Evidence of POGO Transposable Elements in the Atlantic Bay Nettle (Chrysaora chesapeakei) Genome

Jonathan C. Medina
Montclair State University

Follow this and additional works at: https://digitalcommons.montclair.edu/etd

Part of the Biology Commons

Recommended Citation
https://digitalcommons.montclair.edu/etd/143

This Thesis is brought to you for free and open access by Montclair State University Digital Commons. It has been accepted for inclusion in Theses, Dissertations and Culminating Projects by an authorized administrator of Montclair State University Digital Commons. For more information, please contact digitalcommons@montclair.edu.
ABSTRACT

The Atlantic Bay Nettle, *Chrysaora chesapeakei*, is a Scyphozoan found commonly in the bays and brackish waters of estuaries of the eastern United States. Research has shown that there has been a significant increase in jellyfish populations over the past decade, likely the result of factors such as global climate change, eutrophication, overfishing, and the explosive growth of hardened surfaces for polyp attachment.

Transposable elements (TEs) are mobile genetic elements that are widespread and conserved throughout the biological world. Although TEs often comprise a large portion of eukaryotic genomes, their exact function is uncertain but they may provide a mechanism for genetic diversity and recombination. Although previous research has suggested the presence of *Tc1* and *Mariner* DNA transposons within *Hydra* (Class Hydrozoa), the presence of TEs in other Cnidarians has not previously been examined. Based on RNA-seq and direct DNA sequence analysis of gDNA, I have discovered the presence of a member of the *Tc1-Mariner* superfamily, *POGO*, within *C. chesapeakei*. This is the first definitive evidence of TEs in a member of the Class Scyphozoa.

Two putative consensus sequences, TR1 and TR2, were generated averaging 1,028 bp using two different DNA templates. Analysis of the putative translation products of TR1 and TR2 (BLASTx) indicates modest conservation (38% homology) for the length of each fragment, however, analysis of regions confined to conserved domains were upwards of 60% homologous. Furthermore, two variable regions, VR1 and VR2,
were identified within our consensus sequences. VR2, in particular, showed a higher degree of variability with indels, SNPs, and five heterozygosities found within the sequences directly flanking the region perhaps suggesting varying copies within genomes of this element. The fact that these TEs were first identified from RNA-Seq libraries of *Chrysaora chesapeakei* verifies that this element is transcriptionally active in this jellyfish.

Bioinformatic analysis shows that the overwhelming majority of BLASTx matches corresponded to *POGO transposable elements with KRAB domains* (Krüppel-associated boxes). *POGK* is one of many genes to be derived from transposable elements and previously believed to be confined to humans and other mammals. Our data suggests that we have partially cloned a homolog of this gene from *Chrysaora chesapeakei*. Completion of the intact TE will likely require additional amplification of gDNA using inverse PCR.
Evidence of POGO Transposable Elements in the Atlantic Bay Nettle (Chrysaora chesapeakei) Genome

By

Jonathan C. Medina

A Master's Thesis Submitted to the Faculty of
Montclair State University
In partial fulfillment of the requirements
For the degree of Master of Science
May 2018

College of Science and Mathematics
Department of Biology

Thesis Committee:
Thesis Sponsor/ Dr. John J. Gaynor

Committee Member/ Dr. Paul A.X. Bologna

Committee Member/ Dr. Chenguang Du
Evidence of POGO Transposable Elements in the Atlantic Bay Nettle

(Chrysaora chesapeakei) Genome

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of Science

Molecular Biology

By

JONATHAN C. MEDINA

Montclair State University

Montclair, NJ

2018
Acknowledgments

Above all, I would like to express my sincere gratitude toward my advisor and mentor Dr. John J. Gaynor. His talent for inspiring, cultivating, and nurturing the type of curious and inquisitive mindset needed for science and research is unparalleled.

From initially inspiring me towards pursuing research, to sharing of his laboratory and supplies, continuous patience and support, close guidance, and immense knowledge, none of this would have been possible. I am truly grateful for all you have done; thank you Dr. Gaynor.

Also, I would like to kindly thank my thesis committee Dr. Paul A.X. Bologna and Dr. Chunguang Du. I am very appreciative of the time, effort, and insightful feedback they provided while serving on my thesis committee.

Finally, I would like to extend gratitude towards my friends and family for their support throughout this process. To my mother - thank you for always supporting me, leading by example, and being the voice that underpinned my perseverance towards completing my graduate degree and master’s thesis.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td>SIGNATURE PAGE</td>
<td>3</td>
</tr>
<tr>
<td>TITLE PAGE</td>
<td>4</td>
</tr>
<tr>
<td>COPYRIGHT</td>
<td>5</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>6</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>7</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>9</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>I. Chrysaora Biology</td>
<td>11</td>
</tr>
<tr>
<td>II. Discovery of Jumping Genes and the Mobilome</td>
<td>14</td>
</tr>
<tr>
<td>III. The Tcl/mariner Superfamily</td>
<td>19</td>
</tr>
<tr>
<td>i. Structure of the Transposon</td>
<td>19</td>
</tr>
<tr>
<td>ii. The Transposase</td>
<td>20</td>
</tr>
<tr>
<td>a. DNA-Binding Domain</td>
<td>21</td>
</tr>
<tr>
<td>b. Nuclear localization Site (NLS)</td>
<td>21</td>
</tr>
<tr>
<td>c. Catalytic Domain</td>
<td>22</td>
</tr>
<tr>
<td>iii. Mechanism of Transposition</td>
<td>22</td>
</tr>
<tr>
<td>a. Cut-Site Recognition and Cleavage</td>
<td>22</td>
</tr>
<tr>
<td>b. Excision by Dimer</td>
<td>23</td>
</tr>
<tr>
<td>c. Target Site Integration</td>
<td>23</td>
</tr>
<tr>
<td>IV. Research Objectives</td>
<td>24</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>25</td>
</tr>
<tr>
<td>I. RNA-Seq Analysis - Chrysaora chesapeakei Transcriptome</td>
<td>25</td>
</tr>
<tr>
<td>II. Primer Design</td>
<td>27</td>
</tr>
<tr>
<td>III. DNA Extraction</td>
<td>27</td>
</tr>
<tr>
<td>IV. DNA Isolation via Chelex 100</td>
<td>29</td>
</tr>
<tr>
<td>V. PCR Amplification of Predicted POGO Element</td>
<td>29</td>
</tr>
<tr>
<td>VI. Verification of Product by Agarose Gel Electrophoresis</td>
<td>30</td>
</tr>
<tr>
<td>VII. Sequencing and Bioinformatic Analysis</td>
<td>30</td>
</tr>
<tr>
<td>RESULTS</td>
<td>32</td>
</tr>
<tr>
<td>I. Identification of POGO transposable elements in RNA-Seq Library</td>
<td>32</td>
</tr>
<tr>
<td>II. Primers</td>
<td>33</td>
</tr>
<tr>
<td>III. PCR Amplification of Putative POGO Transposable Elements from C. chesapeakei gDNA using POGO Primers</td>
<td>36</td>
</tr>
<tr>
<td>IV. Sequence Analysis of C. chesapeakei POGO Amplicons</td>
<td>39</td>
</tr>
<tr>
<td>i. TR1 Consensus</td>
<td>44</td>
</tr>
<tr>
<td>ii. TR2 Consensus</td>
<td>46</td>
</tr>
<tr>
<td>iii. Variable Regions</td>
<td>49</td>
</tr>
<tr>
<td>a. Variable Region I (VR1)</td>
<td>50</td>
</tr>
<tr>
<td>b. Variable Region II (VR2)</td>
<td>51</td>
</tr>
</tbody>
</table>
V. Homology and Conserved Domains of Putative POGO........................................54
VI. Evidence for the presence of POGO TEs in other Cnidarians.........................58
DISCUSSION....................................................................................................................60
I. Bioinformatic analysis and Evidence for the presence of a partial POGO element.................................................................60
   i. Confirmation of Target Amplicon by Size Validation.............................60
   ii. Reliability of Sequence Data.................................................................60
   iii. Comparison to Existing POGK Genes.................................................61
II. Sequence Variation and Variable Regions........................................................62
   i. Single Nucleotide Polymorphisms..........................................................62
   ii. Heterozygosities....................................................................................62
   iii. Variable Regions................................................................................63
III. Significance and Implications of Conserved Domains....................................65
   i. Brinker DNA Binding Domain...............................................................65
   ii. HTH_Tnp_Tc5 Superfamily..................................................................67
IV. POGK, Genome Evolution, and Transposase Integration as means for Survival.................................................................67
   i. Genome evolution...............................................................................67
   ii. Domestication and Exaptation.............................................................68
   iii. POGO derived elements.................................................................68
V. Future Direction....................................................................................69
LITERATURE CITED......................................................................................................72
LIST OF FIGURES

Figure 1: Scyphozoan Anatomy.................................................................12
Figure 2: Scyphozoan Life Cycle...............................................................13
Figure 3: Class I and II Transposons.......................................................16
Figure 4: Structure of Tc1/mariner transposase......................................20
Figure 5: RNA-Seq Library, Keyword: POGO matched .........................34
Figure 6: Conserved Domain Database (CDD) View of Contig 22506........34
Figure 7: Nucleotide sequence of Contig 22506.....................................35
Figure 8: Agarose Gel of TR1 Samples....................................................37
Figure 9: Agarose Gel of TR2 Samples....................................................38
Figure 10: Geneious overview (raw): Initial import and alignment of unedited sequence data..............................................................40
Figure 11: Geneious overview (edited): Combined POGO consensus (TR1 and TR2).................................................................41
Figure 12: Blast2n combined POGO consensus vs. Contig 22506.............43
Figure 13: TR1 Consensus .....................................................................44
Figure 14: BLAST2n Contig 22506 vs. TR1............................................46
Figure 15: TR2 Consensus .....................................................................47
Figure 16: BLAST2n Contig 22506 vs. TR2............................................49
Figure 17: Variable Region 1 (VR1).........................................................50
Figure 18: Variable Region 2 (VR2).........................................................51
Figure 19: Clustal Omega of Contig 22506, TR1, and TR2..........................54

