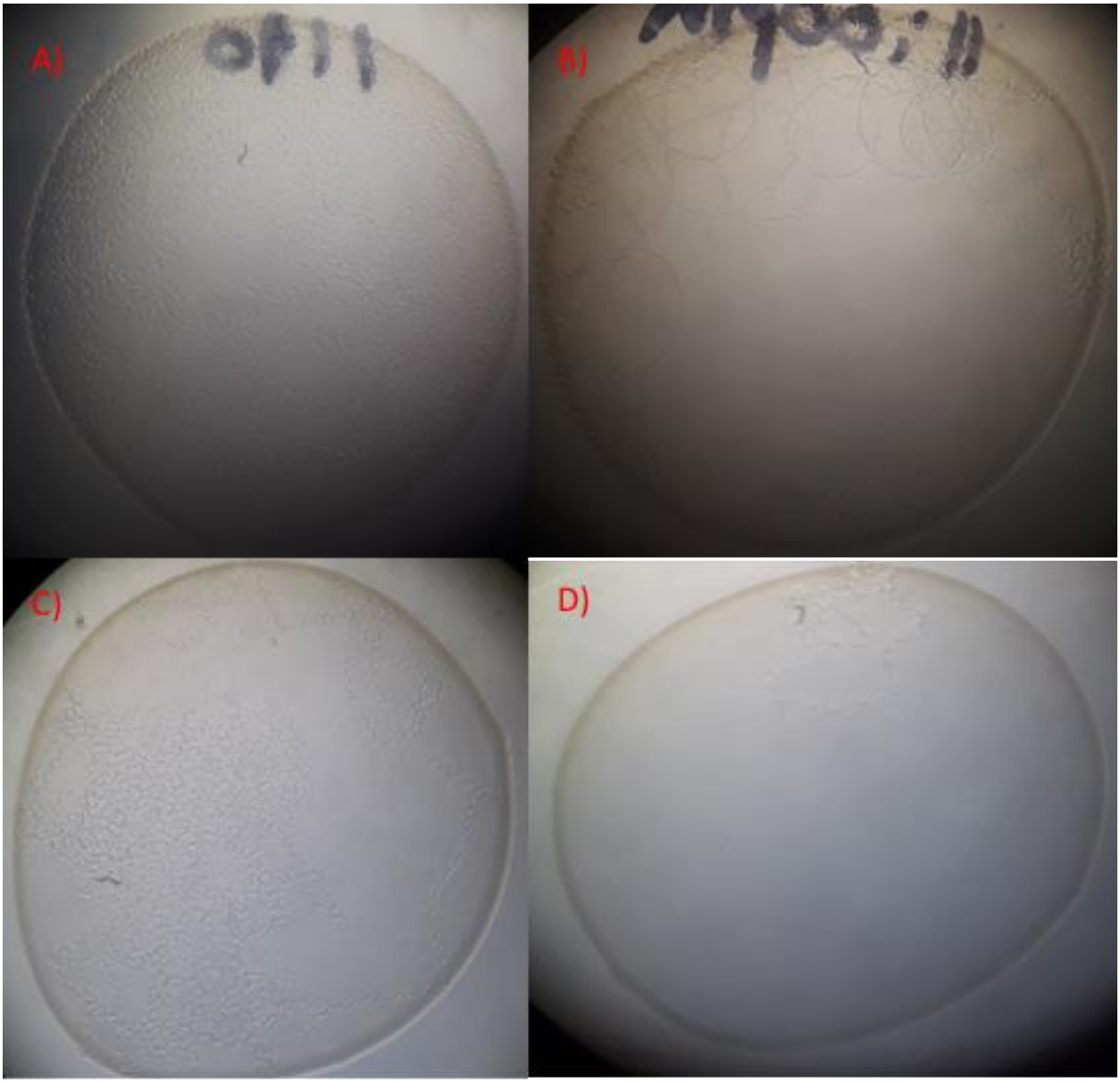

Figure 19. Images of roaming in young L4 worms after 2 hours. Wild type roams the bacteria lawn at a higher rate (A) compared to VC40591 (B) and backcrossed strains 1A (C) and 1E (D).

Chemotaxis Assay:

To determine whether the roaming defect observed in our backcrossed strains was due to a sensory defect, we assessed the worms' ability to detect and avoid low pH or high osmotic pressure in their environment. We used a pipette to place a small drop of repellent near the tail end of the worm to activate its sensory organs, which allowed us to measure its avoidance index (A.I). We used Ethanol as our treatment control because its avoidance effects in worms is well-described (Davies et al., 2004). Analysis of our data using a two-way ANOVA revealed that there was no significant difference between the treatments, ethanol and acetic acid ($F(1, 39)= 1.5$, $p = 0.2250$ or ethanol and glycerol ($F (1 , 39)= 1.492$, $p= 0.2292$). On average, the treatment of ethanol on the wild type worms showed an A.I of 0.73, but A.I decreased to 0.64 when treatment change to acetic acid (Figure 20a). We saw the opposite effect occur when we treated the wild type worms with glycerol (Figure 20b). The *che-13*'s A.I was very low, which showed that this strain of worms does not seem to sense ethanol, glycerol or acetic acid. These results confirmed that our assay procedures were working as we expected. We used the comparison between ethanol and the two other treatments to compare the A.I behavior against our backcrossed strains tested under acetic acid and glycerol.

The application of acetic acid on our backcrossed mutants showed an increase in avoidance behavior. A one-way ANOVA revealed that that acetic acid had a significant effect [$F (4,25) =4.36$, $p=0.0082$] on A.I. Bonferroni's multiple comparison test revealed that there was a significant difference in *che-14* and the 1E backcrossed strain relative to N2 wild type worms ($p=0.0094$ and $p=0.0264$, respectively). Strain 1E showed an A.I of 0.31, which revealed that when acetic acid was applied these worms were not repelled by the chemical (Figure 20b). We analyzed other backcrossed mutants and saw that strain 1A exhibited a high avoidance index,

0.5714, which indicated that it was behaving similarly to the wild type. This result comports with the wild type roaming behavior of this strain.

There is a decrease in the A.I of all backcrossed strains after the application of glycerol. A one-way ANOVA revealed that that acetic acid had a significant effect [$F(4,24) = 11.79$, $p<0.0001$] on A.I. Wild type worms strongly avoided glycerol, with an average A.I. of 0.9048 (Figure 21). In contrast, *che-13* worms avoided ethanol and glycerol similarly, with an average avoidance index of 0.17 and 0.19, respectively. Bonferroni's multiple comparison test revealed that there was a significant difference in *che-14* and the 1B backcrossed strain relative to N2wild type worms ($p<0.0001$ and $p=0.0010$, respectively). Two of the backcrossed mutant strains avoided glycerol at the same rate as did wild-type worms. Specifically, strain 1A had an average avoidance index of 0.6571 and the A.I. of strain 1E was 0.7381 (Figure 21). This finding came as a surprise to us because in the acetic acid assay and roaming/dwelling experiment, strain 1E behaved very similarly to strain 1B.

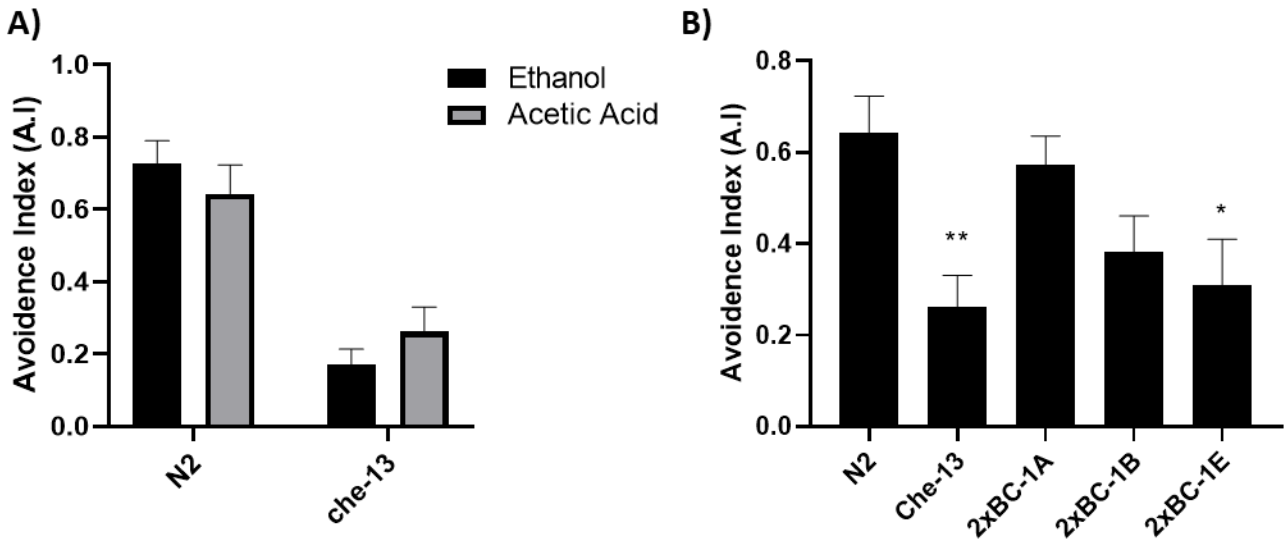

Figure 20. Low pH treatment alters the avoidance response in mutants. (A) Acetic acid induces similar effects in wild type worms and *che-13* worms. (B) A drop of 0.1 M acetic acid was applied to a worm tail every two minutes and their forward or backward movement was recorded. Backcrossed mutant, 1E, have a significantly reduced response to acetic acid compared to N2 worms. In acetic acid, our negative control (*che-13* worms) shows a higher avoidance response to the test treatment. Avoidance index is expressed as the number of avoidance responses, backward movements, by the number of trials (drops delivered) Error bars = +SEM; n=10-15/set. *P<0.05 and ***P<0.01*, compared to the wild type (N2).

