The evolution of heat shock 70 proteins in sea anemones

Undergraduate Research Thesis

Presented in Partial Fulfillment of the Requirements for graduation “with Honors Research Distinction in Zoology” In the undergraduate colleges of The Ohio State University

by
Annelise Del Rio

The Ohio State University
May 2015

Project Advisor: Professor Marymegan Daly, Department of Evolution Ecology and Organismal Biology
Abstract

Heat shock protein 70 (Hsp70) is a ubiquitous protein that functions as a molecular chaperone to repair damaged proteins. Found in all organisms from archaeabacteria to humans, Hsp70 is important in the response of organisms to a variety of environmental stressors such as elevated temperature and oxidation. As the environmental conditions of our oceans are subjected to the impacts of climate change, understanding the stress response of marine organisms is becoming increasingly important to predict how marine organisms may respond to environmental stressors. Cnidarians, including sea anemones and reef building corals, are extremely important marine species which are susceptible to environmental stressors brought on by climate change. To better understand the potential stress responses of these cnidarians, characterizing Hsp70 diversity is necessary. To study how this gene has evolved, RNA was extracted from 20 species of sea anemones and partial transcriptomes were sequenced. Bioinformatic methods were then used to identify and extract candidate Hsp70 sequences from each transcriptome. The extracted sequences for each species were then aligned and used to create a tree of Hsp70 gene families. Across the focal taxa we observed variation in sequence copy number and amino acid composition. Although variation in copy number was coupled with the quality of the assembly, there was apparent concerted /birth-death gene family evolution observed in the Hsp70 genes. Additionally, Hsp70 was highly conserved throughout the tree indicating potential functional constraints or retained phylogenetic signal. The Hsp70 is just one of many genes that can play a key role in a species resilience against climate change. Our findings contribute to understanding the repertoire of genomic resilience across these species of sea anemones.
Introduction

As the environmental conditions of our oceans are subjected to the impacts of climate change, understanding the stress response of marine organisms is becoming increasingly important. For many organisms stress may be interpreted in a variety of ways. Our focus is specifically the condition in which the homeostasis of an organism is threatened or inhibited by external or internal stimuli, commonly referred to as stressors (Wendelaar Bonga, 1997). Organisms often have three primary mechanisms to cope with environmental stress. They can modify their behavior and movement, develop a greater resistance and phenotypic plasticity, or activate recovery mechanisms to repair damages caused by the stressor (Huey et al., 2000).

Cnidarians, including sea anemones and reef building corals, are extremely important marine species that are particularly susceptible to environmental stressors brought on by climate change. For sessile cnidarians, stressors may be in the form of changing temperature, pH, salinity, or other conditions they cannot escape, making physiological repair a primary survival mechanism. If the stress or stressors persist without the aid of a stress genomic repertoire, these sessile organisms may experience effects of reduced fitness including illness or death.

In order to better predict how various environmental stressors may impact marine organisms, we must first understand the natural responses these organisms can employ. Heat shock protein gene families provide insight into intracellular mechanisms these animals utilize to survive in stressful and changing environmental conditions. These genes and their products are organized into families based on their molecular weight and sequence homology. Heat shock protein 70 (Hsp70), is the most conserved protein in evolution given the deep evolutionary time scale in which this gene family exists (Daugaard et al., 2007). Hsp70 is a ubiquitous, highly conserved protein found in all organisms, from archaeabacteria to humans that functions as a
molecular chaperone to repair damaged and reversibly denatured proteins in cells (Feder and Hofmann, 1999, Mayer and Bukau, 2004, Lindquist, 1986). It is also the most abundant heat shock protein produced after a temperature increase (Black and Bloom, 1984, Lindquist, 1986). Hsp70 gene products vary in their amino acid sequence, level of expression, and their location within a cell, but serve similar functions due to the conservation of their domain structures and the majority of amino acid sequences (Daugaard et al., 2007).