Figure 20: BLASTx TR1 Consensus Sequence............................................55

Figure 21: Conserved domain view of TR1 consensus..............................56

Figure 22: Agarose Gel of Multiple Cnidarian Templates............................58

Figure 23: Conserved Domains of known POGK gene.............................61

Figure 24: ClustalX alignment of Contig 22506, TR1, and TR2..................64

Figure 25: BLASTx putative BrkDBD of TR1 consensus............................66

LIST OF TABLES

Table 1: Primer sets 4 and 5.................................................................36
INTRODUCTION

I. Chrysaora Biology

*Chrysaora chesapeakei*, more commonly known as the Bay Sea Nettle, is a species of jellyfish found along the Eastern Coast spanning from southern New England to the Gulf of Mexico (Bayha *et al.*, 2017). *Chrysaora chesapeakei* belongs to the phylum Cnidaria. Members of this phylum include not only jellyfish, but also hydrozoans, sea anemones, and corals, all of which exhibit radial symmetry (Barnes, 1994). Like that of many animals, Cnidarians are known to possess a gut cavity that is lined by endoderm; however, due to its role in circulation and digestion it is referred to as a gastrovascular cavity (Barnes, 1994). Within cnidarians, evaginations of the body wall produce a circle of tentacles. The body wall typically consists of three layers which include an outer epithelium (epidermis), inner epithelium (gastrodermis), and an extracellular layer sandwiched between these two layers known as a mesoglea (Barnes, 1994).

In *Hydra* and many other hydrozoans, the mesoglea is simple and thin with a non-cellular basal lamina (Brusca *et al.*, 2016). In other Cnidarians, such as *Scyphomedusae*, the mesoglea can be a thick, fibrous and jelly-like connective tissue scattered with cells (Brusca *et al.*, 2016). Due to the presence of only two germ layers, cnidarians are known to be diploblastic.
*Chrysaora chesapeakei* is a “true” jellyfish and is a member of the class Scyphozoa. The medusoid body of Scyphozoans is reminiscent of a bell or umbrella (the iconic shape typically associated with jellyfish). The mouth is located on the concave underside of the medusa and tentacles typically hang down from the margin of the bell (Brusca *et al.*, 2016). The bell for this species typically varies in shape and has a scalloped margin to form lobes called lappets; the bell can grow up to 250-mm wide and exhibits semicircular tongue-shaped lappets (Kramp, 1961). Coloration of scyphozoans typically entails gonads and internal structures that are deeply colored relative to a
delicately tinted or transparent bell (Barnes, 1994). Coloration of *Chrysaora chesapeakei* varies between white, colorless, or red/brown (Bayha *et al.*, 2017). Members of this species have an average of 24 tentacles but may vary between individuals and are accompanied by four lappets per octant (Bayha *et al.*, 2017). Cnidocytes, the organelles that are used for ensnaring and immobilizing prey, can be found along the length of the oral arms and the fishing tentacles.

Figure 2. Scyphozoan Life Cycle (Image credit: Deretsky, National Science Foundation)

There are two major life stages which are known to exist and alternate within the class Scyphozoa. These include a sessile polyp form and a free-swimming medusa (Brusca *et al.*, 2016). It is generally believed that the medusae population dies off annually, whereas the scyphistoma (sessile polyp) may be active perennially (Calder, 1972). Including both the polyp and medusoid forms, there are a six distinct stages of the
scyphozoan life cycle (Figure 2). The medusa is generally dioecious and, upon maturation, males and females release their gametes from their gastrovascular cavity and exit through the mouth into surrounding water (Brusca et al., 2016). Fertilization results in larva formation known as a planula. The planula use cilia to navigate through the water and after three to five days, the planula settles upon a hard surface and develops into the polyp stage scyphistoma (Littleford, 1939). This scyphistoma is capable of reproduction by both asexual budding and podocyst formation (Calder, 1972). It is able perpetuate its life cycle and produce future progeny by the process known as transverse fission or strobilation; this latter process occurs at the polyp’s oral end (Brusca et al., 2016). In Chrysaora chesapeakei, medusae are formed by strobilation of the growing scyphistoma (now called a strobila) and stacked like saucers at the oral end of the structure. Once released, these stacked saucer-like discs are known as ephyrae (Barnes, 1994; Brusca et al., 2016). The newly liberated ephyrae of Chrysaora chesapeakei is measured to be 0.84-mm in diameter and will eventually grow into the adult medusa (Littleford, 1939).

II. Discovery of Jumping Genes and the Mobilome

A. Barbara McClintock and the Early Days at Cold Spring Harbor

Transposable elements, often used interchangeably with the term transposons, were first identified and documented by Barbara McClintock in 1950 (McClintock, 1950). McClintock’s work at the Cold Spring Harbor Laboratory during the years of 1944 through 1950 was primarily focused on that of the self-pollinating corn
plant, *Zea mays* (Makalowski *et al.*, 2012). She had observed and documented a phenomenon by which loci appeared to be unstable and mutable within the genome and as a result, one or more plant characters were shown to be directly affected by the movement of those loci (McClintock, 1950).

She had made an important observation – cultures of the self pollinating plants had shown approximately 40 mutable loci, whereas the parents of these plants showed no evidence of such a high number of mutable loci (McClintock, 1950). She believed that there was a modification mechanism and necessary factor associated with the phenomenon. This proposal was later supported and became what is known as the Ac-Ds transposable element in her famous 1953 publication *Induction of instability at selected loci in maize* (McClintock, 1953).

Unfortunately, the concept of mutable elements within the genome was highly antithetical to established dogma and McClintock’s work was dismissed by many and poorly received at the 1951 Cold Spring Harbor Symposium (Malakowski, 2012). It was not until her findings were validated by others, especially molecular confirmation of these elements in the genome, that McClintock’s work gained recognition. McClintock was awarded the Nobel Prize of Medicine for her findings in 1983.

**B. Selfish DNA**

For the past few decades, the phrase “selfish DNA” has been cited repeatedly throughout literature to describe transposable elements. This nickname is fitting considering transposable elements do not reliably increase the fitness or survival of their
host genome, rather, they exist solely to replicate their own genetic material.

Since its initial discovery, much has been learned about the Ac-Ds system in maize as well as many other types of mobile genetic elements. McClintock’s initial research represented just the tip of the metaphorical iceberg when it came to the mobilome. The term mobilome has been proposed to represent the sum of all mobile genetic elements (MGEs) within a genome (Frost et al., 2005). This includes transposons, insertion sequences, group I and II introns, as well as other mobile elements. Within transposons a major distinction has been made categorizing what are known as Class I and Class II transposable elements (Wicker, 2007).

**Class I:**

![Class I Transposon Structure](image)

**Class II:**

![Class II Transposon Structure](image)

**Figure 3.** Structure of Class I and II Transposons (Munos-Lopez and Garcia-Perez, 2010)

Class I transposable elements are RNA transposons which utilize a replicative mechanism, otherwise known as copy-and-paste, to produce an RNA intermediate via reverse transcription (Munos-Lopez and Garcia-Perez, 2010). Because of their mechanism of replication, class I transposable elements are often referred to as
retrotransposons. There are two open reading frames (ORF’s) located within these elements, named ORF1 and ORF2 respectively (see Figure 3). ORF1 encodes a nucleic acid binding protein, whereas, ORF2 encodes a protein with endonuclease and reverse transcriptase (RT) activity (Munos-Lopez and Garcia-Perez, 2010). If the retroelements’ main body is flanked by Long Terminal Repeats (LTRs), they can be further grouped into LTR or non-LTR retrotransposons. LTR-containing retrotransposons are of special interest because they exhibit a structure and lifecycle similar to that of retroviruses (Munos-Lopez and Garcia-Perez, 2010).

Class II transposons, in contrast, are DNA transposons. These work by a cut-and-paste mechanism in which the transposons are excised from one location and reintegrated into another (Munos-Lopez and Garcia-Perez, 2010). Molecular structure of DNA transposons involves a single transposase gene that is flanked on each side by terminal inverted repeats (TIRs) as seen in Figure 3 (Wicker, 2007). The TIRs serve as recognition sites for the transposase to perform excision and transposition of the element to a new location. A target site duplication (TSD) is made upon insertion of the element to its target; this duplication is a key characteristic of DNA Class II transposons. Class II transposons are further subclassified based upon their target sites, length and sequence of TIRs, and structural motifs within their transposase. Members of the Subclass I DNA transposons include the Tc1/mariner, PIF/Parbinger, Mutator, Transib, Merlin, hAT, CACTA, piggyback, and P element (Wicker, 2007). Subclass II transposons include Helitron and Maverick which lack TSDs and are replicated, yet do not induce breakage of
dsDNA upon insertion and subsequently do not produce TSDs (Feschotte and Pritham, 2007; Du et al., 2009).

The discoveries of miniature inverted-repeat transposable elements (MITEs), as well as copy-and-paste transposable elements without an RNA intermediate, have challenged the two class system of transposable elements (Wicker, 2007). These MITEs include transposable elements that work by a cut-and-paste mechanism yet rely on a separate autonomous transposable element (Wicker, 2007; Munos-Lopez and Garcia-Perez, 2010).

C. Transposable Elements and the Genome

Transposable elements can occupy a high portion of a species’ genome and have been identified in nearly all organisms including prokaryotes and eukaryotes (Munos-Lopez and Garcia-Perez, 2010). Statistically they comprise approximately 10% of several fish species, 45% of the human genome, and upwards of 80% in plants such as maize. Multiple strategies have been developed by the transposable elements to minimize reduction in fitness suffered by the host due to transposition; this is due to the fact that perpetuation of the transposon is tied to host survival. One strategy involves preferential insertion of the element into heterochromatin and non-essential regions within the genome, allowing for the transposable element to reduce its deleterious impact (Dimitri et al., 1997; Ikeda et al., 2007). Element activity during the germ-line or embryonic stage will allow only non-deleterious or mildly-deleterious insertions to occur as they are selected against during development (Kano et al., 2009). Host genomes have also
developed defense mechanisms to reduce transposon activity. DNA-methylation, which reduces the expression of transposable elements, RNA interference of the germ line, and inactivation of transposon activity by specific proteins have all been effective in doing so (Munos-Lopez and Garcia-Perez, 2010).