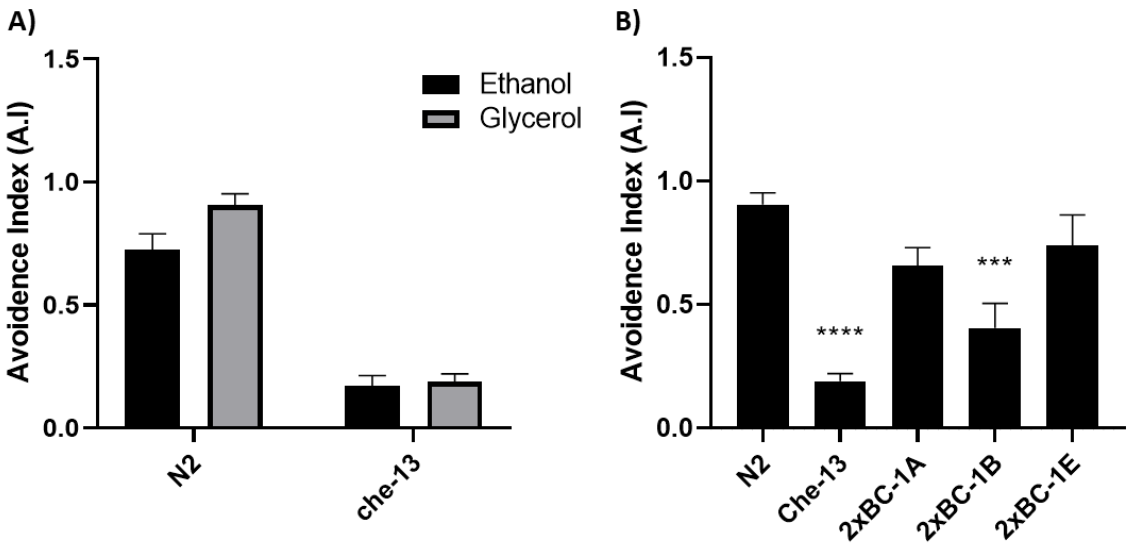

Figure 21. Glycerol does significantly changes an osmotic avoidance response in mutant 1B. (A) Glycerol induces similar effects in wild type worms and *che-13* worms. (B) A drop of 1M glycerol was applied to a worm tail every two minutes and their forward or backward movement was recorded. Backcrossed mutant 1B have a significantly reduced response to glycerol. Positive control shows a lower avoidance when treated with control, ethanol, and then when the test treatment is applied. Our negative control shows a significantly reduced response to both the control and the test treatment. Avoidance index is expressed as the number of avoidance responses, backward movements, by the number of trials (drops delivered) Error bars = +SEM; n=10-15/set. ***P<.001 and ****P<0.0001 compared to wild type (N2).

Wormtracker

To evaluate the movement of our worm strains we used the Wormtracker. Multiple variables were measured using this software. However, the software presents dwelling data in a misleading way, because the software confinements and limitations measure the "paralysis" of the worm, assessing the worm's ability to move. The computer records the worm's movement under a grid, micron size, after the worm moves. Taking the limitation to this assay, we noticed that, on average, strains 1B, 1E, and *che-13* behave similarly (Figure 22). The time these three strains were paused was on overage 18.72 seconds. The data points for the wild type strain seemed counterintuitive at first, because it seemed like the wild type worm was not moving. However, it was not pausing enough because it spent its time roaming and actively exploring the

bacteria lawn. Strain VC40591 showed a lot of variability, on average it did not move for 32.11 seconds. This supported our dwelling assay, because this strain is not as motile. VC40591 showed a significant difference compared to strains 1E, N2, and *che-13*. This shows us that although 1B looked very similar to 1E there is no significantly difference to VC40591.

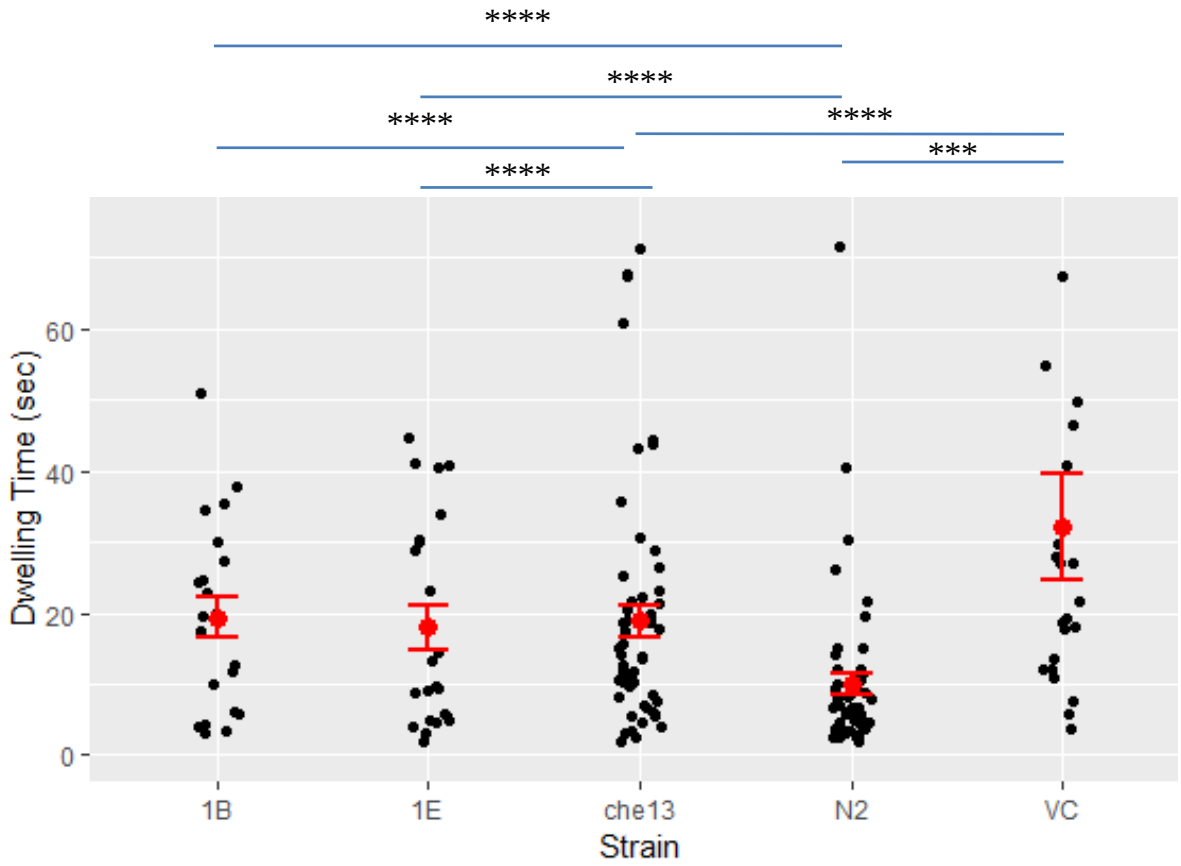

Figure 22. Dwelling time of twice-backcrossed strains. WormTracker was used to compare the difference in how frequently each worm moved within a unit square whose diameter equaled the average width of the worm. Each worm was tracked for a total of 10 mins. A scatter plot represents the distribution of the data; error bars reflect the standard error. (***P<.001 and ****P<0.0001 relative to *che-13*, wild type, and VC40591)

To measure any abnormalities in the bending movement of the different types of strains the software calculates the number of bends of the skeleton of the worm. Qualitatively looking at the graph, we observe that there was an obvious trend where the backcrossed strains show fewer bend counts than do *che-13*, N2, and VC40591 (Figure 23). Strains 1B and 1E display worm

bending about 3.27 times, while VC40591 had the highest amount of bends with 3.72 bends. The mutants' movement was significantly different from all the other strains. This is an interesting variable to measure because under the microscope it did not seem like the mutants had any problems with their bending. However, we did notice from early in this project that VC40591 did have certain abnormal bend movement and now we are able to refute that initial observation with this data set.