It is likely that Hsp70 has co-occurred with proteins early on, aiding organisms in a variety of regulatory functions in addition to stress response processes necessary in maintaining homeostasis (Feder, 1999). Members of the Hsp70 family have both stress-induced and regulatory functions that are expressed independently (Daugaard et al., 2007). Stress-inducible heat shock proteins respond to stressors that cause abnormal protein conformations including changes in ions, osmolytes, toxic substances, and gases, in addition to extreme temperatures (Feder and Hofmann, 1999). Most species have at least two inducible Hsp70 gene products, which in their protective function are induced at elevated temperatures with the temperature at induction typically representing the upper limit of the organism’s natural growth range (Lindquist, 1986). Most species also have multiple regulatory Hsp70 genes that are expressed constitutively in non-stressed cells and function to transport proteins within a cell, modify protein folding conformations, and control additional regulatory proteins. Mice lacking stress-inducible Hsp70 proteins are viable, but more sensitive to sepsis and radiation and prone to genomic instability, which highlights the synergistic effects and significance of multiple Hsp70 genes (Daugaard et al., 2007).

Among Eukaryotes, sessile marine organisms such as bivalves have been shown to exhibit a rapid radiation in their HSP70 diversity (Fabbri et al. 2008, Pantartzki et al., 2014).
Most organisms have multiple copies of Hsp70 genes. Studies in *Drosophila* have found that a higher number of Hsp70 gene copies increases Hsp70 expression as well as overall thermotolerance, however, individuals with more copies of the gene also have a higher metabolic rate during heat stress, indicative of a tradeoff in physiological performance (Feder et al., 1996, Hoekstra and Montooth, 2013). Although the number of Hsp70 genes throughout Cnidaria is not known, there have been several studies highlighting the role of Hsp70 in cnidarian stress responses, particularly with stressors related to climate change. Previous studies have focused on the expression of Hsp70 in a few select species exposed to elevated temperatures in both laboratory and field experiments (Black and Bloom, 1984, Choresh et al., 2001, Sharp et al., 1996, Snyder and Rossi, 2004), which have indicated that Hsp70 proteins are involved in the physiological response of cnidarians to increasing ocean temperatures. The role of HSPs in sessile corals in various scenarios has been the focus of these investigations, particularly in reef building corals (Souter et al., 2011, Edge et al., 2012, Downs et al., 2009). While it is clear that Hsp70 is involved in the stress response of cnidarians based on patterns of up-regulated expression in relation to the level of environmental stress, it is unknown to what extent this protein varies across sea anemones and to what extent gene family evolution may have influenced Hsp70 function.

To better understand the potential stress response of sea anemones it is necessary to characterize Hsp70 diversity. Here we use combined phylogenetics and gene family reconstructions to study the evolution of Hsp70 variation across multiple species of sea anemones (Actiniaria). When compared to the established general phylogeny of sea anemones it provides evolutionary insight to the diversification of this protein. In this study Hsp70 gene sequences extracted from transcriptome data were compared across 20 species of ecologically
diverse sea anemones. There was variation in the sequence copy number and amino acid composition along with evident concerted/birth-death gene family evolution observed in the Hsp70 genes. While Hsp70 is one of many genes that are important in species resilience, our findings contribute a better understanding of the repertoire of genomic resilience across these species of sea anemones.

Methods

Data Collection

Focal taxa for this study represent 20 species found across Actiniaria, with an approximately equal number of species from the superfamilies Actinioidea and Metridioidea and one from Edwardsioidea (Table 1). 9 of these species are found in tropical regions and 7 inhabit intertidal zones. Total RNA was extracted from either fresh or RNA preserved tissues following a combined flash freezing, maceration, and extraction protocol (Macrander et al., 2015). Fresh tissues and those stored in RNALater were flash frozen in liquid nitrogen and macerated in a bead beater using ceramic beads. After homogenization the total RNA was extracted following the RNeasy Mini Kit (QIAGEN) protocol. The conditions in which specimens were prepared prior to RNA extraction and subsequent sequencing varied, making it difficult to control for differential expression of any stress response genes. Therefore, an examination of differential expression of HSP70 would be inappropriate and we instead focus on HSP70 gene family evolution.

Once the total RNA was extracted, samples were quantified on a Qubit 2.0 Fluorometer (Life Technologies) using the Qubit RNA BR Assay kit (Life Technologies). Additionally, RIN scores were calculated using the Agilent RNA 6000 Nano kit (Agilent Technologies) on the
BioAnalyzer (Agilent Technologies). Sample preparation for sequencing, including first strand synthesis and library construction, were conducted at the Nucleic Acid Shared Resource – Illumina Core, The Ohio State University, Columbus, OH, USA.