III. Tc1/mariner Superfamily

The Tc1/mariner superfamily of transposable elements, justly named after its two best-studied members, is believed to include the most widely distributed transposable elements within nature. They are represented in several taxa including rotifers, fungi, plants, fish, and mammals (Munos-Lopez and Garcia-Perez, 2010). Given their ability to transpose within a wide range of species, members of this superfamily have become a prospective candidate to be used as tools for genetic manipulation (Plasterk, 1999; Ivics et al., 2009). Of the elements within this expansive superfamily, all but ten are known to harbor mutations which render the element inactive; however, there are four which have been reconstructed including: Sleeping Beauty from salmonid-type fish, Frog Prince from Rana pipiens, Himar1 from the Horn Fly, and Hsmar1 from Homo sapiens (Munos-Lopez and Garcia-Perez, 2010).

i. Structure of the Transposon

The length of Tc1/mariner elements typically ranges between 1 and 2.4 kb and includes two terminal inverted repeat (TIR) regions varying from 17 to 1100 bp flanking an encoded transposase (Munos-Lopez and Garcia-Perez, 2010). The predicted size of
the encoded polypeptide was initially believed to be 272 amino acid residues (Rosenzweig, 1993). A range between 272-345 amino acids has been cited by some sources (Munos-Lopez and Garcia-Perez, 2010), however, cDNA analysis and in vitro studies (Vos et al., 1993; Vos et al., 1996) have suggested a transcript of approximately 345 amino acids which coincides with the majority of literature.

As is characteristic of all class II DNA transposons, the element must encode the transposase responsible for mediating its transposition. Transposases may or may not be reliant on host factors for effective transposition, however, most reports claim that transposition of the transposon is possible without host assistance. This is the case for transposases of both the Tc1 element from Caenorhabditis elegans and the mariner element from Drosophila mauritania, as they alone, have been shown to be sufficient for transposition in vitro (Vos et al., 1996; Lamp 1996).

ii. The Transposase

Although the sequences for transposases of members belonging to the Tc1/mariner superfamily may differ between subfamilies and the species in which they reside, the transposases of Tc1/mariner elements are comprised of three distinct domains, two of which, are conserved throughout the clade (Plasterk, 1999). These domains include a DNA-binding domain, a nuclear localization signal domain (NLS), and a catalytic domain (Ivicz, 1996).
Figure 4. Structure of Tc1/mariner transposase (Munos-Lopez and Garcia-Perez, 2010)

a. DNA Binding Domain

The major structure-function analysis of the transposase genes has been focused on the N-terminal DNA-binding domain (Plasterk, 1999). This domain is nestled within the amino-terminal region and functions to recognize terminal inverted repeats (TIRs) flanking the gene (Munos-Lopez and Garcia-Perez, 2010). Sequence analysis and prediction of secondary structures has led to the proposal of two helix-turn-helix (HTH) motifs (Pietrokovski and Henikoff, 1997) in Tc1 and mariner elements and a single HTH motif in POOG elements (Wang, 1999). The bipartite binding domain of Tc1 and mariner is comprised of two subdomains corresponding to the two HTH motifs found in these elements: the first is a paired domain and the second - a homeobox domain (van Pouderoyen et al., 1997).

b. Nuclear Localization Signal Domain

The nuclear localization signal (NLS) domain is the only domain not conserved in the Tc1/mariner superfamily and found only within Tc1 and mariner elements. The NLS motif is needed for transport across the nuclear envelope and a motif thought to be linked to the interaction of transposase monomers known as the WVPHEL motif (Bouuaert et al. 2014; Munos-Lopez and Garcia-Perez, 2010). The NLS itself has been shown to partially overlap the binding domain (specifically, the C-terminal end of the homeobox domain) in Tc1 and mariner transposases (Ivicz 1996).
c. Catalytic Domain

The third domain of the transposase is a carboxy-terminal domain that harbors a catalytic motif consisting of three conserved amino acid residues; these are DDE (Asp, Asp, Glu) in Tc1-like elements and, alternatively, DDD (Asp, Asp, Asp) in mariner and POGO elements (Plasterk, 1999). Given its role in DNA cleavage and joining reactions along with the presence of these motifs in other transposases and recombinases (Doak, 1994), this domain has become the putative catalytic domain. Moreover, site-directed mutagenesis of these residues, as described in van Leunen et al. (1994) and Lohe et al. (1997), support this claim by demonstrating inactivation of the transposase.

iii. Mechanism of Transposition

Members of the Tc1/mariner superfamily are Class II DNA transposons and, thus, mobilization of their elements is not replicative and works by a cut-and-paste process. There are four steps required for the mobilization of a Tc1/mariner transposable element.

a. Cut-site Recognition and Cleavage

The first step involves recognition and binding of TIRs by the HTH motifs of two transposase molecules; bound regions form what are known as the SECs (Single-End-Complex) (Munos-Lopez and Garcia-Perez, 2010). Secondly, through hydrolysis of a phosphodiester bond, the 5’ ends of both TIRs are cleaved to liberate the non-transferred
strands which do not participate within the transposition process (Munos-Lopez and Garcia-Perez, 2010).

b. *Excision by a Dimer*

Thirdly, both transposase molecules interact, drawing the ends together and forming a transposase dimer known as the “Paired End Complex” or PEC. Simultaneously, hydrolysis of the phosphodiester bonds on the 3’-ends produce the transferred strands (Munos-Lopez and Garcia-Perez, 2010).

c. *Target Site Integration*

The last portion of this four-step mechanism involves binding of the PEC to the target site to form the *Target Capture Complex*; it is at this complex where insertion takes place. The dinucleotide TA serves as the insertion site for the transposase, thus, any TA dinucleotide within the genome may be selected at random for insertion of the element (Rosenzweig, 1983). The 5’-end of target DNA undergoes nucleophilic attack by the 3’ OH group of the transferred strand and gaps are filled in by the host which generate target site duplications that flank the newly inserted transposon (Munos-Lopez and Garcia-Perez, 2010).
IV. Research Objectives

Transposable elements are continuously being discovered in many different genomes, either deliberately or accidentally during sequencing studies (Robertson, 1997). PCR amplification and sequencing of these elements allows for comparison of related transposable elements through homologous sequences. Sequence studies have been done extensively with the Ac-Ds transposable element family of maize, with 903 elements shown to exist within the corn genome (Du et al., 2011). Members of the Tc1/mariner superfamily are of particular interest based upon their relatively small size, broad distribution in many organisms, and initial reports of their presence in the cnidarian Hydra (Robertson, 1997).

The aim of my thesis research is two-fold:

1. To analyze an RNA-seq library recently created for Chrysaora chesapeakei for the presence of Tc1/mariner superfamily transposons.
2. To use these putative transcripts to design PCR primers which will permit me to amplify, sequence, and verify the gDNA of these transposable elements in C. chesapeakei.
MATERIALS AND METHODS

I. RNA-Seq Analysis - *Chrysaora chesapeakei* Transcriptome

i. Isolation of Total RNA from *Chrysaora chesapeakei*.

Total RNA was isolated from the tentacles of a single medusa collected from the Cattus Island region of Barnegat Bay (collected August 10, 2013). This individual was transported back to the laboratory and washed several times in sterile artificial seawater (19 ppt). It was kept alive for 2 days to allow time for all gut contents to be expelled. It was then rinsed again with artificial seawater to remove any other (non-jellyfish) DNA/RNA. Tentacles were frozen in liquid nitrogen and ground to a fine powder with a homogenizer. Total RNA was isolated using the Qiagen RNaseasy Plus MicroKit (Cat No./ID: 74034) following the manufacturer’s instructions.

ii. Preparation of NGS Library.

Library preparation was performed by GeneWiz, Inc. (South Plainfield, NJ) and included separating out poly A+ RNA (to eliminate or minimize the inclusion of rRNA and tRNA), construction of a cDNA (complementary DNA) library by reverse transcription, and shearing of cDNAs to produce fragments ranging from 100 to 200 bp in length. Ends of dsDNA were repaired and adaptors ligated to ends to permit multiplexing of samples.
iii. **NGS Sequencing.**

DNA was sequenced on an Illumina HiSeq 2500 platform using 2 x 100 paired ends. Approximately 380,000,000 reads were generated from this run from triplicate samples.

iv. **Contig Assembly.**

Raw sequence data were processed by eliminating sequences with low quality scores, removal of adaptor sequences, and then assembling using CLC Workbench to generate a file of 87,600 contigs (JG01-CQTTotalRNA-Contigs.fasta). The data were organized as a series of fasta files, with the first line indicating the contig number and the approximate coverage of the assembled sequence.

v. **BLAST Search.**

This file of assembled contigs was BLASTed against the nr database of Genbank (this is the complete Genbank collection of all known sequences) and the best hit (highest score match or lowest E or Expect value) was recorded in a second file (rna.nr.best.hit.complete.xlsx). Alternatively, BLASTn, BLASTx, and BLASTp were utilized in analyzing this data set. In addition, alignment of multiple nucleic acid or protein sequences was accomplished by CLUSTAL Omega (Sievers *et al.* 2011; http://www.ebi.ac.uk/Tools/msa/clustalo/).
II. Primer Design

The PrimerQuest Tool (https://www.idtdna.com/PrimerQuest/Home/Index) was used to generate prospective oligos; to optimize amplification coverage, two sets of overlapping primers generating the largest predicted amplicon sizes were chosen. The OligoAnalyzer Tool (https://www.idtdna.com/calc/analyzer) was utilized to ensure any secondary structures such as hairpins and potential dimers were within acceptable range in order to preserve yield efficacy of desired product.

Lyophilized primers were resuspended to 100 μM concentrations according to the manufacturer’s protocol. From this, a 10 μM working stock was created using sterile water and stored at minus 20°C for use in the PCR reactions described in later sections.