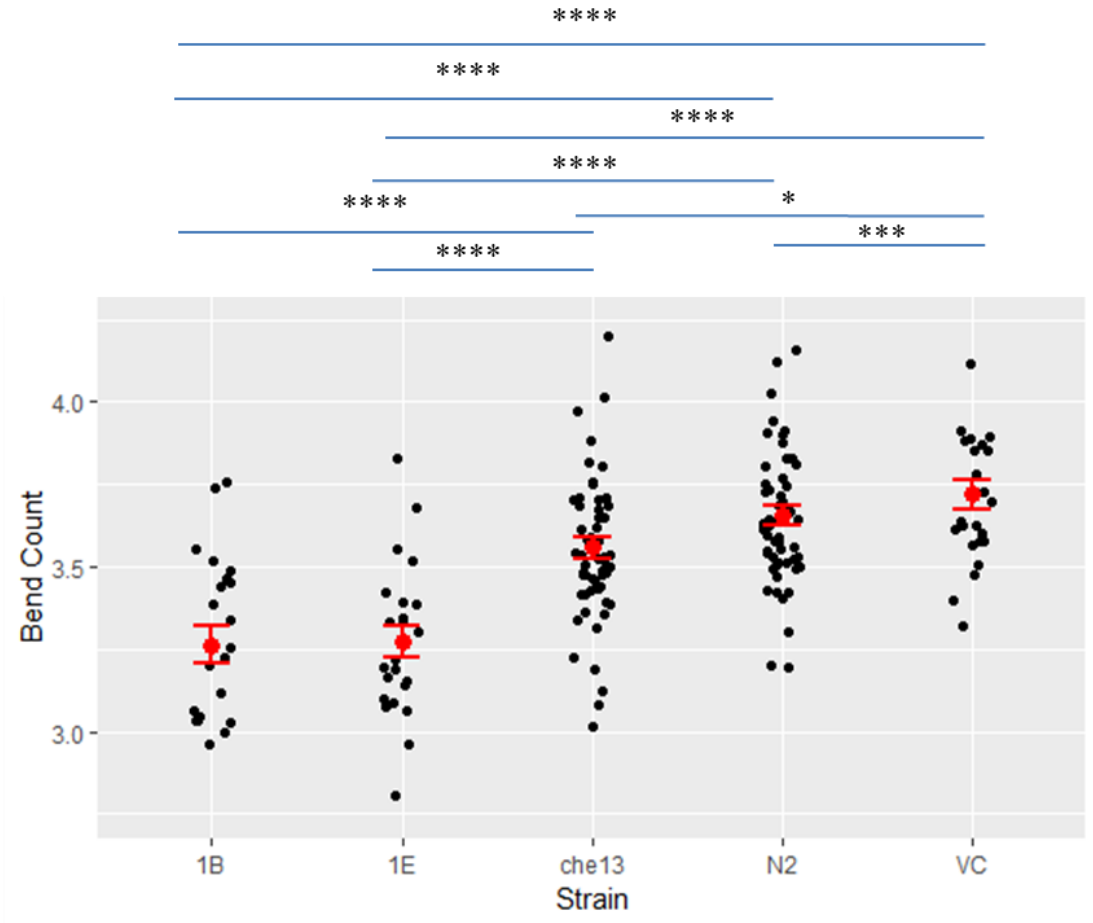

Figure 23. Bend count of twice-backcrossed strains. WormTracker was used to compare the difference in the bend movement of each worm. Each worm was tracked for a total of 10 mins. A scatter plot represents the distribution of the data; error bars reflect the standard error. (*P<0.05, ***P<.001, and ****P<0.0001 relative to che-13, wild type, and VC40591

The WormTracker records the foraging speed by measuring the angular speed of the nose of each worm. There was not a notable difference between any of the strains tested (Figure 24). We did see that there was a lot of variability in the data for all the strains. There was a significant difference in the foraging behavior between the backcrossed mutants, 1E and 1B. This shows that the backcrossed strains continue to show similar phenotypic behavior without any similarity between any of the other strains.

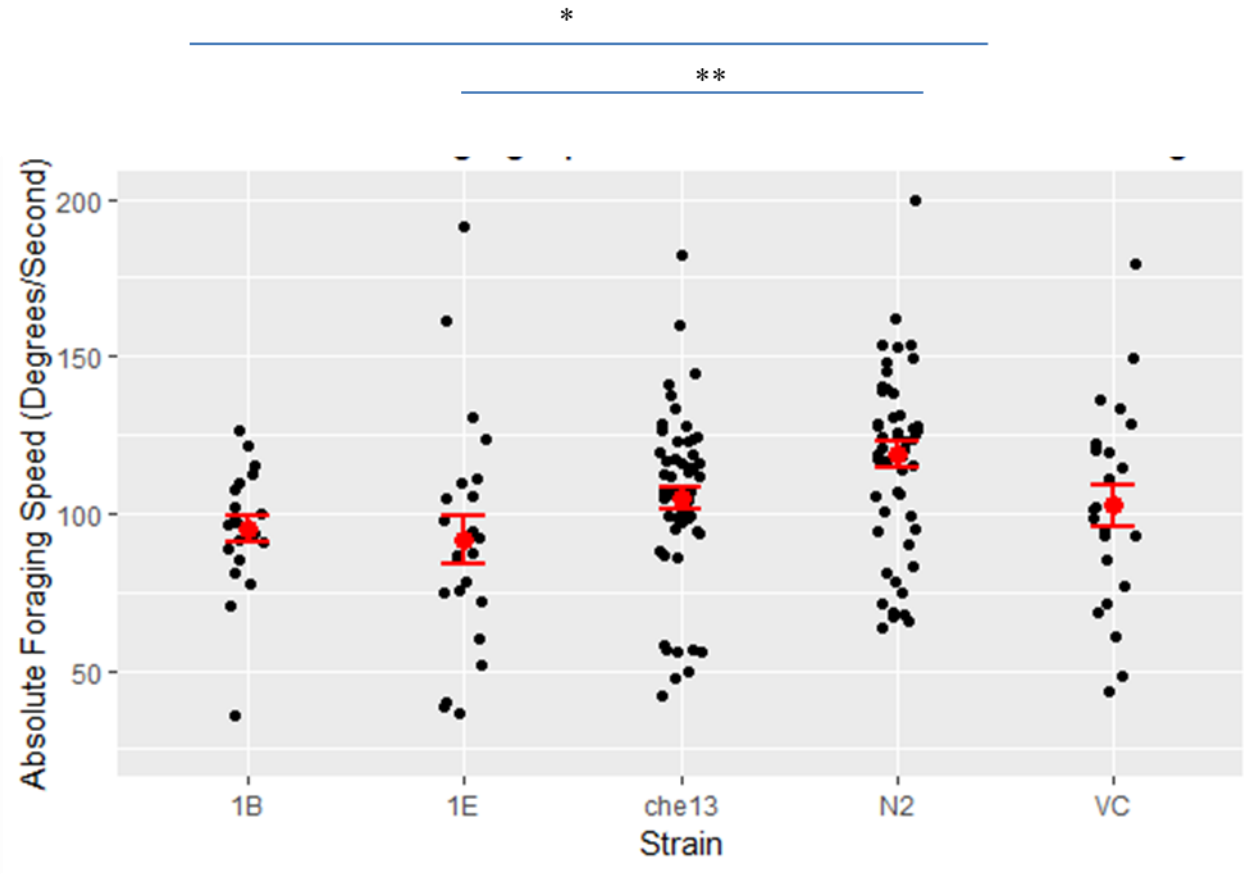

Figure 24. Foraging speed of twice-backcrossed strains. WormTracker was used to compare the difference in foraging speed of each worm. Each worm was tracked for a total of 10 mins. A scatter plot represent the distribution of the data; error bars reflect the standard error. (*P<.05 and **P<0.001, relative to wild type)

We also measured the amplitude of the worm bends using the tracking software. Due to the decrease in bend counts shown in Figure 23, it makes sense that the mutants would have an increase in amplitude compared to the rest of the strains. Just like a shorter wavelength would have a higher frequencies and higher amplitudes, a shorter bend would produce a higher amplitude. However, we would have expected that VC40591 would follow the same pattern and have a decrease in its amplitude because it had a higher bend count, but for some reason the VC40591 strain also had an average an amplitude of 20 microns, similar to strains 1B and 1E (Figure 25).