The raw reads were prepared as previously described (Macrander et al., 2015). Briefly, the program Trimmomatic (Bolger and Giorgi) was used to remove adapters, low quality bases at the leading and trailing ends (using a sliding window greater than 3 bases), reads with less than 36 bases, and raw reads with an average quality score less than 15 using a four base sliding window. Once cleaned, the data was visualized in FastQC (Andres) to ensure that low quality reads and regions were removed. The resulting base paired ends were then assembled using the program Trinity (Grabherr et al., 2011) using default parameters. The resulting transcriptomes were investigated using bioinformatic methods to extract candidate Hsp70 sequences using the program blastx to compare the transcriptomes to a known Hsp70 sequence on GenBank (ID: 729765, 462324)

**Gene Family Tree Construction**

These candidate sequences were visually inspected in the program Geneious, version 7 (http://www.geneious.com, Kearse et al., 2012). The final 277 extracted sequences were aligned in Geneious using a MAFFT (Katoh, 2013) translation alignment. Variable sequence ends were trimmed to include only sequence regions which overlapped with previously published cnidarian Hsp70 sequences. The resulting alignment had sequences ranging from 1581 to 3435 base pairs in length. The aligned sequences were used to build a phylogenetic tree in the program MEGA using maximum likelihood analysis with 1000 bootstrap replicates (Tamura et al., 2013). The species *E. lineata*, belonging to the superfamily Edwardsioidea, was chosen as the outgroup
based on sea anemone superfamily phylogeny, along with a reference Hsp70 sequence from the coral *Stylophora pistillata* available on GenBank (ID:6465981).

Clades with bootstrap values over 50 were selected for further analysis. Each species in the selected clades was labeled according to the superfamily to which it belongs to better compare the phylogeny of Hsp70 sequences to the accepted general phylogeny of sea anemones (Rodriguez et al., 2014). Within our larger Hsp70 alignment, groupings with high support values were visually inspected to determine whether there was variation in sequence alignments within these gene trees that is correlated to functional regions within the genes or functionally different Hsp70 subtypes. The sequences with variation were then analyzed using the program BLAST and blastx to identify and compare known functions of the proteins (Altschul et al., 1990).

**Results**

The resulting Hsp70 gene family tree grouped the twenty species into nine general gene clusters (Figure 1). The three clades with bootstrap support values above 50 were selected for further analysis. These groupings varied in size from 7 to 52 branches, often times representing multiple Hsp70 genes of the same species within a single clade. The Hsp70 phylogeny showed mixed results compared to the general sea anemone phylogeny. Some clades (Figure 1B and 1.2) have groups of species that are somewhat reflective of their phylogenetic relationships with clear separation of the superfamilies they represent. Other clades (Figure 1C) resulted in groupings with superfamilies intermixed and no apparent relation to the general phylogeny.

Within the Hsp70 gene, the largest gene cluster (Figure 1A – 1D) was examined more closely in order to identify any sequences or functional regions that deviated from other associated branches that could relate to functional differences. There appeared to be distinct
groupings of conserved sequences within the gene clusters. Searches were conducted using blastx to search a protein database using translated Hsp70 nucleotide queries. Within clades, sequences that represented the general diversity of the groupings of conserved sequences were compared using the blastx search to identify their putative function. There were not clear results as to how the proteins differed in function and what effects the variations in sequence produce due to a lack of data regarding specific protein function.

<table>
<thead>
<tr>
<th>Table 1. Taxon Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
</tr>
<tr>
<td>A discipulorum</td>
</tr>
<tr>
<td>A elegantissima</td>
</tr>
<tr>
<td>A equina</td>
</tr>
<tr>
<td>A pallida</td>
</tr>
<tr>
<td>A sulcata</td>
</tr>
<tr>
<td>B annulata</td>
</tr>
<tr>
<td>B cavernata</td>
</tr>
<tr>
<td>B globulifera</td>
</tr>
<tr>
<td>C parasitica</td>
</tr>
<tr>
<td>D leucolena</td>
</tr>
<tr>
<td>D lineata</td>
</tr>
<tr>
<td>E prolifera</td>
</tr>
<tr>
<td>E quadricolor</td>
</tr>
<tr>
<td>Haloclava</td>
</tr>
<tr>
<td>E lineata</td>
</tr>
<tr>
<td>M doreensis</td>
</tr>
<tr>
<td>M griffithsi</td>
</tr>
<tr>
<td>M senile</td>
</tr>
<tr>
<td>S coccinea</td>
</tr>
<tr>
<td>S elegans</td>
</tr>
</tbody>
</table>

Table 1. List of all species analyzed, the tissue RNA was extracted from, the super family and family to which they belong, and the number of Hsp70 gene copies each species had based on the number of times it appeared on the tree.
Figure 1. The resulting Hsp70 gene family tree from maximum likelihood analysis with gene clusters of high support values highlighted. These highly supported clusters are represented individually in figures 1A-D, 1.2, and 1.3. The colors correspond to the super family the species belong to. Green represents Edwardsioidea, red Metridioidea, and blue Actinioidea. Bootstrap support values of 50 or greater are shown on the tree along with the sequence references for each species.