III. DNA Extraction from Chrysaora chesapeakei

i. Collection and storage of Jellyfish Samples

Specimens were kept alive for a period of 2 days prior to storage giving adequate time for expulsion of gut contents. Jellyfish were rinsed with artificial seawater (19 ppt) to remove any unwanted non-jellyfish DNA/RNA. Tentacles were preferentially used for production of genomic DNA as to avoid contamination by any gut cavity contents not previously expelled.

ii. DNA Isolation via CTAB

Glycosaminoglycans and polyphenolic proteins may interfere with processing of nucleic acids and thus reduce their quality for use in restriction-endonuclease digestion,
cloning, and PCR (Winnepenninckx et al., 1993). A protocol for extracting high molecular weight DNA from mollusks, Winnepenninckx et al. (1993), was adapted for jellyfish genomic DNA isolation. Details for the preparation and use of isolation buffer can be found in Walsh et al. (1991) and Restaino (2013).

iii. Purification of DNA

Following incubation, 0.5 mL of chloroform:isoamyl alcohol (24:1) was added to each sample and then gently mixed for 2 minutes by inverting tubes. Tubes were spun for 10 minutes at maximum speed (14,000 x g) in a microcentrifuge at 4°C. The upper aqueous phase was transferred into a new 1.5 mL eppendorf tube being careful not to transfer any of the solid material at the interphase. One μL of RNase A (10 mg/mL) was added to the tubes containing supernatant and then incubated for 30 minutes at 37°C. Following incubation, 2/3 volume of isopropanol was added to each tube and then inverted gently to mix.

iv. Elution and Washing of DNA

Tubes were allowed to sit at room temperature for 2 hours and then spun at 14,000 x g at 4°C to pellet DNA. Being careful not to disturb the pellet, supernatant was removed and samples were washed twice with ethanol (500 μL of 70% EtOH added to each tube and spun for 15 minutes at 14,000 x g at 4°C). Any remaining supernatant was removed and pellets were dried briefly (5 min) in the Speed-Vac without heating.

v. Resuspension of DNA

DNA pellets were resuspended in a minimum volume (20 μL) of TE buffer (10 mM TRIS, 1mM EDTA, pH 8.0). Concentration and purity of DNA samples were
checked by UV absorption with the NanoDrop ND-1000 spectrophotometer. A small aliquot of each sample was run on a 1.0 % (w/v) agarose gel to check for quality and size of DNA fragments. Samples were stored at -20°C.

IV. DNA Isolation via Chelex 100

In some instances, a relatively quick and crude method for DNA extraction was also employed using the chelating agent Chelex-100 (Walsh et al., 1991). Tentacle samples were homogenized in sterile 1.5 mL eppendorf tubes using a micropesle. 100 μL of 5% (w/v) Chelex-100 was added to each tube and then placed in a hot water bath and boiled for 10 minutes. Samples were then vortexed and placed on an ice bath for 2 minutes. Tubes were centrifuged (14,000 x g) for 10 minutes and the resulting supernatants were transferred to 1.5 mL sterile tubes and then stored (-20°C) until used in PCR reactions. DNA concentration and quality was assessed by NanoDrop analysis as specified above.

V. PCR Amplification

PCR amplifications were performed using the Veriti Thermal Cycler (Applied Biosystems, Inc.). Unless otherwise specified, parameters for PCR were as follows: 95°C for 2 min (1X); 95°C for 30 s, 52°C for 60 s, and 72°C for 75 s (40X); 72°C for 10 min (1X); with samples held at 4°C. Reactions were prepared using ChoiceTaq Master Mix (Denville Scientific, Denville, NJ; http://www.denvillescientific.com), however, suggested reaction volumes (50 μL) were scaled to accommodate a 20 μL reaction.
Typically, 10 μL ChoiceTaq (2X stock), 7 μL of sterile ddH₂O, 1 μL of forward primer (10 μM), 1 μL reverse primer (10 μM), and 1 μL template DNA were combined into sterile 200 μL dome-capped PCR tubes.

VI. Agarose Gel Electrophoresis

Confirmation of purity and size of amplicons was determined by agarose gel electrophoresis. Samples were run on 1% (w/v) agarose gels in 1X TAE (40 mM Tris-Acetate, 1 mM EDTA) buffer. DNA was visualized by incorporation of 1X SYBR Safe (10,000X concentrate in DMSO [Invitrogen]) into gel with visualization by blue light (470 nm).

VII. Sequencing and Bioinformatic Analysis

i. Automated Dideoxy Sanger Sequencing

Amplicons consisting of single bands of sufficient intensity were submitted for DNA sequence analysis. DNA sequencing was performed using an ABI 3130 Genetic Analyzer (Applied Biosystems Inc, Foster City, CA 94404) in conjunction with the BigDye Terminator Cycle Sequencing Kit Version 3.1 (Applied Biosystems Inc, Foster City, CA 94404) with reactions diluted to 1/16 of the manufacturer’s recommended protocol. Samples were separated on a 36-cm column array and NANOPOPTM 7 polymer (MCLAB, South San Francisco, CA 94080, NP7-100; http://www.mclab.com). Removal of unincorporated ddNTP dye terminators from sequencing reactions was performed using EdgeBio Performa DTR Gel Filtration Cartridges (Gaithersburg, MD;
Sequence calls were made using the KB Basecaller.

ii. Preparation of samples for sequencing

Ten μL samples were prepared for sequencing by combining 7 μL ddH₂O, 1.5 μL of either forward or reverse primer, and 1.5 μL of template DNA. Both a forward and reverse reaction was run for each template. In cases where an amplicon was judged to contain a high concentration of DNA, as verified by bright band intensities, 1:20 dilutions of each sample was used as template.

iii. Bioinformatic Analysis

Chromatograms were visualized using the 4 Peaks Software package (Nucleobytes, http://nucleobytes.com/4peaks/idex.html). Forward and reverse sequences were aligned using the BLAST2Seq function of the BLAST (http://blast.ncbi.nlm.nih.gov) search algorithm (Altschul et al., 1990) and then manually proofread and edited by use of chromatogram peak data. Individually edited sequences were then aligned by use of Geneious R10 (www.geneious.com) bioinformatic software; any remaining inconsistencies were edited, and final contig consensus sequence was constructed from overlapping segments. Alignment of multiple nucleic acid or protein sequences was accomplished by CLUSTAL Omega (Sievers et al. 2011; http://www.ebi.ac.uk/Tools/msa/clustalo/).
RESULTS

I. Identification of POGO transposable elements in RNA-Seq Library of *Chrysaora chesapeakei*.

Assembly of the raw sequence data generated by the RNA-Seq library using CLC Workbench generated a file of 87,600 contigs (JG01-CQTTotalRNA-Contigs.fasta). BLASTx analysis of these assembled contigs against the nr database of Genbank generated the best hit (highest score match or lowest E or Expect value) for each contig. Of the original 87,600 contigs in this transcriptome, 30,817 (35.18%) had BLASTx hits with significant e values ($<10^{-4}$) and these data were recorded in a second file (rna.nr.best.hit.complete.xlsx). Interestingly, this implies that nearly 65% of the *C. chesapeakei* transcriptome is unknown. A total of 6 matches (0.02%) was found to POGO transposable elements with KRAB-like domains (Krüppel Associated Box Domain; see Figure 5), including two chordates (*Danio rerio*, *Odobenus rosmarus divergens*), an arthropod (*Metaseiulus occidentalis*), a mollusc (*Crassostrea gigas*), a sponge (*Amphimedon queenslandica*), and a cnidarian (*Hydra vulgaris*). This domain represents a highly conserved motif related to the Krüppel protein of *Drosophila* (Schuh *et al.*, 1986). KRAB was initially characterized in N-terminus of more than 300 zinc-finger proteins in humans and later to be present in approximately one-third of all zinc finger proteins and acts as a transcriptional repressor (*Bellefroid et al.*, 1991; Witzgall *et al.*, 1994). A BLASTx search of Contig 22506 identified several significant CDD (Conserved Domain Database; Marchler-Bauer *et al.*, 2010)
hits to proteins found in POGO transposable elements (Figure 6). As demonstrated in Figure 6, significant translational matches are seen to HTH_Tnp_Tc5 (Helix-Turn-Helix Tc5 transposase DNA-binding domain), CENPB (the putative DNA-binding domain in centromere protein B, mouse jerky and transposases), and BrkDBD (the Brinker DNA-binding domain).

II. Primers

The largest contig assembled in this group from *C. chesapeakei* (contig 22506) was 1463 nt long and matched the POGO element from the mite, *Metaseiulus occidentalis*. To verify that this contig was valid, primers for the amplification and sequencing of this putative POGO element were designed and synthesized by Integrated DNA Technologies (www.idtdna.com). Primers were designed using the nucleotide sequence of contig 22506 from the RNASeq library as a template.
Figure 5. Shown are matches in the RNA-Seq library of *Chrysaora chesapeakei* found by searching keyword: POGO. Contig 70530 (boxed in red) is a match to *Hydra vulgaris* - a hydrozoan also in the phylum Cnidaria. Contig 22506 (boxed in blue) was used as the scaffold for creating primers as it represented the largest nucleotide sequence (1463 nt) from this dataset.

Figure 6. Identification of Conserved Domain Database (CDD) hits with BLASTx search of Contig 22506 from *Chrysaora chesapeakei*. Significant matches to HTH_Tnp_Tc5 (Tc5 transposase DNA-binding domain), CENPB (the putative DNA-binding domain in centromere protein B, mouse jerky and transposases), and BrkDBD (the Brinker DNA-binding domain).
**Figure 7.** Contig 22506 derived during generation of the RNA-Seq library. There are 1463 nt of sequence derived with an average coverage of 31.85X. Visually represented are both overlapping sets of primers. Depicted in yellow is primer set 4F and 4R and in green, primer set 5F and 5R. Predicted amplicon size for primer sets 4 and 5 are 787 bp and 505 bp, respectively.