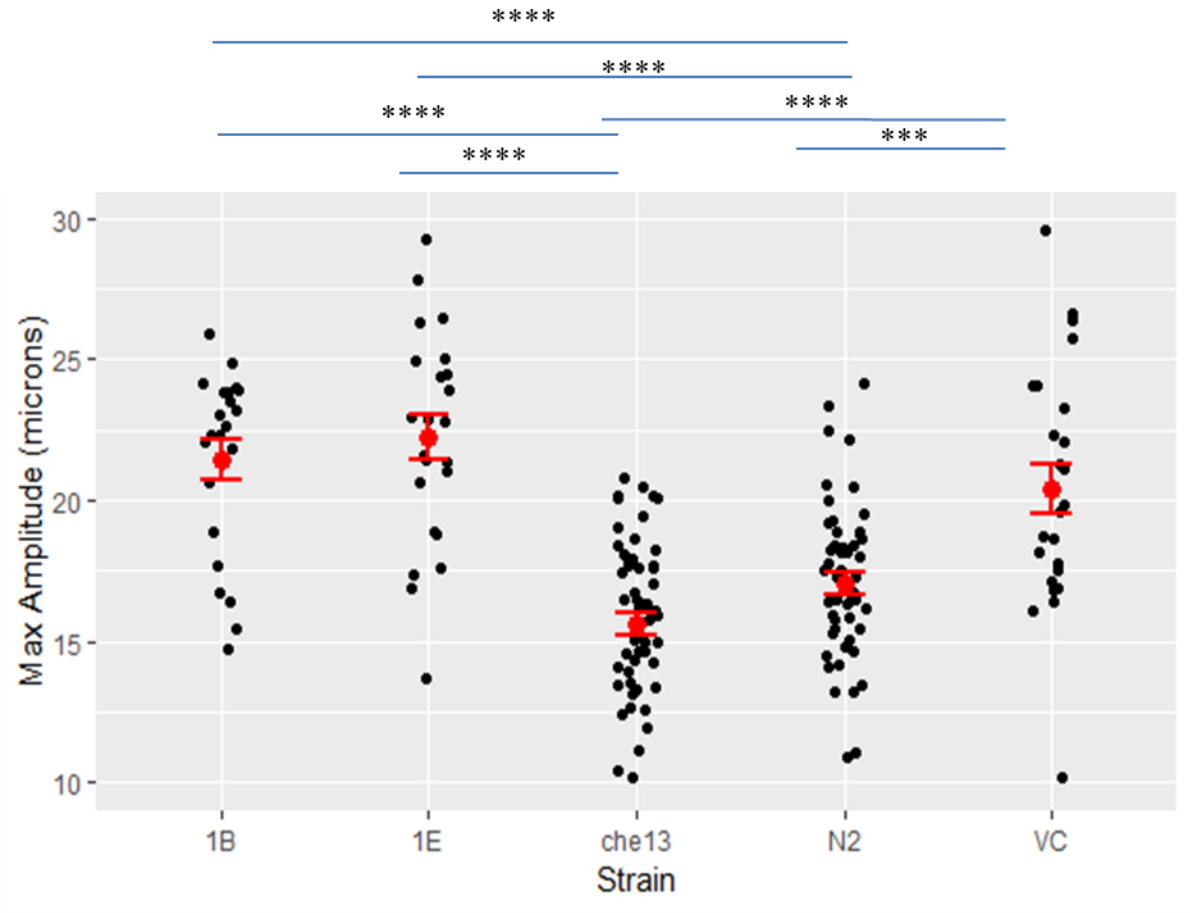

Figure 25. Max amplitude of twice-backcrossed strains. WormTracker was used to compare the difference in the amplitude of each worm's bend. Each worm was tracked for a total of 10 mins. A scatter plot represent the distribution of the data; error bars reflect the standard error. (***P<.001 and ****P<0.0001 relative to che-13, wild type, and VC40591

Neuronal Assessment

To study the neuronal properties of our worms of each strain of our worms, we conducted a dye-filling assay. A subset of sensory neurons in the head and two tail phasmid neurons normally take up lipophilic fluorescent dye if these ciliated neurons properly reach through the cuticle of the worm. Our results revealed that the worms from our backcrossed strains and VC40591 did not dye fill normally in the tail, and sometimes neurons in both the tail and head did not dye fill. The dye-filling defect at multiple age stages was pattern that stood out, but what we found interesting was that at the adult stage all worms for all strains were dye filled appropriately (Table 26). Our assumption was that wild type would dye fill at each life stage, but we observed that an L3 stage there were some worms that did not dye fill in their tail (Figure 26, 28). When comparing the VC40591 strain to the backcrossed strains we noticed that the backcross had a higher percentage of non-dye filling tail neurons, and the VC40591 had more worms dye fill in the tail than non-dye filling. In L4/young adults of VC40591, 13.64% of the population dye filled in both the head and tail but in backcrossed strains as many as 71.62% of the worms did not dye fill in the head nor in the tail (Figure 28). There were more worms that did not dye fill in the head and tail than those that did not dye fill in the head in the backcrossed strains when we are comparing the L4/young adults (Figure 27).

Strain	Life Stage	n=	Head^{+} (%)	Head^{-} (%)	Tail^{+} (%)	Tail^{-} (%)	Head & Tail^{-} (%)
VC40591	Adults	150	100.00	0	96.67	3.33	0
	L4/Young Adults	22	86.36	0	54.55	31.82	13.64
	L3	47	25.53	4.26	14.89	14.89	74.47
	L2	8	25.00	0	25.00	0	75.00
1B	Adults	139	100.00	0	79.14	20.86	0.00
	L4/Young Adults	74	28.38	0	1.35	29.73	71.62
	L3	55	23.64	0	0	23.64	76.36
	L2	N/A	N/A	N/A	N/A	N/A	N/A
1E	Adults	92	100.00	0	86.96	13.04	0
	L4/Young Adults	79	58.23	0	5.06	53.16	41.77
	L3	77	25.97	0	0	25.97	74.03
	L2	N/A	N/A	N/A	N/A	N/A	N/A
Wildtype	Adults	24	100	0	100	0	0
	L4/Young Adults	15	100	0	100	0	0
	L3	8	100	0	0	100	0
	L2	N/A	N/A	N/A	N/A	N/A	N/A

Table 2.0 Dye fill assessment at multiple stages for different strains. VC40591 and backcrossed mutants show there is an irregularity in dye filling at multiple stages of the worm, however wild type follows a constant pattern.

Figure 26. Dye-filling in tail phasmid neurons in adult, wildtype hermaphrodite. Adults dye-fill normally in PHA and PHB. In younger stage, L3, there is no expression of these neurons, which are indicated by the blue circles.

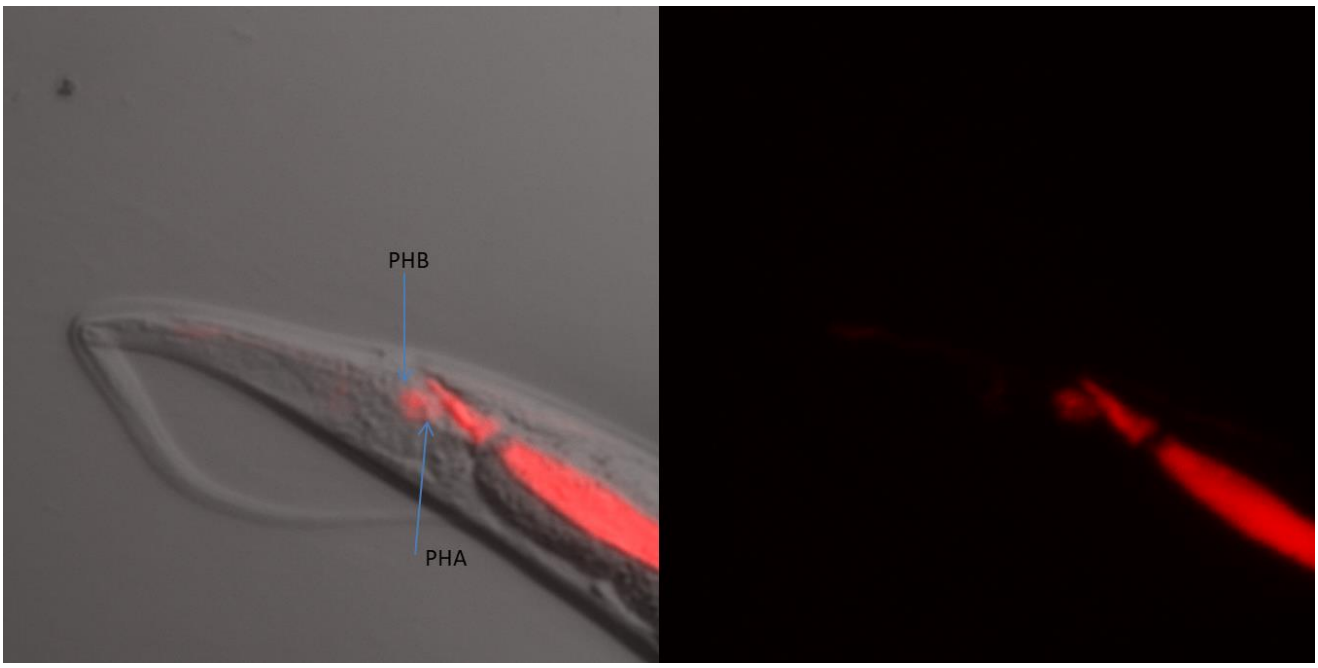

Figure 27. Dye-filling phasmid neurons in an L4 VC40591 hermaphrodite. This L4 dye-filled normally in PHA and PHB.

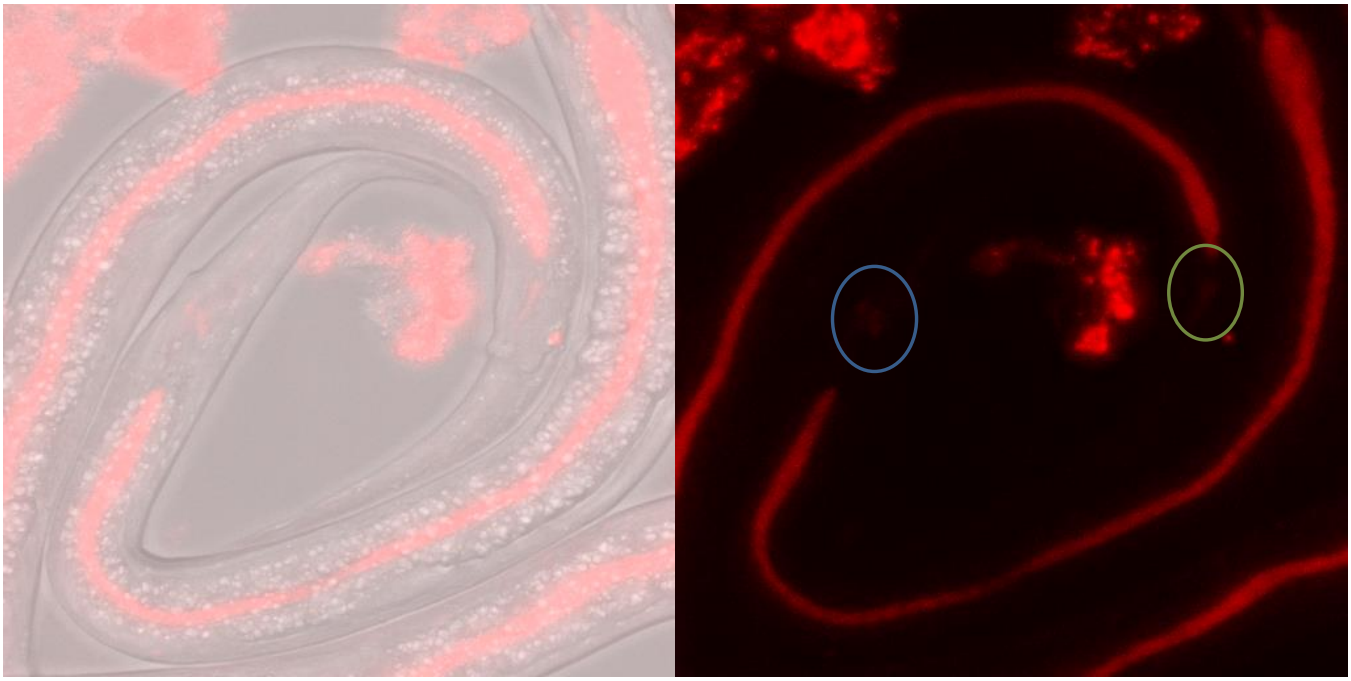

Figure 28. Attempted dye-filling of phasmid neurons in L3, 1E backcrossed (CE*xap5*) hermaphrodite. From a dyefill assay our hermaphrodite shows very weak proper dye filling in the phasmid neurons of one L3 worm, in blue, and absence of neurons, in green.