Discussion

The resulting Hsp70 gene family tree depicting the diversity of heat shock protein 70 genes provides insight into the evolution of this gene family in sea anemones. Overall there were a high number of Hsp70 gene copies across taxa. Most species were represented in 4-6 of the gene clusters and the number of times each species was represented in all clusters ranged from 3 to 25 times. Species with lower gene copy numbers also had incomplete transcriptome data (Haloclava, E. prolifera) whereas the species with complete transcriptomes had higher copy numbers (S. coccina, M. griffithsi). The high copy number likely resulted from several gene duplications reflecting the conserved aspects of this protein family. For some species such as A. discipulorum, there were several copies of identical sequences in the same clade that may represent potential isoforms. These sequences were later condensed on the tree and the number of identical copies appears in parentheses after the species names. Heat shock protein gene families have developed through both concerted and birth-and-death evolution in other species (Nei and Rooney, 2005). However, the possible concerted evolution may simply be a byproduct of the assembly program. The raw reads flanking the exons may have sequencing or other errors that assign them to different genes rather than reassemble them in the same contig (Grabherr et al., 2011). The pattern of diversification here indicates mechanisms of birth-and-death evolution may have contributed to the variability in the type and number of gene copies each species has.
Each major gene cluster typically contained three to four groups of similar Hsp70 sequences. These sequences varied in composition sometimes by just a few amino acids, and in other cases they were very different in length and sequence. There have been limited studies conducted thus far to determine the function of each Hsp70 gene product and the functional effects of the variations in amino acids. In some cases it appeared that the diversity within and between clusters could be due to differences in regulatory and inducible proteins where the proteins may have diverged based on different selective pressures on function, however there is a lack of data in this area to support a comprehensive analysis of these patterns.

The evolutionary patterns of Hsp70 from our results closely follow the established phylogeny for sea anemones in some clades, while other clades show no reflection of the relationship within or between sea anemone super families. This may be due to the presence of both regulatory and stress-inducible Hsp70 gene products. Selection may act more strongly on stress-inducible proteins causing more diversification based on individual species, whereas the regulatory gene products could retain a more highly conserved function. Differences between regulatory and stress-inducible Hsp70 genes may also explain the birth-and-death evolutionary patterns.

There are no clear relationships between the ecology of these species and their Hsp70 gene families. Tropical and temperate species showed no distinct differences in the number or type of gene families they possess or the number of times they appear on the tree. Although intertidal organisms express heat shock proteins on a daily basis to react to changing tidal and ambient conditions, intertidal species did not have a different Hsp70 gene family composition when compared to species typically found in deeper waters. Hsp70 gene families usually have fewer stress-inducible genes than regulatory genes. If selection is acting more strongly on those
few stress-inducible genes it could explain the lack of overall correlation between Hsp70 gene families and sea anemone ecology. Another possibility is that there is more variation in the expression levels of these proteins than in the gene copies. Variation in the expression of Hsp70 has been shown to influence thermotolerance but here we did not analyze levels of expression, only the gene copies.

Sea anemones possess a diverse array of Hsp70 proteins. It is not yet clear how this diversity contributes to the ability to withstand environmental changes or how this diverse gene family varies functionally. More data is needed to understand the specific role of each Hsp70 gene product and how they contribute to the stress response of the organism.
Acknowledgements
I would like to thank Jason Macrander and Marymegan Daly for their assistance with this project. Funding was provided by the National Science Foundation.

References
Andres S: FastQC. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
*Drosophila melanogaster* increases energetic demand. BMC Evolutionary Biology. 13:68.


Macrander, J., M. R. Brugler, and M. Daly. 2015. A RNA-seq approach to identify putative toxins from acrorhagi in aggressive and non-aggressive *Anthopleura elegantissima* polyps. BMC Genomics. 16:221