As indicated in the materials and methods, primers were generated using the PrimerQuest tool from IDT DNA. Two primer sets were chosen based upon several factors including the length of their predicted amplicons, matching GC content, and position relative to the contig used as a scaffold. Melting temperature ($T_m$) was also a factor in selection as a way preserving conditions between them when optimizing their cycling parameters. Overlapping primer sets also provided the opportunity for different combinations, which in turn, gave rise to amplicons of varying lengths to be used in sequence analysis.
Table 1. Primers used for PCR amplification and DNA sequence analysis in this study. G/C Content and T_m values were obtained from the datasheet provided with each primer. Primers were synthesized by Integrated DNA Technologies (Coralville, IA). Start and stop positions refer to orientation of the 5’ and 3’ ends of a primer relative to contig 22506 (see Figure 7).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>G/C Content</th>
<th>T_m</th>
<th>Start</th>
<th>Stop</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>4F</td>
<td>5’ TGACAAGAGAGGTGCGTACAATAG 3’</td>
<td>43.5%</td>
<td>54.4°C</td>
<td>253</td>
<td>276</td>
<td>23 bp</td>
</tr>
<tr>
<td>4R</td>
<td>5’ TGTCTCAGGGCAAGTTAATCA 3’</td>
<td>40.9%</td>
<td>54.2°C</td>
<td>1018</td>
<td>1040</td>
<td>22 bp</td>
</tr>
<tr>
<td>5F</td>
<td>5’ CAAGAGAAAGCCACAGAAATGAG 3’</td>
<td>43.5%</td>
<td>54.1°C</td>
<td>824</td>
<td>847</td>
<td>23 bp</td>
</tr>
<tr>
<td>5R</td>
<td>5’ GAAATCCTACAACACTCCCTAGAA 3’</td>
<td>41.7%</td>
<td>53.9°C</td>
<td>1305</td>
<td>1329</td>
<td>22 bp</td>
</tr>
</tbody>
</table>

III. PCR Amplification of Putative POGO Transposable Elements from C. chesapeakei gDNA using POGO Primers.

The results of the PCR amplification of POGO transposable elements from C. chesapeakei gDNA (TR1) are shown in Figure 8. Lanes 1 and 10 correspond to the HiLo DNA Ladders included for band size prediction. Bright amplicons in lanes 2, 4, 6, and 8 correspond to template DNA pairing with varying combinations of primers (4F + 4R, 4F + 5R, 5F + 5R, and 5F + 4R). Each amplification containing a unique pairing of primers was also run with a corresponding No Template Control (NTC) in the neighboring lane. As can be seen from the gel, all of the NTC’s are clean suggesting no contamination of template DNA in reagents or buffers. The faint band seen migrating at 50 bp is presumed
to be primers. The fact that there is no primer band in lanes where amplicons are produced is expected since primers have been consumed in the PCR reaction.

**Figure 8**: PCR of *C. chesapeakei* DNA (TR1) samples run on 1% (w/v) agarose gel. Lanes 1 and 10 are Hi-Lo ladders. Following each positive lane, a NTC (no template control) was run utilizing the same primer pair. Primer pairs are as follows - Lane 2: 4F+4R. Lane 4: 4F+5R. Lane 6: 5F+5R. Lane 8: 5F+4R.
Figure 9: PCR of *C. chesapeakei* DNA (TR2) samples run on 1% (w/v) agarose gel. Lanes 1 and 10 are Hi-Lo ladders. Primer pairs are as follows - Lane 2: 4F+4R. Lane 5: 4F+5R. Lane 6: 5F+5R. Lane 8: 5F+4R. There was no amplification in lane 8.

Figure 9 shows the results of the PCR reactions of the *C. chesapeakei* DNA (TR2) sample with the POGO primers. Again, lanes 1 and 10 correspond to the HiLo DNA Ladders included for band size prediction. Each DNA sample was again run with a NTC in the lane beside it, however, lane pair 8 and 9 lack any banding. This suggests that this one sample did not amplify. Lane pair 4+5 are reversed; in this case lane 4 is the NTC and lane 5 is the positive (using primers 4F + 5R).
Bands are visible, however, overall band intensity is low in comparison to Gel 1. This was most likely due to insufficient SYBR safe incorporation into gel or issues in capturing a digitized image of the gel. In all cases, however, the amplicons produced matched the predicted size based on the primer positions in the original contig (22506) used to design putative POGO primers.

IV. Sequence Analysis of *C. chesapeakei* POGO Amplicons

All amplicons produced using the sets of POGO primers were subjected to sequencing by automated Sanger Dideoxy sequencing on an ABI 3130 platform as described in Materials and Methods. Amplicons were only used if they produced a clean band (no other contaminating bands present), of sufficient quantity, and of the correct size for the primers used. In all cases both forward and reverse strands were sequenced and used to edit a final, correct sequence of the amplicon.
Figure 10: Initial alignment of sequence data generated by various combinations of the POGO 4F, 4R, 5F and 5R primers. Electropherograms shown above were those of sequences prior to being edited. Sequences were imported and aligned combining overlapping sequences using the de novo assembly tool of Geneious to generate a full consensus sequence of 1,106 bp. A large proportion of the contigs are highlighted in blue indicating that the base calls are of a high quality.

Figure 10 illustrates a baseline reference. Sequences were later aligned and mapped to a consensus that was assembled based upon regions where sequences had sufficient overlap to one another. The consensus, shown at the top of the figure, utilized a quality color scheme that assigns a shade of blue to each base based upon its quality. As per the Geneious user manual: Dark blue for confidence < 20, blue for 20 - 40 and light blue for > 40. There is a large portion near the start of the generated consensus sequence that was dark blue representing low scoring. Sequences were edited improve consensus scoring quality.

All forward and reverse generated sequences were aligned using the BLAST2Seq tool in Genbank (NCBI). Sequences were then edited manually using 4Peaks and base
call edits were marked directly on the electropherograms and saved prior to importing edited sequences to Geneious for alignment and assembly.

There were several instances in electropherograms where individual peaks were perfectly overlapping. It was assumed that these represented heterozygosities in the amplified gDNA sequences. This might come about by having more than one copy of the putative POGO transposable element in the genome with alleles that have SNP variants at particular positions (data not shown). These potential heterozygosities were only seen in samples of the TR2 template.

Figure 11: Alignment of edited POGO amplicon consensus. Sequences from TR1 and TR2 were all edited pairwise in 4Peaks and then re-aligned in Geneious to create an updated consensus. New consensus was highlighted in teal for the majority of contig indicating a higher degree of reliability. Gaps and individual nucleotide differences may be hard to visualize based upon image size. Electropherograms have been color reduced due to size.
As seen in Figure 11, confidence of nucleotide sequence appeared to be quite high post-editing as indicated by the teal colored bar of the full POGO consensus sequence seen at the top of the figure. There were however, various nucleotide differences interspersed amongst the sequences that were overlooked and not accounted for, simply because of the software’s quality threshold algorithm. Moreover, the software will assign a nucleotide at a given position if at least half of the electropherograms include a base there, making it non ideal for identifying deletion or insertion mutations between members of a given population. I will address those concerns shortly.

To assess the quality of this initial assembly of amplicon sequences of *C. chesapeakei* POGO, I aligned this consensus assembly (generated by combining amplicon sequences of both TR1 and TR2) to the *C. chesapeakei* contig 22506 that was generated from the original RNA-Seq library. As can be seen in Figure 12, the homology between the two sequences was very high (98.3%), with 18 mismatches and 11 gaps in 1,067 bp using the BLAST2Seq alignment tool of Genbank. In an effort to generate the most accurate sequence data, I manually edited the electropherograms in Geneious for the TR1 and TR2 samples separately since I suspected that some of this variability may be due to SNP’s present in these different DNA sources.
**Figure 12:** Blast2Seq alignment of the combined *POGO* consensus sequence against the nucleotide sequence of RNA-Seq Contig 22506.
i. **TR1 Consensus Sequence**

The figure below (Figure 13) shows the electropherogram alignments in Geneious generated from the TR1 template. Ends of the contigs were trimmed to generate the highest quality sequence. These regions are represented as red bars in Figure 13. This edited TR1 consensus is teal for the entirety of its length representing a high quality consensus.

![Figure 13](image)

**Figure 13:** Alignment of amplicon sequences 1 through 8 generated from TR1. Highlighted in yellow are two conserved domains: BrkDBD and HTH_Tnp_Tc5 found in all POGO transposable elements. Potential ORF’s are indicated by the orange bars underneath each chromatogram. Electropherograms have been color reduced and compressed to include the entire assembly.

The TR1 consensus was aligned to the original RNA-Seq POGO contig (22506) using the BLAST2Seq algorithm (Figure 14). Nucleotide differences are highlighted in blue. This shows better identity between these two sequences than seen previously in
Figure 12. Identities are 99% with 11 gaps found in two clusters, one a 5 nt and a second a 6 nt insertion, found in the TR1 consensus sequence.