Genotyping

In the process of trying to genotype for the gene in the wild type, VC40591, and backcrossed strains there were multiple troubleshooting problems that impeded us from getting quality data to present. For that reason we decided to explain what we would have expected, if this part of our project worked. First, we identified the location of the mutation and designed primers to make multiple copies of the region through PCR (Figure 29, 30). The gene we were expecting to isolated would be 720 bp (Figure 31). We selected for restriction enzymes that would make unique cuts in the mutant and wild type gene. Once we had amplified our gene we

would use restriction enzymes to cut the gene to make verify if the backcrossed strains were true carriers of CE*xap5*.

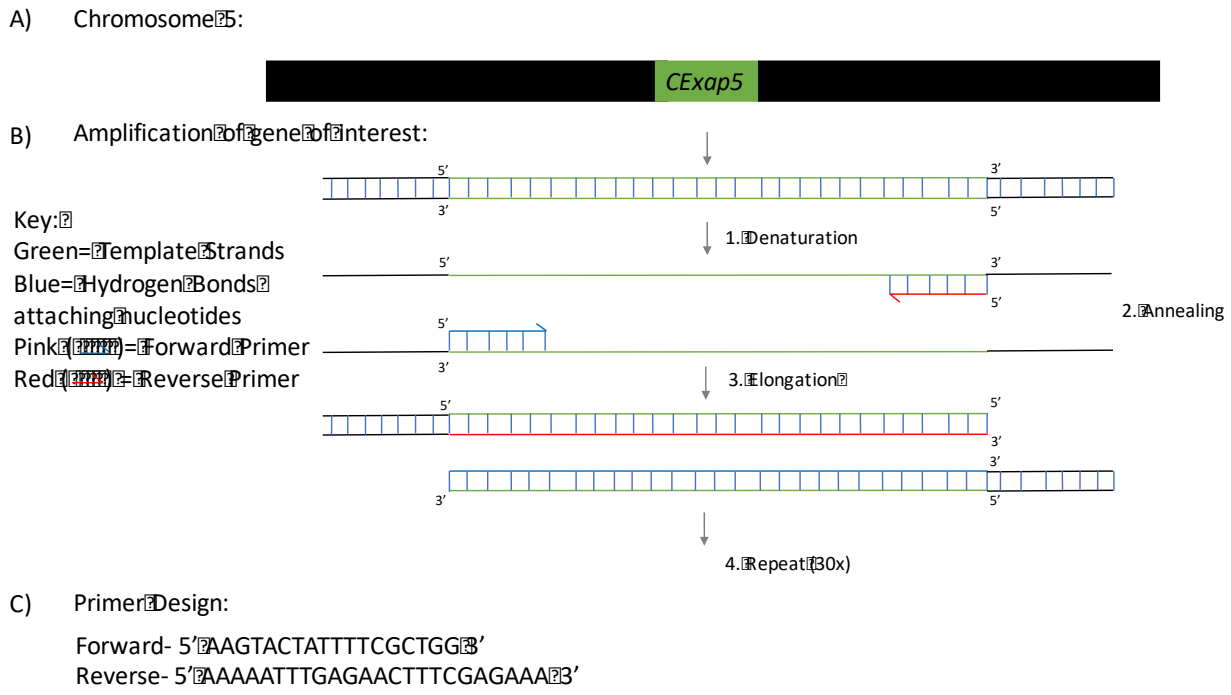

Figure 29. Visualizing the molecular process used to amplify CE47E8.4, referred CE*xap5* (A). We selected an allele of CE*xap5* (*gk709587*), on chromosome 5 (B). A more in depth look of that region shows the double stranded DNA that is stabilized by hydrogen bonds (C). Using molecular techniques, we denatured the DNA by heating it until double stranded DNA are separated into two individual strands. Following denaturation, two fragments of DNA, 18-22, are introduced to attach and function as a starting point for DNA synthesis to begin on each strand. DNA polymerase recognizes the primer and adds new nucleotides sequentially, as it replicates the template strand. This process is repeated and the number of copies of the gene becomes amplified exponentially. The design of primers must take into consideration the template strand to which to attach to and the direction DNA polymerase will read the primer, since DNA polymerase can only add nucleotides to 3' ends.

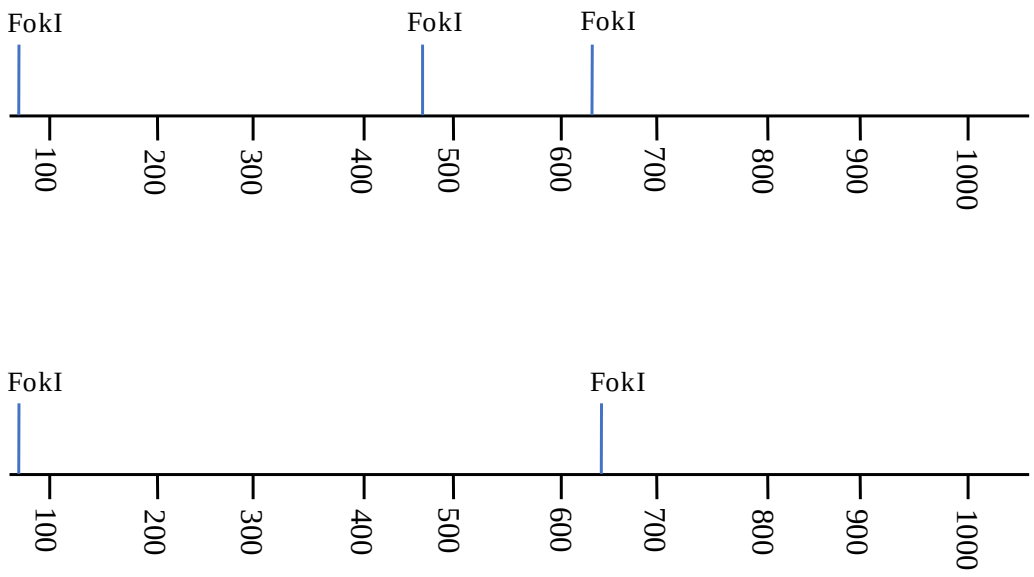

Figure 30. Restriction sites in the amplified region of CE*xap5*. Wildtype and mutant DNA segments are long before PCR amplification. An enzyme was needed to visibly show a difference between DNA amplified from mutant and wildtype worms. FokI can cut in three sites in the wildtype worm (A), but can only cut twice in DNA from CExap5 (*gk709587)* mutant worm (B). This restriction enzyme will provide us with the ability to distinguish a wildtype and mutant DNA.

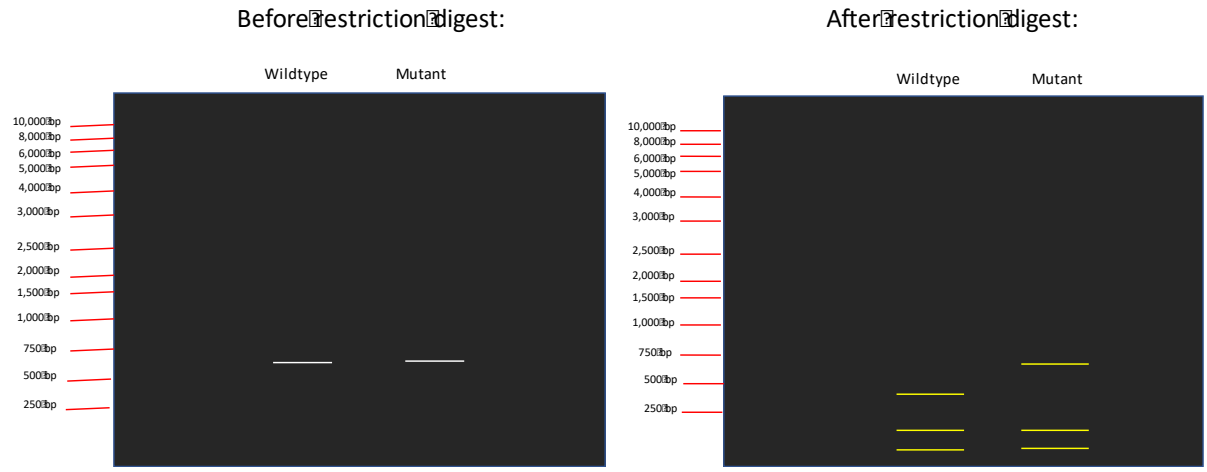

Figure 31. Expected results of gel electrophoresis with wildtype and mutant worms. The amplified fragment from both wildtype and mutant will look the same, with a size of 720 bp (A). After, restriction enzyme is used to the DNA, the bands will look different. The expected size of fragments for wildtype DNA are: 29bp, 77 bp, 156 bp, 449 bp and for mutants we expect: 28bp, 77bp, 605bp

DISCUSSION

The aim of this research project was to study the role of a *xap5* ortholog in an multicellular organism. We wanted to see whether there was a conservation in function of the XAP5 protein between a unicellular and multicellular organism. Our first finding showed that there was a shared promoter between *daf-21* and CE*xap5.* Previous studies demonstrated that the directionality of gene pairs in prokaryotes, co-oriented and divergent, correlate with operons and bi-directional promoters, respectively, and are useful in predicting functionality between gene pairs (Kensche et al., 2008; Dandekar et al., 1998; Korbel et al., 2004). Divergent transcription from a promoter is a pervasive feature in mammalian genes too, but our understanding of whether this is a general feature for all eukaryotic organisms remains unclear (Wu and Sharp, 2013; Seila et al., 2008; Ibrahim et al., 2018). Recent evidence, using deep-sequencing assays, suggests that this divergent transcription is abundant in *C. elegans* and other eukaryotic organisms (Ibrahim et al., 2018). Another study showed that directionality of divergently transcribed genes is highly conserved in eukaryotes, and they show that divergent genes with a high level of conservation in directionality among orthologous groups show a high level of co-expression linked with a likelihood of being functionally similar (Kensche et al., 2008). The finding from this scientific literature adds a new dimension to our study, because *daf-21* and *xap-5* could potentially be co-expressed and might even have related functions.

The percent identity from protein, coding, and promoter sequences indicate that there is a structural similarity and possibly a similar function between CE-XAP5 and *Chlamydomonas'* XAP5. Our bioinformatics results led us to believe that there is a high level of conservation in directionality between the orthologous groups, which increases the chance of co-expression between *daf-21* and *xap-5* as suggested by (Kensche et al., 2008). Under the assumption that CE-

XAP5 has the same function as *Chlamydomonas*' XAP5, we would expect the normal CE-XAP5 gene to regulate the expression of ciliary genes. Thus, in a strain with a loss-of-function allele of CE-XAP5, we would expect to see aberrant chemosensory behaviors. As of now, it is believed that *daf-21* regulates chemosensory behavior in *C. elegans*, but from the evidence we have collected, we hypothesize that this observation may be due to co-expression with CE*xap5* (Birnbv et al., 2000).

Our behavioral analysis of CE*xap5*(*gk709587*) showed that this allele confers cilia-deficient phenotypes, which were seen in the WormTracker, roaming/dwelling assay, and a chemotaxis (osmotic and pH avoidance) assay. Our behavioral assays revealed an over-dwelling pattern that remained consistent across our selected backcrossed mutants. This phenotype is also seen in mutant worms that lack particular isoforms of the DAF-19 transcription factor, also known to regulate cilia gene expression (Senti & Swobda, 2008; De Stasio et al., 2018). This similarity suggests a ciliated sensory neuron defect. The variables assessed in the WormTracker quantified a significant abnormal motility pattern in VC40591 and our backcrossed strains, which provided further evidence for a locomotor change that could be neuronal in origin. Our chemotaxis and dye fill assays also supported the idea of cilia-related phenotypic behaviors being due to neuronal change. We also noted that there was a difference between the phenotypic behavior of VC40591 and the backcrossed strains. We interpreted these differences to be a result of "cleaning" the VC40591 genome from other mutations.

Evolution has selected olfaction (a form of chemosensation) to detect many volatile chemicals, such as repellents or attractants (Yoshida et al., 2012; Bargmann et al., 1993) and our mutant worms appear defective in chemosensation. Our drop test assessed the avoidance response of a population to water soluble repellents. This type of assay is predicated to activate

ASE, ASH, ASK, ASL, PHA, PHB neurons, because they function in detecting multiple types of chemicals as well as osmotic change (Bargmann, 2006). Once activated, these neurons elicit a response in the worm, which makes them move forward, a signal of attraction, or move in reverse, a signal of repulsion. With the application of an acidic chemical to a worm, literature shows that wild type worms will be repulsed by the stimulus (Wakabayashi et al., 2015; Hilliard et al., 2005). This corresponded to what we saw, the wild type N2 worms avoided acetic acid when it was applied, but backcrossed strains, VC40591, and *che-13* worms were not avoiding the stimulus as much was as the wild type. Similarly, in an osmotic avoidance assay, our results match those of previous studies where the wild type *C. elegans* were strongly repelled by glycerol (Culotti and Russell, 1978). However, our mutant strains were not as strongly affected by the application of glycerol. Both of these assays indicated that there is a chemosensory defect where there is a deficiency in the neuronal circuitry. This best explains why backcrossed strains were not able to sense the stimuli by two well characterized repellants.

These two chemical assays suggested the possibility that behavioral defects might be due to a cilia deficiency of some sort. To confirm or disprove these observations, we conducted a dye filling assay. Previous studies show that there are dye filling-defective worms that take in lipophilic dye in the intestine but do not absorb it into their ciliated sensory neurons due to impairments in their cilia. The difference can be easily visualized under a microscope; wildtype worms will fluoresce in particular regions of the head and tail while ciliary-deficient worms will not. We did not see any changes when we looked at the sensory neurons in the head in control and CE*xap5* mutant worms, but we did see notable differences in dye-filling in the tail PHA and PHB neurons, so that is where we focused our most attention. When comparing the VC40591 strain to the backcrossed strains in our dye-filling assessment, we noticed that our backcrossed

strain would dye fill less frequently at L3, L4, and young adult stages. At first, this phenotype seemed very similar to the one described by Piasecki et al. (2017) in which a mutation in the catalytic site of MAPK-15 domain affected dye-filling in PHA and PHB tail neurons only. However, Piasecki et al. (2017) report an absence of dye-filling in tail neurons in both the hermaphrodites and the males without any recuperation of dye filling cilia at a later life stage. In our study, it seemed like there was a slower development of the cilia in our backcrossed and VC40591 strains. From our dye filling assay results we hypothesize that there might be a problem that is affecting proper development or timing of development in the cilia conferred by CE*xap5*.

We are proposing that CE*xap5* is being co-expressed with *daf-21* and is the driving force for the phenotypes seen in *daf-21* mutant worms (Birnby et al., 2000). Since the time Birnby et al., 2000 published their work, we learned that DAF-19, a transcription factor, is responsible for regulating gene expression in ciliated sensory neurons in *C. elegans* (Senti & Swoboda, 2008; De Stasio et al., 2018). The study assessing *daf-21* concluded there was chemosensory deficiency, which is similarly seen in *daf-19c* deficient worms (Senti and Swoboda, 2008). However, no prior relationship between the DAF-19 and *daf-21* has been established.

Future studies:

Our results have set the foundation for a project with a lot of potential. It would be a good idea to compare the upstream control regions of known target genes of XAP5 in *Chlamydomonas* with sequences from orthologous genes in *C. elegans*. Then if there are any similarities we should determine whether there is an X box motif or the same DNA sequence motif found in Li et al. (2018) to verify a direct regulation by CE-XAP5. Due to the complexity of some of the

previous X box analyses, we did not have the algorithm that conditioned the sequences to pick up on the motif (Efimenko et al., 2005). If there could be an analysis performed on the orthologous' promoter of target genes we could hypothesize whether there is a direct or indirect relationship between DAF-19 and CE-XAP5. For future experiments, the twice-backcrossed strains should be further backcrossed to up to six more times, to further replace mutant chromosomes with wildtype counterparts while maintaining the CE*xap5* mutation. Studying development of the worm neurons would be of great interest to continue measuring an appropriate correlation of the mutant worm behavior. We should also aim to get the genotyping to work to verify the phenotypes are corresponding to the gene we are studying.

ACKNOWLEDGMENT

Le quiero dedicar este proyecto a mi familia por siempre apoyarme en todo. Gracias a su esfuerzo y su ejemplo he podido esforzarme, luchar por mis sueños y desafiar cualquier obstáculo. También le quiero agradecer a mi hermano por ser una importante parte de mi vida.

This project could not have been possible without the love, support, and encouragement of many people in my life. First, I would like to thank my research advisor and mentor Beth De Stasio for always being willing to work with me to help me become a better scientist. I am so grateful for all that you do for me, it has been an honor working in your lab. I would like to thank all the other professors and faculty that have supported me throughout my four years. Special thanks to Dr. Piasecki and Ricky Peña for being great people to be around in lab, but a section of my thesis would not be possible if it was not for their WormTracker. I would also like to thank Dr. Ramos and Dr. Hicks for helping me with statistical analysis, and providing me with a lot of support during their time here. I am also grateful for all of my friends who have supported me

throughout this process. Thank you to Sam Bader and Maya Guenther for providing constructive criticism on my paper. Thank you to JoAnn Stamm. and Wayne Kruegar for providing me with many worm plates and reagents that were needed for this experiment.