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1799 bits(1994)</td>
<td>0.0</td>
<td>1015/1027(99%)</td>
<td>11/1027(1%)</td>
<td>Plus/Plus</td>
</tr>
</tbody>
</table>

Query 135 TGAATCTACAACACTCCCTAGAAGCATATATGTTTCAAACAGGGTTATGGGAGTTT 194
Sbjct 1 TGAATCTACACTCCCTAGAAGCATATATGTTTCAAACAGGGTTATGGGAGTTT 60

Query 195 TTACACCTACAGTCAGCCAGATTCAATAAGAGTTTAAAAGTGTTACAATGTCATATACATA 254
Sbjct 61 TTACACCTACAGTCAGCCAGATTCAATAAGAGTTTAAAAGTGTTACAATGTCATATACATA 120

Query 255 TATACAGTTTTTAAGTTTACACTTATAACGAAGCACTCTTCTTGTAAATGCTTGCAGTG 314
Sbjct 121 TATACAGTTTTTAAGTTTACACTTATAACGAAGCACTCTTCTTGTAAATGCTTGCAGTG 180

Query 315 TGAATTTGTTTTGAGCTGAATA----TGATACAAATGTTTGAAGTTTACATTCTATCATGAGT 369
Sbjct 181 TGAATTTGTTTTGAGCTGAATA----TGATACAAATGTTTGAAGTTTACATTCTATCATGAGT 240

Query 370 TTGGTAAAGACTTGTTACAGTGCTCGCCACCTTGACAGCATTTGCATATGCTCT 429
Sbjct 241 TTGGTAAAGACTTGTTACAGTGCTCGCCACCTTGACAGCATTTGCATATGCTCT 300

Query 430 CAGGGCAAGTTAAATCATCCACTCCTAAAAATCCTACCATACATTGTTTGGTTTCCCATTTTGATTTT 489
Sbjct 301 CAGGGCAAGTTAAATCATCCACTCCTAAAAATCCTACCATACATTGTTTGGTTTCCCATTTTGATTTT 360

Query 490 CATTTTCAGTATTACCTTGTTGGTGGTGGCTGACACTTCAATGTAGCACAATTACTCTGCC 549
Sbjct 361 CATTTTCAGTATTACCTTGTTGGTGGTGGCTGACACTTCAATGTAGCACAATTACTCTGCC 420

Query 550 GCTTTAAGAACTCTCCACCCGGAAGTTTATAAGCAGTGAAACTTACCATTTGGGAGAGAG 609
Sbjct 421 GCTTTAAGAACTCTCCACCCGGAAGTTTATAAGCAGTGAAACTTACCATTTGGGAGAGAG 480

Query 610 CAGTACACTTATTTCTGTTCGTTTCTCTGTATTACTGATTCGATTTTCTCTCTCT 669
Sbjct 481 CAGTACACTTATTTCTGTTCGTTTCTCTGTATTACTGATTCGATTTTCTCTCTCT 540

Query 670 CTCTTAACATCATCATTAAAACCCACCTCCAAAAAACAAATTGTTTAAAATGTATATAGCTAC 729
Sbjct 541 CTCTTAACATCATCATTAAAACCCACCTCCAAAAAACAAATTGTTTAAAATGTATATAGCTAC 600
Figure 14: BLAST2Seq alignment of Contig 22506 obtained from the RNA-Seq library with the TR1 consensus. Nucleotide differences are highlighted in blue. Note the two large blue gaps that suggest insertions (5 nt and 6 nt) in TR1 DNA. Other than these two gaps, there is only a single SNP present at position 485 (T replacing A) in the TR1 consensus sequence.

ii. TR2 Consensus Sequence

The figure below (Figure 15) shows the electropherogram alignments in Geneious generated from the TR2 template. Ends of the contigs were trimmed to generate the highest quality sequence. These regions are represented as red bars in Figure 15. This edited TR2 consensus is teal for the entirety of its length representing a high quality
The consensuses produced from TR1 and TR2 sequences were compared with Contig 22506 separately and can be found in figures 14 and 15.

**Figure 15:** Alignment of sequences 9 through 14 amplified from *C. chesapeakei* TR2 DNA. Highlighted in yellow are the two conserved POGO domains: BrkDBD and HTH_Tnp_Tc5. ORF’s are indicated by the orange bars underneath each chromatogram. Electropherograms have been color reduced and compressed to include the entire assembly.
<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1748 bits (1938)</td>
<td>0.0</td>
<td>995/1013 (98%)</td>
<td>7/1013 (0%)</td>
<td>Plus/Plus</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query</th>
<th>Sbjct</th>
</tr>
</thead>
<tbody>
<tr>
<td>136</td>
<td>GAAATCCCTACAACACTCCCTCAAGAAAGCATATAGTTTTCAACAAAGGTTTATGGGAGTTTT195</td>
</tr>
<tr>
<td>196</td>
<td>TACACCTACAGTACCGACATCAATAGGTTTAAAGGTGTTTAACTATACATAT255</td>
</tr>
<tr>
<td>61</td>
<td>TACACCTACAGTACCGACATCAATAGGTTTAAAGGTGTTTAACTATACATAT120</td>
</tr>
<tr>
<td>256</td>
<td>ATACAGTTTTTTAAGTTTACCTTCAATTCGAAGCGAGCTCTCTTCTTGTAAATGCTTGCAGTGCT315</td>
</tr>
<tr>
<td>121</td>
<td>ATACAGTTTTTTAAGTTTACCTTCAATTCGAAGCGAGCTCTCTTCTTGTAAATGCTTGCAGTGCT180</td>
</tr>
<tr>
<td>316</td>
<td>GAAATGGTTTTAGAGCTCAAGATCATGTTTTGTTGTTGTTTCAAACCTTTGGGATGTTTT240</td>
</tr>
<tr>
<td>60</td>
<td>GAAATGGTTTTAGAGCTCAAGATCATGTTTTGTTGTTGTTTCAAACCTTTGGGATGTTTT</td>
</tr>
<tr>
<td>241</td>
<td>AGACTGTTAGCTGCTGCATTCTCGCCACTGAGGCATTGGCATCATATGTCTCAGGGC435</td>
</tr>
<tr>
<td>181</td>
<td>GAATTTGTTTTAGAAATATAGCTCAATAGGTTTAAAGGTGTTTAACTATACATAT240</td>
</tr>
<tr>
<td>361</td>
<td>AGACTGTTAGCTGCTGCATTCTCGCCACTGAGGCATTGGCATCATATGTCTCAGGGC300</td>
</tr>
<tr>
<td>301</td>
<td>GAAATTTAATCATCCACTATATACATGTTTTGTTGTTGTTTCAAACCTTTGGGATGTTTT240</td>
</tr>
<tr>
<td>421</td>
<td>AGACTGTTAGCTGCTGCATTCTCGCCACTGAGGCATTGGCATCATATGTCTCAGGGC480</td>
</tr>
<tr>
<td>496</td>
<td>CAGTTATTACCTGTGTTGCAATGCAATAAAAATATACATAGTCAATATGCTTTA555</td>
</tr>
<tr>
<td>361</td>
<td>CAGTTATTACCTGTGTTGCAATGCAATAAAAATATACATAGTCAATATGCTTTA</td>
</tr>
<tr>
<td>556</td>
<td>AGAATCTCTCCAACCAGGTTTATAAGAGCTCAACATTTTCTTCTGGGGGAGAGCAGTAA615</td>
</tr>
<tr>
<td>421</td>
<td>AGAATCTCTCCAACCAGGTTTATAAGAGCTCAACATTTTCTTCTGGGGGAGAGCAGTAA</td>
</tr>
<tr>
<td>616</td>
<td>CACTCATTTCTGCGCTTCCTTTCTTCTGTAATCTGCTGCTGATACTCTTCTCCTTA675</td>
</tr>
<tr>
<td>481</td>
<td>CACTCATTTCTGCGCTTCCTTTCTTCTGTAATCTGCTGCTGATACTCTTCTCCTTA</td>
</tr>
<tr>
<td>676</td>
<td>ACATCATCAACCCCACTCAAATACCAATTGATTTCTAAATGTTTAAATACTCAGCTTCTC735</td>
</tr>
<tr>
<td>541</td>
<td>ACATCATCAACCCCACTCAAATACCAATTGATTTCTAAATGTTTAAATACTCAGCTTCTC</td>
</tr>
<tr>
<td>736</td>
<td>TTCTCGACCCTCCATATTCTTTTTCTCTTCTTGGAGACGATAAACAAGGTTTATTTTT795</td>
</tr>
<tr>
<td>601</td>
<td>TTCTCGACCCTCCATATTCTTTTTCTCTTCTTGGAGACGATAAACAAGGTTTATTTTT</td>
</tr>
<tr>
<td>796</td>
<td>TCTTCCATTCTCTTATTCTTTTTATCTGAAACATTAATATGTTTCTGGGTCTGGCTGTGT855</td>
</tr>
<tr>
<td>661</td>
<td>TCTTCCATTCTCTTATTCTTTTTATCTGAAACATTAATATGTTTCTGGGTCTGGCTGTG720</td>
</tr>
</tbody>
</table>
iii. Variable Regions Within *C. chesapeakei* POGO Elements

A closer look at the aligned sequences revealed a very distinct pattern that may have been easily overlooked amongst them. In all occurrences of nucleotide differences, three of six exhibited one genotype, while others, exhibited another. An interesting observation was made regarding this phenomena: these variations correlated precisely with their respective DNA templates TR1 and TR2. Differences between the sequences of both TR1 and TR2 samples were recorded as single nucleotide polymorphisms (SNPs), deletions, and heterozygosities (confined to TR2).

Two regions of polynucleotide variability were identified during manual screening of sequences and are shown in the Figures 17 and 18. The inconsistencies of
the BLAST2Seq of the full *POGO* consensus against RNASeq Contig 22506 coincide with these variable regions.

a. *Variable Region I (VR1)*

**Figure 17:** Putative variable region 1 (VR1). From the top: Chromatograms 1, 3, and 6 are sequences generated from TR2. Chromatograms 2, 4, and 5 are sequences generated from TR1.
b. *Variable Region 2 (VR2)*

![Chromatogram Image]

**Figure 18:** Putative variable region 2 (VR2). From the top: Chromatograms 1, 3, and 5 are sequences generated from TR1. Chromatograms 2, 4, and 6 are sequences generated from TR2.
Variable Region 1 (VR1) includes a 6 nucleotide region, with the 5th base being a conserved adenine (A) for all sequences.

VR2 is much larger than that of VR1. Labeled in a white rectangle at the top of Figure 18 are six nucleotides 5’ TCAATG 3’ which are present in sequences generated from both TR1 and TR2, but absent in the RNASeq Contig 22506. The enlarged thumbnail depicts a small nineteen nucleotide sequence that has a total of six positions where the bases differ. Moreover, four of these six locations show potential heterozygosities with overlapping peaks. This variable region (TR2) shows multiple positions at which base calls vary between chromatograms as well as several located within the upstream and downstream flanking sequences (see Figure 19).

As noted above, TR1 and TR2 sequences were separated and individual consensus sequences were derived for each and edited. The sequence of the RNA-Seq Contig 22506 was aligned to the new consensuses separately, and for TR2 showed only one region of dissimilarity. (Figure 16). A Clustal Omega alignment of all three sequences was created to visualize differences among them. The alignment shows homology between the Contig 22506, TR1, and TR2 for the majority of their sequences (Figure 19), with the exception of two regions; these correspond with variable regions 1 and 2 that were identified.
<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TR2consensus</td>
<td>TGTATTATAAGAACTTTTCCTTTTAATATCTGTGCTTGTCCCTACCTCTGCTTCAGGTTT</td>
<td>820</td>
</tr>
<tr>
<td>Contig22506</td>
<td>TGTATTATAAGAACTTTTCCTTTTAATATCTGTGCTTGTCCCTACCTCTGCTTCAGGTTT</td>
<td>955</td>
</tr>
<tr>
<td>TR1consensus</td>
<td>TGTATTATAAGAACTTTTCCTTTTAATATCTGTGCTTGTCCCTACCTCTGCTTCAGGTTT</td>
<td>826</td>
</tr>
<tr>
<td></td>
<td>**************************************************************************</td>
<td></td>
</tr>
<tr>
<td>TR2consensus</td>
<td>TATATTTATAGCTGYAGTAGTTGTCTGATTATTATGATCAATGATCACATTTTTTGGGGG</td>
<td>880</td>
</tr>
<tr>
<td>Contig22506</td>
<td>TATATTTATAGCTGTAGTAGTTGTCTGATTATTATG------AATCACATTTT---TGGGGG</td>
<td>1008</td>
</tr>
<tr>
<td>TR1consensus</td>
<td>TATATTTATAGCTGTAGTAGTTGTCTGATTATTATGATCAATGATCACATTTT-TGGGGG</td>
<td>885</td>
</tr>
<tr>
<td></td>
<td>***********************</td>
<td></td>
</tr>
<tr>
<td>TR2consensus</td>
<td>CAGGAATAGCATCATCATCATCATAATTTAATCCATATTTTTCATGATAAGCTTGYCTYA</td>
<td>940</td>
</tr>
<tr>
<td>Contig22506</td>
<td>CTGGAATAATATTATCATCATCATAATTTAATCCATATTTTTCATGATAAGCTTGCCTTA</td>
<td>1068</td>
</tr>
<tr>
<td>TR1consensus</td>
<td>CTGGAATAATATTATCATCATCATAATTTAATCCATATTTTTCATGATAAGCTTGCCTTA</td>
<td>945</td>
</tr>
<tr>
<td></td>
<td>***********************</td>
<td></td>
</tr>
<tr>
<td>TR2consensus</td>
<td>ACAAATCACTCCGTTTTCKCCATTCATCAATATCTACAACATTCAAATGAAATGCTTTTG</td>
<td>1000</td>
</tr>
<tr>
<td>Contig22506</td>
<td>ACAAATCACTCCGTTTTCCCCATTCATCAATATCGACAACATTCAAATGAAATGCTTTTG</td>
<td>1128</td>
</tr>
<tr>
<td>TR1consensus</td>
<td>ACAAATCACTCCGTTTTCCCCATTCATCAATATCGACAACATTCAAATGAAATGCTTTTG</td>
<td>1005</td>
</tr>
<tr>
<td></td>
<td>**************************************************************************</td>
<td></td>
</tr>
<tr>
<td>TR2consensus</td>
<td>CTATATGTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
<td>1013</td>
</tr>
<tr>
<td>Contig22506</td>
<td>CTATATGTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
<td>1188</td>
</tr>
<tr>
<td>TR1consensus</td>
<td>CTATATGTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
<td>1028</td>
</tr>
</tbody>
</table>

**Figure 19.** A Clustal Omega alignment comparing both TR1 and TR2 *POGO* consensus sequences generated with Contig 22506 from the RNA-Seq library. As can be seen, the sequences share very high homology between all three along the length of the element. Conservation between sequences is highlighted - green indicates homology between Contig 22506 and TR2, blue indicates homology between Contig 22506 and TR1, and yellow indicates homology between TR1 and TR2. Red was used to highlight the nucleotide differences. Heterozygosities notated using IUPAC designations (GDR, 1984).

V. Homology and Conserved Domains of Putative POGO Element

BLASTx analysis (Figure 20) of the POGO consensus sequence shows modest (38%) but significant homology (Expect values of 1e⁻¹¹) to dozens of POGO transposable elements with KRAB domains, zinc finger proteins, and the Centromere Binding Protein.

B. The phylogenetic distribution is also quite diverse ranging from the whipworm to sea anemones, fish, and various other vertebrates. Although these hits were produced using the combined POGO consensus that was generated prior to separating, both TR1 and TR2
consensus sequences produce identical results when run in a BLASTx analysis (data not shown).
**Figure 21:** BLASTx analysis identifies two putative conserved domains within ORF -2 for the TR1 and Contig 22506 as well as RF -1 for TR2 sequences (data not shown).

Additionally, the BLASTx of both the TR1 and TR2 consensus sequences show that they contain two conserved domain superfamilies, BrkDBD and HTH_Tnp_Tc5, that are conserved in POGO transposable elements.
a. **BrkDBD**

The first conserved domain superfamily is the Brinker DNA Binding domain (Sivasankaran *et al.*, 2000). This domain consists of 113 nucleotide and 38 amino acid residues. Alignment of the two domains shows 50% homology and a deletion of five amino acid residues at positions 22 through 26.

b. **HTH_Tnp_Tc5**

This superfamily includes two putative domains, both of which are found within the TR1 and TR2 consensus sequences as well as Contig 22506: the Centromere Protein B (CENPB) and the Tc5 transposase DNA-binding domains. Shown in Figure 21 above, is the putative conserved Tc5 transposase domain. This DNA binding domain is 174 nucleotides long, 58 amino acid residues in length and conserved at 40% (23 out of 58 residues).

The putative CENPB domain is conserved at 40% (23 of 58 residues) and spans the same stretch of residues as the Tc5 transposase, however, also includes a serine (S) residue at position 32 and a lysine (K) residue at position 41 not found in the known consensus of this motif.
VI. Evidence for the presence of POGO transposable elements in other Cnidarians

Figure 22. PCR of multiple Cnidarian species; samples run on 1% (w/v) agarose gel. Lanes 1 and 13 are Hi-Lo ladders. Lanes 6, 11, 17, and 22 were NTCs done utilizing all four primers and sterilized water as a template. Multiple bands can be seen in reactions prepared using *Tamoya haplonema* and *Gonionemus vertens*.
The gel (Figure 22) suggests that POGO transposable elements may be present in the genomes of other Cnidarians, including *Tamoya haplonema* (TH) and *Gonionemus vertens* (GV).

Bands of expected sizes were seen each lane, however, were much fainter than those produced by *C. chesapeakei*. A reaction sample of particular interest was one of *Gonionemus vertens* which showed a band of high intensity (lane 15, Figure 22) correlating with the predicted band size for the primer pair used.

The absence of banding in the NTC lanes 6, 11, 17, and 22 discredit the possibility that the multiple bands observed were due to contamination. The multiple banding observed in reactions of TH and GV are likely related to the fact the primers used were created using a *Chrysaora chesapeakei* as a template.
DISCUSSION

I. Bioinformatic analysis and evidence for the presence of a partial POGO element in *C. chesapeakei* genomic DNA.

The sequence data obtained suggests the presence of at least one active POGO transposable element in the *C. chesapeakei* genome. This claim is supported in several different ways by analysis of the sequences and data obtained.

i. Confirmation of Target Amplicon by Size Validation

Sizes of each potential amplicon were calculated based upon the positions of each primer relative to the RNA-Seq contig. The bands on all agarose gels correspond precisely with the expected fragment lengths of different primer combinations, confirming that the target sequences were successfully amplified. More importantly, it confirms the validity of the assembly method used for the contigs in the RNA-Seq library. Additionally, it confirms that this element does not contain any introns, at least between the primer sets used in these experiments.

ii. Reliability of Sequence Data

With the exception of the largest predicted amplicon (1,076 bp for fragments amplified by the 4F and 5R primers), the amplified sequence lengths for each fragment was very close to their expected size. The longest amplicon obtained was 911 bp, however, this was good considering the performance specifications for the ABI 3130 Genetic Analyzer is cited to produce a maximum read length of 950 bp.
The sequences can be assumed correct given that the peaks of individual chromatograms were clean and defined with no background noise. In all cases sequences were run multiple times and both forward and reverse sequences were generated to facilitate calling of difficult regions and to verify final reads. The base call values were acceptable with values typically between 150-250 with the lowest being 90. There were several instances of peaks that overlapped but are likely to represent endogenous SNPs; this hypothesis was confirmed during the multiple sequence alignment in Geneious. These nucleotide differences appeared to occur at specific positions not only amongst different samples but also between different individuals. This phenomena may be indicative of variation amongst members of the Barnegat Bay population of *C. chesapeakei*.

### iii. Comparison to Existing POGK Genes

Figure 23: The protein blast (BLASTP) of a transcript produced by the POGK gene from *Homo sapiens* (NP_060012.3). As can be seen are four conserved domains of POGO transposable elements: KRAB, BrkDBD, HTH_Tnp_Tc5, and DDE_1.

The domain view of the POGK gene from *Homo sapiens* (accession NP_060012.3) shows a similarity to that of the putative POGO sequence isolated from *C. chesapeakei*. Two of the four domains within the putative contig match that within the known sequence (Figure 23). The presence of both the Brinker-binding and Tc5
transposase domains, in conjunction with the list of related sequences, suggests that our sequence represents part of the whole POGK gene.

Additionally, the assembled putative contigs directly correspond with the sequence of the RNA-Seq Contig 22506 of which the primers used in this experiment were based upon. Lack of apparent intronic sequence indicates that not only is this transcript actively expressed in our specimen, but also that only a partial portion of the element has been amplified.