REFERENCE

Adams, G. M., Huang, B., Piperno, G., & Luck, D. J. (1981). Central-pair microtubular complex of Chlamydomonas flagella: polypeptide composition as revealed by analysis of mutants. *The Journal of Cell Biology*, *91*(1), 69–76. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7028763

Anver, S., Roguev, A., Zofall, M., Krogan, N. J., Grewal, S. I. S., & Harmer, S. L. (2014). Yeast X-chromosome-associated protein 5 (Xap5) functions with H2A.Z to suppress aberrant transcripts. *EMBO Reports*, *15*(8), 894–902. https://doi.org/10.15252/embr.201438902

Bae, Y.-K., & Barr, M. M. (2008). Sensory roles of neuronal cilia: cilia development, morphogenesis, and function in C. elegans. *Frontiers in Bioscience : A Journal and Virtual Library*, *13*, 5959–5974. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18508635

Bainton, R. J., Tsai, L. T.-Y., Singh, C. M., Moore, M. S., Neckameyer, W. S., & Heberlein, U. (2000). Dopamine modulates acute responses to cocaine, nicotine and ethanol in Drosophila. *Current Biology*, *10*(4), 187–194. https://doi.org/10.1016/S0960-9822(00)00336-5

Bangs, F., & Anderson, K. V. (2017). Primary Cilia and Mammalian Hedgehog Signaling. *Cold Spring Harbor Perspectives in Biology*, *9*(5), a028175. https://doi.org/10.1101/cshperspect.a028175

Bargmann, C. (2006). Chemosensation in C. elegans. *WormBook*. https://doi.org/10.1895/wormbook.1.123.1

Beales, P., & Jackson, P. K. (2012). Cilia - the prodigal organelle. *Cilia*, *1*(1), 1. https://doi.org/10.1186/2046-2530-1-1

Berbari, N. F., O'Connor, A. K., Haycraft, C. J., & Yoder, B. K. (2009). The Primary Cilium as a Complex Signaling Center. *Current Biology*, *19*(13), R526–R535. https://doi.org/10.1016/j.cub.2009.05.025

Birnby, D. A., Link, E. M., Vowels, J. J., Tian, H., Colacurcio, P. L., & Thomas, J. H. (2000). A transmembrane guanylyl cyclase (DAF-11) and Hsp90 (DAF-21) regulate a common set of

chemosensory behaviors in caenorhabditis elegans. *Genetics*, *155*(1), 85–104. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10790386

Butcher, R. A., Fujita, M., Schroeder, F. C., & Clardy, J. (2007). Small-molecule pheromones that control dauer development in Caenorhabditis elegans. *Nature Chemical Biology*, *3*(7), 420–422. https://doi.org/10.1038/nchembio.2007.3

Caswell-Chen, E. P. (2005). Revising the Standard Wisdom of C. elegans Natural History: Ecology of Longevity. *Science of Aging Knowledge Environment*, *2005*(40), pe30-pe30. https://doi.org/10.1126/sageke.2005.40.pe30

Choksi, S. P., Lauter, G., Swoboda, P., & Roy, S. (2014). Switching on cilia: transcriptional networks regulating ciliogenesis. *Development*, *141*(7), 1427–1441. https://doi.org/10.1242/dev.074666

Corsi, A. K., Wightman, B., & Chalfie, M. (2015). A Transparent Window into Biology: A Primer on Caenorhabditis elegans. *Genetics*, *200*(2), 387–407. https://doi.org/10.1534/genetics.115.176099

Culotti, J. G., & Russell, R. L. (1978). Osmotic avoidance defective mutants of the nematode Caenorhabditis elegans. *Genetics*, *90*(2), 243–256. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/730048

Dandekar, T. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *Trends in Biochemical Sciences*, *23*(9), 324–328. https://doi.org/10.1016/S0968-0004(98)01274-2

De Stasio, E. A., Mueller, K. P., Bauer, R. J., Hurlburt, A. J., Bice, S. A., Scholtz, S. L., … Swoboda, P. (2018). An Expanded Role for the RFX Transcription Factor DAF-19, with Dual Functions in Ciliated and Nonciliated Neurons. *Genetics*, *208*(3), 1083–1097. https://doi.org/10.1534/genetics.117.300571

Efimenko, E. (2005). Analysis of xbx genes in C. elegans. *Development*, *132*(8), 1923–1934. https://doi.org/10.1242/dev.01775=

Falk, N., Lösl, M., Schröder, N., & Gießl, A. (2015). Specialized Cilia in Mammalian Sensory Systems. *Cells*, *4*(3), 500–519. https://doi.org/10.3390/cells4030500

Fawcett, D. W. (1954). THE STUDY OF EPITHELIAL CILIA AND SPERM FLAGELLA WITH THE ELECTRON MICROSCOPE. *The Laryngoscope*, *64*(7), 557???567. https://doi.org/10.1288/00005537-195407000-00002

Herman, R. (2005). Introduction to sex determination. *WormBook*. https://doi.org/10.1895/wormbook.1.71.1

Hilliard, M. A., Apicella, A. J., Kerr, R., Suzuki, H., Bazzicalupo, P., & Schafer, W. R. (2005). In vivo imaging of C. elegans ASH neurons: cellular response and adaptation to chemical repellents. *The EMBO Journal*, *24*(1), 63–72. https://doi.org/10.1038/sj.emboj.7600493

Ibrahim, M. M., Karabacak, A., Glahs, A., Kolundzic, E., Hirsekorn, A., Carda, A., … Ohler, U. (2018). Determinants of promoter and enhancer transcription directionality in metazoans. *Nature Communications*, *9*(1), 4472. https://doi.org/10.1038/s41467-018-06962-z

Inglis, P. (2006). The sensory cilia of Caenorhabditis elegans. *WormBook*. https://doi.org/10.1895/wormbook.1.126.1

Ishikawa, H., & Marshall, W. F. (2011). Ciliogenesis: building the cell's antenna. *Nature Reviews Molecular Cell Biology*, *12*(4), 222–234. https://doi.org/10.1038/nrm3085

Jenkins, P. M., McEwen, D. P., & Martens, J. R. (2009). Olfactory Cilia: Linking Sensory Cilia Function and Human Disease. *Chemical Senses*, *34*(5), 451–464. https://doi.org/10.1093/chemse/bjp020

Kensche, P. R., Oti, M., Dutilh, B. E., & Huynen, M. A. (2008). Conservation of divergent transcription in fungi. *Trends in Genetics*, *24*(5), 207–211. https://doi.org/10.1016/j.tig.2008.02.003

Korbel, J. O., Jensen, L. J., von Mering, C., & Bork, P. (2004). Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nature Biotechnology*, *22*(7), 911–917. https://doi.org/10.1038/nbt988

Lee, T. I., & Young, R. A. (2013). Transcriptional Regulation and Its Misregulation in Disease. *Cell*, *152*(6), 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014

Li, L., Tian, G., Peng, H., Meng, D., Wang, L., Hu, X., … Hu, Z. (2018). New class of transcription factors controls flagellar assembly by recruiting RNA polymerase II in Chlamydomonas. *Proceedings of the National Academy of Sciences*, *115*(17), 4435–4440. https://doi.org/10.1073/pnas.1719206115

Mao, S., Shah, A. S., Moninger, T. O., Ostedgaard, L. S., Lu, L., Tang, X. X., … Welsh, M. J. (2018). Motile cilia of human airway epithelia contain hedgehog signaling components that mediate noncanonical hedgehog signaling. *Proceedings of the National Academy of Sciences*, *115*(6), 1370–1375. https://doi.org/10.1073/pnas.1719177115

Marshall, W. F., & Nonaka, S. (2006). Cilia: Tuning in to the Cell's Antenna. *Current Biology*, *16*(15), R604–R614. https://doi.org/10.1016/j.cub.2006.07.012

Martin-Tryon, E. L., & Harmer, S. L. (2008). XAP5 CIRCADIAN TIMEKEEPER Coordinates Light Signals for Proper Timing of Photomorphogenesis and the Circadian Clock in Arabidopsis. *THE PLANT CELL ONLINE*, *20*(5), 1244–1259. https://doi.org/10.1105/tpc.107.056655

Mazzarella, R., Pengue, G., Yoon, J., Jones, J., & Schlessinger, D. (1997). Differential Expression of XAP5, a Candidate Disease Gene. *Genomics*, *45*(1), 216–219. https://doi.org/10.1006/geno.1997.4912

Perkins, L. A., Hedgecock, E. M., Thomson, J. N., & Culotti, J. G. (1986). Mutant sensory cilia in the nematode Caenorhabditis elegans. *Developmental Biology*, *117*(2), 456–487. https://doi.org/10.1016/0012-1606(86)90314-3

Piano, F., Schetter, A. J., Morton, D. G., Gunsalus, K. C., Reinke, V., Kim, S. K., & Kemphues, K. J. (2002). Gene Clustering Based on RNAi Phenotypes of Ovary-Enriched Genes in C. elegans. *Current Biology*, *12*(22), 1959–1964. https://doi.org/10.1016/S0960-9822(02)01301-5

Piasecki, B. P., Burghoorn, J., & Swoboda, P. (2010). Regulatory Factor X (RFX)-mediated transcriptional rewiring of ciliary genes in animals. *Proceedings of the National Academy of Sciences*, *107*(29), 12969–12974. https://doi.org/10.1073/pnas.0914241107

Piasecki, B. P., Sasani, T. A., Lessenger, A. T., Huth, N., & Farrell, S. (2017). MAPK-15 is a ciliary protein required for PKD-2 localization and male mating behavior in Caenorhabditis elegans. *Cytoskeleton*, *74*(10), 390–402. https://doi.org/10.1002/cm.21387