II. Sequence Variation and Variable Regions

i. Single Nucleotide Polymorphisms

As was shown in all comparisons of analyzed data (Geneious, Clustal, and BLAST alignments), the presence of genetic differences, predominantly single point mutations, were not uncommon between different amplified fragments.

SNPs were found between the three different DNA templates utilized (RNASeq, TR1, and TR2). It is suspected that these differences are directly correlated with variation between individuals of the jellyfish population given that each one had 2-3 additional amplicons produced from the same DNA template confirming the nucleotide of interest at the given position.

ii. Heterozygositites

Aside from point differences between different sequences, randomly interspersed throughout individual sequences were occurrences of overlapping peaks. The likelihood that these have arisen as the sole result of incorrect nucleotide addition by the polymerase
is extremely low, given the high similarity in peak height. Rather these overlaps may be indicative of variation within the individual fragment population as a result of multiple copies of the target sequence being amplified within the genome. It was noted that heterozygosities were confined primarily to the TR2 sample template and not present in the template used to create Contig 22506 nor the TR1 template.

iii. Variable Regions

In addition to the SNPs found previously, two regions of variability, VR1 and VR2, were isolated. They have been deemed variable regions, based upon having multiple consecutive and non-consecutive differences between DNA templates, that are found within proximity of one another at specific loci.

Within VR2, nucleotide differences are present between different DNA templates, signal overlap within sequences using the same template, as well as a polynucleotide indel found within two-of-three DNA templates. The large quantity and high degree of variability found within such a short sequence makes VR2 of especial interest. Additionally, it was observed that a higher concentration of variation is also found in the sequences flanking VR2 relative to the entire length of the amplicon.

The reason for the presence of these regions of variability is unknown, however, they are conserved in three different DNA templates. One explanation may be the lack of selective pressure at this particular locus resulting in higher rates of mutagenesis in these regions. Both VR1 and VR2 are located within protein coding regions, thus influencing translation. Because VR2 is located outside of any conserved domain this may suggest
that this specific region of the DNA is non-essential to survival of the species. The 6 nt difference of VR2 results in loss of two amino acids and has a minimal effect on the overall protein. Alternatively, the 5 nt insertion of VR1 causes a frame-shift mutation altering the protein’s sequence and length. Furthermore, the lack of conservation within these regions may go hand in hand with inactivation of the transposase after loss of activity.

**Figure 24**: ClustalX alignment of ORF found within the sequence. The 5 nt insertion of VR1 shifts the reading frame resulting in a different peptide sequence and premature termination. Mutations such as this one may be responsible for decay of transposase activity over time.

Initially thought to be one species of jellyfish, *Chrysaora quinquecirrha*, recent research suggests that there are two distinct *Chrysaora* species of the Atlantic Sea Nettle. Genetic and morphological analysis has revealed the presence of two distinct *Chrysaora* species by geographical distribution (Bayha *et al.*, 2017). They are the Atlantic Sea Nettle, *Chrysaora quinquecirrha*, found within coastal ocean waters versus the newly classified estuarine Atlantic Bay Nettle, *Chrysaora chesapeakei* (Bayha *et al.*, 2017). There are undoubtedly variable regions as shown by the sequence data. For VR2, it is uncertain whether or not this particular point in the genome is variable for other jellyfish
within the same genus, or simply representative of variation within a given population of the same species.

III. Significance and Implications of Observed Conserved Domains

The presence of both the Brinker-binding and Tc5 transposase-binding domains is especially significant given their structure-function profiles in relation to the known structure of transposons. Van Pouderoyn *et al.* (1997) describes two separate HTH (Helix-Turn-Helix) motifs of the transposase binding domain, wherein the first is a homeobox domain, and the second a paired domain.

i. Brinker DNA Binding Domain

Brinker is a protein that has been found to have a profound effect on the Decapentaplegic (Dpp) morphogen pathway of *Drosophila*. More specifically, this pathway plays an integral role in embryonic development and is modulated by the sequence specific binding of Brinker to Dpp controlled genes and subsequent transcriptional repression of those genes (Sivasankaran *et al.*, 2000; Cordier *et al.*, 2006).

Structurally, the Brinker protein is comprised of four alpha helices with an HTH motif found within the first 44-99 amino acid residues of the N-terminal region (Sivasankaran *et al.*, 2000).

My *C. chesapeakei* POGO sequence shows homology to the sequence of the putative Brinker DNA binding domain (BrkDBD) generated. This domain found within the sequence is noted as approximately 38 amino acid residues (43 with the inclusion of
the 5 residue discrepancy shown in Figure 21); these findings are not far from the expected domain length.

**Figure 25:** A more focused BLASTx analysis of the 113 nucleotide sequence of only the Brinker DNA binding domain shows a much higher level of conservation. The closest and highest conserved sequence is from *Stylophora pistillata*, a species of coral and member of the phylum Cnidaria, matching at 68% identity.

The presence of this domain is significant as it likely represents the first HTH motif of the greater bipartite DNA binding domain of the *POGO* element. Of the two, it is hypothesized that this conserved domain is the homeobox domain based upon what is
known of the Dpp and Brinker/BrkDBD pathways. A BLASTx of just the 113 nucleotides of the Brinker DNA binding domain (Figure 25) shows a much higher degree of conservation relative to the rest of the amplicon, perhaps suggesting that at one point in time during the evolution of this species’ genome, there was a functional need for the protein produced by this gene.

ii. **HTH-Tnp-Tc5 Superfamily**

This superfamily includes two homologous domains which are the Tc5 transposase and the centromere binding protein B (CENBP). As a whole, presence of the domain superfamily is significant as it directly corresponds to the known binding domain for the Tc5 transposase. This domain was the fifth Tc element isolated from *C. elegans* and a related member of the Tc1/mariner superfamily of transposons (Collins and Andersen, 1994).

IV. **POGK, Genome Evolution, and Transposase Integration as means for Survival**

*Genome Evolution*

Genomic landscapes and sizes have been directly impacted by the presence of transposable elements (Kidwell, 2002). Within humans alone, 44% of our genome is occupied by transposons (Mills, 2007; Lander *et al.*, 2001). In their 2007 publication Feschotte and Pritham (2007) outline three potential mechanisms by which transposons may impact genome evolution:
“(i) via alterations of gene function through insertion; (ii) through the induction of chromosomal rearrangements; (iii) as a source of coding and noncoding material that allows for the emergence of genetic novelty (such as new genes and regulatory sequences).” – Feschotte and Pritham, 2007

**Domestication and Exaptation**

Furthermore, various properties such as palindromic structures and inherent functions of TEs make them good candidates for domestication or exaptation (Brosius and Gould, 1992) into host genomes. This notion of derivation and co-evolution has been repeatedly suggested and supported as more and more genetic data has become available, linking functional genes to transposable element origins (Kidwell and Lisch, 2001; Volff, 2006; Feschotte and Pritham, 2007).

The relationship between transposons and DNA repair / replication factors may account for a higher propensity of transposase domestication. This would also justify the correlation between many transposase-derived proteins having recombination (Jones and Gellert, 2004), cell cycle control (Walisko *et al.*, 2006), and other chromosome-related functions.

**POGO derived elements**

These integration events are not isolated and are observed with members of the *POGO* family of transposons. Conservation of the Proliferating Cell Nuclear Antigen (PCNA) binding domain has been shown within the *POGO* transposases of *D.*
melanogaster and human Tigger1 (Warbick et al., 1998), as well as the POGO-like transposase Lemi1 from Arabidopsis (Feschotte and Mouches, 2000). POGO-like transposases have also been observed in the centromere binding protein (CENP-B) of fission yeast and mammals and thought to have been convergently domesticated (Casola et al., 2008).

POGK is another known transposase-derived gene. With this in mind, it is plausible to believe that the sequences generated within this experiment represent a homolog of the POGK gene found within humans and other mammals. The function of the POGK gene is not well studied, however, it is believed to play a role in transcription due to fusion with a KRAB domain, a known transcriptional repressor (Margolin et al., 1994).

V. Future Directions

Although I have discovered the presence of POGO transposable elements in a common Scyphozoan, there is still work to be done. Comparison of the sequence data obtained has elucidated that the fragment amplified only represents the middle portion (approximately 56%) of what is presumed to be the POGK gene. This claim was supported by the near identical homology of the putative TR1 and TR2 consensus sequences to that of Contig 22506 (Figure 18). It is apparent that the ends of this gene have not yet been sequenced as the presence of an intron would have suggested otherwise. Continuation of this project would need to employ a method for obtaining the sequence of the full element. Inverse PCR overcomes the hurdle by providing a method
for amplifying DNA outside of a known sequence and has long been employed in identifying the flanking regions and insertion sites of transposable elements (Ochman et al., 1988; 1990).

Southern blot analysis of this gene would provide potentially useful information such as the location of the gene within this species’ genome as well as information regarding the number of copies present. Class II transposons such as the POGO element are not known to be replicative as their mechanism of transposition would dictate, however, transposition during the cell’s S Phase would result in the production of a duplicated gene. Although the likelihood may be low given the specific conditions needed and incidence rate not well known, it cannot be discounted as a potential explanation for multiple copies of Class II transposons in a species’ genome. Heterozygositites were found within the TR2 consensus sequence; such observations may suggest multiple copies/variations of the gene were present in the sample and subsequently amplified.

Prevalence of the POGK gene is widespread throughout different taxa (Casola et al., 2008). Its presence within Cnidaria may be the result of lateral gene transfer of an ancestral element over millions of years, or a convergent domestication event described above. Following cloning of the full gene (using methods previously described), determining the presence of this fragment within other closely related Cnidarian species would be the next logical pursuit. It is unclear whether the variable regions found within this experiment can also be found within other species and identifying the underlying
mechanism for their appearance may be of interest. Preliminary experiments conducted have produced promising results using the primers designed within this experiment in conjunction with various DNA templates from other Cnidarians (data not included). This data would serve to help build upon what is known regarding the evolutionary origin of this gene and its chronological time of domestication.
LITERATURE CITED


McClintock, B. (1953). Induction of instability at selected loci in maize. *Genetics, 38*(6), 579.