Reiter, J. F., & Leroux, M. R. (2017). Genes and molecular pathways underpinning ciliopathies. *Nature Reviews Molecular Cell Biology*, *18*(9), 533–547. https://doi.org/10.1038/nrm.2017.60

Riddle, D. L., Blumenthal, T., Meyer, B. J., & Priess, J. R. (1997). *Introduction to C. elegans*. *C. elegans II*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21413243

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering, Design and Selection*, *12*(2), 85–94. https://doi.org/10.1093/protein/12.2.85

Satir, P., & Christensen, S. T. (2008). Structure and function of mammalian cilia. *Histochemistry and Cell Biology*, *129*(6), 687–693. https://doi.org/10.1007/s00418-008-0416-9

Satir, P., Pedersen, L. B., & Christensen, S. T. (2010). The primary cilium at a glance. *Journal of Cell Science*, *123*(4), 499–503. https://doi.org/10.1242/jcs.050377

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., … Sharp, P. A. (2008). Divergent Transcription from Active Promoters. *Science*, *322*(5909), 1849–1851. https://doi.org/10.1126/science.1162253

Senti, G., & Swoboda, P. (2008). Distinct Isoforms of the RFX Transcription Factor DAF-19 Regulate Ciliogenesis and Maintenance of Synaptic Activity. *Molecular Biology of the Cell*, *19*(12), 5517–5528. https://doi.org/10.1091/mbc.e08-04-0416

Swoboda, P., Adler, H. T., & Thomas, J. H. (2000). The RFX-Type Transcription Factor DAF-19 Regulates Sensory Neuron Cilium Formation in C. elegans. *Molecular Cell*, *5*(3), 411–421. https://doi.org/10.1016/S1097-2765(00)80436-0

Todeschini, A.-L., Georges, A., & Veitia, R. A. (2014). Transcription factors: specific DNA binding and specific gene regulation. *Trends in Genetics*, *30*(6), 211–219. https://doi.org/10.1016/j.tig.2014.04.002

Veland, I. R., Awan, A., Pedersen, L. B., Yoder, B. K., & Christensen, S. T. (2009). Primary Cilia and Signaling Pathways in Mammalian Development, Health and Disease. *Nephron Physiology*, *111*(3), p39–p53. https://doi.org/10.1159/000208212

Venkatesh, D. (2017). Primary cilia. *Journal of Oral and Maxillofacial Pathology*, *21*(1), 8. https://doi.org/10.4103/jomfp.JOMFP_48_17

Wakabayashi, T., Kitagawa, I., & Shingai, R. (2004). Neurons regulating the duration of forward locomotion in Caenorhabditis elegans. *Neuroscience Research*, *50*(1), 103–111. https://doi.org/10.1016/j.neures.2004.06.005

Wang, J., Schwartz, H. T., & Barr, M. M. (2010). Functional Specialization of Sensory Cilia by an RFX Transcription Factor Isoform. *Genetics*, *186*(4), 1295–1307. https://doi.org/10.1534/genetics.110.122879

Wu, X., & Sharp, P. A. (2013). Divergent Transcription: A Driving Force for New Gene Origination? *Cell*, *155*(5), 990–996. https://doi.org/10.1016/j.cell.2013.10.048

Xu, Y.-J., Lei, Y., Li, R., Zhang, L.-L., Zhao, Z.-X., Zhao, J.-H., … Wang, W.-M. (2017). XAP5 CIRCADIAN TIMEKEEPER Positively Regulates RESISTANCE TO POWDERY MILDEW8.1–Mediated Immunity in Arabidopsis. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.02044

Yoshida, K., Hirotsu, T., Tagawa, T., Oda, S., Wakabayashi, T., Iino, Y., & Ishihara, T. (2012). Odour concentration-dependent olfactory preference change in C. elegans. *Nature Communications*, *3*(1), 739. https://doi.org/10.1038/ncomms1750

Zimmermann, K. W. (1898). Beiträge zur Kenntniss einiger Drüsen und Epithelien. *Archiv Für Mikroskopische Anatomie*, *52*(3), 552–706. https://doi.org/10.1007/BF02975837

APPENDICES

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10
	Consensus$equence8	NCBI-ConservedDomainDatabaseID(CDD)	Position	Score	E-Value		Pfam	Position	E-Value
C.briggsae	CTPENLMFVKEDLIPHFTTFQDFVTKAMGKTGPLFVFDSASDVRIRQD		309862 4..5	85	7.70E-15		XAP5	4..46	7.70E-15
	YGESHPAKVLRSWYEKNKHIYPASRWEPFVPSKKYGRNFD		309862 3..38	74.5	3.30E-16		XAP5	3..39	3.30E-16
	LAAEWRVKQNTEKNEEITVAYAYWDGSSHRKNNMKVKKGNTISQCLARAIE		309862 1..5	65.3	8.00E-15		XAP5	1..48	2.80E-13
							NETI	34..46	0.1
	CKVGITTKFTANYETMEESVKSKTYGLVSLDDMKNIQKNESNRDLQVAR		309747 3..33	26.7	0.54		Ub-RnfH	16..47	0.9
	EHVAKHTQKRVLSFAFEDEEEEEDVVPIPKKRVGMDPTVDTSFLPDKER		309862 5..50	42.6	2.00E-06		-	-	
C.brenneri	CTPENLMFVKEDLIPHFTTFQDFVTKAMGKETGPLFVFDSASDVRIRQD								
	YGESHPAKVLRSWYEKNKHIYPASRWEPFVPSKKYGRNFD	Same as *C.briggsae*	308537 1..30	27.2	2.80E-01				
	LAAEWRVKQNTEKNEEITVAYAYWDGSSHRKNNMKVKKGNTISQCLARAIE	Same as *C.briggsae*							
	CKVGITTKFTANYETMEESVKSKTYGLVSLDDMKNIQKNESNRDLQVAR	Same as *C.briggsae*							
C.elegans	CTAENLMFVKEDLIPHFTTFQDFVTKAMGKSGPLFNFDVHDDVRLSD		309862 1..50	64.5	2.00E-14		XAP5	1..48	6.70E-13
			215396 2..43	27.5	0.35				
	YGESHPAKVLRSWYEKNKHIYPASRWEPFVPSKKYGRNFD	Same as *C.briggsae*							
	LAAEWRVKQDAEKNEEITVAYAYWDGSSHRKNNMKKKGNTISQCLGRAIE		335937 1..50	63.8	3.00E-14		XAP5	1..48	6.70E-13
	CRVGITNKFMANYETVEEVVKSKTYGLVSLDDMKNIQKNEISNRDLQVAR		235292 10..35	29.31	1.10E-01		-	-	
	EHVAKHTQKRFLSFAFDDEEDEDAAPIPKKRVGMDPTVDTSFLPDKER		309862 5..5	45.7	1.00E-07				
C.remanei	CTTENLMFVKEDLIPHFTTFQDFVTKAMGKTGPLFVFDSASDVRIRQD		309862 4..5	86.5	2.00E-22	-	XAP5	3..46	3.91
	YGESHPAKVLRSWYEKNKHIYPASRWEPFVPSKKYGRNFD	Same as *C.briggsae*	-	-	-	-	-	-	
	LAAEWRVKQNTEKNEEITVAYAYWDGSSHRKNNMKKKGNTISQCLARAIE	Same as *C.briggsae*	-	-	-	-	-	-	
	CRVGITNKFTANYETMEESVKSKTYGLVSLDDMKNIQKNESNRDLQVAR	Same as *C.elegans*	-	-	-	-	-	-	
	EHVAKHTQKRVLSFAYEDEEEEEDAVPIPKKRVGMDPTVDTSFLPDKER		309862 5..5	44.1	5.00E-07	-	-	-	
Chlamydomonas									
	VGIDNLMYKKEDLIMPHHHTFYEUISKARGKSGPLFDFTVKDDRITND		309862 1..50	93.4	4.00E-25		XAP5	1..46	5.20E-17
			319379 8..34	25	0.55				
	KQDTHAGKVVERHWTDKNKHIFPASRWEYTDPEKKYGEYTI		309862 1..41	87.6	4.00E-23		XAP5	1..40	1.10E-19
	LREWQLAQDVIKAEPLTITYSYWNGTGHRRNVSVKKGDSIGGFLKAVRE		309862 1..50	80.3	3.00E-20		XAP5	1..49	2.40E-15
H.sapiens									
	AGVEQLMYIKEDLIPHHHSYDFIVTKARGKSGPLFNFDVHDDVRLLSD		309862 1..41	92.6	6.00E-25		-	-	
			198362 5..37	27.4	0.34				
	KDEHAGKVVLRSWYEKNKHIFPASRWEPYDPEKKWDKYTI		309862 1..41	92.6	6.00E-25		-	-	
			198362 5..37	27.4	3.40E-01		XAP5	1..49	3.20E-20
	LRQEWEAKGEKIKSEEEIETFSYWDGSGHRRTVKKMRKGNTMQQFLQKALE		309862 1..5	95.3	1.00E-25		XAP5	1..49	2.80E-19
			314402 28..49	25.5	4.90E-01		TUG-UBL1	28..49	0.12
			310218 3..42	26.1	9.50E-01		DUF2653	6..44	0.11
							RBD	29..50	0.38
	MKSNIDKFSAHYDAVEAELKSSTVGLVTLNDMKAKQEALVKERERQLAK	309862 1..50	65.3	8.00E-15			XAP5	1..48	2.80E-13
							NETI	34..47	0.1
							Ub-RnfH	16..47	0.9

Appendix A. NCBI Motif IDs for protein sequences.